

### Applying the rescaling bootstrap under imputation for a multistage sampling design

Bruch, Christian

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Bruch, C. (2022). Applying the rescaling bootstrap under imputation for a multistage sampling design. *Computational Statistics*, 37(3), 1461-1494. <https://doi.org/10.1007/s00180-021-01164-6>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:  
<https://creativecommons.org/licenses/by/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:  
<https://creativecommons.org/licenses/by/4.0>



# Applying the rescaling bootstrap under imputation for a multistage sampling design

Christian Bruch<sup>1</sup>

Received: 15 February 2021 / Accepted: 5 October 2021 / Published online: 18 November 2021  
© The Author(s) 2021

## Abstract

In this paper, we propose a method that estimates the variance of an imputed estimator in a multistage sampling design. The method is based on the rescaling bootstrap for multistage sampling introduced by Preston (Surv Methodol 35(2):227–234, 2009). In his original version, this resampling method requires that the dataset includes only complete cases and no missing values. Thus, we propose two modifications for applying this method to nonresponse and imputation. These modifications are compared to other modifications in a Monte Carlo simulation study. The results of our simulation study show that our two proposed approaches are superior to the other modifications of the rescaling bootstrap and, in many situations, produce valid estimators for the variance of the imputed estimator in multistage sampling designs.

**Keywords** Rescaling bootstrap · Variance estimation · Nonresponse · Imputation · Multistage sampling designs · Monte Carlo simulation study

## 1 Introduction

The estimation of standard errors is an important topic for social surveys and official statistics. Standard errors are used to evaluate the quality of point or parameter estimations, for example, in confidence intervals or hypothesis testing. The basis of a valid standard error estimation is an accurate variance estimation. This estimation may be highly challenging to achieve, for example, due to a complex sampling design, such as a multistage sampling, the occurrence of nonresponse, and the use of imputation methods to compensate for nonresponse. In practice, it may be very difficult or even impossible to consider all these influences in the variance estimation. Thus, frequently,

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00180-021-01164-6>.

✉ Christian Bruch  
Christian.Bruch@gesis.org

<sup>1</sup> GESIS Leibniz Institute for the Social Science, B6, 4-5, 68159 Mannheim, Germany

not all stages of the sampling design are considered, or the variance estimator for simple random sampling is used even when another sampling design has been applied. Furthermore, imputed values often are treated as observed, and the nonresponse as well as the imputation process are ignored in the variance estimation. Such procedures may lead to a heavily biased variance estimation that results in a biased standard error estimation. Hence, conclusions, e.g., from hypothesis testing and confidence intervals may be wrong. Thus, the aim of this paper is to present an approach to a variance estimation that includes all stages of a sampling design and that considers the imputation process.

## 1.1 Previous studies and research

There exists some previous studies in the literature that cover the variance estimation under imputation or the variance estimation in a multistage sampling design. Some of the studies also cover both issues.

Frequently, for both issues so called resampling methods (see, for example, Shao and Tu 1995) are applied, particularly, bootstrap procedures (Efron 1979). A good overview of bootstrap procedures is given in Mashreghi et al. (2016), in which procedures for multistage sampling or imputation methods are also described.

The basics for considering the imputation process in the bootstrap procedures are presented in Shao (2002), Shao and Sitter (1996) and Efron (1994). Further proposals of a bootstrap under imputation are, for example, presented in Mashreghi et al. (2014) or Saigo et al. (2001). Additional explanations of the application of bootstrap procedures under imputation are given in Haziza (2009, 2010). Moya et al. (2020) examine a rescaled bootstrap for confidence intervals in the presence of spikes in the distribution that may also be a result of certain imputation procedures. In this paper, we focus on single imputation procedures. In case of multiple imputation, two bayesian bootstrap procedures for multistage sampling designs are, for example, proposed in Zhou et al. (2016).

A bootstrap procedure for multistage sampling designs for complete observations that does not cover imputation is proposed in Rao et al. (1992) and that is an extension of the bootstrap procedure proposed in Rao and Wu (1988). Also, Saigo (2010) compares four different bootstrap procedures for stratified multistage sampling designs and complete data sets in a simulation study. Particularly, he considers the Bernoulli bootstrap for stratified multistage designs of Funaoka et al. (2006) and extensions of the bootstrap procedures described in Rao et al. (1992), Sitter (1992a) and Sitter (1992b). Further proposals of bootstrap procedures for multistage designs are presented in Haziza (2009), Wolter (2007) and Preston (2009). The latter one is the resampling method of interest in this paper.

To estimate the variance of an imputed estimator in a multistage design, we combine the rescaling bootstrap for multistage design of Preston (2009) with procedures that allows for a consideration of the imputation procedure in the resampling method. We choose the rescaling bootstrap of Preston (2009), since it allows for the inclusion of all stages of the sampling design in the variance estimation. Furthermore, the rescaling bootstrap considers finite population corrections in the weight adjustment that needs

to be included in the variance estimation for the important case of without replacement random sampling (at the different stages of the sampling design). Thus, in this paper, we propose modifications of the rescaling bootstrap for multistage design of Preston (2009) to obtain a valid variance estimator in a multistage without replacement sampling design and under single imputation.<sup>1</sup> We present the rescaling bootstrap of Preston (2009) in the next section.

## 1.2 Rescaling bootstrap for a three-stage sampling design and complete response

In multistage sampling designs in practice, the variance estimators often consider only the first stage, since the first stage variance estimator covers more than the actual variance of the point estimator at stage one (Särndal et al. 1992; Lohr 1999; Bruch et al. 2011; Bruch 2016). Under some parameter constellations, such as a small sampling fraction at the first stage or a homogeneous composition of sampling elements within the units of the first stage (primary sampling units; PSUs)<sup>2</sup> it is even possible to cover (almost) the entire variance by using only the variance estimator of the first stage. However, as shown by Särndal et al. (1992), Lohr (1999), Bruch et al. (2011); Bruch (2016), the variance may be heavily underestimated when only the variance estimator of the first stage is applied under parameter constellations, such as an increasing sampling fraction at the first stage or a heterogeneous composition of sampling elements within the PSU. Using such a variance estimate, for example, as a basis to compute a standard error estimate for hypothesis testing or confidence intervals, may lead to false conclusions. Thus, in some situations, it is important that the variance estimator includes all stages of the sampling design. A resampling method that takes all the sampling stages into account in the variance estimation is the multistage rescaling bootstrap proposed by Preston (2009), which is an extension of the rescaling bootstrap of Chipperfield and Preston (2007) for multistage sampling designs.<sup>3</sup> For this resampling method, subsamples are drawn independently without replacement at each stage of the sampling design. The rescaling bootstrap is applied in the case of a three-stage sampling design as follows [for the following explanations see Preston (2009)].

