Discussion Paper

On Kalman Filtering, Parameter Uncertainty and Variances after Single and Multiple Imputation

Paul Knottnerus

February 14, 2022

Contents

Summary: When observations in a survey are missing, one may use single imputation (SI) or multiple imputation (MI) for filling in the missing data. Using Kalman equations, we derive straightforward formulas for the total imputation variance for several imputation methods commonly used in regression analysis and (un)equal probability sampling without replacement. From these formulas it emerges that, in general, there is no necessity for taking parameter uncertainty into account. The paper proposes a new MI method which provides improved efficiency compared to the standard MI method for combining variances. Examples of the regression estimator for the unemployment rate and a simulation study are carried out to illustrate the theoretical results. The paper also discusses the important role of the underlying data model. It is shown that for some data patterns MI yields confidence intervals with undercoverage when estimating a ratio while SI and the new MI method produce valid confidence intervals. Obviously, these results are relevant for external users of imputed datasets as well. By providing several applications a better understanding of MI and SI may arise, including the three conditions for proper MI.
Key words: Confidence intervals; deterministic ratio imputation; g-weights; missing covariates; regression estimator; self-efficiency; unequal probability sampling.

Acknowledgements
I wish to thank Jeroen Pannekoek and Kees Zeelenberg for their comments on earlier drafts of the paper. I am also grateful to Sander Scholtus and Ton de Waal for their reviews and useful suggestions that have led to improvements of the paper. Besides, special thanks must go to Sander Scholtus who was so kind to carry out the simulations in section 6 and to read various drafts. The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

1 Introduction

Since nonresponse in many surveys is increasing, imputation methods are important for making correct statistical inferences. In particular, this holds for national statistical offices (NSOs), that have a long tradition with imputation [e.g., Deville and Särndal (1994) and De Waal et al. (2011)]. Comprehensive overviews of imputation methods are given in Murray (2018) and Chen and Haziza (2019). The main aim of this paper is to examine the impact of single imputation (SI) and multiple imputation (MI) on the variances of commonly used estimators by using explicit variance formulas, in particular for the variance of the widely used regression estimator of a population mean. This kind of insight is important not only for statistical agencies but also for external users (or analysts) of imputed datasets.

The outline of the paper is as follows. Section 2 introduces some notation and describes the MI method, including when sampling from a finite population. Furthermore, the three essential requirements for proper MI are explained in frequentist terms. Using Kalman equations, section 3 derives straightforward variance formulas applicable to several imputation methods, including hotdeck imputation. These formulas are in line with the MI results in Schenker and Welsh (1988), Wang and Robins (1998) and Kim (2004) who examined more specific situations. Section 4 examines SI methods as hotdeck which is important for discrete variables. It emerges that for other imputation methods than Rubin's (1987) MI there are no frequentist reasons for accounting for uncertainty of estimated parameters. Further, we propose a new MI method with improved efficiency. To illustrate the merits of SI and MI, section 5 compares the 95% confidence intervals for several SI and MI methods. Section 6 derives variance formulas for the regression estimator after using imputations. To illustrate the formulas, the section gives several examples of the regression estimator of the unemployment rate after using MI, hotdeck and the new MI method. Besides, a simulation study is conducted. Section 6 also pays attention to unequal probability sampling without replacement, and g-weights in the case of nonresponse. Section 7 discusses MI when estimating a ratio and the important role of the underlying data model. It emerges that for some data patterns Rubin's MI yields confidence intervals with undercoverage while SI and the new MI method work well. Section 7 also examines deterministic ratio (regression) imputation for the Horvitz-Thompson (HT) estimator as proposed by Deville and Särndal (1994) and shows that the ratio (regression) estimator has a smaller variance in such cases. Section 8 briefly discusses missing covariates and section 9 summarizes the main conclusions.

2 MI from a Frequentist Perspective

2.1 Notation and a Single MI Step

To explain the Bayesian MI method from a frequentist point of view, consider a sample s of n independent drawings $Y_1, \dots, Y_n$ from the normal distribution $N(\mu, \sigma^2)$. Suppose that a simple random subsample (say s_r) of p drawings is available and that q drawings are missing ($p + q = n$); this is often called missing completely at random (MCAR). The quantities p and q are fixed. For more notational convenience, denote the p observations by $y_1, \dots, y_p$ and the missing values by $y_{p+1}^{mis}, \dots, y_n^{mis}$. Let $\overline{y}_l$ and s_{yl}^2 be defined by $\overline{y}_l = \sum_{i=1}^{l} y_i/l$ and $s_{yl}^2 = \sum_{i=1}^{l} \left(y_i - \overline{y}_l\right)^2 /(l-1)$ $(l = 2, 3, \dots)$. Now almost similarly to the Bayesian approach of Rubin (1987, 83), consider the following q imputations

$$y_i \equiv \overline{y}_p + v + u_i \quad (i = p+1, \dots, n), \tag{2.1}$$

where v and the u_i are mutually independent drawings with $v \sim N(0, s_{yp}^2/p)$ and $u_i \sim N(0, s_{yp}^2)$. In MI terms, v reflects the uncertainty of $\overline{y}_p$. Note that from a frequentist point of view the random v is desirable to avoid that imputations are centered too much at $\overline{y}_p$ instead of μ leading to a biased estimator s_{yn}^2 for σ^2. Moreover, writing $y_i = \mu + \varepsilon_i$ $(i \le p)$ and $y_i^{mis} = \mu + \varepsilon_i$ $(i > p)$ where $\varepsilon_i \sim N(0, \sigma^2)$, we have $y_i^{mis} = \overline{y}_p - \overline{\varepsilon}_p + \varepsilon_i$.

Further, it follows from (2.1) that the (unbiased) estimator $\overline{y}_n$ for μ can be written as

$$\overline{y}_n = \overline{y}_p + e, \tag{2.2}$$

where $e \equiv q(v + \overline{u}_q)/n$ with $\overline{u}_q \equiv \sum_{i=p+1}^{n} u_i/q$. The variance of $\overline{y}_n$ is

$$\begin{aligned} \text{var}\left(\overline{y}_n\right) &= \text{var}\, E\left(\overline{y}_n \mid s_r\right) + E\, \text{var}\left(e \mid s_r\right) \\ &= \text{var}\left(\overline{y}_p\right) + E\left\{\frac{q^2}{n^2}\left(\frac{1}{p} + \frac{1}{q}\right) s_{yp}^2\right\} = \frac{\sigma^2}{p} + \frac{q\sigma^2}{np}. \end{aligned} \tag{2.3a}$$

When comparing SI and MI, it is useful to write (2.3a) as

$$\text{var}\left(\overline{y}_n\right) = \frac{\sigma^2}{n} + \frac{2q\sigma^2}{np} = \frac{\sigma^2}{n} + 2\text{var}(e). \tag{2.3b}$$

where we used $1/p = 1/n + q/np$. In addition, we get $E(s_{yn}^2) = \sigma^2$ because

$$\begin{aligned} E\{(n-1)s_{yn}^2\} &= E\left\{(p-1)s_{yp}^2 + \sum_{i=p+1}^{n} \left(y_i - \overline{y}_p\right)^2 - n\left(\overline{y}_n - \overline{y}_p\right)^2\right\} \\ &= (p-1)\sigma^2 + q\left(\frac{1}{p} + 1\right)\sigma^2 - \frac{q^2}{n}\left(\frac{1}{p} + \frac{1}{q}\right)\sigma^2 = (n-1)\sigma^2, \end{aligned} \tag{2.4}$$

where we used $\overline{y}_n - \overline{y}_p = e$; see (2.2). An unbiased estimator for var $\left(\overline{y}_n\right)$ can be obtained from (2.3b) by replacing σ^2 by s_{yp}^2 or s_{yn}^2. Hence, the 95% confidence interval for μ is $\overline{y}_n \pm 1.96 s_{yn}\sqrt{(1 + 2q/p)/n}$ almost just as in a complete-data situation $(p, n \gg 1)$; in the remainder of this paper we assume that there exists a c such that $p/n > c > 0$. When in practice the y_i are not normal, the u_i in (2.1) can be drawn with replacement from $\{y_1 - \overline{y}_p, \dots, y_p - \overline{y}_p\}$. This does not really affect the above results provided that p is sufficiently large so that $\overline{y}_p$ is normal; also see section 4. Finally, it should be noted that throughout this paper $E(.)$, var(.) and plim(.) are frequentist operators with respect to the sampling design and additional drawings in the case of missing observations.

2.2 The MI Method and Rubin's Rule for Combining Variances

Now we explain the major differences between a single MI step and the MI method:

A. Instead of s_{yp}^2 in the above imputations Rubin (1987, 83) uses s_{yp}^{*2} defined by $s_{yp}^{*2} = s_{yp}^2(p-1)/r_{p-1}$ where r_{p-1} is a drawing from a χ^2-distribution with $p-1$ degrees of freedom. This specific modification in his Bayesian approach is a consequence of the fact that his prior distribution for (μ, σ^2) has density proportional to σ^{-2}. From a Bayesian perspective the imputations $y_{p+1}, \dots, y_n$ thus obtained can be seen as drawings from the posterior distribution of the missing data.

B. The most important difference is that the above imputation procedure is to be repeated m times, with new independent drawings from the aforementioned distributions, including r_{p-1}. This leads to new sample means (say $\overline{y}_{nj}$), sample variances (say s^2_{ynj}) and imputation noise (say e_j) in the jth step ($j = 1, \dots, m$). In Rubin's notation this leads to estimators $\widehat{Q}_j$ $(= \overline{y}_{nj} = \overline{y}_p + e_j)$ for μ and the associated customary variance estimators $\widehat{U}_j$ $(= s^2_{ynj}/n)$.

In practice, the average (say $\overline{Q}_m$) of the $\widehat{Q}_j$ can be seen as the MI estimator for μ. Following Rubin (1987, 91), the variance of $(\mu - \overline{Q}_m)$ can be estimated by

$$T_m \equiv \overline{U}_m + (1 + 1/m)B_m, \tag{2.5}$$

where the within-imputation variance $\overline{U}_m$ and the between-imputation variance B_m are defined by $\overline{U}_m = \sum_{j=1}^m \widehat{U}_j/m$ and $B_m = \sum_{j=1}^m (\widehat{Q}_j - \overline{Q}_m)^2/(m-1)$. From a Bayesian perspective, $\overline{Q}_\infty$ and T_∞ are the mean and variance of the posterior distribution of μ given $y_1, \dots, y_p$ provided that the prior distribution of (μ, σ^2) has density proportional to σ^{-2}. Noting that $\widehat{Q}_j$, $\widehat{U}_j$ and B_m can also be seen as random variables in a frequentist sense, it holds that $E\left(\widehat{U}_j \mid s^{*2}_{ypj} = \sigma^2\right) = \sigma^2/n$ and since $\widehat{Q}_j - \overline{y}_p = e_j = q(v_j + \overline{u}_{qj})/n$, we get

$$E\left(B_m \mid s^{*2}_{ypj} = \sigma^2\right) = \text{var}\left(e_j \mid s^{*2}_{ypj} = \sigma^2\right) = \frac{q\sigma^2}{np} = \left(\frac{1}{p} - \frac{1}{n}\right)\sigma^2. \tag{2.6}$$

By (2.6), the expectation of the variance in (2.5) conditional on $s^{*2}_{ypj} = \sigma^2$ is

$$E\left(T_m \mid s^{*2}_{ypj} = \sigma^2\right) = \left(\frac{1}{n} + \frac{m+1}{mnp}q\right)\sigma^2 = \left(\frac{1}{p} + \frac{q}{mnp}\right)\sigma^2, \tag{2.7}$$

which equals the variance in (2.3) when $m = 1$. However, noting that $E(s^{*2}_{yp} \mid s_r) > \sigma^2$ because according to Jensen's inequality $E(r^{-1}_{p-1} \mid s_r) > 1/E(r_{p-1} \mid s_r)$, a practical disadvantage of Rubin's approach appears to be that the variance estimator in (2.5) may yield a substantial overestimate of the actual variance of $\overline{Q}_m$ in small samples.

Example 2.1. In this example, we look more closely at the consequences of the drawings from a χ^2-distribution when p is small. Noting that s^{*2}_{yp}/σ^2 follows Fisher's $F_{a,b}$-distribution with $a = b = p - 1$, we have $E\left(s^{*2}_{yp}\right) = \sigma^2(p-1)/(p-3)$. Now choosing, for example, $p = 4$ and $q = 1$, we get $E\left(s^{*2}_{yp}\right) = 3\sigma^2$. Consequently, formula (2.4) changes into $E\{(n-1)s^2_{yn}\} = (p - 1 + 3q)\sigma^2 = 6\sigma^2$. Hence, $E(s^2_{y5}) = 1.5\sigma^2$ and $E(\widehat{U}_j) = 6\sigma^2/20$. Since

$$E(B_m) = E\,E\left(B_m \mid s^{*2}_{yp}\right) = E\,\text{var}\left(e_j \mid s^{*2}_{yp}\right) = 3q\sigma^2/np = 3\sigma^2/20,$$

we get $E\,(T_m) = E(\overline{U}_m + B_m) = 9\sigma^2/20$ $(m \gg 1)$ which means an overestimation of $\text{var}(\overline{Q}_\infty) = \sigma^2/4$ by 80%; note that we used $\overline{Q}_\infty = E(\widehat{Q}_j \mid s_r) = \overline{y}_p$. For $p = q = 4$, it is almost 160%. To avoid this kind of problem, it should hold that $p, n \gg 1$ so that $E(s^{*2}_{yp}) \approx \sigma^2$; $a \approx b$ indicates that a/b tends to unity as $n \to \infty$.

This example with a very small n also gives insight into some important properties of the Bayesian approach. Although the prior density used here is convenient for calculations in the Bayesian approach, this example seems to require a different prior. Further, the prior density becomes irrelevant when more data become available. Also, note that $\text{var}(s^{*2}_{yp} \mid s_r) > \text{var}(s^2_{yp} \mid s_r)$. Next, we conclude this section with some general remarks:

1. Since $\overline{Q}_m = \overline{y}_p + \overline{e}_m$ and $\text{plim}(\overline{e}_m) = 0$, we have $\text{plim}(\overline{Q}_m) = \overline{y}_p$ as $m \to \infty$. That is, the efficiency of $\overline{Q}_m$ is increasing as m increases. Also recall that the t-value is decreasing (Rubin,

1987, 77) so that confidence intervals become shorter. Note that MI yields valid confidence intervals only if the posterior variance T_∞ of $(\mu - \overline{Q}_\infty)$ is in expectation equal to the frequentist variance of $\overline{Q}_\infty$ $(= \overline{y}_p)$. That is, $E\ (T_\infty) = \sigma^2/p\ (p, n \gg 1)$.

2. Several years after the introduction of MI Rubin and Schenker (1986) proposed to include in (2.5) the additional variance component B_m/m as a small sample correction; for a Bayesian justification of this component, see Rubin (1987, 89). However, from a frequentist viewpoint B_m/m just stands for an estimate of var$(\overline{e}_m)$; recall $\overline{Q}_m = \overline{y}_p + \overline{e}_m$.
3. Apart from $E(e_j) = 0$ and $E(s^2_{ynj}) = \sigma^2$ a third essential requirement for *proper* multiple imputation according to Rubin (1996) in the present notation is var$(e_j) = q\sigma^2/np$ $(p, n \gg 1)$. Denote these three requirements by R_1, R_2 and R_3, respectively. R_3 can be derived in a frequentist manner as follows. By definition, $E(T_\infty) = E(\overline{U}_\infty) + E(B_\infty)$. As we have seen in remark 1, $E(T_\infty) = \sigma^2/p$ and by R_2, $E(\widehat{U}_j) = \sigma^2/n$. Hence, $E(B_\infty) = q\sigma^2/np$. Since by construction, $E(B_\infty) =$ var(e_j), we get var$(e_j) = q\sigma^2/np$. So from a frequentist viewpoint the underlying parameter B to be estimated by B_m can be defined by $B \equiv$ var$(\overline{y}_p) -$ var$(\overline{y}_s) = q\sigma^2/np$ and R_3 amounts to $E(B_m) = B$; $\overline{y}_s$ stands for the sample mean in the complete-data case. More generally, when θ is the parameter to be estimated and $E(\widehat{\theta}_p) = E(\widehat{\theta}_s) = \theta$, B is defined similarly by $B \equiv$ var$(\widehat{\theta}_p) -$ var$(\widehat{\theta}_s)$ and R_3 becomes var$(\widehat{Q}_j \mid s_r) = B$; R_1 and R_2 become $E(\widehat{Q}_j \mid s_r) = \widehat{\theta}_p$ and $E(\widehat{U}_j) =$ var$(\widehat{\theta}_s)$, respectively.
4. B_m is an asymptotically unbiased estimator for var(e_j) as $p, n \to \infty$ $(m \gg 1)$. However, $q\overline{U}_m/p$ is also asymptotically unbiased for var(e_j) because $E(\overline{U}_m) = \sigma^2/n$. In later sections we will see that in more complicated situations similar results can be derived.
5. There are no frequentist reasons for drawing $r_{p-1,j}$ from a χ^2-distribution $(j = 1, \dots, m)$. Moreover, for small p and large m, T_m substantially overestimates in expectation var$(\overline{Q}_m)$ leading to inaccurate confidence intervals while by omitting drawings $r_{p-1,j}$, T_m becomes unbiased. That is, in MI terms, imputations with drawings $r_{p-1,j}$ are *improper* for small p while plim$(r_p/p) = 1$ as $p \to \infty$. Furthermore, these drawings don't have any evident practical advantage. Since in most cases $p \gg 1$, we omit this kind of drawing in the remainder of this paper.
6. Since $E(\overline{U}_m) = \sigma^2/n$, $B = q\sigma^2/np$, $E(T_\infty) = \sigma^2/p$ and $E(T_m - T_\infty) = B/m$, the requirement $E(T_m) < (1 + \delta)E(T_\infty)$ for a given $\delta > 0$ implies $m > q/n\delta$. So when $E(T_m)$ and its lower bound $E(T_\infty)$ should differ less than 10% and $q/n = 0.4$, then $m > 4$. This is a convenient rule of thumb for finding a suitable value for m.