First, a subsample of size  $l^* = \lfloor l/2 \rfloor$  is drawn from the  $l$  PSUs in the sample at the first stage. The indicator variable  $\delta_d$  takes the value 1 if PSU  $d$  is in the subsample, and zero if not. Afterwards, new design weights  $w_{1d}^*$  for the original design weights  $w_{1d} = L/l$  ( $L$  is the number of PSUs in the population) are computed at the first stage by:

$$w_{1d}^* = \left( 1 - \lambda + \lambda \cdot \frac{l}{l^*} \cdot \delta_d \right) \cdot w_{1d}, \quad \text{where } \lambda = \sqrt{l^* \cdot \frac{(1 - f_1)}{(l - l^*)}}, \quad (1)$$

<sup>1</sup> In the following we use the term imputation for single imputation since multiple imputation procedures are not covered in this study.

<sup>2</sup> When the sampling elements in the PSU are similar with respect to the variable of interest  $y$ , a large part of the variance of the point estimator results from the first-stage sampling. In contrast, when sampling elements are very different regarding  $y$  in the PSUs, most of the variance results from the sampling at subsequent stages.

<sup>3</sup> For simplification, we use the term *rescaling bootstrap* for both variance estimators independent of the sampling design.

where  $f_1$  is the sampling fraction at the first stage.

A subsample of size  $m_d^* = \lfloor m_d/2 \rfloor$  is drawn from the  $m_d$  secondary sampling units (SSUs) within each PSU in the original sample at the second stage. The indicator variable  $\delta_{dq}$  takes the value 1 if SSU  $q$  in PSU  $d$  is in the subsample, and zero if not. Afterwards, new design weights  $w_{2dq}^*$  for the original design weights  $w_{2dq} = M_d/m_d$  ( $M_d$  is the number of SSUs in the PSU  $d$  in the population) are computed at the second stage by:

$$w_{2dq}^* = w_{2dq} \cdot \frac{w_{1d}}{w_{1d}^*} \cdot \left(1 - \lambda + \lambda \cdot \frac{l}{l^*} \cdot \delta_d - \lambda_d \cdot \sqrt{\frac{l}{l^*}} \cdot \delta_d + \lambda_d \cdot \sqrt{\frac{l}{l^*}} \cdot \delta_d \cdot \frac{m_d}{m_d^*} \cdot \delta_{dq}\right), \text{ where } \lambda_d = \sqrt{m_d^* \cdot f_1 \cdot \frac{(1 - f_{2d})}{(m_d - m_d^*)}}, \quad (2)$$

where  $f_{2d}$  is the sampling fraction in a PSU at the second stage.

A subsample of size  $n_{dq}^* = \lfloor n_{dq}/2 \rfloor$  is drawn from the  $n_{dq}$  ultimate sampling units (USUs) within each SSU in the original sample at the third stage. The indicator variable  $\delta_{dqi}$  takes the value 1 if USU  $i$  in SSU  $q$  is in the subsample, and zero if not. Afterwards, new design weights  $w_{3dqi}^*$  for the original design weights  $w_{3dqi} = N_{dq}/n_{dq}$  ( $N_{dq}$  is the number of USUs in SSU  $q$  in PSU  $d$  in the population) are computed at the third stage by:

$$w_{3dqi}^* = w_{3dqi} \cdot \frac{w_{1d}}{w_{1d}^*} \cdot \frac{w_{2dq}}{w_{2dq}^*} \cdot \left(1 - \lambda + \lambda \cdot \frac{l}{l^*} \cdot \delta_d - \lambda_d \cdot \sqrt{\frac{l}{l^*}} \cdot \delta_d + \lambda_d \cdot \sqrt{\frac{l}{l^*}} \cdot \delta_d \cdot \frac{m_d}{m_d^*} \cdot \delta_{dq} - \lambda_{dq} \cdot \sqrt{\frac{l}{l^*}} \cdot \delta_d \cdot \sqrt{\frac{m_d}{m_d^*}} \cdot \delta_{dq} + \lambda_{dq} \cdot \sqrt{\frac{l}{l^*}} \cdot \delta_d \cdot \sqrt{\frac{m_d}{m_d^*}} \cdot \delta_{dq} \cdot \frac{n_{dq}}{n_{dq}^*} \cdot \delta_{dqi}\right),$$

$$\text{where } \lambda_{dq} = \sqrt{n_{dq}^* \cdot f_1 \cdot f_{2d} \cdot \frac{(1 - f_{3dq})}{(n_{dq} - n_{dq}^*)}}, \quad (3)$$

where  $f_{3dq}$  is the sampling fraction in a SSU at the third stage.

After computing the adjusted design weights at each stage, the overall adjusted design weights  $w_{dqi}^*$  are computed by

$$w_{dqi}^* = w_{1d}^* \cdot w_{2dq}^* \cdot w_{3dqi}^*, \quad (4)$$

combined to vector  $\mathbf{w}^*$ . The point estimator of the rescaling bootstrap in each simulation run  $b$  is computed by:

$$\hat{\theta}^* = T(\mathbf{y}_s, \mathbf{w}^*), \quad (5)$$

where  $T$  is the function to compute the statistic of interest and  $\mathbf{y}_s$  is the vector with sample realisations.

The process beginning with the subsampling is repeated  $B$  times and the point estimator of the rescaling bootstrap is computed in each run  $b$ . Thus, it results the estimators  $\hat{\theta}_1^* \dots \hat{\theta}_b^* \dots \hat{\theta}_B^*$ , and the variance estimate is obtained by:

$$\hat{V}_{resc}(\hat{\theta}) = \frac{1}{B-1} \cdot \sum_{b=1}^B \left( \hat{\theta}_b^* - \frac{1}{B} \cdot \sum_{v=1}^B \hat{\theta}_v^* \right)^2. \quad (6)$$

The procedure shows that the rescaling bootstrap considers all elements of the original sample in the computation of the rescaling point estimator. This procedure is hugely different from classical resampling methods, in which only elements drawn for the subsample are considered in the computation of the point estimator of the resampling method. However, weight adjustments for the rescaling bootstrap are performed with respect to the underlying subsample. Hence, it can be seen if USUs as well as their superordinate units are drawn for the subsample, the influence of these USUs on the variance estimation increases due to the weight adjustments compared to the elements that are not drawn.