2.3 MI and Unequal Probability Sampling from a Finite Population

Using the above results, we examine in this section proper MI for a probability sample s of size n with replacement from a population of size N. Let y_i be the study variable and let population mean $\overline{y}_N$ or, for short, $\overline{Y}$ be the parameter to be estimated. Each unit in s is drawn from the units of the population with probabilities $p_1, \dots, p_N$ $(p_1 + \dots + p_N = 1)$. Defining Z_i by $Z_i = y_i/Np_i$ $(i = 1, \dots, N)$ and denoting the Z-values of the n units in s by $z_1, \dots, z_n$, the Hansen-Hurwitz estimator (say $\widehat{\overline{Y}}^n_{HH}$) is $\widehat{\overline{Y}}^n_{HH} = \overline{z}_s = \sum_{i=1}^n z_i/n$. Its variance is var$(\widehat{\overline{Y}}^n_{HH}) = \sigma^2_{zN}/n$ where $\sigma^2_{zN} = \sum_{i=1}^N p_i(Z_i - \overline{Y})^2$. The variance can be estimated by s^2_{zs}/n where the (complete) sample variance s^2_{zs} is defined by $s^2_{zs} = \sum_{i=1}^n (z_i - \overline{z}_s)^2/(n-1)$. Assuming as before that p drawings are available and that a simple random subset of q drawings is missing $(p + q = n)$, it holds that $\overline{z}_p$ and s^2_{zp} are unbiased for their counterparts of the population. That is, $E(\overline{z}_p) = E\{E(\overline{z}_p \mid s)\} = E(\overline{z}_s) = \overline{Y}$ and $E(s^2_{zp}) = E\{E(s^2_{zp} \mid s)\} = E(s^2_{zs}) = \sigma^2_{zN}$. Next, consider the imputations $z_i = \overline{z}_p + v + u_i$ $(i = p + 1, \dots, n)$, where v and the u_i are independent drawings with $v \sim N(0, s^2_{zp}/p)$ and $u_i \sim N(0, s^2_{zp})$. Subsequently, the corresponding imputations for the y_i are

$$y_i = Np_i(\overline{z}_p + v + u_i) \quad (i = p+1, \dots, n). \tag{2.8}$$

In analogy with section 2.2 we can write the resulting estimator as $\widehat{\overline{Y}}_{HH}^{imp} = \overline{z}_n = \overline{z}_p + e$, where e is as defined in (2.2), and var$(e \mid s_r) = qs_{zp}^2/np$. Next, taking into account the sampling design, the variance of $\overline{z}_n$ is

$$\text{var}\left(\widehat{\overline{Y}}_{HH}^{imp}\right) = \left(\frac{1}{p} + \frac{q}{np}\right)\sigma_{zN}^2 = \left(\frac{1}{n} + \frac{2q}{np}\right)\sigma_{zN}^2. \tag{2.9}$$

As in section 2.2, it can be shown that $E(s_{zn}^2) = \sigma_{zN}^2$. An unbiased estimator for the variance can be obtained from (2.9) by replacing σ_{zN}^2 by s_{zp}^2 or s_{zn}^2. Note that in an MI procedure ($m > 1$) we have $E(B_m) = \text{var}(e) = q\sigma_{zN}^2/np = \text{var}(\overline{z}_p) - \text{var}(\overline{z}_s) = B$. For similar results for sampling without replacement, see Appendix A.1. To our best knowledge these imputation formulas for MI are not mentioned elsewhere in the literature.

3 Kalman Filtering and MI

In this section we look from a frequentist point of view at the MI method for estimating a vector of regression coefficients; also see Rubin (1987, 167). Using the Kalman filter, the derivations of the corresponding SI and MI variance formulas are relatively straightforward, including when $m = 1$. Consider the linear model (say ξ)

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad E(\varepsilon_i^2) = f_i\sigma^2 \quad \text{and} \quad E(\varepsilon_i\varepsilon_j) = 0 \ (i \neq j), \tag{3.1}$$

where $\mathbf{x}_i$ is a $k \times 1$ vector of known explanatory variables associated with the ith unit. Further, suppose that $\mathbf{x}_i$ contains f_i. For a given $l \times 1$ data vector $\mathbf{y}_l$ ($l \geq k$), the generalized least squares (GLS) estimator $\mathbf{b}_l$ of $\boldsymbol{\beta}$ is defined by $\mathbf{b}_l \equiv \mathbf{V}_l\mathbf{X}_l'\mathbf{F}_l^{-1}\mathbf{y}_l$ where $\mathbf{X}_l = (\mathbf{x}_1, \dots, \mathbf{x}_l)'$, $\mathbf{F}_l = \text{diag}(f_1, \dots, f_l)$ and $\mathbf{V}_l = \left(\mathbf{X}_l'\mathbf{F}_l^{-1}\mathbf{X}_l\right)^{-1}$; usually $\mathbf{y}_l$ consists of observations, but $\mathbf{y}_l$ may also contain imputed values. Further, define $s_{el}^2 = \mathbf{e}_l'\mathbf{F}_l^{-1}\mathbf{e}_l/(l-k)$ with $\mathbf{e}_l = \mathbf{y}_l - \mathbf{X}_l\mathbf{b}_l$. Recall from regression theory that $\overline{e}_l = 0$ (Greene, 2003, 24) when $\mathbf{x}_i$ contains f_i, and that $\mathbf{b}_l$ is the best linear unbiased (BLU) estimate of $\boldsymbol{\beta}$ in model ξ in (3.1) with var$(\mathbf{b}_l) = \sigma^2\mathbf{V}_l$ provided that $\mathbf{y}_l$ does not contain imputations; when the $\varepsilon_i/\sqrt{f_i}$ are iid $N(0, \sigma^2)$, $\mathbf{b}_l$ is the maximum likelihood (ML) estimate. Consider now an arbitrary partition $(\mathbf{y}_{p'}', \mathbf{y}_{q'}')'$ of an arbitrary $n \times 1$ vector $\mathbf{y}_n$ in $\mathbb{R}^n$ ($p' + q' = n;\ p' \geq k$). An interesting question in the present context is if there exists an explicit expression for the difference $(\mathbf{b}_n - \mathbf{b}_{p'})$. The answer is in the affirmative and is given by the GLS recursions or Kalman's (1960) update equations. Partitioning $\mathbf{X}_n$ as $\mathbf{X}_n = (\mathbf{X}_{p'}', \mathbf{X}_{q'}')'$, the GLS recursions are

$$\mathbf{b}_n = \mathbf{b}_{p'} + \mathbf{K}_{q'}\left(\mathbf{y}_{q'} - \mathbf{X}_{q'}\mathbf{b}_{p'}\right) \tag{3.2}$$

$$\mathbf{V}_n = \left(\mathbf{I}_k - \mathbf{K}_{q'}\mathbf{X}_{q'}\right)\mathbf{V}_{p'} \tag{3.3}$$

$$\mathbf{K}_{q'} \equiv \mathbf{V}_{p'}\mathbf{X}_{q'}'\left(\mathbf{X}_{q'}\mathbf{V}_{p'}\mathbf{X}_{q'}' + \mathbf{F}_{q'}\right)^{-1}, \tag{3.4}$$

where $\mathbf{I}_k$ is the identity matrix of order k. For the derivations of (3.2) and (3.3) based on Kalman filtering, see Appendix A.2. For other derivations with $f_i = 1$, see Plackett (1950) and Harvey (1990, 53) who made the additional assumption $q' = 1$. In fact, they used different versions of the somewhat laborious but useful matrix inversion lemma

$$(\mathbf{A} + \mathbf{BCB}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}'\mathbf{A}^{-1}. \tag{3.5}$$

For a simplification of (3.4), see section 4. From (3.2) it is seen that $(\mathbf{b}_n - \mathbf{b}_{p'})$ is a linear function of $\mathbf{e}_{q'} \equiv \mathbf{y}_{q'} - \mathbf{X}_{q'}\mathbf{b}_{p'}$, often called *'prediction error'* in Kalman filtering.

Next, suppose as before that p observations are available and that q observations are missing $(p + q = n)$. Following Rubin (1987, 162-167), we make the assumption of ignorable nonresponse, sometimes also referred to as missing at random (MAR) or ignorably missing. That is, the distribution function $F(y_i \mid \mathbf{x}_i)$ does not depend on the response mechanism for a given s. Assuming that p is sufficiently large, the MI imputations under model ξ are in Rubin's (1987) notation

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta}^* + u_i \quad (i = p + 1, \dots, n), \tag{3.6a}$$

where $\boldsymbol{\beta}^* \sim N(\widehat{\boldsymbol{\beta}}, s_{ep}^2\mathbf{V}_p)$, $u_i \sim N(0, f_i s_{ep}^2)$, and $\widehat{\boldsymbol{\beta}} = \mathbf{b}_p$. Now defining $\mathbf{v} = \boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}$ and replacing $\boldsymbol{\beta}^*$ in (3.6a) by $\widehat{\boldsymbol{\beta}} + \mathbf{v} = \mathbf{b}_p + \mathbf{v}$, (3.6a) can be rewritten as

$$y_i^* = \mathbf{x}_i'(\mathbf{b}_p + \mathbf{v}) + u_i, \tag{3.6b}$$

where, conditional on s_r, $\mathbf{v} \sim N(0, s_{ep}^2\mathbf{V}_p)$. An important advantage of (3.6b) is that it gives an explicit expression for the total imputation noise in y_i^*, namely $\mathbf{x}_i'\mathbf{v} + u_i$. Further, note that (3.6b) reduces to (2.1) when $\mathbf{x}_i = f_i = 1$ for all i. In analogy with section 2.1, the u_i in (3.6) can be drawn from $\{e_1\sqrt{f_i/f_1}, \dots, e_p\sqrt{f_i/f_p}\}$ when required. For expository purposes, we write the q imputations in (3.6b) in obvious matrix notation as

$$\mathbf{y}_q = \mathbf{X}_q\mathbf{b}_p + \mathbf{X}_q\mathbf{v} + \mathbf{u}_q, \tag{3.7}$$

where we skipped the asterisk because subscript q $(q > 0)$ already indicates that $\mathbf{y}_q$ is a vector of q imputed values; also see section 2. So defining $\mathbf{y}_n$ by $\mathbf{y}_n = (\mathbf{y}_p', \mathbf{y}_q')'$ where $\mathbf{y}_q$ is given in (3.7), the prediction error for $\mathbf{y}_q$ given $\mathbf{y}_p$ is $\mathbf{e}_q = \mathbf{X}_q\mathbf{v} + \mathbf{u}_q$. Subsequently, under the assumptions of model (3.1), we can prove the following theorem.

Theorem 3.1. Let $\mathbf{y}_n$ be defined by $\mathbf{y}_n = (\mathbf{y}_p', \mathbf{y}_q')'$ where $\mathbf{y}_q$ is given in (3.7). Define the GLS estimates $\mathbf{b}_n = \mathbf{V}_n\mathbf{X}_n'\mathbf{F}_n^{-1}\mathbf{y}_n$, $\mathbf{b}_p = \mathbf{V}_p\mathbf{X}_p'\mathbf{F}_p^{-1}\mathbf{y}_p$, $\mathbf{e}_n = \mathbf{y}_n - \mathbf{X}_n\mathbf{b}_n$, $\mathbf{e}_q = \mathbf{y}_q - \mathbf{X}_q\mathbf{b}_p$ and $s_{en}^2 = \mathbf{e}_n'\mathbf{F}_n^{-1}\mathbf{e}_n/(n-k)$. Let this estimation step be repeated m times, each time with new imputations $\mathbf{y}_{qj}$ in (3.7) $(j = 1, \dots, m)$. Let $\overline{\mathbf{b}}_{nm}$, $\overline{\mathbf{e}}_{qm}$ and $\overline{\mathbf{u}}_{qm}$ denote the averages of the $\mathbf{b}_{nj}$, $\mathbf{e}_{qj}$ and $\mathbf{u}_{qj}$. Consider MI estimator $\overline{\mathbf{b}}_{nm}$. Let $\mathbf{T}_m$, $\mathbf{B}_m$ and $\overline{\mathbf{U}}_m$ be as defined in section 2.2 but now in multivariate form. Then,

(*i*) requirements R_1, R_2 and R_3 are met and

$$\mathbf{b}_n = \mathbf{b}_p + \mathbf{K}_q\mathbf{e}_q \quad (\mathbf{e}_q = \mathbf{X}_q\mathbf{v} + \mathbf{u}_q) \tag{3.8}$$

$$\mathbf{K}_q \equiv \mathbf{V}_p\mathbf{X}_q'\left(\mathbf{X}_q\mathbf{V}_p\mathbf{X}_q' + \mathbf{F}_q\right)^{-1} \tag{3.9}$$

$$E(\mathbf{T}_m) = \sigma^2\{\mathbf{V}_p + (\mathbf{V}_p - \mathbf{V}_n)/m\} = \text{var}(\overline{\mathbf{b}}_{nm}); \tag{3.10}$$

(*ii*) assuming $\mathbf{v} = \mathbf{0}$, $n\mathbf{V}_n = O(1)$ and $\mathbf{X}_q'\mathbf{F}_q^{-1}\mathbf{X}_q/q = O(1)$ as $n \to \infty$, it holds for such a zero $\mathbf{v}$ (ZV) and the corresponding $\overline{\mathbf{b}}_{nm}$ (say $\overline{\mathbf{b}}_{nmZV}$) that

$$\overline{\mathbf{b}}_{nmZV} = \mathbf{b}_p + \mathbf{K}_q\overline{\mathbf{u}}_{qm} \tag{3.11a}$$

$$E\{s_{en}^2(\mathbf{V}_p + \mathbf{K}_q\mathbf{F}_q\mathbf{K}_q'/m)\} = \text{var}(\overline{\mathbf{b}}_{nmZV})\{1 + O(1/n)\}. \tag{3.11b}$$

Proof. Equation (3.8) follows from (3.2) and (3.7). By (3.8), $E(\mathbf{b}_n) = E(\mathbf{b}_p) = \boldsymbol{\beta}$ so that R_1 is met. Since $E(s_{en}^2) = \sigma^2$, proved in Appendix A.3, R_2 is met and $E(\overline{\mathbf{U}}_m) = \sigma^2\mathbf{V}_n$. Repeating the

GLS regression m times, it follows from (3.8) that

$$E(\hat{\mathbf{B}}_m) = E\left\{\sum_{j=1}^{m}(\mathbf{b}_{nj} - \bar{\mathbf{b}}_{nm})(\mathbf{b}_{nj} - \bar{\mathbf{b}}_{nm})'\right\}/(m-1)$$

$$= \mathrm{var}(\mathbf{K}_q\mathbf{e}_q) = \sigma^2\mathbf{K}_q(\mathbf{X}_q\mathbf{V}_p\mathbf{X}_q' + \mathbf{F}_q)\mathbf{K}_q'$$

$$= \sigma^2\mathbf{K}_q\mathbf{X}_q\mathbf{V}_p = \sigma^2(\mathbf{V}_p - \mathbf{V}_n) \equiv \mathbf{B}, \tag{3.12}$$

where we used the (implicit) definition of $\mathbf{K}_q'$ in (3.9). In (3.12) we used (3.3). Hence, R_3 is met. Since $\bar{\mathbf{b}}_{nm} = \mathbf{b}_p + \mathbf{K}_q\bar{\mathbf{e}}_{qm}$, (3.10) follows from (3.12) and R_2. When $\mathbf{v} = \mathbf{0}$, (3.11) follows directly from (3.8) and $E(s_{en}^2) = \sigma^2\{1 + O(1/n)\}$, proved in Appendix A.3. □

In addition, **B** can be estimated by $s_{en}^2(\mathbf{V}_p - \mathbf{V}_n)$, without MI just as we saw in section 2. Expressions for $E(\mathbf{B}_m)$ are also given in Schenker and Welsh (1988), Wang and Robins (1998) and Kim (2004) for special cases but not for the general linear model (3.1) with unspecified ε_i. Further, their derivations do not use the convenient GLS recursion (3.8) and therefore somewhat more algebra is required. Since we also use Kalman filtering in sections 4 and 6 for cases with categorical variables, $\mathbf{v} = \mathbf{0}$, $E(\varepsilon_i^2) = f_i\sigma^2$ $(f_i > 0)$ and sampling without replacement, this paper can be seen as a further extension of their work.

4 Consequences of Ignoring Parameter Uncertainty and a New MI Method

In this section we examine in more detail the consequences of omitting the random v in the imputations discussed so far. As a point of departure we consider again drawings from a normal distribution as described in section 2. The imputations without v are

$$y_i = \bar{y}_p + u_i \quad (i = p+1, \dots, n), \tag{4.1}$$

where the u_i are mutually independent drawings from $N\left(0, s_{yp}^2\right)$. Imputations with a zero v (ZV) lead to the following estimator $\bar{y}_{nZV} = \bar{y}_p + e$ with $e = q\bar{u}_q/n$. Its variance is

$$\mathrm{var}\left(\bar{y}_{nZV}\right) = \left(\frac{1}{p} + \frac{q}{n^2}\right)\sigma^2. \tag{4.2}$$

In addition, we now get for the sample variance based on the imputations in (4.1)

$$E\left\{(n-1)s_{yn}^2\right\} = E\left\{(p-1)s_{yp}^2 + \sum_{i=p+1}^{n}\left(y_i - \bar{y}_p\right)^2 - n\left(\bar{y}_{nZV} - \bar{y}_p\right)^2\right\}$$

$$= (p-1)\sigma^2 + q\sigma^2 - q\sigma^2/n = (n-1-q/n)\sigma^2$$

and hence,

$$E(s_{yn}^2) = \left\{1 - \frac{q}{n(n-1)}\right\}\sigma^2 = \sigma^2\{1 + O(1/n)\}. \tag{4.3}$$

It follows from (4.3) that omitting v leads to a negligible decrease of $E(s_{yn}^2)$. In contrast, a maybe somewhat counterintuitive result now is that the variance of e $(= q\bar{u}_q/n)$ is decreasing dramatically from $q\sigma^2/np$ (see section 2) to $q\sigma^2/n^2$, due to omitting v. That is, var(e) is decreasing by $(q/n) \times 100\%$ and by construction, this also holds for $E(B_m)$. For example, when the nonresponse is 40%, the decrease of var(e) and $E(B_m)$ is 40% as well. This means an important increase of the efficiency of the single ZV method compared to a single MI step. Moreover, for MI the random v is indispensable. Without the random v MI fails because the third requirement for *proper* imputation var$(e) = B \equiv q\sigma^2/np$ might be severely violated especially for high nonresponse rates. Therefore, Rubin (1987, 122-123) stresses the role of the uncertainty of the estimated parameters in his MI procedure.

A related estimator (say $\bar{y}_{nSR}$), discussed in Rubin and Schenker (1986), is defined by $\bar{y}_{nSR} = (p\bar{y}_p + q\bar{y}_q)/n$, where $\bar{y}_q$ is the sample mean of a simple random sample of size q *with* replacement (SR) from $s_r = \{y_1, \dots, y_p\}$. In fact, it is a kind of hotdeck (HD) imputation, often used for the imputation of binary data. Its variance is

$$\begin{aligned}\text{var}(\bar{y}_{nSR}) &= \text{var}\, E(\bar{y}_{nSR} \mid s_r) + E\,\text{var}(\bar{y}_{nSR} \mid s_r)\\ &= \text{var}(\bar{y}_p) + \frac{q^2}{n^2}E\frac{(p-1)s_{yp}^2}{pq} = \sigma^2\left(\frac{1}{p} + \frac{q\{1+O(p^{-1})\}}{n^2}\right) \qquad (4.4)\end{aligned}$$

as $p \to \infty$. The variance in (4.4) is asymptotically equal to var $(\bar{y}_{nZV})$ in (4.2) and $E\left(s_{yn}^2\right) = \sigma^2\{1 + o(1)\}$.

In analogy with (4.2), similar formulas can be derived in the case of estimating a regression vector $\boldsymbol{\beta}$ when $\mathbf{v} = \mathbf{0}$ as we have seen in section 3 under model (3.1). By (3.11) in Theorem 3.1 with $m = 1$, we get

$$\mathbf{b}_{nZV} = \mathbf{b}_p + \mathbf{K}_q\mathbf{u}_q \quad \text{and} \quad \text{var}\,(\mathbf{b}_{nZV}) = \sigma^2(\mathbf{V}_p + \mathbf{K}_q\mathbf{F}_q\mathbf{K}_q') \qquad (4.5)$$

with

$$\begin{aligned}\mathbf{K}_q &= \mathbf{V}_p\left(\mathbf{I}_k - \mathbf{V}_q^{-1}\mathbf{V}_n\right)\mathbf{X}_q'\mathbf{F}_q^{-1} \qquad (4.6a)\\ &= \mathbf{V}_n\mathbf{X}_q'\mathbf{F}_q^{-1}, \qquad (4.6b)\end{aligned}$$

where (4.6a) follows from applying (3.5) to (3.9); recall that $\mathbf{V}_q = (\mathbf{X}_q'\mathbf{F}_q^{-1}\mathbf{X}_q)^{-1}$ and

$$\mathbf{V}_n = (\mathbf{X}_n'\mathbf{F}_n^{-1}\mathbf{X}_n)^{-1} = (\mathbf{X}_p'\mathbf{F}_p^{-1}\mathbf{X}_p + \mathbf{X}_q'\mathbf{F}_q^{-1}\mathbf{X}_q)^{-1} = (\mathbf{V}_p^{-1} + \mathbf{V}_q^{-1})^{-1}. \qquad (4.7)$$

Replacing $\mathbf{V}_q^{-1}$ in (4.6a) by $\mathbf{V}_n^{-1} - \mathbf{V}_p^{-1}$ yields (4.6b). Applying imputations with $\mathbf{v} = \mathbf{0}$ m times yields the following ZVMI versions $(m \geq 1)$ of the variances in (4.2) and (4.5), say

$$\text{var}\left(\bar{y}_{nmZV}\right) = \left(\frac{1}{p} + \frac{q}{mn^2}\right)\sigma^2 \qquad (4.8)$$

$$\begin{aligned}\text{var}\left(\bar{\mathbf{b}}_{nmZV}\right) &= \sigma^2(\mathbf{V}_p + \mathbf{K}_q\mathbf{F}_q\mathbf{K}_q'/m)\\ &= \sigma^2(\mathbf{V}_p + \mathbf{V}_n\mathbf{V}_q^{-1}\mathbf{V}_n/m). \qquad (4.9)\end{aligned}$$

In (4.9) we used (4.6b). Note that (4.9) is of the same form as (4.8) and likewise by (3.12),

$$E(\mathbf{B}_m) = \sigma^2(\mathbf{V}_p - \mathbf{V}_n) = \sigma^2\mathbf{V}_n\mathbf{V}_q^{-1}\mathbf{V}_p = \sigma^2\mathbf{V}_p\mathbf{V}_q^{-1}\mathbf{V}_n \qquad (4.10)$$

is of the same form as (2.6); recall from (4.7) $\mathbf{V}_q^{-1} = \mathbf{V}_n^{-1} - \mathbf{V}_p^{-1}$. By Theorem 3.1, σ^2 in (4.8) can be estimated by s_{yn}^2 and in (4.9) by s_{en}^2. Since $\mathbf{v} = \mathbf{0}$, the variances in (4.8) and (4.9) are smaller than Rubin's MI variances $(m > 1)$. Moreover, one may choose as t-value $t = 1.96$ for a 95% confidence interval provided $p, n \gg 1$. This yields much more accurate confidence intervals than the MI method with its higher t-values for small values of m.

To the best of our knowledge formula (4.5) is not mentioned elsewhere in the literature for a model as (3.1). For some related ML results, see Wang and Robins (1998). However, their ML results are of little use here because the ε_i in model (3.1) are unspecified. Besides, ML ignores finite population corrections and the authors assume that each unit is *observed* with equal probability π. Hence, their results cannot be applied to the widely used general regression estimator; see the examples in sections 6 and 7. The following corollary states that ZVMI has an improved efficiency compared to MI.

Corollary 4.1. Let $\overline{\mathbf{b}}_{nm}$ and $\overline{\mathbf{b}}_{nmZV}$ be as defined in Theorem 3.1. Then under the assumptions of model (3.1), var$(\overline{\mathbf{b}}_{nm})$ exceeds var$(\overline{\mathbf{b}}_{nmZV})$ by a positive definite matrix. That is,

$$\text{var}\left(\overline{\mathbf{b}}_{nm}\right) - \text{var}\left(\overline{\mathbf{b}}_{nmZV}\right) = \sigma^2 \mathbf{V}_n \mathbf{V}_q^{-1} \mathbf{V}_p \mathbf{V}_q^{-1} \mathbf{V}_n / m. \tag{4.11}$$

Proof. The proof follows directly from (3.10), (4.9) and (4.10). □

Remark 4.1. It emerges that there are no frequentist reasons for taking **v** into account. So we can set $\mathbf{v} = \mathbf{0}$. Only requirements R_1 and R_2 are relevant for SI and ZV. Further, it is noteworthy that $\mathbf{u}_q$ $(= \mathbf{e}_q)$ can be adjusted so that the adjusted ZV (AZV) method is as efficient as MI with $m = \infty$. Suppose $\mathbf{u}_q = \mathbf{F}_q^{.5} \boldsymbol{\omega}_q$ with $\boldsymbol{\omega}_q \sim N(\mathbf{0}, s_{ep}^2 \mathbf{I}_q)$. The adjustment follows from $\min_{\boldsymbol{\omega}} f(\boldsymbol{\omega})$ subject to $\mathbf{A}\boldsymbol{\omega} = \mathbf{0}$, with $f(\boldsymbol{\omega}) = (\boldsymbol{\omega} - \boldsymbol{\omega}_q)'(\boldsymbol{\omega} - \boldsymbol{\omega}_q)$ and $\mathbf{A} = \mathbf{K}_q \mathbf{F}_q^{.5}$; $\mathbf{K}_q$ is given in (4.6). The solution (say $\boldsymbol{\omega}^*$) of this optimization problem is $\boldsymbol{\omega}^* = \mathbf{M}\boldsymbol{\omega}_q$ with $\mathbf{M} = \mathbf{I}_q - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}$. The proof is as follows. Using the Lagrangian saddle plane $L = f(\boldsymbol{\omega})/2 + \boldsymbol{\lambda}'\mathbf{A}\boldsymbol{\omega}$, we get $\partial L/\partial \boldsymbol{\omega} = \boldsymbol{\omega} - \boldsymbol{\omega}_q + \mathbf{A}'\boldsymbol{\lambda} = \mathbf{0}$ or $\boldsymbol{\omega} = \boldsymbol{\omega}_q - \mathbf{A}'\boldsymbol{\lambda}$. Since $\mathbf{A}\boldsymbol{\omega} = \mathbf{0}$, this yields $\boldsymbol{\lambda} = (\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}\boldsymbol{\omega}_q$ and thus also $\boldsymbol{\omega}^* = \mathbf{M}\boldsymbol{\omega}_q$; note $\mathbf{K}_q \mathbf{u}_q^* = \mathbf{K}_q \mathbf{F}_q^{.5} \boldsymbol{\omega}^* = \mathbf{A}\mathbf{M}\boldsymbol{\omega}_q = \mathbf{0}$ because $\mathbf{A}\mathbf{M} = \mathbf{0}$. Hence, by (3.8) $\mathbf{b}_{nAZV} = \mathbf{b}_p$ $(= \overline{\mathbf{Q}}_\infty)$. Since $\mathbf{M}$ is a symmetric and idempotent $q \times q$ matrix ($\mathbf{M}\mathbf{M} = \mathbf{M}$) with rank $q - k$, we get $E\left(\mathbf{u}_q^{*\prime} \mathbf{F}_q^{-1} \mathbf{u}_q^* \mid s_{ep}^2\right) = E\left(\boldsymbol{\omega}^{*\prime} \boldsymbol{\omega}^* \mid s_{ep}^2\right) = (q - k) s_{ep}^2$ (Greene, 2003, 49). Hence, under model (3.1), $E_\xi(s_{en}^2) \approx \sigma^2$.

5 Comparison of 95% Confidence Intervals after SI, MI and ZVMI

In order to get a first indication of the performances of the various SI and MI methods, we consider iid observations from a normal distribution $N(\mu, \sigma^2)$ with response rates between 40% and 90%; in sections 6 and 7 we examine more general models. Note that the results in this section also hold for large samples with non-normal data as stated by Rubin and Schenker (1986). It is also demonstrated that general statements in the literature such as that SI systematically underestimates the variance [e.g., Enders (2010, 55) and Rubin (2003)] are incorrect and misleading; for a similar discussion, see Särndal (1992). Assuming $n, p \gg 1$, the drawings r_{p-1} from a χ^2-distribution can be omitted because plim$(r_p/p) = 1$ as $p \to \infty$. For different situations, we calculate the variance of $\overline{y}_n$ $(\overline{Q}_m)$, including the corresponding t-value. In the case of MI, the degrees of freedom (say κ) of Student's t-distribution can be approximated by using Rubin's (1987, 77) formula

$$\kappa = \left\{1 + \left(\frac{m}{m+1}\right)\frac{\overline{U}_m}{B_m}\right\}^2 (m-1) = \left(\frac{T_m}{T_m - \overline{U}_m}\right)^2 (m-1), \tag{5.1}$$

where $\overline{U}_m$ and B_m are replaced by their expectations σ^2/n and $q\sigma^2/np$, respectively, that is $\overline{U}_m/B_m = p/q$. Without loss of generality, we set $\sigma^2/n = 1$. Further, for SI and ZVMI we may choose $t_p = 1.96$ for sufficiently large p. Since 95% confidence intervals for μ are of the form $\widehat{\mu} \pm t\sqrt{\widehat{\text{var}}(\widehat{\mu})}$, we give in Table 5.1 an overview of $t^2\text{var}(\widehat{\mu})$ for various response rates r and some SI and MI methods as an indication for the lengths of the intervals.

Table 5.1 Values of $t^2\text{var}(\widehat{\mu})$ for different MI and SI methods and various response rates.*

	method	$r = 90\%$	$r = 80\%$	$r = 70\%$	$r = 60\%$	$r = 50\%$	$r = 40\%$
1	MI, $m = 1$	4.7 (1.96)	5.8 (1.96)	7.1 (1.96)	9.0 (1.96)	11.5 (1.96)	15.4 (1.96)
2	ZV, $m = 1$	4.7 (1.96)	5.6 (1.96)	6.6 (1.96)	7.9 (1.96)	9.6 (1.96)	11.9 (1.96)
3	SR/HD	4.7 (1.96)	5.6 (1.96)	6.6 (1.96)	7.9 (1.96)	9.6 (1.96)	11.9 (1.96)
4	MI, $m = 2$	4.7 (2.01)	6.4 (2.15)	9.5 (2.40)	15.4 (2.78)	29.4 (3.43)	57.5 (4.21)
5	ZVMI, $m = 2$	4.5 (1.96)	5.2 (1.96)	6.1 (1.96)	7.2 (1.96)	8.6 (1.96)	10.8 (1.96)
6	MI, $m = 3$	4.5 (1.98)	5.5 (2.04)	7.1 (2.13)	9.7 (2.26)	13.9 (2.44)	21.4 (2.67)
7	ZVMI, $m = 3$	4.4 (1.96)	5.1 (1.96)	5.9 (1.96)	6.9 (1.96)	8.3 (1.96)	10.4 (1.96)
8	MI, $m = 5$	4.4 (1.96)	5.2 (1.99)	6.3 (2.03)	7.8 (2.08)	10.2 (2.15)	14.0 (2.24)
9	ZVMI, $m = 5$	4.3 (1.96)	5.0 (1.96)	5.7 (1.96)	6.7 (1.96)	8.1 (1.96)	10.1 (1.96)
10	MI, $m = 10$	4.3 (1.96)	4.9 (1.96)	5.8 (1.99)	7.0 (2.01)	8.7 (2.04)	11.3 (2.07)
11	ZVMI, $m = 10$	4.3 (1.96)	4.9 (1.96)	5.6 (1.96)	6.6 (1.96)	7.9 (1.96)	9.8 (1.96)
12	MI, $m = \infty$	4.3 (1.96)	4.8 (1.96)	5.5 (1.96)	6.4 (1.96)	7.7 (1.96)	9.6 (1.96)

* Between parentheses the used t-values are mentioned

Rows 1-3 deal with SI. That is, row 1 gives the results for MI with $m = 1$ as described in section 2; see (2.3a). Row 2 deals with single ZV imputations ignoring the uncertainty of $\overline{y}_p$, that is $v = 0$; see (4.2). In row 3 we have included the simple random (SR) imputation method, which can be seen as a form of hotdeck (HD); see section 4 and Rubin and Schenker (1986). Since p is large, its variance can be approximated by formula (4.2) for ZV imputations with $v = 0$ so that the difference between rows 2 and 3 vanishes. Rows 4, 6, 8, 10 and 12 in Table 5.1 deal with MI as discussed in section 2; see (2.7). When κ resulting from (5.1) was not an integer, we used a simple linear interpolation for estimating the t-value (say t_κ). In rows 5, 7, 9 and 11 we applied the ZVMI method for $m \in \{2,\ 3,\ 5,\ 10\}$; see (4.8). Although today's advice for m might be much higher, in their seminal paper Rubin and Schenker (1986) paid much attention to the case with $m = 2$ in their comparison with SI where data were generated from a normal distribution. Moreover, they recommended $m = 2$ in a number of cases while Rubin (1996) recommended $m = 3$ or $m = 5$. Therefore, we also consider such low values for m.

For $m = \infty$, MI yields an estimator which is as efficient as $\overline{y}_p$; recall that the variance of MI is decreasing as m increases. So row 12 can be seen as a row of lower bounds. Recall from (2.7) and (4.8) that the incremental variances due to imputation are $q\sigma^2/npm$ and $q\sigma^2/n^2m$ for MI and ZVMI, respectively and hence the ratio of the increments is n/p irrespective of the value of m. So in the case of 50% nonresponse, the additional imputation variance of MI is twice that of ZVMI. Further, it is striking that the MI method with $m = 2$ may lead to larger confidence intervals than each of the three SI methods, especially for low response rates leading to high t-values for MI. These results seem to deviate from other literature on MI and SI. For instance, Little (2011, 167) states (in our notation): "The quantity $(1 + 1/m)B_m$ in (13) estimates the contribution to the variance from imputation uncertainty, missed (i.e., set to zero) by single imputation methods". In addition, comparing MI and SR, Rubin and Schenker (1986, 366 and 370) state: "Using $m = 2$ imputations per missing value gives accurate coverages of the resulting intervals in common cases and is clearly superior to single imputation ($m = 1$) in all cases" and "multiple imputation methods that reflect uncertainty due to parameter estimation are clearly superior to SR imputation, although the effect becomes quite modest for high response rates". In order to understand these differences, we have a closer look at Table 1 of Rubin and Schenker (1986). For instance, consider the entry in their Table 1 (95%) with $m = 1$ and response rate $r = 0.5$ which shows a coverage rate of only 78% for SR. From a frequentist viewpoint, their analysis runs here as follows. Because the between-imputation variance B_1 cannot be calculated for $m = 1$, the authors set $B_1 = 0$ so that $T_1 = s_{yn}^2/n$ which underestimates the actual var $\left(\overline{y}_{nSR}\right)$ by 60% as can be seen from the above formula (4.4) with $p = q = n/2$ ($p \gg 1$). Hence, the resulting 95% confidence interval is 36.8% too short leading to a coverage rate of only 78.5%; recall $P(|Z| > 0.632t_\infty) = 0.2152$ with $Z \sim N(0, 1)$ and $t_\infty = 1.96$ so that the coverage rate becomes only 78% as found by the authors. Similar results can be obtained for other response rates. That is, for $r = i/10$ ($i = 4, 6, 7, 8, 9$) the coverage rates derived in this way for SR are 73.4%, 82.7%, 86.4%, 89.6% and 92.5%, respectively, which are also fully in line with the coverages (in two digits) in their Table 1. Hence, the only point the authors demonstrate here is that the combination of MI and SR does not work because using SR in each of the m steps of the MI procedure leads to underestimation of parameter $B = q\sigma^2/np$ by $100g$% with $g = q/n$ ($m > 1$) as we have seen in section 4, and even by 100% for $m = 1$. That is, in MI terms, SR and HD imputations are indeed not *proper* for MI purposes because R_3 is not met; also see Rubin (1987, 122). But if the authors had used the correct variance formula (4.4), the coverage rates for SR would have been 95%. So from a statistical viewpoint their simulation results are certainly not a correct justification for the above incorrect statements. A correct statement in this context would be that in the case of imputed values the use of the customary software for n observations leads to an underestimation of the variance.