Nonetheless, elements not drawn for the subsample can also have an influence on the variance estimation. This finding is, for example, shown in the simulation study by Bruch (2019) with respect to a single-stage design whereby elements not drawn for the subsample have a larger influence in the scenarios with larger sampling fractions than in the scenarios with smaller sampling fractions.

For the rescaling bootstrap in a multistage sampling design, influences of elements not drawn for the subsample on the variance estimation can also be larger or smaller, depending on the sampling fractions at the several sampling stages. However, this relationship is not as straightforward as for the single stage design. Elements can be drawn or not drawn at several stages and more parameters, such as the sampling fractions at a higher stage, also have an effect on the adjusted weights at certain stages in a multistage design as illustrated in (1), (2), and (3). The influence of elements drawn for the subsample and not drawn for the subsample on the variance estimation using the rescaling bootstrap is an important point. Thus, we examine this influence in more detail in the Appendix A.1 and the simulation study.

Due to subsampling at each stage of the sampling design and adjusting the corresponding weights, the rescaling bootstrap of Preston (2009) considers all stages of the sampling design in the variance estimation. However, the rescaling bootstrap of Preston (2009) is based on complete observations. Thus, the method has to be modified when the dataset consists of missing values, and imputation is applied as compensation. In the next section, we describe the variance decomposition of an imputed

estimator and its implication for a multistage design that builds the starting point to include the imputation process in the rescaling bootstrap of Preston (2009).

### 1.3 Variance under imputation and nonresponse

For reasons of clarity, we first assume a deterministic imputation procedure to illustrate the proposed modifications in their fundamental form.<sup>4</sup> In general, in the presence of nonresponse and when a deterministic imputation procedure is applied as compensation for the nonresponse, the variance of an imputed estimator  $\hat{\theta}_I$  can be decomposed as follows [Mashreghi et al. (2014), see also the reversed framework of Fay (1991) and Shao and Steel (1999)]:

$$V(\hat{\theta}_I) = \underbrace{E_\eta V_s(\hat{\theta}_I|\mathbf{z})}_{V_1} + \underbrace{V_\eta E_s(\hat{\theta}_I|\mathbf{z})}_{V_2}. \quad (7)$$

The expected value and the variance according to the nonresponse mechanism are indicated by the subscript  $\eta$ . The expected value and the variance with respect to the sampling are indicated by the subscript  $s$ .

Thus, the variance component  $V_1$  includes  $V_s(\hat{\theta}_I|\mathbf{z})$ , which is the sampling variance of the imputed estimator  $\hat{\theta}_I$ , given the response vector  $\mathbf{z}$ . According to Haziza (2010), calculating  $V_s(\hat{\theta}_I|\mathbf{z})$  is enough to estimate the component  $V_1$ . Variance component  $V_2$  results from the treatment of the nonresponse mechanism as a random process. The variance component arises in a simulation study when a new response vector is generated in each simulation run (which we call in the following sections a *varying* response vector) by using a random model to reproduce the nonresponse mechanism. The variance component does not arise when the same response vector is used in each simulation run. Such a response vector may be called *fixed* (Mashreghi et al. 2014; Bruch 2019).

Referring to Mashreghi et al. (2014), the variance component  $V_2$  is only significant if the overall sampling fraction  $f$  is large. If the overall sampling fraction is negligible, the variance component  $V_2$  also is negligible, and it is sufficient to estimate the variance component  $V_1$  to cover the whole variance of the imputed estimator.

When applying a multistage sampling design, the variance component  $V_2$  often may be small or negligible. High survey costs and low survey budgets prevent a large overall sample size in relation to the population size. Multistage sampling designs are often used with respect to population surveys, for example, when the ultimate sampling units such as households or persons cannot drawn directly (i.e. when lists such as central population registers are not available, see Lynn et al. 2007; Häder 2004), when other constraints of the sampling frames such as the assignment of interviewers (European Social Survey 2018) are presented or due to geographic decisions with regard to cost-effectiveness are made (comparatively lower costs of sampling superordinate units, Lohr 1999; European Social Survey 2018). Thus, a comparatively small sample of USUs is faced with a large population. As a result, the overall sample fraction  $f$  may

<sup>4</sup> However, in the second part of the simulation study, the proposed modifications are also evaluated with respect to a stochastic imputation procedure since such imputation methods are often applied in practice.

be rather small whereas this does not preclude that the sampling fraction is not large at a particular sampling stage. The sampling fractions at the different stages have an important influence on the variance estimation via the rescaling of Preston (2009), as seen in the previous section.

However, since the overall fraction may often be rather small in multistage sampling designs in practice, it may therefore often be enough to derive a valid variance estimator of  $V_1$  to obtain a valid variance estimator for the whole variance  $V(\hat{\theta}_I)$ . As a result, this component is of central interest in this paper. However, in the simulation study in Sect. 4, we also examine to what extent it is sufficient to estimate  $V_1$  to cover the whole variance or whether any parameter constellations exist that additionally require the estimation of  $V_2$ .

### 1.3.1 Reimputation of imputed values

With respect to the explanations of the previous section, deriving a variance estimator for imputed estimator  $\hat{\theta}_I$  in a multistage sampling design especially requires a valid variance estimator for variance component  $V_1$  and it is enough to estimate  $V_s(\hat{\theta}_I|\mathbf{z})$  to cover variance component  $V_1$  (Mashreghi et al. 2014; Haziza 2010). Thus, we intend to find a variance estimator that considers the sampling variability of the imputed estimator, given the response vector  $\mathbf{z}$ . By doing so, we receive a valid variance estimator for the first term  $V_1$  in (7) (Mashreghi et al. 2014).