The above Table 5.1 shows that also compared to $m = 3$ the single SR, HD and ZV imputations may lead to more accurate confidence intervals than MI provided correct variance formulas are used unless the response rate is much more than 70%. Then MI might be slightly more accurate. Furthermore, it emerges that ZVMI leads to much shorter intervals than the corresponding MI methods for $m \in \{2,\ 3,\ 5\}$. For instance, the difference between row 7 of ZVMI ($m = 3$) and row 12 (of lower bounds) in Table 5.1 is less than 10% for any r whereas the difference between row 6 of MI ($m = 3$) and row 12 is increasing to 123% as r decreases to 40%. So ZVMI is much more stable than MI. This is due to smaller variances and lower t-values. Of course, the differences are vanishing as m increases.

6 On the Regression Estimator after MI, ZVMI and SI

6.1 Introduction

In this section we examine MI, ZVMI and SI for the widely used regression estimator of a population mean $\overline{Y}$. Consider a simple random sample s without replacement (SRS) of size n from a population of size N. Now first assuming $q = 0$, the familiar regression estimator (say $\widehat{\overline{Y}}_{RGn}$) of $\overline{Y}$ and its design based variance approximation are

$$\widehat{\overline{Y}}_{RGn} = \overline{y}_n + \mathbf{b}_n' \left(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_n\right) \tag{6.1}$$

$$\text{var}\left(\widehat{\overline{Y}}_{RGn}\right) \approx \left(\frac{1}{n} - \frac{1}{N}\right) s_{eN}^2, \tag{6.2}$$

where, quite generally, $s_{el}^2 = \mathbf{e}_l'\mathbf{e}_l/(l-k)$, $\mathbf{e}_l = \mathbf{y}_l - \mathbf{X}_l\mathbf{b}_l$ and $\mathbf{b}_l = (\mathbf{X}_l'\mathbf{X}_l)^{-1}\mathbf{X}_l'\mathbf{y}_l$ for $l \geq k$; $\mathbf{b}_l$ is the ordinary least squares (OLS) estimate from a regression of $\mathbf{y}_l$ on $\mathbf{X}_l$. It is assumed that $\bar{\mathbf{x}}_N$ and the $\mathbf{x}_i$ are available from registers. The variance in (6.2) can be estimated by replacing s_{eN}^2 by s_{en}^2 [e.g., Cochran (1977, 195) and Särndal et al. (1992, 279)]. Recall that in variance approximation (6.2) the random character of $\mathbf{b}_n$ can be ignored for large n. In the next three subsections we examine SI, MI and ZVMI for the regression estimator based on model (3.1) with $f_i = 1$, $f_i \neq 1$, and unequal π_i, respectively, with special attention for the finite population correction and parameter uncertainty. In subsection 6.5 a simulation study is conducted to investigate several variance formulas.

6.2 SI, MI and ZVMI for the Regression Estimator when $f_i = 1$

Assume that there are q ignorably missing observations and that the data obey model ξ in (3.1) with $f_i = 1$ unless stated otherwise. Recall that the regression results based on the first p (available) observations can be written as

$$y_i = \bar{y}_p + \mathbf{b}_p'(\mathbf{x}_i - \bar{\mathbf{x}}_p) + e_i \quad (1 \leq i \leq p), \tag{6.3}$$

where the e_i are the residuals from the underlying OLS regression. Now consider the imputations y_i from (3.6b)

$$y_i = \mathbf{x}_i'(\mathbf{b}_p + \mathbf{v}) + u_i \quad (i = p+1, \dots, n); \tag{6.4}$$

note that for notational convenience, we dropped in (6.4) the asterisk of y_i^* because in the present context with $q > 0$ it is clear that each y_i is an imputed value if $p < i \leq n$. Now the actual regression estimator (say $\hat{Y}_{RGn}^{imp}$) can be written as

$$\hat{Y}_{RGn}^{imp} = \bar{y}_n + \mathbf{b}_n'(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_n), \tag{6.5}$$

where as in sections 2 and 3, $\bar{y}_n$ and $\mathbf{b}_n$ are based on imputations. Defining h_i by $h_i = e_i$ if $1 \leq i \leq p$ and $h_i = \mathbf{v}'\mathbf{x}_i + u_i$ if $p < i \leq n$, we can write y_i given in (6.3) and (6.4) as

$$y_i = \bar{y}_p + \mathbf{b}_p'(\mathbf{x}_i - \bar{\mathbf{x}}_p) + h_i \quad (i = 1, \dots, n).$$

Recall that $\bar{e}_p = \bar{y}_p - \mathbf{b}_p'\bar{\mathbf{x}}_p = 0$ and $\bar{h}_p = \bar{e}_p = 0$ so that $\bar{h}_n = q(\mathbf{v}'\bar{\mathbf{x}}_q + \bar{u}_q)/n$. Subsequently, we get $\bar{y}_n = \bar{y}_p + \mathbf{b}_p'(\bar{\mathbf{x}}_n - \bar{\mathbf{x}}_p) + \bar{h}_n$. Substituting this into (6.5) yields

$$\begin{aligned}\hat{Y}_{RGn}^{imp} &= \bar{y}_p + \mathbf{b}_p'(\bar{\mathbf{x}}_n - \bar{\mathbf{x}}_p) + \bar{h}_n + \mathbf{b}_n'(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_n) \\ &= \bar{y}_p + \mathbf{b}_p'(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_p) + \bar{h}_n + (\mathbf{b}_n - \mathbf{b}_p)'(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_n) \\ &= \hat{\hat{Y}}_{RGp} + \bar{h}_n + O_p(1/n).\end{aligned} \tag{6.6}$$

Repeating this m times, it can be seen from (6.6) that $E(B_m \mid s_r) \approx \text{var}(\bar{h}_n) = q^2\text{var}(\mathbf{v}'\bar{\mathbf{x}}_q + \bar{u}_q)/n^2$. Now suppose $\mathbf{x}_i = (1, \mathbf{z}_i')'$ and $n, p \gg 1$ so that $(p-1)/p \approx 1$. Further, let $\mathbf{0}_k$ denote a vector of k zeroes. Then defining $\bar{\mathbf{z}}_{2p} = \sum_{i=1}^{p} \mathbf{z}_i\mathbf{z}_i'/p$, $\mathbf{s}_{zzp} = \bar{\mathbf{z}}_{2p} - \bar{\mathbf{z}}_p\bar{\mathbf{z}}_p'$, and using the inversion lemma for a block matrix, it can be seen and verified that for $f_i = 1$,

$$\mathbf{V}_p = \frac{1}{p}\begin{pmatrix} 1 & \bar{\mathbf{z}}_p' \\ \bar{\mathbf{z}}_p & \bar{\mathbf{z}}_{2p} \end{pmatrix}^{-1} = \frac{1}{p}\begin{pmatrix} 1 + \bar{\mathbf{z}}_p'\mathbf{s}_{zzp}^{-1}\bar{\mathbf{z}}_p & -\bar{\mathbf{z}}_p'\mathbf{s}_{zzp}^{-1} \\ -\mathbf{s}_{zzp}^{-1}\bar{\mathbf{z}}_p & \mathbf{s}_{zzp}^{-1} \end{pmatrix}; \tag{6.7}$$

note that e.g. $[\mathbf{X}_p'\mathbf{X}_p\mathbf{V}_p]_{12} = (p, p\bar{\mathbf{z}}_p')(-\mathbf{s}_{zzp}^{-1}\bar{\mathbf{z}}_p/p, \mathbf{s}_{zzp}^{-1}/p)' = \mathbf{0}_{k-1}'$. Using (6.7), we get

$$\begin{aligned} E(B_m \mid s_r) \approx \text{var}(\bar{h}_n) &= q^2\left\{\bar{\mathbf{x}}_q'\mathbf{V}_p\bar{\mathbf{x}}_q s_{ep}^2 + E(\bar{u}_q^2 \mid s_r)\right\}/n^2 \\ &= q^2 s_{ep}^2\left\{(1 + A_{pq})/p + 1/q\right\}/n^2 && (6.8) \\ &= q s_{ep}^2(1 + qA_{pq}/n)/np. && (6.9) \\ A_{pq} &= (\bar{\mathbf{z}}_p - \bar{\mathbf{z}}_q)'\mathbf{s}_{zzp}^{-1}(\bar{\mathbf{z}}_p - \bar{\mathbf{z}}_q). && (6.10) \end{aligned}$$

More generally, A_{pq} can be written as $A_{pq} = (\bar{\mathbf{x}}_p - \bar{\mathbf{x}}_q)'\mathbf{s}_{xxp}^{+}(\bar{\mathbf{x}}_p - \bar{\mathbf{x}}_q)$ where $\mathbf{s}_{xxp}^{+}$ is the Moore-Penrose inverse of $\mathbf{s}_{xxp}$.

Next, we examine $B \equiv \text{var}(\hat{\bar{Y}}_{RGp}) - \text{var}(\hat{\bar{Y}}_{RGn})$. Since $\bar{Y} = \mathbf{b}_N'\bar{\mathbf{x}}_N$, it holds under ξ that

$$\text{var}_\xi\left(\hat{\bar{Y}}_{RGp} - \bar{Y}\right) = \text{var}_\xi\left(\mathbf{b}_p'\bar{\mathbf{x}}_N - \mathbf{b}_N'\bar{\mathbf{x}}_N\right). \tag{6.11}$$

Note that $E_\xi(.)$ and $\text{var}_\xi(.)$ are (frequentist) operators with respect to the random ε_i $(= y_i - \boldsymbol{\beta}'\mathbf{x}_i)$ according to model ξ. Using Kalman filtering, it is shown in Appendix A.2 (also for $f_i \neq 1$) that

$$\text{var}_\xi\left(\mathbf{b}_p - \mathbf{b}_N\right) = \sigma^2\left(\mathbf{V}_p - \mathbf{V}_N\right), \tag{6.12}$$

where $\mathbf{V}_l$ is as defined in section 3 for $l \in \{p, N\}$. So assuming $\bar{\mathbf{x}}_N = (1, \bar{\mathbf{z}}_N')'$ and $f_i = 1$, we get from (6.11) and (6.12) in analogy with (6.8)

$$\begin{aligned} \text{var}_\xi\left(\hat{\bar{Y}}_{RGp} - \bar{Y}\right) &= \sigma^2\left\{(1 + A_{pN})/p - 1/N\right\} && (6.13) \\ &\approx \sigma^2\left(p^{-1} - N^{-1} + q^2A_{pq}/pn^2\right); && (6.14) \end{aligned}$$

note $A_{NN} = 0$. In (6.14) we used $\bar{\mathbf{z}}_N \approx \bar{\mathbf{z}}_n = (p\bar{\mathbf{z}}_p + q\bar{\mathbf{z}}_q)/n$ so that $A_{pN} \approx A_{pn} = q^2A_{pq}/n^2$, where $a \approx b$ indicates that a/b tends (element-wise) to unity as $N, n \to \infty$. Subsequently, by (6.2) and (6.14),

$$B \approx \left(\frac{q}{np} + \frac{q^2A_{pq}}{n^2p}\right)\sigma^2 \approx \left(\frac{q}{np} + \frac{A_{pN}}{p}\right)\sigma^2. \tag{6.15}$$

From (6.9) and (6.15) it follows that $E(B_m) \approx B$. Hence, for $m \geq 1$ the model-based variance of the (proper) MI regression estimator (say $\hat{\bar{Y}}_{RGm}^{imp}$) can be approximated by

$$\begin{aligned} \text{var}_\xi\left(\hat{\bar{Y}}_{RGm}^{imp} - \bar{Y}\right) \approx E_\xi(T_m) &\approx \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2 + \frac{m+1}{m}B \\ &\approx \left(\frac{1}{p} - \frac{1}{N} + \frac{q}{npm} + \frac{(m+1)A_{pN}}{pm}\right)\sigma^2. && (6.16) \end{aligned}$$

This variance approximation can be estimated unbiasedly by replacing σ^2 by s_{en}^2, proved in Appendix A.3, including when $m = 1$. Note that (6.13) with p replaced by n can be seen as a generalization of the model-based variance formula (5.9) in Chambers and Clark (2012, 55) where one auxiliary variable is involved.

Some additional remarks are now in order. First, in the case of MCAR with $E(\bar{\mathbf{z}}_p) = \bar{\mathbf{z}}_N$ the quantity A_{pN} in (6.13) and (6.16) is of a negligible order $1/n$ and in fact, (6.13) amounts to the customary variance formula for the regression estimator with p observations. Second, note that use of formula (6.16) leads to lower t-values than using B_m and hence, to more accurate confidence intervals since $p, n \gg 1$. Third, setting $\mathbf{v} = \mathbf{0}$, we get $\bar{h}_n = q\bar{u}_q/n$ and hence, in analogy with (6.6), $\hat{\bar{Y}}_{RGZV}^{imp} = \hat{\bar{Y}}_{RGp} + q\bar{u}_q/n + O_p(1/n)$. Subsequently, using (6.13), its variance can be approximated by

$$\text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGZV} - \overline{Y}\right) \approx \left(\frac{1 + A_{pN}}{p} - \frac{1}{N} + \frac{q}{n^2}\right) s^2_{eN}, \tag{6.17a}$$

where A_{pN} is given in (6.10) with q replaced by N. Repeating these imputations m times, we get the following variance formula ($m \geq 1$)

$$\text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGmZV} - \overline{Y}\right) \approx \left(\frac{1 + A_{pN}}{p} - \frac{1}{N} + \frac{q}{mn^2}\right) s^2_{eN}, \tag{6.17b}$$

which is smaller in expectation than Rubin's T_m. Since $E_\xi(s^2_{en}) = \sigma^2\{1 + O(1/n)\}$, proved in Appendix A.3, the variances in (6.17a) and (6.17b) can be estimated by replacing s^2_{eN} by s^2_{en}. Finally, note that var($\mathbf{v} \mid s_r$) used above does not depend on N.

6.3 General Case with $f_i \neq 1$, G-weights and Applications

Assuming $E(\varepsilon_i^2) = \sigma^2$, it can be seen from (6.16) that the finite population correction (say fpc) is $-\sigma^2/N$. When $f_i \neq 1$ it follows from (6.11) and (6.12) that $fpc = -\sigma^2 \overline{\mathbf{x}}'_N \mathbf{V}_N \overline{\mathbf{x}}_N$ with $\mathbf{V}_l = (\mathbf{X}'_l \mathbf{F}_l^{-1} \mathbf{X}_l)^{-1}$ for $l \in \{p, n, N\}$. However, also now we have $fpc = -\sigma^2/N$ provided $\mathbf{x}_i$ contains f_i and $\overline{f}_N = 1$. To show this, let $\boldsymbol{\iota}_N$ denote a vector of N ones. Further, define $\mathbf{f}_N$ by $\mathbf{f}_N = \mathbf{F}_N \boldsymbol{\iota}_N$ and the projection matrix $\mathbf{P}_l$ by $\mathbf{P}_l = \mathbf{X}_l \mathbf{V}_l \mathbf{X}'_l \mathbf{F}_l^{-1}$ for $l \in \{p, n, N\}$; recall $\hat{\mathbf{y}}_p \equiv \mathbf{X}_p \mathbf{b}_p = \mathbf{P}_p \mathbf{y}_p$ and $\mathbf{P}_p \mathbf{X}_p = \mathbf{X}_p$. Under the above assumptions, we get

$$\overline{\mathbf{x}}'_N \mathbf{V}_N \overline{\mathbf{x}}_N = \boldsymbol{\iota}'_N \mathbf{X}_N \mathbf{V}_N \mathbf{X}'_N \mathbf{F}_N^{-1} \mathbf{F}_N \boldsymbol{\iota}_N / N^2 = \boldsymbol{\iota}'_N \mathbf{P}_N \mathbf{f}_N / N^2 = \boldsymbol{\iota}'_N \mathbf{f}_N / N^2 = 1/N. \tag{6.18}$$

The following corollary presents some general results on the improved efficiency of the ZVMI method compared to the standard MI method when using the regression estimator.

Corollary 6.1. Let MI estimator $\hat{\overline{Y}}^{imp}_{RGm}$ be defined by $\hat{\overline{Y}}^{imp}_{RGm} = \overline{\mathbf{x}}'_N \overline{\mathbf{b}}_{nm}$ and ZVMI estimator $\hat{\overline{Y}}^{imp}_{RGmZV}$ by $\hat{\overline{Y}}^{imp}_{RGmZV} = \overline{\mathbf{x}}'_N \overline{\mathbf{b}}_{nmZV}$ where $\overline{\mathbf{b}}_{nm}$ and $\overline{\mathbf{b}}_{nmZV}$ are as defined in Theorem 3.1. Then under the assumptions of model (3.1),

$$\text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGmZV} - \overline{Y}\right) = \sigma^2 \left\{\overline{\mathbf{x}}'_N (\mathbf{V}_p + \mathbf{V}_n \mathbf{V}_q^{-1} \mathbf{V}_n / m) \overline{\mathbf{x}}_N - 1/N\right\} \tag{6.19}$$

$$\text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGm} - \overline{Y}\right) = \sigma^2 \left\{\overline{\mathbf{x}}'_N (\mathbf{V}_p + \mathbf{V}_n \mathbf{V}_q^{-1} \mathbf{V}_p / m) \overline{\mathbf{x}}_N - 1/N\right\} \tag{6.20}$$

$$\text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGm} - \overline{Y}\right) - \text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGmZV} - \overline{Y}\right) = \sigma^2 \overline{\mathbf{x}}'_N \mathbf{V}_n \mathbf{V}_q^{-1} \mathbf{V}_p \mathbf{V}_q^{-1} \mathbf{V}_n \overline{\mathbf{x}}_N / m > 0 \tag{6.21}$$

and when $f_i = 1$ then, defining $A_{pN} = (\overline{\mathbf{x}}_p - \overline{\mathbf{x}}_N)' \mathbf{s}^+_{xxp} (\overline{\mathbf{x}}_p - \overline{\mathbf{x}}_N)$,

$$\text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGm} - \overline{Y}\right) - \text{var}_\xi\left(\hat{\overline{Y}}^{imp}_{RGmZV} - \overline{Y}\right) \approx \frac{\sigma^2}{m}\left(\frac{q^2}{n^2 p} + \frac{A_{pN}}{p}\right) > 0. \tag{6.22}$$

Proof. Both (6.19) and (6.20) consist of three terms. Each second term can be seen as the additional imputation variance, where we used (4.9) in (6.19), and (3.12) and (4.10) in (6.20). The first and third term in both equations follow from $\overline{Y} = \overline{\mathbf{x}}'_N \mathbf{b}_N$, (6.12) and (6.18). Next, (6.21) follows from (6.19), (6.20) and (4.10). Finally, (6.22) follows from (6.16) and (6.17b). □

Remark 6.1. Denoting the regression estimator in the complete-data case by $\hat{\hat{\overline{Y}}}_{RGn}$, the above corollary states that within the class of stochastic regression imputations under the (approximate) restrictions $E(\widehat{U}_j) \approx \text{var}(\hat{\hat{\overline{Y}}}_{RGn})$ ($j = 1, \ldots, m$), MI is less efficient than ZVMI. Obviously, deterministic imputations $y_i = \mathbf{x}'_i \mathbf{b}_p$ ($i = p + 1, \ldots, n$) do not satisfy these restrictions and lead to an estimator (say $\widehat{\overline{Y}}{}^{imp}_{RGdt}$) with a smaller variance, i.e. var$_\xi(\widehat{\overline{Y}}{}^{imp}_{RGdt} - \overline{Y}) = \sigma^2 \overline{\mathbf{x}}'_N(\mathbf{V}_p - \mathbf{V}_N)\overline{\mathbf{x}}_N$. However, a serious drawback of deterministic imputations is that they do not preserve the distributions of the missing data. Further, ZVMI imputations can easily be adjusted for $m = 1$ resulting in fully efficient imputations with $\mathbf{b}_{nAZV} = \mathbf{b}_p$ as pointed out in remark 4.1.