With respect to Mashreghi et al. (2014), one possibility for deriving a valid variance estimator for term  $V_s(\hat{\theta}_I|\mathbf{z})$  is to apply bootstrap procedures by taking paired subsamples from the elements in  $\mathbf{y}_I$  and their corresponding response status  $\mathbf{z}$  of the original sample to which the so-called *reimputation* (Shao 2002; Shao and Sitter 1996; Rao and Shao 1992) is applied. This approach considers the imputation process in a resampling method by reimputing the imputed values that are drawn for the subsample on the basis of the subsample's observed values using the same imputation procedure—especially the same imputation model—as in the original sample. The reimputation of imputed values is appropriate for deterministic imputation procedures. However, it may lead to an overestimation of the variance if the imputation procedure includes random components, and if the subsample size is not the same size as the original sample (Shao 2002). In such cases, rather the adjustment of imputed values as described in Shao (2002) should be applied but this procedure requires a metric variable to be imputed.<sup>5</sup> For other imputation methods that do not allow the application of the reimputation or the adjustment of imputed values, other adjustments may be applied. However, in the following, we assume that an imputation method is applied that allows for a consideration of the imputation process in the rescaling bootstrap via reimputation or the adjustment of imputed values.

It is also possible to use the rescaling bootstrap of Chipperfield and Preston (2007) with the extension to multistage sampling designs proposed by Preston (2009) and to make use of subsampling process of the rescaling bootstrap to derive a variance estimator for  $V_s(\hat{\theta}_I|\mathbf{z})$ . The important concern is how to apply the reimputation for the rescaling bootstrap. Usually, studies using the reimputation of imputed values,

<sup>5</sup> The computation of differences has to be possible.



such as Shao and Sitter (1996), focus on classical resampling methods, in which only elements drawn for the subsample are considered in the computation of the point estimator of the resampling method. However, as described in the Sect. 1.2, the rescaling bootstrap includes all elements of the original sample in the computation of its point estimators. This is an important point, particularly, when elements not drawn for the subsample affect the variance estimation due to their influence on the rescaling bootstrap point estimators. Thus, all elements of the original sample need to be considered in the reimpuation process (Bruch 2016, 2019). In the next section, we describe a modification of the rescaling bootstrap that includes the reimpuation process for a single stage design and that can be used as a basis to modify the rescaling bootstrap in a multistage sampling with respect to imputation.

### 1.3.2 Rescaling bootstrap in a single-stage sampling design and imputed values

A modification of the rescaling bootstrap with respect to imputed values for the single-stage sampling design simple random sample is proposed by Bruch (2016, 2019). To derive this modification, different possibilities to apply the reimpuation are evaluated that result in different requirements that have to be met.

One possibility is to conduct the reimpuation regarding all the imputed values of the whole sample, without distinguishing between the imputed values drawn for the subsample and the imputed values not drawn for the subsample. By doing so, all the elements of the original sample will be considered in the reimpuation process. However, in the case of a deterministic imputation procedure and when the reimpuation is applied on the basis of the whole sample, the reimpuated values are identical to the imputed values of the original sample. Thus, no variation can occur due to the reimpuation process when computing the variance over all rescaling bootstrap point estimators (Bruch 2016). In this case, the estimator matches the variance estimator that treats the imputed values as actually observed. With respect to Shao and Sitter (1996), since this estimator does not consider the effect of missing values and imputation, the variance may be heavily underestimated (Bruch 2019).

Thus, Bruch (2016, 2019) rather proposes to reimpuate the imputed values of the original sample drawn for the subsample and the imputed values of the original sample not drawn for the subsample separately. The necessity for such a procedure results from the comparison with the rescaling bootstrap for which the reimpuation is done only for the imputed values drawn for the subsample and not for the imputed values not drawn for the subsample. Bruch (2016, 2019) shows that neglecting the reimpuation of the imputed values not drawn for the subsample leads to an unintentional variance component, which results in the overestimation of the variance. The unintentional variance only occurs for increasing sampling fractions in a single-stage sampling design, since elements not drawn for the subsample only have a small impact when sampling fractions are small. We give a more detailed explanation for the appearance of this component in Appendix A.2.

However, the procedure of reimpuating separately imputed values drawn and not drawn for the subsample only pertains to a single-stage sampling design. The question is how to apply the method in a multistage sampling design, when more stages of the sampling design must be considered. Thus, in the next section, we propose two mod-

ifications of the rescaling bootstrap to estimate the variance of an imputed estimator in a multistage design. We evaluate these modifications in a Monte Carlo simulation study. We describe this study in Sect. 3, and present the results in Sect. 4. We discuss and summarize the results in Sect. 5 and provide an outlook for future research.

## 2 Rescaling bootstrap under nonresponse and imputation

The fundamentals of these approaches and this research are developed in the PhD thesis of Bruch (2016) with respect to a two-stage sampling design. However, in practice, multistage sampling designs often include more than two stages. Thus, in this paper, we apply the modifications in more realistic situations with more stages. The proposed methods are basically described for three-stage sampling designs, but they can be modified for sampling designs with more than three stages.

In the next section, we describe the sampling design and the estimator to derive the proposed modification and that build the basis for the Monte Carlo simulation study. Afterwards, we derive the proposed modifications as a combination of the rescaling bootstrap of Preston (2009) and variance estimation methods that consider the imputation. A further derivation and explanation of our proposed modifications to be a valid variance estimator, particularly under which conditions, is given in the Appendix A.3. Our modifications aim to estimate the variance component  $V_s(\hat{\theta}_I|\mathbf{z})$  and thus aim to cover variance component  $V_1$  in (7). As written in Sect. 1.3, it may often be enough to estimate only variance component  $V_1$  in practice when multistage sampling designs are applied. However, in Sect. 2.3 we discuss shortly the case when  $V_2$  has a strong contribution.

### 2.1 Imputed estimation for a multistage sampling design

To derive our modifications of the rescaling bootstrap and for their evaluation in the simulation study, we assume that a sample is drawn from the population  $U$  using a multistage sampling design with three stages. We use the notation introduced in Sect. 1. The sample sizes at the different stages are described by  $l$ ,  $m_d$  and  $n_{dq}$ . The corresponding population sizes are indicated by  $L$ ,  $M_d$  and  $N_{dq}$ . Overall, the population consists of  $N$  elements, and the complete sample  $S$  is size  $n$ .

The complete design weights  $w_{dqi}$  are computed by:

$$w_{dqi} = w_{1d} \cdot w_{2dq} \cdot w_{3dqi} = \frac{L}{l} \cdot \frac{M_d}{m_d} \cdot \frac{N_{dq}}{n_{dq}}, \quad (8)$$

which are combined to the vector  $\mathbf{w}$ .

The sampling fraction  $f_1$  at the first stage, the sampling fraction  $f_{2d}$  in a PSU at the second stage, and the sampling fraction  $f_{3dq}$  in a SSU at the third stage are defined by  $f_1 = l/L$ ,  $f_{2d} = m_d/M_d$  and  $f_{3dq} = n_{dq}/N_{dq}$ . For simplicity, in the first part of the simulation study in Sect. 3.1.1, we assume that the sampling fraction  $f_{2d}$  is equal in each PSU with  $f_{2d} = f_2$ , and the sampling fraction  $f_{3dq}$  is equal in each SSU

with  $f_{3dq} = f_3$ . Thus, we can use  $f_2$  and  $f_3$  to indicate unambiguously the sampling fraction at the corresponding stage and to define unambiguously the scenarios in the first part of the simulation study. However, with respect to the proposed methods, it is not necessary that the sampling fractions  $f_{2d}$  in each PSU and  $f_{3dq}$  in each SSU are all equal. Thus, the proposed methods can be applied in the same way for different sampling fractions  $f_{2d}$  in the PSUs and different sampling fractions  $f_{3dq}$  in the SSUs. In the second part of the simulation study in Sect. 3.2, the sampling fractions  $f_{2d}$  and  $f_{3dq}$  are not equal.

The variable of interest is described by  $y$  with sample realisation  $y_i$  for element  $i$  with  $i = 1 \dots n$  and that are combined to vector  $\mathbf{y}_s$ .

We assume that this variable consists of missing values that only appear at the last stage with respect to certain USUs. The missing values divide the sample  $S$  into a subset of elements with observed values  $i \in R$  and a subset of elements with missing values defined by  $i \in G$  with  $S = R \dot{\cup} G$ . The size of  $R$  is  $r$ , and the size of  $G$  is  $g$ , where  $n = r + g$ . We assume that the dataset contains a response indicator  $\mathbf{z}$ , which takes the value 1 if an element responds and 0 if not. We use a certain single imputation method to impute missing values, and our imputation model includes two auxiliary variables  $x_1$  and  $x_2$  with realisations  $x_{1i}$  and  $x_{2i}$  for element  $i$ . We assume that the auxiliary variables have no missing values. After the imputation of sample nonresponse, it results the vector  $\mathbf{y}_I$  with respect to the variable of interest  $y$  that contains imputed as well as observed values. In detail, the values  $y_{I,i}$  of  $\mathbf{y}_I$  have the following manifestations:

$$y_{I,i} := \begin{cases} y_i, & \text{if } i \in R, \\ \tilde{y}_i, & \text{if } i \in G, \end{cases}$$

where  $\tilde{y}_i$  is the imputed value for a missing value  $\forall i \in G$ .

The statistic of interest is indicated by  $\theta$ . This quantity is estimated by using the estimator  $\hat{\theta}_I$  on the basis of imputed vector  $\mathbf{y}_I$  and including the design weights  $w_{dqi}$  from (8) with vector  $\mathbf{w}$ :

$$\hat{\theta}_I = T(\mathbf{y}_I, \mathbf{w}),$$

where  $T$  is the function for the computation of the statistic of interest.

In this paper, we focus on the total value  $\theta = \tau$ . The total value  $\tau$  is estimated on the basis of the imputed data and under the multistage sampling design by:

$$\hat{\tau}_I = \sum_{i \in S} w_{dqi} \cdot y_{I,i}. \quad (9)$$

## 2.2 Rescaling bootstrap under imputations for a multistage design

When applying the rescaling bootstrap in a multistage design under imputations, three requirements have to be met. The first two requirements are a result of the single stage application described in Sect. 1.3.2, while the third requirement is essential for the adjustment to a multistage design.

**Table 1** Categorization of USUs with respect to the subsampling process

Category	First stage: $\delta_d$	Second stage: $\delta_{dq}$	Third stage: $\delta_{dqi}$
C1	0	0	0
C2	1	0	0
C3	0	1	0
C4	1	1	0
C5	0	0	1
C6	1	0	1
C7	0	1	1
C8	1	1	1

1. Not just a subset of the original sample, but all of the imputed values have to be reimputed to avoid the unintentional variance component described in Sect. 1.3.2. This variance component arises when not reimputed imputed values have an influence on the variance estimation.
2. For the reimputation of imputed values, elements drawn and not drawn have to be separated. Otherwise, in case of deterministic imputation procedures, the reimputed values are the same as the imputed values of the original sample. Hence, the imputation process is not considered in the variance estimation.
3. The variance estimator should capture  $V_s(\hat{\theta}_I|\mathbf{z})$ , the sampling variance of the imputed estimator, given the response vector  $\mathbf{z}$ . As described in Sect. 1.2, to capture the variance of an estimator in a multistage sampling design, all stages of the sampling design have to be considered in the variance estimation (at least when the parameter constellations does not allow to consider only the first stage). To ensure the correct estimation of the sampling variance in the multistage design with the rescaling bootstrap, subsampling from the original sample is applied independently at each stage of the sampling design. However,  $V_s(\hat{\theta}_I|\mathbf{z})$  concerns the sampling variance of the imputed estimator. Even when the imputations are done only with respect to the missing values at the last stage, the imputations affect the variance of the imputed estimator at all stages of the sampling design. This is a consequence of the imputed values being included in the computation of total values or mean values of the PSUs and the SSUs, which determine the variance at the corresponding stage. Thus, to capture  $V_s(\hat{\theta}_I|\mathbf{z})$ , the reimputation of imputed values has to consider all stages of the sampling design and has to be conducted with respect to the subsampling process of the rescaling bootstrap.

To derive a modification of the rescaling bootstrap that takes the three requirements into account, we present the subsampling process of the rescaling bootstrap that is described in Sect. 1.2 in another form as illustrated in Table 1.

The table shows the possible combinations for categorizing USUs with respect to the subsample process. A value of 1 in the last column indicates that the category includes the USUs that are drawn for the subsample, and a value of 0 means that the category includes the USUs that are not drawn for the subsample. A value of 1 in the third column indicates that the category includes the USUs whose superordinate SSU

is drawn for the subsample, and a value of 0 means that the category includes the USUs whose superordinate SSU is not drawn for the subsample. A value of 1 in the second column specifies that the category consists of the USUs whose superordinate PSU is drawn for the subsample, and a value of 0 means that the category consists of the USUs whose superordinate PSU is not drawn for the subsample. The category C2, for example, includes the USUs whose superordinate PSU is drawn for the subsample, but these USUs and their superordinate SSUs are not drawn for the subsample. In contrast, category C8 includes all the USUs that are drawn for the subsample, as well as their corresponding PSU and SSU. By doing so, all the USUs of the original sample can be divided with respect to these eight categories of subsampling process. The categories are mutually exclusive, and each USU can be assigned only to one of the categories.