Example 6.1. Let U be a population of N persons with two strata (regions) U_1 and U_2 of sizes $N_1 = 4500$ and $N_2 = 5500$. The study variable y_i is defined by $y_i = 1$ if person i is unemployed and $y_i = 0$ otherwise ($i = 1, \dots, N$). Let s^2_{yNh}, N_h, n_h, p_h, q_h, B_h, $\overline{y}_{ph}$, and $\overline{\mathbf{x}}_{Nh}$ denote the stratum counterparts ($h = 1, 2$) of the corresponding population quantities. For expediency, suppose that the unemployment rates in U_1 and U_2 are 0.4 and 0.2, respectively. Hence, the stratum variances are $s^2_{yN1} = 0.24$ and $s^2_{yN2} = 0.16$. In fact, for categorical data heteroscedasticity is the rule rather than the exception. In an SRS sample of size $n = 1000$, the response rate in U_1 is 50% and in U_2 90%. Now conditioning on $n_h = N_h/10$, we compare the variances of the regression estimator of $\overline{Y}$ for several imputation methods; so far the (artificial) data description.

Consider a regression of y_i on two dummy variables x_{ih} ($h = 1, 2$) defined by $x_{ih} = 1$ if $i \in U_h$ and $x_{ih} = 0$ otherwise. Further, define $\mathbf{V}_{fl} = (\mathbf{X}'_l \mathbf{F}_l^{-1} \mathbf{X}_l)^{-1}$ where $\mathbf{F}_l$ is a $l \times l$ diagonal matrix with $f_{l,ii} = s^2_{yNh}$ for $i \in U_h$ ($l \geq 2$) so that $\mathbf{V}_{fp} = \text{diag}(s^2_{yN1}/p_1,\ s^2_{yN2}/p_2)$; note that $\sigma^2 = 1$, $p_1 = 225$ and $p_2 = 495$. Using GLS, we get $\mathbf{b}_{pGLS} = \mathbf{V}_{fp}\mathbf{X}'_p\mathbf{F}_p^{-1}\mathbf{y}_p = (\overline{y}_{p1}, \overline{y}_{p2})'$ and var$(\mathbf{b}_{pGLS}) = \mathbf{V}_{fp}$. Hence, the resulting regression estimator can be written as

$$\widehat{\overline{Y}}_{RGp} = \overline{\mathbf{x}}'_N \mathbf{b}_{pGLS} = \overline{x}_{1N}\overline{y}_{p1} + \overline{x}_{2N}\overline{y}_{p2} = W_1\overline{y}_{p1} + W_2\overline{y}_{p2} = \widehat{\overline{Y}}_{PSp},$$

where $W_h = N_h/N$ and $\widehat{\overline{Y}}_{PSp}$ stands for the poststratification (PS) estimator of $\overline{Y}$ based on p observations. Using the PS variance formula [e.g., Cochran (1977, 134)], we get var$(\widehat{\overline{Y}}_{PSp}) = 0.000294$ (≡100%); later outcomes are expressed as a percentage of this PS outcome. Next, using (6.11) and (6.12) for calculating $V_{RGp} \equiv$ var$(\widehat{\overline{Y}}_{RGp})$ with $\overline{\mathbf{x}}_N = (0.45,\ 0.55)'$ and $\sigma^2\mathbf{V}_l$ replaced by $\mathbf{V}_{fl}$ for $l \in \{p, N\}$, we get the same outcome. That is,

$$V_{RGp} \approx \text{var}_\xi\left(\widehat{\overline{Y}}_{RGp} - \overline{Y}\right) = \overline{\mathbf{x}}'_N\left(\mathbf{V}_{fp} - \mathbf{V}_{fN}\right)\overline{\mathbf{x}}_N = 100\%.$$

Now we first examine MI per stratum with $m = 1$; for MI methods for discrete data, see Rubin and Schenker (1986). Using (2.6) for calculating the between imputation variances B_h ($h = 1, 2$) and combining the results, it is seen that the additional imputation variance is $B = \sum_{h=1}^{2} W_h^2 q_h s^2_{yNh}/n_h p_h = 40.0\%$. On the other hand, it follows from $\widehat{\overline{Y}}_{RGp} = \overline{\mathbf{x}}'_N\mathbf{b}_{pGLS}$ and formula (3.12) with $\sigma^2\mathbf{V}_l$ replaced by $\mathbf{V}_{fl}$ for $l \in \{p, n\}$ that the increase of the MI variance for the corresponding regression estimator is $\overline{\mathbf{x}}'_N(\mathbf{V}_{fp} - \mathbf{V}_{fn})\overline{\mathbf{x}}_N = 40\%$ as well.

Next, consider hotdeck imputation in each stratum as described in section 4 without accounting for the uncertainty of the parameters. On the one hand, it follows from (4.4) that var$(\widehat{\overline{Y}}_{PSp})$ is now increasing only by $\sum_{h=1}^{2} W_h^2 q_h s^2_{yNh}/n_h^2 = 21.3\%$. On the other hand, this hotdeck imputation is equivalent to stochastic regression imputation where the imputed residuals in U_h are random drawings with replacement from the actual residuals in U_h ($h = 1, 2$) from the regression on x_{i1} and x_{i2}. So, by (6.19), the variance of $\widehat{\overline{Y}}_{RGp}$ ($= \overline{\mathbf{x}}'_N\mathbf{b}_{pGLS}$) is increasing by $\overline{\mathbf{x}}'_N\mathbf{V}_{fn}\mathbf{V}_{fq}^{-1}\mathbf{V}_{fn}\overline{\mathbf{x}}_N = 21.3\%$. In summary, applying m times Rubin's MI in this case with discrete data leads to an increase of var$_\xi(\widehat{\overline{Y}}_{RGp} - \overline{Y})$ [= var$(\widehat{\overline{Y}}_{PSp})$] by $40.0\%/m$ while applying m times ZVMI leads to an increase of only $21.3\%/m$.

Example 6.2. In this example we look more closely at the relationship between A_{pN} and the g-weights [e.g., Särndal et al. (1992, 235)]. Suppose that the unemployment rates in U_1 and U_2 are now 0.7 and 0.3, respectively, so that $s^2_{yN1} = s^2_{yN2} = s^2_{eN} = s^2_{ep} = 0.21$ and $f_i = 1$. For simplicity, consider the regression of y_i on a constant and $z_i \equiv x_{i2}$. Noting that $p = 720$, $p_2 = 495$, $\bar{z}_p = p_2/p = 0.6875$, $\bar{z}_N = \bar{x}_{N2} = 0.55$ and $s_{zzp} = \bar{z}_p(1 - \bar{z}_p) = 0.215$, we obtain $A_{pN} = 0.088$; see (6.10) with q replaced by N. Both variance approximation (6.13) and the PS variance formula yield 0.000296 (=100%). In contrast, using the standard variance formula (6.2) with n replaced by p leads to a negative bias of 8.7%. To remove this bias, one may use the variance estimator based on g-weights (say $\hat{V}_{gp}$). That is, writing $\hat{\bar{Y}}_{RGp} = \bar{\mathbf{x}}'_N\mathbf{b}_p = \mathbf{w}'_p\mathbf{y}_p$ with $\mathbf{w}_p = \mathbf{X}_p\mathbf{V}_p\bar{\mathbf{x}}_N$, the g-weights g_i are defined by $g_i = pw_i$ $(i = 1, \dots, p)$. Since $\mathbf{w}_p = \mathbf{X}_p\mathbf{V}_p\bar{\mathbf{x}}_N$, we get $\mathbf{w}'_p\mathbf{w}_p = \bar{\mathbf{x}}'_N\mathbf{V}_p\bar{\mathbf{x}}_N = (1 + A_{pN})/p$ [see derivation of (6.13)] so that $\mathbf{g}'_p\mathbf{g}_p/p = 1 + A_{pN} = 1.088$. Hence, $s^2_{gep} = 1.088s^2_{ep} = 0.2285$ $(ge_i \equiv g_ie_i)$; recall $s^2_{ep1} = s^2_{ep2} = 0.21$. Estimating var$(\hat{\bar{Y}}_{RGp})$ by $(1/p - 1/N)s^2_{gep}$ gives $\hat{V}_{gp} = 99\%$. Using a slightly modified version (say $\hat{V}^{md}_{gp}$) yields $\hat{V}^{md}_{gp} \equiv s^2_{gep}/p - s^2_{ep}/N = 100\%$. Further, since $\bar{g}_p = 1$ and $\mathbf{g}'_p\mathbf{g}_p/p = 1 + A_{pN}$, the coefficient of variation (say C_{gp}) equals $C_{gp} \equiv s_{gp}/\bar{g}_p = \sqrt{A_{pN}}$ $(p \gg 1)$. Using $C_{gp} = C_{wp}$, we get $1 + A_{pN} = 1 + C^2_{wp}$. Maybe somewhat surprisingly, this is exactly the factor $1 + L$ used by Kish (1992) to indicate the increase of a variance of a weighted mean compared to that of an unweighted mean (s^2/n in his notation). Finally, note that a regression on x_{i1} and x_{i2} gives $A_{pN} = (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_p)'\mathbf{s}^+_{xxp}(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_p) = 0.088$ as well. This follows from $(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_p) = 0.1375\mathbf{r}$, $\mathbf{r} = (1, -1)'$, $\mathbf{s}_{xxp} = 0.215\mathbf{R}$, $\mathbf{R} = \mathbf{r}\mathbf{r}'$ and $\mathbf{R}^+ = \mathbf{R}/4$; recall $\bar{z}_p = 0.6875$, $\bar{z}_N = 0.55$ and $s_{zzp} = 0.215$.

Example 6.3. In this example we look at g-weights in the case of $f_i \neq 1$. Consider the same data as in Example 6.1. Applying OLS yields $\mathbf{b}_p = (\mathbf{X}'_p\mathbf{X}_p)^{-1}\mathbf{X}'_p\mathbf{y}_p = (\bar{y}_{p1}, \bar{y}_{p2})'$ just as GLS. Estimating the standard OLS variance formula (6.2) with p observations yields $(1/p - 1/N)s^2_{ep} = 81\%$; note that $s^2_{eN} = W_1s^2_{yN1} + W_2s^2_{yN2} = 0.196$ while $s^2_{ep} = 0.185$ (based on $f_i = 1$). Since for OLS and GLS, $g_i = 1.44$ if $i \in U_1$ and $g_i = 0.8$ if $i \in U_2$, we get $s^2_{gep} = 0.226$ and $\hat{V}_{gp} = (1/p - 1/N)s^2_{gep} = 99.0\%$ which is a substantial improvement. The modification yields even $\hat{V}^{md}_{gp} \equiv s^2_{gep}/p - s^2_{ep}/N = 100\%$.

The next theorem presents some results on g-weights in the case of MAR nonresponse.

Theorem 6.1. Consider model ξ in (3.1). Further, suppose that $\mathbf{x}_i$ contains f_i and $\bar{f}_N = 1$. Define $\mathbf{V}_l = (\mathbf{X}'_l\mathbf{F}^{-1}_l\mathbf{X}_l)^{-1}$ for $l \in \{p, N\}$, $\mathbf{b}_l = \mathbf{V}_l\mathbf{X}'_l\mathbf{F}^{-1}_l\mathbf{y}_l$, $\hat{\bar{Y}}_{RGp} = \bar{y}_p + \mathbf{b}'_p(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_p)$, $\mathbf{g}'_p = p\bar{\mathbf{x}}'_N\mathbf{V}_p\mathbf{X}'_p\mathbf{F}^{-1}_p$, $s^2_{ep} = \mathbf{e}'_p\mathbf{F}^{-1}_p\mathbf{e}_p/(p - k)$ and $s^2_{gep} = \sum_{i=1}^{p} g^2_i e^2_i/(p - k)$. Let $l\mathbf{V}_l$, $\mathbf{x}_i$ and $1/f_i$ be $O(1)$ as $n, N \to \infty$. Let $V_{p\xi}$ denote the variance of $(\hat{\bar{Y}}_{RGp} - \bar{Y})$ under model ξ in (3.1). Then under MAR nonresponse,

$$V_{p\xi} = \sigma^2\bar{\mathbf{x}}'_N(\mathbf{V}_p - \mathbf{V}_N)\bar{\mathbf{x}}_N \tag{6.23}$$

$$= \sigma^2(\bar{\mathbf{x}}'_N\mathbf{V}_p\bar{\mathbf{x}}_N - 1/N); \tag{6.24}$$

$$E_\xi\left\{s^2_{ep}(\bar{\mathbf{x}}'_N\mathbf{V}_p\bar{\mathbf{x}}_N - 1/N)\right\} = V_{p\xi}; \tag{6.25}$$

$$E_\xi\left(s^2_{gep}/p - s^2_{ep}/N\right) = V_{p\xi} + O(1/p^2). \tag{6.26}$$

Proof. In analogy with (6.11) and (6.12), we get (6.23). Next, (6.24) follows from (6.18). Recall from regression theory that $E_\xi(s^2_{ep}) = \sigma^2$, from which (6.25) follows. To prove (6.26), define $\mathbf{E}_p = \text{diag}(\mathbf{e}_p)$ so that we can write s^2_{gep} as $s^2_{gep} = \mathbf{g}'_p\mathbf{E}^2_p\mathbf{g}_p/(p - k)$. Using $\mathbf{g}'_p = p\bar{\mathbf{x}}'_N\mathbf{V}_p\mathbf{X}'_p\mathbf{F}^{-1}_p$ and $E_\xi(\mathbf{E}^2_p) = \sigma^2\mathbf{F}_p(1 + O(p^{-1}))$, we obtain $E_\xi(s^2_{gep}) = (1 + O(p^{-1}))\sigma^2p^2\bar{\mathbf{x}}'_N\mathbf{V}_p\bar{\mathbf{x}}_N/(p - k)$

from which (6.26) follows. Recall that $\mathbf{e}_p = \mathbf{F}_p^{.5}\mathbf{M}_p\boldsymbol{\varepsilon}_p^*$ with $\mathbf{M}_p = \mathbf{I}_p - \mathbf{F}_p^{-.5}\mathbf{X}_p\mathbf{V}_p\mathbf{X}_p'\mathbf{F}_p^{-.5}$ and $\boldsymbol{\varepsilon}_p^* = \mathbf{F}_p^{-.5}\boldsymbol{\varepsilon}_p$ (Greene, 2003, 49) so that $\text{var}_\xi(\mathbf{e}_p) = \sigma^2\mathbf{F}_p^{.5}\mathbf{M}_p\mathbf{F}_p^{.5} = \sigma^2(\mathbf{F}_p - \mathbf{X}_p\mathbf{V}_p\mathbf{X}_p')$ where we used $\mathbf{M}_p\mathbf{M}_p = \mathbf{M}_p$, $\mathbf{M}_p' = \mathbf{M}_p$ and $\text{var}_\xi(\boldsymbol{\varepsilon}_p^*) = \sigma^2\mathbf{I}_p$. Hence, $E_\xi(\mathbf{E}_p^2) = \sigma^2\mathbf{F}_p(1 + O(p^{-1}))$. □

Remark 6.2. Another estimator for σ^2 is $s_{ep2}^2 \equiv \mathbf{e}_p'\mathbf{e}_p/\overline{f}_p(p-k)$. Noting that $E(\boldsymbol{\varepsilon}_p'\boldsymbol{\varepsilon}_p/p) = \overline{f}_p\sigma^2$, it follows that s_{ep2}^2 is consistent for σ^2; also see Särndal (1992).

6.4 Unequal Probability Sampling

Let π_i denote the first order inclusion probability of the ith unit ($i = 1, \dots, N$). Further, assume that π_i is a function of $\mathbf{x}_i$ so that π_i is available for each unit in s. Now first assuming $q = 0$, the HT estimator of $\overline{Y}$ can be written as $\hat{\overline{Y}}_{HT} = \overline{y}_{*n}$ where $\overline{y}_{*n} = \sum_{i=1}^n y_{*i}/n$ with $y_{*i} = ny_i/N\pi_i$; also see Appendix A.1. Similarly, we can write the HT estimator of $\overline{\mathbf{x}}_N$ as $\hat{\overline{\mathbf{x}}}_{N,HT} = \overline{\mathbf{x}}_{*n}$. Noting that under model (3.1) we have $Y_{*i} = \mathbf{x}_{*i}'\boldsymbol{\beta} + \varepsilon_{*i}$, where $Y_{*i} = nY_i/N\pi_i$, $E_\xi(\varepsilon_{*i}^2) = f_{*i}\sigma^2$ ($f_{*i} = n^2 f_i/\pi_i^2 N^2$) and using the GLS estimator of $\boldsymbol{\beta}$ (say $\mathbf{b}_{*n}$), the regression estimator in this case can be written as

$$\hat{\overline{Y}}_{RG*n} = \overline{y}_{*n} + \mathbf{b}_{*n}'(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_{*n}) \quad \left[\mathbf{b}_{*n} = \left(\mathbf{X}_{*n}'\mathbf{F}_{*n}^{-1}\mathbf{X}_{*n}\right)^{-1}\mathbf{X}_{*n}'\mathbf{F}_{*n}^{-1}\mathbf{y}_{*n}\right]. \tag{6.27}$$

Next, a number of remarks are in order. First, replacing $\mathbf{F}_{*n}$, $\mathbf{X}_{*n}$ and $\mathbf{y}_{*n}$ by their definitions yields $\mathbf{b}_{*n} = \mathbf{V}_n\mathbf{X}_n'\mathbf{F}_n^{-1}\mathbf{y}_n$ ($\equiv \mathbf{b}_n$); also see section 3. That is, given model ξ both $\mathbf{b}_{*n}$ and $\mathbf{b}_n$ can be seen as the GLS estimator of $\boldsymbol{\beta}$ given the n observations irrespective of the π_i. Second, assuming that $\mathbf{x}_{*i}$ contains f_{*i} or, equivalently, $\mathbf{x}_i$ contains f_i/π_i, we have $\overline{e}_{*n} = \overline{y}_{*n} - \mathbf{b}_{*n}'\overline{\mathbf{x}}_{*n} = 0$ so that $\hat{\overline{Y}}_{RG*n} = \mathbf{b}_{*n}'\overline{\mathbf{x}}_N = \mathbf{b}_n'\overline{\mathbf{x}}_N = \hat{\overline{Y}}_{RGn}$ provided $\mathbf{x}_i$ contains f_i so that $\overline{e}_n = 0$ as well. Third, replacing $\mathbf{b}_{*n}$ ($= \mathbf{b}_n$) in (6.27) by $\mathbf{b}_{\pi n} \equiv (\mathbf{X}_n'\mathbf{F}_n^{-1}\boldsymbol{\Pi}_n^{-1}\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{F}_n^{-1}\boldsymbol{\Pi}_n^{-1}\mathbf{y}_n$ with $\boldsymbol{\Pi}_n = \text{diag}(\pi_1, \dots, \pi_n)$ leads to another estimator (say $\hat{\overline{Y}}_{RG\pi n}$). However, the difference is negligible since

$$\hat{\overline{Y}}_{RG*n} - \hat{\overline{Y}}_{RG\pi n} = (\mathbf{b}_{*n} - \mathbf{b}_{\pi n})'(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_{*n}) = O_p(1/n);$$

note $E_\xi\, E(\mathbf{b}_{\pi n}) \approx E_\xi(\mathbf{b}_N) = \boldsymbol{\beta}$. Options $\mathbf{b}_n$ and $\mathbf{b}_{\pi n}$ are also discussed in Deville and Särndal (1994) in a somewhat different context. Fourth, since under model ξ in the case of q ($q > 0$) ignorably missing observations $E_\xi(B_m) = \sigma^2\overline{\mathbf{x}}_N'(\mathbf{V}_p - \mathbf{V}_n)\overline{\mathbf{x}}_N = B$ when $\mathbf{b}_n$ is used, option $\mathbf{b}_n$ is to be preferred when applying MI. Finally, in a sample with unequal π_i we can use the same kind of MI, ZVMI or SI as in an SRS sample with $\pi_i = n/N$ provided that $\mathbf{b}_n$ ($= \mathbf{b}_{*n}$) is used and $\mathbf{x}_i$ contains f_i and f_i/π_i. In the next example it is shown that MI may fail when x_i is a scalar and $b_{\pi n}$ and $b_{\pi p}$ are used; $b_{\pi p}$ is defined as $b_{\pi n}$ with n replaced by p when $q > 0$.

Example 6.4. Consider model ϑ: $Y_i \sim N(\beta x_i, \sigma^2)$ ($i = 1, 2, \dots, N$). Suppose $\sum_{i=1}^N x_i = 1$, $\pi_i = nx_i$ and $n \ll N$. Then we get $b_{\pi p} = \overline{y}_p/\overline{x}_p$ and $\text{var}_\vartheta(b_{\pi p} \mid s_r) = s_{ep}^2/p\overline{x}_p^2$; recall $b_{\pi p} = \beta + \overline{\varepsilon}_p/\overline{x}_p$ ($\varepsilon_i = y_i - \beta x_i$). In the spirit of Rubin (1987), we use the imputations

$$y_i = (b_{\pi p} + v)x_i + u_i, \quad (i = p+1, \dots, n) \tag{6.28}$$

where $v \sim N(0, s_{ep}^2/p\overline{x}_p^2)$ and $u_i \sim N(0, s_{ep}^2)$. Using $\overline{y}_p - b_{\pi p}\overline{x}_p = 0$, it follows from (6.28) that $\overline{y}_n = b_{\pi p}\overline{x}_n + q(v\overline{x}_q + \overline{u}_q)/n$. Next, using $b_{\pi n} = \overline{y}_n/\overline{x}_n$, $\hat{\overline{Y}}_{RG\pi n}^{imp}$ can be written as

$$\hat{\overline{Y}}_{RG\pi n}^{imp} = \overline{x}_N b_{\pi n} = \overline{x}_N\left\{b_{\pi p} + \frac{q}{n\overline{x}_n}\left(v\overline{x}_q + \overline{u}_q\right)\right\}; \tag{6.29}$$

recall that the intercept is zero. Repeating this m times, it follows from (6.29) that

$$E(B_m \mid s_r) = \text{var}\left\{\frac{q\overline{x}_N\left(v\overline{x}_q + \overline{u}_q\right)}{n\overline{x}_n} \mid s_r\right\} = \left(\frac{\overline{x}_q^2}{p\overline{x}_p^2} + \frac{1}{q}\right)\frac{q^2\overline{x}_N^2 s_{ep}^2}{n^2\overline{x}_n^2}. \tag{6.30}$$

Now choosing $\overline{x}_p = 0.6\overline{x}_N$, $\overline{x}_q = 1.4\overline{x}_N$, and $p = q = n/2$, formula (6.30) yields $E(B_m \mid s_r) = (5.44q/n + p/n)qs_{ep}^2/np = 3.22s_{ep}^2/n$. Because var$_\vartheta(\hat{\hat{\bar{Y}}}_{RG\pi p}) = \overline{x}_N^2\sigma^2/p\overline{x}_p^2 = 5.56\sigma^2/n$ and var$_\vartheta(\hat{\hat{\bar{Y}}}_{RG\pi n}) \approx s_{eN}^2/n$, we get $B \approx 4.56\sigma^2/n$. Hence, B_m has a negative bias of about 29% in a model-based sense. So in the case of MAR one should be careful using $b_{\pi p}$ in an MI procedure. For more examples, see section 7.

6.5 Simulation Study

Let the superpopulation models for generating the Y_i and the response indicators R_i of a population of size $N = 2000$ be given by

$$Y_i = 5 + x_i + 10d_{2i} + \sqrt{f_i}\varepsilon_i, \quad x_i = a_i + 10d_{2i} \quad \text{and}$$
$$\varphi_i \equiv P(R_i = 1) = (45 + x_i/2 - 10d_{2i})/100 \quad (i = 1, \dots, N),$$

where a_i follows a uniform distribution $U(15,\ 55)$, $d_{2i} = 1$ if $N/2 < i \leq N$ and $d_{2i} = 0$ otherwise, and $\varepsilon_i \sim N(0,\ 126)$; note that $\overline{x}_N \approx 40$, $\overline{d}_{2N} = 0.5$, $\overline{Y} \approx 50$ and $\overline{\varphi}_N \approx 60\%$. For f_i we considered two scenarios, that is $f_i = 1$ and $f_i = x_i/\overline{x}_N$. We also considered the scenario with ε_i replaced by $\varepsilon_i^* = \delta_i - E(\delta_i)$ where ln$(\delta_i) \sim N(3,\ 0.22305)$ so that $E(\varepsilon_i^*) = 0$ and var$(\varepsilon_i^*) = 126$. From 100,000 generated populations an SRS sample of size $n = 400$ was drawn; vector $\mathbf{u}_q$ in (3.7) was drawn from $N(0, s_{ep2}^2\mathbf{F}_q)$. Several estimators were examined for $m = 1$ under four scenarios. Table 6.1 summarizes the averages of 100,000 Monte Carlo (MC) means of three variance estimates (say Vp, VMI and VZVMI) of $\hat{\hat{\bar{Y}}}_{RGp}$, $\hat{\bar{Y}}_{RGm}^{imp}$ and $\hat{\bar{Y}}_{RGmZV}^{imp}$ given in (6.24), (6.20) and (6.19), respectively, with $m = 1$ and σ^2 replaced by s_{ep2}^2; MC mean of s_{ep2}^2 (s_{ep}^2) is 126.1 (126.0). Between parentheses their biases (in %) are mentioned or, in fact, the differences (in %) with the MC variances of those estimators. For calculating the coverages (in %) of the 95% confidence intervals we used $t_\infty = 1.96$ since p is large. All coverages are slightly less than 95%.

Table 6.1 Variance estimates for $m = 1$ under four scenarios for $\hat{\hat{\bar{Y}}}_{RGp}$, $\hat{\bar{Y}}_{RGm}^{imp}$ and $\hat{\bar{Y}}_{RGmZV}^{imp}$ with biases (in %) between parentheses, and coverages (in %).

scenario	Vp		VMI		VZVMI	
ε_i and $f_i = 1$	0.472 (0.2)	94.9	0.698 (-0.7)	94.8	0.600 (-0.3)	94.9
ε_i and $f_i = x_i/\overline{x}_N$	0.461 (-0.0)	95.0	0.668 (-0.0)	94.9	0.580 (0.2)	95.0
ε_i^* and $f_i = 1$	0.473 (0.0)	94.7	0.699 (-0.9)	94.6	0.601 (-0.6)	94.7
ε_i^* and $f_i = x_i/\overline{x}_N$	0.461 (-0.0)	94.6	0.670 (-0.2)	94.8	0.581 (0.1)	94.7

It emerges that all variance biases are small, including when the disturbances are lognormal. The additional imputation variance of MI exceeds that of ZVMI considerably by about $75\%/m$ ($m \geq 1$); note (VMI-Vp)/(VZVMI-Vp) ≈ 1.75. This again confirms that there is no reason for using MI when applying the regression estimator. A nonzero **v** leads to an unnecessary, sizable increase of the MI variance unless m is sufficiently large. Further, note that in practice MI intervals based on T_m are larger because t_κ in (5.1) is larger than $t_\infty = 1.96$ used in Table 6.1. We also considered an NMAR case with ε_i, $f_i = 1$ and $\varphi_i = (45 + 0.4Y_i - 10d_{2i})/100$ so that $\overline{\varphi}_N \approx 60\%$. The three biases of $\hat{\hat{\bar{Y}}}_{RGp}$, $\hat{\bar{Y}}_{RGm}^{imp}$ and $\hat{\bar{Y}}_{RGmZV}^{imp}$ ($m = 1$) are now 0.85 resulting in the following (under)coverages 76%, 82% and 80%, respectively. The variances are 0.468 (-0.1%), 0.685 (-0.3%) and 0.594 (-0.4%); biases (between parentheses) are fairly small.

7 MI, SI and ZVMI in the Case of Estimating a Ratio

In the case of an SRS sample with $1 \ll p = n \ll N$, the variance of the ratio estimator $\widehat{\overline{Y}}_{Rn}$ $(= \overline{y}_n \overline{X}/\overline{x}_n)$ of a population mean can be estimated by $\widehat{\text{var}}(\widehat{\overline{Y}}_{Rn}) = s_{en}^2/n$ where $s_{el}^2 = \mathbf{e}_l' \mathbf{e}_l/(l-1)$, $\mathbf{e}_l = \mathbf{y}_l - \widehat{R}_l \mathbf{x}_l$ and $\widehat{R}_l = \overline{y}_l/\overline{x}_l$ for $l \in \{p, n, N\}$ [e.g., Cochran (1977, 155) and Rubin (1987, 19)]. Note that $\overline{X}$ is defined by $\overline{X} = \sum_{i=1}^{N} x_i/N = \overline{x}_N$. The ratio estimator is important because it is often used in business surveys. Moreover, the estimator $\widehat{R}_n$ of the ratio $R = \overline{Y}/\overline{X}$ is a relevant quantity in its own right. For instance, in a business survey where y stands for the turnover of an establishment and x for the number of employees the ratio $R = \overline{Y}/\overline{X}$ is quite relevant. In this section we examine the ratio estimator for different data patterns and for several imputation methods.

As before we assume that p $(p < n)$ observations are available and that the nonresponse is ignorable. It is shown that the classical MI procedure yields biased variance estimators for some data patterns while SI and ZVMI work well. First, assuming that the data obey ratio model ψ: $Y_i/\sqrt{x_i} \sim N(\beta\sqrt{x_i}, \sigma^2)$, the OLS estimate of β equals the familiar sample ratio $\widehat{R}_n = \overline{y}_n/\overline{x}_n$ when $q = 0$ [e.g., Särndal et al. (1992, 247 and 535) and Rubin (1987, 47)]. Hence, in the case of missing observations, one may apply MI to the ratio estimator of a population mean as we have seen in sections 3 and 6. Second, consider model ϑ: $Y_i \sim N(\beta x_i, \sigma^2)$ $(i = 1, 2, \dots, N)$ as in Example 6.4. With this kind of data we show that MI may fail here as well when using the ratio estimator in the case of an SRS sample. Since $\widehat{R}_p = \overline{y}_p/\overline{x}_p = b_{\pi p}$, we get here the same imputations as in (6.28). That is, in the present notation,

$$y_i = \left(\widehat{R}_p + v\right) x_i + u_i \quad (i = p+1, \dots, n), \tag{7.1}$$

where $v \sim N(0, s_{ep}^2/p\overline{x}_p^2)$ and $u_i \sim N(0, s_{ep}^2)$. Hence, $\widehat{\overline{Y}}_{Rn}^{imp} \equiv \widehat{R}_n \overline{X} = \widehat{\overline{Y}}_{RG\pi n}^{imp}$ with $\widehat{R}_n = \overline{y}_n/\overline{x}_n$ and $\overline{y}_n = \widehat{R}_p \overline{x}_n + q(v\overline{x}_q + \overline{u}_q)/n$ so that $E(B_m \mid s_r)$ is as given in (6.30). Choosing again $\overline{x}_p = 0.6\overline{X}$, $\overline{x}_q = 1.4\overline{X}$, and $p = q = n/2$ as in Example 6.4, the negative bias of B_m is about 29%. In the next case we show that imputations also may lead to a positive bias of B_m. Assume $\overline{x}_p = 1.6\overline{X}$, $\overline{x}_q = 0.4\overline{X}$, and $p = q = n/2$. In analogy with Example 6.4 we now get $E(B_m \mid s_r) = 0.53 s_{ep}^2/n$, $\text{var}_\vartheta(\widehat{\overline{Y}}_{Rp}) = \overline{X}^2 \sigma^2/p\overline{x}_p^2 = 0.78\sigma^2/n$, $\text{var}(\widehat{\overline{Y}}_{Rn}) \approx s_{eN}^2/n$ and hence, $B \approx -0.22\sigma^2/n$; $\widehat{\overline{Y}}_{Rn}$ stands for the complete-data estimator. Because $B < 0$ [i.e., $\text{var}_\vartheta(\widehat{\overline{Y}}_{Rp}) < \text{var}(\widehat{\overline{Y}}_{Rn}) \approx \sigma^2/n < E(T_\infty)$] any form of MI fails when estimating $\overline{Y}/\overline{X}$ even when the ML/OLS estimate b_p of β is used for imputation. In the words of Meng (1994), the complete-data estimator $\widehat{\overline{Y}}_{Rn}$ in this case is not *self-efficient*. However, SI and ZVMI work well under model ϑ as shown in the following somewhat more general example.