This division can be used as a basis for the reimpuation to take all three requirements into account: The reimpuation of all imputed values, the differentiation between elements drawn for the subsample and elements not drawn for the subsample, as well as to consider the different stages of the sampling design and the subsampling process respectively. We propose two different modifications of the rescaling bootstrap that use this division as a basis for the reimpuation. In the first modification, the reimpuation is done separately in all eight categories (with corresponding estimator *Resc\_Mod1*, see Table 2), whereas the second modification (with corresponding estimator *Resc\_Mod2*, see Table 2) combines categories with the same weight adjustments of the original design weights with respect to the formulas (1), (2), and (3). For example, the categories C3, C5, and C7 lead to same weight adjustments<sup>6</sup> as category C1, since even if the USUs or the SSUs are drawn for the subsample, their corresponding PSU is not drawn for the subsample.<sup>7</sup> In this case,  $\delta_d$  in (1), (2), and (3) takes the value 0 that results in same weight adjustments. Furthermore, the USUs in the categories C2 and C6 have the same weight adjustments, but these weight adjustments are different in comparison to the weight adjustments in the categories C1, C3, C5, and C7.

Both procedures have their advantages. When a reimpuation is done separately in all eight categories, they are more equal in size, which may make the variance estimation more stable. Furthermore, this procedure makes more use of the subsampling process of the rescaling bootstrap. The second procedure, in which categories with the same weight adjustment are combined, has the advantage that it makes more use of the weight adjustment process of the rescaling bootstrap.

As a consequence, the two proposed modifications are described as follows, where they differ in Step 2:

1. First, subsamples of size  $l^* = \lfloor l/2 \rfloor$ ,  $m_d^* = \lfloor m_d/2 \rfloor$ , and  $n_{dq}^* = \lfloor n_{dq}/2 \rfloor$  are drawn independently at each stage of the sampling design in each run  $b$ . This procedure corresponds to the rescaling bootstrap of Preston (2009) that is based on complete observations.
2. To conduct the reimpuation, all USUs are assigned to categories that are formed in each run  $b$ . For modification 1, the eight categories in Table 1 are built to include

<sup>6</sup> In this case, weight adjustment refers to the technical adjustment of the original weights via the formulas (1), (2), and (3). However, the adjusted weights of different elements can be different, for example, due to different sample sizes and subsample sizes or different original weights.

<sup>7</sup> Such combinations are possible, since the rescaling bootstrap is applied independently at each stage of the sampling design, as described in Sect. 1.2.

all possibilities to categorize the USUs with respect to the elements drawn for the subsample and the elements not drawn for the subsample at each stage of the sample design. For modification 2, the categories in Table 1 with the same weight adjustments are combined to one category. As explained, this concerns categories C1, C3, C5, and C7 as well as C2 and C6. Four categories remain—the two combined categories and the categories C4 and C8. For both modifications, the categories are formed again in each run  $b$ .

3. The reimputation is conducted separately in each category  $c$ . In each run  $b$ , the imputed values of the original sample in each category  $c$  with  $i \in G_{bc}$  are reimputed on the basis of the observed values in each category  $c$  with  $i \in R_{bc}$ , which results in the vector  $\mathbf{y}_{reimp}^*$  that contains reimputed imputed values as well as observed values.<sup>8</sup>
4. The design weights  $w_{dqi}^*$  are computed in each run  $b$  with respect to (4) on the basis of the adjusted weights at each stage regarding (1), (2), (3), and as described by Preston (2009). The adjusted weights are combined to the vector  $\mathbf{w}^*$ .
5. The point estimator of the rescaling bootstrap is calculated in each run  $b$  by

$$\hat{\theta}_I^* = T(\mathbf{y}_{reimp}^*, \mathbf{w}^*).$$

6. The steps 1 until 5 are repeated  $B$  times. Thus, the estimators  $\hat{\theta}_{I,1}^* \dots \hat{\theta}_{I,b}^* \dots \hat{\theta}_{I,B}^*$  result and the variance is estimated with respect to (6) by

$$\hat{V}_{resc}(\hat{\theta}_I) = \frac{1}{B-1} \cdot \sum_{b=1}^B \left( \hat{\theta}_{I,b}^* - \frac{1}{B} \cdot \sum_{v=1}^B \hat{\theta}_{I,v}^* \right)^2. \tag{10}$$

We include both modifications in the simulation study because of their different advantages and to evaluate their performance. The described procedure can be extended easily to sampling designs with more than three stages. In this case, the number of categories in Table 1 increases to  $2^\phi$ , where  $\phi$  is the number of stages of the sampling design. Weight adjustments also can be extended easily to more stages by using the formula in Preston (2009, p. 229).

### 2.3 Procedures when $V_2$ has a strong contribution

Assuming a deterministic imputation procedure and a variance decomposition as shown by (7), both modifications cover the variance component  $V_1$  and are valid variance estimators for the overall variance, if component  $V_2$  is negligible. Variance component  $V_2$  cannot be covered because this component includes the variance resulting from the treatment of the response mechanism as a random process and, thus, is based on the variance component related to a varying response vector. However, the rescaling bootstrap treats the response vector as fixed (for treatment of response vec-

<sup>8</sup> The adjustment of imputed values with respect to Shao (2002) must be applied for stochastic imputation procedures instead of the reimputation at this step, since the subsample size of the rescaling bootstrap differs from the size of the original sample at the different stages.

tors in resampling methods, see Mashreghi et al. 2014). This issue is due to the fact that the response vector, and thus nonresponse, is not generated in each run  $b$  of the rescaling bootstrap. The same response vector is used, and respondents and nonrespondents maintain their response status in each run  $b$  (Mashreghi et al. 2014). If the variance component  $V_2$  makes a significant contribution to the overall variance, an estimator is required that additionally covers this variance component. In this case, the two proposed modifications of the rescaling bootstrap may be extended by the procedures described in Mashreghi et al. (2014). For example, a new response vector can be generated for each run  $b$  by using a binomial distribution. As for the resampling methods described in Mashreghi et al. (2014), this procedure also may require adjustments of the weighting formulas in (1), (2), and (3) of the Preston's rescaling bootstrap (Preston 2009). Another procedure for considering the variance component  $V_2$  is the usage of correction factors as described in Mashreghi et al. (2014) but modifications of these correction factors regarding multistage sampling designs may be necessary.