Example 7.1. Consider model λ: $Y_i = \beta x_i + \varepsilon_i$, $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \sigma^2$. Let q observations be ignorably missing $(1 \ll p, n \ll N)$. Instead of (7.1) apply now the following imputations $y_i = \widehat{R}_p x_i + u_i$ $(i = p+1, \dots, n)$ where the u_i are iid $N(0, s_{ep}^2)$. This yields $\overline{y}_n = \widehat{R}_p \overline{x}_n + q\overline{u}_q/n$ and $\widehat{\overline{Y}}_{RZV}^{imp} = \widehat{R}_p \overline{X} + q\overline{X}\overline{u}_q/n\overline{x}_n$; recall $\overline{e}_p = 0$. Noting that $\widehat{R}_p = \beta + \overline{\varepsilon}_p/\overline{x}_p$ and $\overline{X}^2/\overline{x}_n^2 \approx 1$, the model-based variance of $\widehat{\overline{Y}}_{RZV}^{imp}$ is $\text{var}_\lambda(\widehat{\overline{Y}}_{RZV}^{imp}) \approx (\overline{X}^2/p\overline{x}_p^2 + q/n^2)\sigma^2$. This can be estimated by replacing σ^2 by s_{ep}^2 or s_{en}^2. Unlike MI this variance estimator is asymptotically unbiased for $\text{var}_\lambda(\widehat{\overline{Y}}_{RZV}^{imp})$ as $n \to \infty$. When the sampling fraction is not negligible, the first variance component of var$_\lambda(\widehat{\overline{Y}}{}^{imp}_{RZV})$ should be replaced by $\overline{X}^2$ var$_\lambda(\widehat{R}_p - \widehat{R}_N)$ where $\widehat{R}_p - \widehat{R}_N = \bar{\varepsilon}_p/\bar{x}_p - \bar{\varepsilon}_N/\overline{X}$. Noting that cov$_\lambda(\bar{\varepsilon}_p, \bar{\varepsilon}_N) = \sigma^2/N$ and defining $a = \overline{X}/\bar{x}_p$, further calculations yield

$$\text{var}_\lambda\left(\widehat{\overline{Y}}{}^{imp}_{RZV} - \overline{Y}\right) = \left\{a^2/p - (2a-1)/N + q/n^2\right\}\sigma^2. \tag{7.2}$$

Repeating this m times gives the ZVMI version and hence, $q\sigma^2/n^2$ in (7.2) should be replaced by $q\sigma^2/mn^2$. Note that for the u_i hotdeck imputation can be used when required. Similar results can be derived for model ξ in (3.1) with $E(\varepsilon_i^2) = f_i\sigma^2$ and $\bar{f}_N = 1$. Recall that now $u_i \sim N(0, f_i s^2_{ep})$ for $p < i \leq n$ and $s^2_{el} = \mathbf{e}'_l\mathbf{F}_l^{-1}\mathbf{e}_l/(l-k)$ for $l \in \{p, n, N\}$ with $k = 1$. Noting that cov$_\xi(\bar{\varepsilon}_p, \bar{\varepsilon}_l) = E_\xi(\sum_{i=1}^p \varepsilon_i^2)/pl = \sigma^2\bar{f}_p/l$ for $l \in \{p, N\}$, it can be derived in analogy with (7.2) that

$$\text{var}_\xi\left(\widehat{\overline{Y}}{}^{imp}_{RZV} - \overline{Y}\right) = \left\{a^2\bar{f}_p/p - (2a\bar{f}_p - 1)/N + q\bar{f}_q/n^2\right\}\sigma^2 \tag{7.3a}$$

$$\text{var}_\xi\left(\widehat{\overline{Y}}{}^{imp}_{RmZV} - \overline{Y}\right) = \left\{a^2\bar{f}_p/p - (2a\bar{f}_p - 1)/N + q\bar{f}_q/mn^2\right\}\sigma^2. \tag{7.3b}$$

Remark 7.1. From (7.3a) it follows that under model ξ with $\bar{f}_N = 1$ deterministic imputations $y_i = \widehat{R}_p x_i$ $(i = p+1, \dots, n)$ lead to estimator (say $\widehat{\overline{Y}}{}^{imp}_{Rdt}$) $\widehat{\overline{Y}}{}^{imp}_{Rdt} = \widehat{R}_p\overline{X}$ with variance (say V_1)

$$V_1 \equiv \text{var}_\xi\left(\widehat{\overline{Y}}{}^{imp}_{Rdt} - \overline{Y}\right) = \left\{a^2\bar{f}_p/p - (2a\bar{f}_p - 1)/N\right\}\sigma^2. \tag{7.4}$$

Under an SRS sampling design and model ξ with $f_i = x_i$, Särndal (1992) proposes the same deterministic imputations for the HT estimator yielding the following estimator (say $\widehat{\overline{Y}}{}^{imp}_{HTdt}$) and variance approximation (say V_2)

$$\widehat{\overline{Y}}{}^{imp}_{HTdt} = \frac{p}{n}\bar{y}_p + \frac{q}{n}\widehat{R}_p\bar{x}_q = \widehat{R}_p\bar{x}_n$$

$$\text{var}\left(\widehat{\overline{Y}}{}^{imp}_{HTdt}\right) \approx V_2 \equiv \left(\frac{1}{n} - \frac{1}{N}\right)s^2_{yN} + \left(\frac{1}{p} - \frac{1}{n}\right)C_1\sigma^2 \quad (C_1 = \bar{x}_n\bar{x}_q/\bar{x}_p). \tag{7.5}$$

Note that $V_2 > V_1$. Under model ξ with $k > 1$, a similar result can be derived for the regression estimator compared to the HT estimator after using deterministic regression imputation. That is, $\widehat{\overline{Y}}{}^{imp}_{HTdt} = \mathbf{b}'_p\bar{\mathbf{x}}_n$; for further details, see Deville and Särndal (1994). To prove $V_2 > V_1$ in this more general case, write $\widehat{\overline{Y}}{}^{imp}_{HTdt} - \overline{Y}$ $(= \mathbf{b}'_p\bar{\mathbf{x}}_n - \mathbf{b}'_N\bar{\mathbf{x}}_N)$ as

$$\widehat{\overline{Y}}{}^{imp}_{HTdt} - \overline{Y} = (\mathbf{b}_p - \mathbf{b}_N)'\bar{\mathbf{x}}_N + \boldsymbol{\beta}'\left(\bar{\mathbf{x}}_n - \bar{\mathbf{x}}_N\right) + (\mathbf{b}_p - \boldsymbol{\beta})'\left(\bar{\mathbf{x}}_n - \bar{\mathbf{x}}_N\right).$$

Since the utmost right term is of negligible order $1/n$, var$(\widehat{\overline{Y}}{}^{imp}_{HTdt})$ can be approximated by

$$V_2 \approx V_1 + (1/n - 1/N)\boldsymbol{\beta}'\mathbf{s}_{xxN}\boldsymbol{\beta} > V_1, \tag{7.6}$$

where V_1 is given in (6.24); note that the covariance is zero because $E_\xi(\boldsymbol{\varepsilon}_N\mathbf{x}'_i) = 0$. Recall that σ^2 in V_1 can be estimated by s^2_{ep} or s^2_{ep2} defined in remark 6.2, and $\boldsymbol{\beta}'\mathbf{s}_{xxN}\boldsymbol{\beta}$ by $\mathbf{b}'_p\mathbf{s}_{xxN}\mathbf{b}_p$. To illustrate that (7.5) and (7.6) are equivalent when $f_i = x_i$ and $\bar{f}_N = \overline{X} = 1$, choose for instance $N = 4000, n = 1000, p = 667, \bar{x}_p = 0.75, \bar{x}_q = 1.5, s^2_{yN} = 10{,}000$ and $\rho = 0.8$ where ρ is the correlation coefficient between x_i and y_i $(i = 1, \dots, N)$. Then we obtain $\sigma^2 = 3600$, $\beta^2 s_{xxN} = 6400$, $C_1 = 2$ and $V_1 = 6.3$ so that $V_2 = 11.1$ according to (7.5) and (7.6).

8 Missing Covariates

In this section we briefly look at an adjustment procedure for the regression estimator in the case of item nonresponse and imputations among the $\mathbf{x}_i$ provided that $\overline{\mathbf{x}}_N$ is known. Suppose that the missing observations in the $p \times k$ matrix $\mathbf{X}_p$ can be estimated by a fully conditional specification (FCS) method [e.g., Murray (2018) and Van Buuren (2018)]. Let $\mathbf{X}_p^{FC}$ denote such an estimated matrix. For expository purposes, suppose that $x_{il} = \mathbf{w}_{il}'\boldsymbol{\tau}_l + \eta_{il}$ where $E(\eta_{il}) = 0$, $E(\eta_{il}\eta_{ig}) = \sigma_{lg}$ $(i = 1, \dots, p; l, g = 1, \dots, k)$, and the missingness of the x_{il} is independent of the random η_{il} and ε_i. Let the variables in $\mathbf{w}_{il}$ be a subset of those in $\mathbf{x}_i$; so some elements in $\mathbf{w}_{il}$ might be missing as well. Further, suppose that the first c records are complete. Subsequently, under model ξ we can write the regression estimator of $\overline{Y}$ adjusted for item nonresponse (say $\widehat{\overline{Y}}_{RGpIN}$) as

$$\widehat{\overline{Y}}_{RGpIN} = \overline{y}_p + \mathbf{b}_c'\left(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_p^{FC}\right), \tag{8.1}$$

where $\mathbf{b}_c$ is the GLS estimate of $\boldsymbol{\beta}$ from the first c records; recall that $\mathbf{b}_p$ based on imputations for the missing x_{il} might be biased (Little, 1992). In order to examine its variance, write $\widehat{\overline{Y}}_{RGpIN} - \overline{Y}$ as

$$\begin{aligned}\widehat{\overline{Y}}_{RGpIN} - \overline{Y} &= \overline{y}_p + \mathbf{b}_p'\left(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_p\right) - \mathbf{b}_N'\overline{\mathbf{x}}_N + \left(\mathbf{b}_c - \mathbf{b}_p\right)'\left(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_p\right) + \mathbf{b}_c'\left(\overline{\mathbf{x}}_p - \overline{\mathbf{x}}_p^{FC}\right) \\ &= \left(\mathbf{b}_p - \mathbf{b}_N\right)'\overline{\mathbf{x}}_N + \left(\mathbf{b}_c - \mathbf{b}_p\right)'\left(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_p\right) + \mathbf{b}_c'\left(\overline{\mathbf{x}}_p - \overline{\mathbf{x}}_p^{FC}\right).\end{aligned} \tag{8.2}$$

Next, define $q_{il} = 1$ if x_{il} is *missing* and $q_{il} = 0$ if x_{il} is *observed*, $q_l = \sum_{i=1}^p q_{il}$, $q_{lg} = \sum_{i=1}^p q_{il}q_{ig}$, $\overline{\eta}_{1l} = \sum_{i=1}^p q_{il}\eta_{il}/q_l$, $\overline{\mathbf{w}}_{1l} = \sum_{i=1}^p q_{il}\mathbf{w}_{il}/q_l$ and $\overline{\mathbf{w}}_{1l}^{FC} = \sum_{i=1}^p q_{il}\mathbf{w}_{il}^{FC}/q_l$; note $q_{ll} = q_l$. Further, note that $x_{il}^{FC} = x_{il}$ if $q_{il} = 0$ and define the deterministic imputations $x_{il}^{FC} = \mathbf{w}_{il}^{FC'}\boldsymbol{\tau}_l^{FC}$ if $q_{il} = 1$. Now using a Taylor approximation, the last term in (8.2) can be decomposed in two summation terms. That is,

$$\begin{aligned}\mathbf{b}_c'\left(\overline{\mathbf{x}}_p - \overline{\mathbf{x}}_p^{FC}\right) &\approx \sum_{l=1}^k \beta_l\left(\overline{x}_{pl} - \overline{x}_{pl}^{FC}\right) \\ &= \sum_{l=1}^k \beta_l q_l \overline{\eta}_{1l}/p + \sum_{l=1}^k \beta_l q_l\left(\overline{\mathbf{w}}_{1l}'\boldsymbol{\tau}_l - \overline{\mathbf{w}}_{1l}^{FC'}\boldsymbol{\tau}_l^{FC}\right)/p.\end{aligned} \tag{8.3}$$

Now noting that $\mathbf{b}_p - \mathbf{b}_N$ and $\mathbf{b}_c - \mathbf{b}_p$ are uncorrelated under model ξ, proved in Appendix A.2, it follows from (8.2) and (8.3) that var$_\xi(\widehat{\overline{Y}}_{RGpIN} - \overline{Y})$ consists of four components (say $V_1, \dots, V_4$). V_1 is given in (6.13). Using (6.12) with (p, N) replaced by (c, p) we get $V_2 = \sigma^2(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_p)'(\mathbf{V}_c - \mathbf{V}_p)(\overline{\mathbf{x}}_N - \overline{\mathbf{x}}_p)$. Since $E(\eta_{il}\eta_{ig}) = \sigma_{lg}$, we get $E(q_l\overline{\eta}_{1l}q_g\overline{\eta}_{1g}) = q_{lg}\sigma_{lg}$ so that $V_3 = \sum_{l,g=1}^k \beta_l q_{lg}\sigma_{lg}\beta_g/p^2$, where β_l can be estimated by b_{cl} and σ_{lg} from the FCS residuals by, say σ_{lg}^{FC}. Repeating this m times $(j = 1, \dots, m)$, we can write each regression estimator (say $\widehat{Q}_j$) as $\widehat{Q}_j = const - \sum_{l=1}^k b_{cl}q_l\overline{\mathbf{w}}_{1lj}^{FC'}\boldsymbol{\tau}_{lj}^{FC}/p$. Hence, the fourth component V_4 of var$(\overline{Q}_m)$ can be estimated by $\widehat{V}_4 = \widehat{B}_m/m$ where $\widehat{B}_m$ is as defined in section 2.2. In analogy with (5.1) it can be derived that the degrees of freedom can be approximated by $\kappa = (\widehat{V}_{1234}/\widehat{V}_4)^2(m - 1)$, where $V_{12} = V_1 + V_2$, $V_{123} = V_{12} + V_3$, and so forth. Advantages of this partial MI (PMI) approach compared to Rubin's MI are its lower t-value and smaller variance. Interesting issues for further research are what happens when c is too small, how can this approach be adjusted for categorical variables and heteroscedastic η_{il}, and also to what extent are $\widehat{Q}_j$, $\widehat{V}_{1234}$, $\boldsymbol{\tau}_l^{FC}$ and σ_{lg}^{FC} asymptotically unbiased. Note that we did not use imputations for $i > p$. In our opinion, there are no frequentist reasons for further computational efforts. Also note that it follows from (8.2) and (8.3) that var$_{\xi}(\hat{\bar{Y}}_{RGpIN} - \bar{Y})$ consists of four components. V_1 stands for the variance in the hypothetical case of deterministic regression imputation and $c = p$. V_2 is due to the bias of $\bar{\mathbf{x}}_p$ provided $c < p$. Finally, V_3 is due to the noise terms η_{il} and V_4 to the imputations x_{il}^{FC} when $q_{il} = 1$.

9 Concluding Remarks

In this paper, we have examined MI from a frequentist point of view, including the three conditions for proper MI. It emerges that drawings from the χ^2-distribution can be omitted. This removes the bias from T_m and leads to a smaller variance. We have shown that SI without accounting for parameter uncertainty leads to valid confidence intervals provided that the correct variance formulas are used. Further, the new ZVMI estimate has a smaller variance than the standard MI estimate when estimating a regression model or a population mean by means of the regression estimator under model (3.1) and MAR nonresponse. It emerges that the bias of $\bar{U}_m$ is of a negligible order $1/n^2$ when $v = 0$. Besides ZVMI leads to a lower t-value which, in turn, also leads to a more accurate confidence interval than MI; using (6.20) instead of T_m yields lower t-values for MI. Similar results can be derived for unequal probability sampling with(out) replacement from a finite population. A somewhat counterintuitive result is that the parameter uncertainty [i.e., var(v)] in an MI procedure is the same for sampling with and without replacement. Further, deterministic ratio (regression) imputation for the HT estimator is less efficient than for the ratio (regression) estimator. In addition, adjusted AZV imputations are fully efficient so that $\mathbf{b}_{nAZV} = \mathbf{b}_p$ while the distributions of the missing data are preserved.

Interesting questions for future research are to what extent the ZVMI method can be used in the case of missing covariates, multistage sampling and generalized linear models. Another issue that needs more attention is that unlike ZVMI MI may fail when estimating a ratio for a data pattern that differs from the ratio model. This suggests that, in general, MI may not work well for parameters that are not directly related to the parameters of the superpopulation model or imputation model. The results in this paper are also relevant for external users of imputed datasets. Provided that imputed values are flagged for identification, those users or analysts can choose their own imputation method for their (regression) analysis and use ZV, AZV or ZVMI formulas proposed in this paper leading to more accurate estimates and lower t-values. Besides, after hotdeck imputation only some minor adjustments of the software are required for OLS and GLS regressions.

Appendix

A.1 Valid MI in Unequal Probability Sampling Without Replacement

For an unequal probability sample s of size n without replacement similar results can be derived as in section 2.3. Let π_i denote the first-order inclusion probability and π_{ij} the second-order inclusion probability ($i, j = 1, \dots, N$); recall $\pi_{ii} = \pi_i$. Next, defining $p_i = \pi_i/n$, $p_{ij} = \pi_{ij}/n(n-1)$ $(i \neq j)$ and $z_i = y_i/Np_i$, the HT estimator (say $\hat{\bar{Y}}^n_{HT}$) of population mean $\bar{Y}$ can be written as $\hat{\bar{Y}}^n_{HT} = \bar{z}_s = \sum_{i=1}^n y_i/N\pi_i$ and its variance is

$$\text{var}\left(\hat{\bar{Y}}^n_{HT}\right) = \sum_{i,j=1}^{N} (\pi_{ij} - \pi_i\pi_j)\frac{y_i y_j}{N^2\pi_i\pi_j}, \tag{A.1}$$

where it is assumed that the units of the population are numbered such that the first n units are sampled. Following Knottnerus (2011), (A.1) can be rewritten as

$$\text{var}\left(\hat{\bar{Y}}^n_{HT}\right) = \{1 + (n-1)\rho_{zn}\}\,\sigma^2_{zN}/n \tag{A.2}$$

$$\rho_{zn} = \sum_{i=1}^{N}\sum_{j\neq i}^{N} p_{ij}(z_i - \bar{Y})(z_j - \bar{Y})/\sigma^2_{zN}.$$

Applying the same q imputations as in (2.8), we obtain $\hat{\bar{Y}}^{imp}_{HT} = \bar{z}_n = \bar{z}_p + e$. Now using $E(s^2_{zp}) = E\{E(s^2_{zp} \mid s)\} = E(s^2_{zs}) = (1-\rho_{zn})\sigma^2_{zN}$, we get in analogy with (2.9),

$$\text{var}\left(\hat{\bar{Y}}^{imp}_{HT}\right) = \sum_{i,j=1}^{N} (\pi^*_{ij} - \pi^*_i\pi^*_j)\frac{y_i y_j}{N^2\pi^*_i\pi^*_j} + \frac{q}{np}(1-\rho_{zn})\sigma^2_{zN}, \tag{A.3}$$

where $\pi^*_i = p\pi_i/n$, $\pi^*_{ij} = p(p-1)\pi_{ij}/n(n-1)$, and $\pi^*_{ii} = \pi^*_i$; recall that missing values are random. The variance in (A.3) can be estimated in a standard manner by

$$\widehat{\text{var}}\left(\hat{\bar{Y}}^{imp}_{HT}\right) = \sum_{i,j=1}^{p}\left(1 - \frac{\pi^*_i\pi^*_j}{\pi^*_{ij}}\right)\frac{y_i y_j}{N^2\pi^*_i\pi^*_j} + \frac{q}{np}s^2_{zp}.$$

When one applies a systematic probability proportional to size (PPS) sample from a randomly ordered list, a more convenient estimator for the variance in (A.3) is

$$\widehat{\text{var}}\left(\hat{\bar{Y}}^{imp}_{PPS}\right) = \{1 + (p-1)\hat{\rho}_{zp}\}\frac{s^2_{zp}}{p} + \frac{q}{np}s^2_{zp} \tag{A.4}$$

$$\hat{\rho}_{zp} = -\frac{\sum_{i=1}^{p} p_i(z_i - \bar{z}_p)^2/\hat{\gamma}(1-2p_i)}{\sum_{i=1}^{p}(z_i - \bar{z}_p)^2} \quad \text{and} \quad \hat{\gamma} = \frac{1}{2} + \frac{1}{2p}\sum_{i=1}^{p}(1-2p_i)^{-1}.$$

For further details, see Knottnerus (2011) and the references therein. Furthermore, since

$$\begin{aligned}\text{var}\left(\hat{\bar{Y}}^p_{HT}\right) &= \text{var}\left(\bar{z}_p\right) = \text{var}\,E(\bar{z}_p \mid s) + E\,\text{var}(\bar{z}_p \mid s) \\ &= \text{var}(\bar{z}_s) + E\left(\frac{1}{p} - \frac{1}{n}\right)s^2_{zs} = \text{var}\left(\hat{\bar{Y}}^n_{HT}\right) + \frac{q}{np}(1-\rho_{zn})\sigma^2_{zN},\end{aligned}$$

we get $B \equiv \text{var}(\hat{\bar{Y}}^p_{HT}) - \text{var}(\hat{\bar{Y}}^n_{HT}) = q(1-\rho_{zn})\sigma^2_{zN}/np$ which equals $\text{var}(e) = E(B_m)$ just as with the HH estimator as we saw in section 2.3; also see (A.3). Hence, the MI procedure is valid and $\text{var}(v \mid s_r)$ or the parameter uncertainty is the same for sampling with and without replacement. Further, recall from Knottnerus (2003) that $\rho_{zn} \geq -(n-1)^{-1}$ and that for an SRS sample, $\rho_{zn} = -1/(N-1)$ for any n.

A.2 Proofs of (3.2), (3.3) and (6.12)

In order to prove (3.2) and (3.3), consider state space model ξ_N: $\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t$ $(t = 0, 1, 2, \ldots)$ where $\boldsymbol{\varepsilon}_t \sim N(0, \sigma^2\mathbf{F}_t)$, $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_s') = \mathbf{0}$ $(t \neq s)$ and $\boldsymbol{\beta}$ is a constant $k \times 1$ state vector $(\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} = \boldsymbol{\beta})$. Further, $\mathbf{y}_t$ and $\boldsymbol{\varepsilon}_t$ are vectors of the same (time-varying) order and $\mathbf{X}_t$ is a design matrix of appropriate order. Let $\mathbf{b}_t$ denote the GLS estimator for $\boldsymbol{\beta}$ given $\mathbf{y}_0, \ldots, \mathbf{y}_t$ and define $\mathbf{P}_t = \text{var}(\mathbf{b}_t)$. Since $\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} = \boldsymbol{\beta}$, the celebrated Kalman equations reduce to

$$\mathbf{b}_t = \mathbf{b}_{t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{X}_t\mathbf{b}_{t-1}) \tag{A.5}$$

$$\mathbf{P}_t = (\mathbf{I}_k - \mathbf{K}_t\mathbf{X}_t)\mathbf{P}_{t-1} \tag{A.6}$$

$$\mathbf{K}_t = \mathbf{P}_{t-1}\mathbf{X}_t'\left(\mathbf{X}_t\mathbf{P}_{t-1}\mathbf{X}_t' + \sigma^2\mathbf{F}_t\right)^{-1} \quad (t = 1, 2, \ldots) \tag{A.7}$$

$$\mathbf{b}_0 = (\mathbf{X}_0'\mathbf{F}_0^{-1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{F}_0^{-1}\mathbf{y}_0 \quad \text{and} \quad \mathbf{P}_0 = \sigma^2(\mathbf{X}_0'\mathbf{F}_0^{-1}\mathbf{X}_0)^{-1}.$$

Recall $\mathbf{P}_t = \sigma^2(\mathbf{Z}_t'\mathbf{D}_t^{-1}\mathbf{Z}_t)^{-1}$ with $\mathbf{Z}_t = (\mathbf{X}_0', \ldots, \mathbf{X}_t')'$ and $\mathbf{D}_t = \text{blkdiag}(\mathbf{F}_0, \ldots, \mathbf{F}_t)$ [e.g. Kalman (1960) and Anderson and Moore (1979, 108)]. Because the arbitrary (partitioned) dataset $(\mathbf{y}_n, \mathbf{X}_n)$ in (3.2) and (3.3) could have been generated by model ξ_N for $t \in \{0, 1\}$ with $\sigma^2 = 1$, we may choose starting values $\mathbf{b}_0 = \mathbf{b}_{p'}$ and $\mathbf{P}_0 = \mathbf{V}_{p'}$ and apply the Kalman equations for $t = 1$ with $\mathbf{y}_1 = \mathbf{y}_{q'}$, $\mathbf{X}_1 = \mathbf{X}_{q'}$ and $\mathbf{F}_1 = \mathbf{F}_{q'}$. Hence, (3.2) follows from (A.5), and (3.3) from (A.6). In addition, note that under model ξ_N the variance of $\mathbf{b}_n$ [say $\text{var}(\mathbf{b}_n; \xi_N)$] would have been equal to *both* sides of (3.3) irrespective of the actual variance of $\mathbf{b}_n$. For readers less familiar with Kalman filtering, we provide a sketch of the algebraic proofs. Applying (3.5) to $\mathbf{V}_n = (\mathbf{V}_{p'}^{-1} + \mathbf{X}_{q'}'\mathbf{F}_{q'}^{-1}\mathbf{X}_{q'})^{-1}$ gives (3.3) and substituting (3.3) into $\mathbf{b}_n = \mathbf{V}_n(\mathbf{X}_{p'}'\mathbf{F}_{p'}^{-1}\mathbf{y}_{p'} + \mathbf{X}_{q'}'\mathbf{F}_{q'}^{-1}\mathbf{y}_{q'}) - \mathbf{K}_{q'}\mathbf{e}_{q'} + \mathbf{K}_{q'}\mathbf{e}_{q'}$ and using (3.4) give (3.2).

In the remainder of this Appendix we consider again the more general (state space) model ξ in (3.1) instead of ξ_N. In order to prove (6.12), define $\mathbf{e}_t \equiv \mathbf{y}_t - \mathbf{X}_t\mathbf{b}_{t-1}$ $[= \mathbf{X}_t(\boldsymbol{\beta} - \mathbf{b}_{t-1}) + \boldsymbol{\varepsilon}_t]$. By (A.5), $\mathbf{b}_t - \mathbf{b}_{t-1} = \mathbf{K}_t\mathbf{e}_t$. In analogy with (3.12), we get $\text{var}(\mathbf{K}_t\mathbf{e}_t) = \mathbf{K}_t\mathbf{X}_t\mathbf{P}_{t-1}$ which equals $\mathbf{P}_{t-1} - \mathbf{P}_t$; see (A.6). Hence, $\text{var}(\mathbf{b}_t - \mathbf{b}_{t-1}) = \mathbf{P}_{t-1} - \mathbf{P}_t$ from which (6.12) follows. Also note that $\mathbf{b}_t - \mathbf{b}_{t-1}$ $(= \mathbf{K}_t\mathbf{e}_t)$ and $\mathbf{b}_s - \mathbf{b}_{s-1}$ $(= \mathbf{K}_s\mathbf{e}_s)$ $(s \neq t)$ are uncorrelated because the so-called innovations or prediction errors $\mathbf{e}_1, \mathbf{e}_2, \ldots$ are uncorrelated.

Furthermore, it is noteworthy that the (frequentist) Kalman equations have a Bayesian interpretation as well. That is, for a given σ^2 $(\mathbf{b}_{t-1}, \mathbf{P}_{t-1})$ can be seen as the mean and variance of the prior distribution of a random $\boldsymbol{\beta}$ while $(\mathbf{b}_t, \mathbf{P}_t)$ can be seen as the mean and variance of the posterior distribution (Harvey, 1981, 106). Moreover, following Knottnerus (1991, 70) and Knottnerus (2003, 52), the optimal Kalman gain $\mathbf{K}_t$ can be seen as a matrix of regression coefficients from a regression of k prior errors on the prediction error $\mathbf{e}_t$ resulting in k posterior errors. That is, in the present context,

$$\boldsymbol{\beta} - \mathbf{b}_{t-1} = \mathbf{K}_t\mathbf{e}_t + \boldsymbol{\beta} - \mathbf{b}_t, \tag{A.8}$$

which is identical to (A.5). Solving the normal equations $E\{(\boldsymbol{\beta} - \mathbf{b}_t)\mathbf{e}_t'\} = \mathbf{0}$ or, equivalently, $E\{(\boldsymbol{\beta} - \mathbf{b}_{t-1} - \mathbf{K}_t\mathbf{e}_t)\mathbf{e}_t'\} = \mathbf{0}$ yields (A.7); recall $\mathbf{e}_t = \mathbf{X}_t(\boldsymbol{\beta} - \mathbf{b}_{t-1}) + \boldsymbol{\varepsilon}_t$. In other words, the residual variance of $\mathbf{b}_t$ $[= \text{tr}\{\text{var}(\mathbf{b}_t)\}]$ in (A.8) attains its minimum for $\mathbf{K}_t$ given in (A.7). Applying Pythagoras's Theorem to regression (A.8), we get

$$\mathbf{P}_{t-1} = \text{var}(\mathbf{K}_t\mathbf{e}_t) + \mathbf{P}_t, \tag{A.9}$$

which is identical to (A.6); recall $\operatorname{var}(\mathbf{K}_t\mathbf{e}_t) = \mathbf{K}_t\mathbf{X}_t\mathbf{P}_{t-1}$ and $\boldsymbol{\beta} - \mathbf{b}_t$ is orthogonal to $\mathbf{e}_t$.

A.3 Additional Proofs for Theorem 3.1

We start proving $E(s_{en}^2) = \sigma^2$. Assume without loss of generality $f_i = 1$. Write $\mathbf{e}_n$ as

$$\begin{aligned}\mathbf{e}_n &= \mathbf{y}_n - \mathbf{X}_n\mathbf{b}_n = \mathbf{y}_n - \mathbf{X}_n\mathbf{b}_p - \mathbf{X}_n\left(\mathbf{b}_n - \mathbf{b}_p\right)\\ &= \begin{pmatrix}\mathbf{e}_p\\0\end{pmatrix} + \begin{pmatrix}-\mathbf{X}_p\mathbf{K}_q\mathbf{e}_q\\(\mathbf{I}_q - \mathbf{X}_q\mathbf{K}_q)\mathbf{e}_q\end{pmatrix},\end{aligned} \tag{A.10}$$

where we used (3.8). Since $\mathbf{e}_n'\mathbf{e}_n = \operatorname{tr}(\mathbf{e}_n\mathbf{e}_n')$, three components are to be examined. First,

$$E(\mathbf{e}_p'\mathbf{e}_p) = (p-k)\sigma^2. \tag{A.11}$$

Second, using (3.12), we get

$$\operatorname{tr}\left\{\operatorname{var}(\mathbf{X}_p\mathbf{K}_q\mathbf{e}_q)\right\} = \operatorname{tr}\left(\mathbf{X}_p\mathbf{K}_q\mathbf{X}_q\mathbf{V}_p\mathbf{X}_p'\right)\sigma^2 = \operatorname{tr}\left(\mathbf{K}_q\mathbf{X}_q\right)\sigma^2, \tag{A.12}$$

where we used $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$. Third,

$$\operatorname{tr} E\left\{(\mathbf{I}_q - \mathbf{X}_q\mathbf{K}_q)\mathbf{e}_q\mathbf{e}_q'(\mathbf{I}_q - \mathbf{X}_q\mathbf{K}_q)'\right\} = \operatorname{tr}\left(\mathbf{I}_q - \mathbf{X}_q\mathbf{K}_q\right)'\sigma^2. \tag{A.13}$$

Note $E\{(\mathbf{I}_q - \mathbf{X}_q\mathbf{K}_q)\mathbf{e}_q\mathbf{e}_q'\} = \sigma^2\mathbf{I}_q$ because $E(\mathbf{e}_q\mathbf{e}_q') = \sigma^2(\mathbf{X}_q\mathbf{V}_p\mathbf{X}_q' + \mathbf{I}_q)$ and hence by (3.9), $E(\mathbf{K}_q\mathbf{e}_q\mathbf{e}_q') = \sigma^2\mathbf{V}_p\mathbf{X}_q'$; recall $\mathbf{F}_q = \mathbf{I}_q$. Combining (A.10)-(A.13) yields $E(\mathbf{e}_n'\mathbf{e}_n) = (n-k)\sigma^2$ and hence, R_2 is met; recall $\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{A}')$. This concludes the proof.

Next, we prove $E(s_{en}^2) = \sigma^2\{1 + O(1/n)\}$ when $\mathbf{v}$ is set to zero. Define $\boldsymbol{\gamma} = (\mathbf{e}_p', \mathbf{u}_q')'$ and $\boldsymbol{\delta} = \mathbf{X}_n\mathbf{K}_q\mathbf{u}_q$. Setting $\mathbf{v} = \mathbf{0}$, we get $\mathbf{e}_q = \mathbf{u}_q$ and it is seen from (A.10) that $\mathbf{e}_n = \boldsymbol{\gamma} - \boldsymbol{\delta}$. Hence, $\mathbf{e}_n'\mathbf{e}_n = \boldsymbol{\gamma}'\boldsymbol{\gamma} + \boldsymbol{\delta}'\boldsymbol{\delta} - 2\boldsymbol{\gamma}'\boldsymbol{\delta}$. Define $\boldsymbol{\gamma}_2 = \mathbf{u}_q$ and $\boldsymbol{\delta}_2 = \mathbf{X}_q\mathbf{K}_q\mathbf{u}_q$. Subsequently, using $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$ and $\mathbf{K}_q = \mathbf{V}_n\mathbf{X}_q'$ [see (4.6b)], we get

$$\begin{aligned}E(\boldsymbol{\gamma}'\boldsymbol{\gamma}) &= E\{(p-k)s_{ep}^2\} + E(\mathbf{u}_q'\mathbf{u}_q) = (n-k)\sigma^2\\ E(\boldsymbol{\delta}'\boldsymbol{\delta}) &= \operatorname{tr} E(\mathbf{u}_q'\mathbf{X}_q\mathbf{V}_n\mathbf{X}_n'\mathbf{X}_n\mathbf{V}_n\mathbf{X}_q'\mathbf{u}_q) = \sigma^2\operatorname{tr}\mathbf{V}_n\mathbf{V}_q^{-1}\\ E(\boldsymbol{\gamma}'\boldsymbol{\delta}) &= E(\boldsymbol{\gamma}_2'\boldsymbol{\delta}_2) = \operatorname{tr} E(\mathbf{u}_q'\mathbf{X}_q\mathbf{V}_n\mathbf{X}_q'\mathbf{u}_q) = \sigma^2\operatorname{tr}\mathbf{V}_n\mathbf{V}_q^{-1}.\end{aligned}$$

Hence, $E(s_{en}^2) = \sigma^2\{1 + O(1/n)\}$; note $\mathbf{V}_n\mathbf{V}_q^{-1} = O(q/n)$ under the assumptions.

References

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. New York: Prentice-Hall.

Chambers, R. L. and R. G. Clark (2012). *An Introduction to Model-Based Survey Sampling with Applications*. Oxford: Oxford University Press.

Chen, S. and D. Haziza (2019). Recent Developments in Dealing with Item Non-Response in Surveys: A Critical Review. *International Statistical Review 87*, S192–S218.

Cochran, W. G. (1977). *Sampling Techniques*. New York: John Wiley & Sons.

De Waal, T., J. Pannekoek, and S. Scholtus (2011). *Handbook of Statistical Data Editing and Imputation*. New York: John Wiley & Sons.

Deville, J. C. and C. E. Särndal (1994). Variance Estimation for the Regression Imputed Horvitz-Thompson Estimator. *Journal of Official Statistics 10*, 381–394.

Enders, C. K. (2010). *Applied Missing Data Analysis*. New York: Guilford Press.

Greene, W. H. (2003). *Econometric Analysis*. New York: Prentice-Hall.

Harvey, A. C. (1981). *Time Series Models*. London: Philip Allan.

Harvey, A. C. (1990). *The Econometric Analysis of Time Series*. London: Philip Allan.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions ASME, Journal of Basic Engineering 82*, 35–45.

Kim, J. K. (2004). Finite Sample Properties of Multiple Imputation Estimators. *Annals of Statistics 32*, 766–783.

Kish, L. (1992). Weighting for Unequal P_i. *Journal of Official Statistics 8*, 183–200.

Knottnerus, P. (1991). *Linear Models with Correlated Disturbances*. New York: Springer.

Knottnerus, P. (2003). *Sample Survey Theory: Some Pythagorean Perspectives*. New York: Springer.

Knottnerus, P. (2011). On the Efficiency of Randomized Probability Proportional to Size Sampling. *Survey Methodology 37*, 95–102.

Little, R. J. A. (1992). Regression with Missing X's: A Review. *Journal of the American Statistical Association 87*, 1227–1237.

Little, R. J. A. (2011). Calibrated Bayes, for Statistics in General, and Missing Data in Particular. *Statistical Science 26*, 162–174.

Meng, X.-L. (1994). Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science 9*, 538–558.

Murray, J. S. (2018). Multiple Imputation: A Review of Practical and Theoretical Findings. *Statistical Science 33*, 142–159.

Plackett, R. L. (1950). Some Theorems in Least Squares. *Biometrika 37*, 149–157.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Rubin, D. B. (1996). Multiple Imputation after 18+ Years. *Journal of the American Statistical Association 91*, 473–489.

Rubin, D. B. (2003). Nested Multiple Imputation of NMES via Partially Incompatible MCMC. *Statistica Neerlandica 57*, 3–18.

Rubin, D. B. and N. Schenker (1986). Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse. *Journal of the American Statistical Association 81*, 366–374.

Schenker, N. and A. H. Welsh (1988). Asymptotic Results for Multiple Imputation. *Annals of Statistics 16*, 1550–1566.

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., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer.

Van Buuren, S. (2018). *Flexible Imputation of Missing Data*. Boca Raton: Chapman & Hall.

Wang, N. and J. M. Robins (1998). Large-Sample Theory for Parametric Multiple Imputation Procedures. *Biometrika 85*, 935–948.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information