However, as described previously, in many situations, the overall sampling fraction  $f$  in multistage sampling designs may not be large enough in order for  $V_2$  to have a strong influence on the overall variance. Thus, the focus in the simulation study in Sect. 4 is more on the examination under which parameter constellations, particularly, the overall sampling fraction, the variance component  $V_2$  makes a significant contribution to the overall variance.

### 3 Design of the simulation study

We use a Monte Carlo simulation study to evaluate the quality of the proposed modifications of the rescaling bootstrap. According to Enderle et al. (2013), comparisons that are made on the basis of Monte Carlo simulation studies ensures an evaluation of such methods with respect to their actual application. This is a major point of this paper, because we aim to compare the proposed methods with variance estimation methods in which important components of the practical application such as the sampling process, the subsampling process or the imputation process are not considered. The effects of the non-consideration of such important components can be examined directly by using the corresponding distributions of the estimators. Furthermore, the Monte Carlo simulation study ensures an analysis of the effects of different parameter constellations. Within the scope of this paper, this is, for example, important with respect to the examination of possible effects of the variance component  $V_2$  on the variance estimation dependent on the overall sampling fraction.

We implement the simulation study by using the statistical software **R** (R Core Team 2021), and we describe this study in the following sections.

The simulation study consists of two parts. To evaluate the modifications in their basics, we draw a population from a multivariate normal distribution in the first part. This population is described in Sect. 3.1.1. In the second part of the simulation study in Sect. 3.2, we consider a more realistic dataset, to analyse the modifications under more complex conditions.

### 3.1 Multivariate normal distribution

#### 3.1.1 Population, sampling design, and scenarios

The population generation process is similar to the population generation process used by Bruch (2016, 2019). In contrast to these studies, the population generation process in the simulation study of this paper is not related to a one or two stage sampling design but to a three-stage random sampling design.

First, we create the structure of the population with a size of  $N = 10,000,000$  by dividing the population in 500 PSUs, 80 SSUs per PSU, and 250 USUs per SSU. This was done by a simple replication of the corresponding IDs of PSUs and SSUs according to their respective sizes.

Given this predefined structure, we draw the variable of interest  $y$  and auxiliary variables  $x_1$  and  $x_2$  from a multivariate normal distribution described by

$$\zeta \sim N(\mu, \Sigma),$$

where parameters  $\mu$  and  $\Sigma$  are defined by

$$\mu = (200, 50, 80)$$

and

$$\Sigma = \begin{pmatrix} 1000 & 170 & 35 \\ 170 & 100 & 5 \\ 35 & 5 & 75 \end{pmatrix}.$$

We apply the function `rmvnorm` from the **R** package `mvtnorm` (Genz et al. 2014; Genz and Bretz 2009) to draw the universe from the multivariate normal distribution. As a result of this random drawing process to generate  $y$ ,  $x_1$  and  $x_2$  given the predefined structure of PSUs and SSUs, these units are heterogeneous in this scenario.

We use the multivariate normal distribution for two reasons. First, the main aim of the paper is to propose the application of variance estimation methods when data is collected using multistage sampling designs and that in this collected data missing values appear that are compensated by using imputation. The proposed resampling method should be examined with regard to this main aim in this part of the simulation study, and the influences resulting, for example, from highly skewed variable distributions should be kept at a minimum. Second, the use of a multivariate normal distribution guarantees that a large proportion of the variance of the estimator appear at the last stage of the sampling design. In Särndal et al. (1992), Lohr (1999), Bruch et al. (2011), Bruch (2016) and in Sect. 1.2 it is described that with variance estimation procedures that only include the first stage of the sampling design, more than the variance of the estimator at the first stage is covered. In case the PSUs and SSUs are homogeneous, it may occur that most of the overall variance could be covered with the first stage's variance estimator. However, in the simulation study, we aim to analyse, if the proposed modifications are able to estimate the variance correctly when



also a large proportion of the variance occurs at the last stage and thus all stages of the sampling design may have to be considered. Thus, the PSUs and SSUs should be heterogeneous in this part of the simulation study. However, the PSU and SSU in the first part of the simulation study may be too heterogeneous for some applications. Thus, in the second part of the simulation study, we also consider a population with a less heterogeneous structure that may be more realistic for practical applications.

The simulation study includes different scenarios that differ with respect to the sampling fractions and the generation process of the response vector. We use the same response vector in each simulation run in the first two scenarios. The variance component  $V_2$  of (7) does not appear in this case. The aim is to show the extent the modifications of the rescaling bootstrap are able to estimate variance component  $V_1$ . In Sect. 1.3, we described that this component may determine the whole variance of an imputed estimator in multistage sampling designs in many situations. Thus, it is of particular interest to find a valid variance estimator for this variance component, which is accomplished by comparing the two scenarios in the simulation study with different sampling fractions.

Sampling fractions are small in the first scenario, and 50 PSUs are drawn at the first stage ( $f_1 = 0.1$ ); 8 SSUs in each drawn PSU are selected at the second stage ( $f_2 = 0.1$ ) and 10 USUs in each drawn SSU are selected at the third stage ( $f_3 = 0.04$ ). The overall sample size is  $n = 4000$  ( $f = 0.0004$ ).

We examine the impact of large sampling fractions in the second scenario, particularly, the impact of missing reimputation of certain elements by which the influence increases with increasing sampling fractions. Thus, 300 PSUs are drawn at the first stage ( $f_1 = 0.6$ ), 40 SSUs in each drawn PSU at the second stage ( $f_2 = 0.5$ ), and 125 USUs in each drawn SSU at the third stage ( $f_3 = 0.5$ ). The overall sample size is  $n = 1,500,000$  ( $f = 0.15$ ).

A new response vector is generated in each simulation run in Scenario 3. Thus, variance component  $V_2$  also may occur. In this scenario, the aim is to examine under which conditions this component has to be considered in a multistage sampling design and up to which point the proposed methods may still be a valid instrument to estimate the whole variance. As mentioned previously, we assume that the variance component  $V_2$  may be negligible in multistage sampling designs in many situations. However, the component is not negligible in the case of a large overall sampling fraction  $f$ . Thus, in this scenario, we choose sampling fractions that are rather high with  $f_1 = 0.25$ ,  $f_2 = 0.25$ , and  $f_3 \approx 0.25$  to obtain a rather high overall sampling fraction  $f$ . Hence, 125 PSUs are drawn at the first stage, 20 SSUs in each drawn PSU are selected at the second stage, and 63 USUs in each drawn SSU are selected at the third stage. The overall sample size is  $n = 157,500$ . Also, the overall sampling fraction  $f \approx 0.016$  may be rather high compared to population surveys that use multistage sampling designs in practice but may not large enough for variance component  $V_2$  to have a large influence. Hence, to examine from which sampling fraction onwards the variance component  $V_2$  has an influence on the overall variance, we increase the sampling fractions up to  $f_1 = 0.6$ ,  $f_2 = 0.5$ , and  $f_3 = 0.5$  with overall sampling fraction  $f = 0.15$ .

### 3.1.2 Nonresponse generation

Nonresponse is generated in the population before sampling (see, for example, Fay 1991 or Shao and Steel 1999) and only for the USUs in our simulation study. The nonresponse is implemented by the procedure described and used in Enderle et al. (2013).<sup>9</sup> Following this procedure, the response vector is generated by using a logistic model and depends on the auxiliary variables  $x_1$  and  $x_2$ . Thus, the missing mechanism is missing at random.

First, the response propensities  $p_i$  for each element in the population  $i \in U$  are computed by

$$p_i = \frac{\exp(\tilde{\alpha} + \tilde{\beta} \cdot x_{1i} + \tilde{\gamma} \cdot x_{2i})}{1 + \exp(\tilde{\alpha} + \tilde{\beta} \cdot x_{1i} + \tilde{\gamma} \cdot x_{2i})}, \quad (11)$$

where  $\tilde{\alpha}$ ,  $\tilde{\beta}$ , and  $\tilde{\gamma}$  are model parameters that determine the response mechanism. Second, random numbers  $u_i$  are selected from a uniform distribution  $U(0, 1)$  for each element  $i$ . The response vector  $\mathbf{z}$  for an element  $i$  takes the value 1 if  $p_i \geq u_i$ , which indicates that element  $i$  has an observed value regarding  $y$ . The response vector  $\mathbf{z}$  takes the value 0 for an element  $i$  if  $p_i < u_i$ , which indicates a missing value regarding  $y$ .

However, the main aim of this paper is to analyse the modifications of the rescaling bootstrap under nonresponse and imputation. Thus, the nonresponse rate has to be large enough. Also, the correlation of the auxiliary variables to the response vector should not be too small to ensure a missing-at-random mechanism. Thus, we set the parameters to generate the response vector to  $\tilde{\alpha} = -20$ ,  $\tilde{\beta} = 0.3$ , and  $\tilde{\gamma} = 0.07$ , which leads to a nonresponse rate of about 43% for each scenario. The correlation of  $x_1$  to the response vector  $\mathbf{z}$  is approximately 0.68, and the correlation of  $x_2$  to the response vector is approximately 0.17.

We apply deterministic regression imputation to compensate for missing values. We apply a regression model in the linear form. The independent variables are the auxiliary variables  $x_1$  and  $x_2$  and the variable to be imputed  $y$  is the dependent variable. The regression imputation is conducted unweighted and all observed units are weighted equally when training the regression model. As a result of the population and sampling design, the ultimate sampling units have the same design weights so that it is not necessary to include them. The corresponding **R**-Code for the regression imputation is based on the **R**-code used by Davison and Sardy (2004).

### 3.2 AMELIA

The proposed procedures should also be evaluated under more realistic conditions. Thus, in a second part, we conduct the simulation study on the synthetic dataset AMELIA (Burgard et al. 2017, 2020; Kolb 2012; Alfons et al. 2011) that is based on the EU-SILC dataset from 2005. Also, the previously mentioned studies with respect to the rescaling bootstrap of Bruch (2016, 2019) used this dataset in their simulation studies.

<sup>9</sup> The procedure is also applied in Bruch (2019).

The analyses are done at the person level and the population includes 10,012,600 persons. The variable of interest is the variable INC of the AMELIA dataset, that is the sum of all income components of a person. As a result, the variable is highly skewed and poses further challenges to the variance estimation. From this population, we draw samples with a size of 30,000 persons using a multistage sampling design. As PSUs we use the districts (variable DIS) whereby we assign districts 1 and 2 to district 3 due to their small population size. The total dataset therefore contains 38 PSUs in total. As SSUs, we use the cities and communities respectively (variable CIT). We assign small communities to larger communities when their population size is below 500.

To realize a sample size of 30,000, we select 10 PSUs, 10 SSUs in each PSU drawn, and 300 USUs in each SSU drawn. To generated nonresponse, we apply again the procedure described in Enderle et al. (2013). To model the imputation process, we use the variables BAS that describe the work status and the variable HHS that indicates the household size. We categorize the latter so that both variables consist of four categories. Since both variables are categorical, instead of (11), the propensities are computed by

$$p_i = \frac{\exp(\tilde{\alpha} + \tilde{\beta}_{BAS,v[i]} + \tilde{\gamma}_{CIT,o[i]})}{1 + \exp(\tilde{\alpha} + \tilde{\beta}_{BAS,v[i]} + \tilde{\gamma}_{CIT,o[i]})}, \quad (12)$$

where the parameters  $\tilde{\beta}_{BAS,v[i]}$  and  $\tilde{\gamma}_{CIT,o[i]}$  depend on the category  $v$  of variable BAS and on the category  $o$  of variable CIT to which an element  $i$  belongs.  $\tilde{\beta}_{BAS,v[i]}$  results in a value of  $-3$  for  $v = 1$ ,  $1$  for  $v = 2$ ,  $1$  for  $v = 3$  and  $0$  for  $v = 4$ .  $\tilde{\gamma}_{CIT,o[i]}$  has the values  $-3$  for  $o = 1$ ,  $1$  for  $o = 2$ ,  $0$  for  $o = 3$  and  $-2$  for  $o = 4$ . The parameter  $\tilde{\alpha}$  is 2.5 in this scenario. The following steps to generate nonresponse are the same as for the previous scenarios. The nonresponse rate is also comparable with previous scenarios with a value of 43%. In this part of the simulation study, we use a varying response vector.

In practice, often a stochastic imputation procedure is used. Thus, in this scenario, we examine the proposed variance estimation procedures for a stochastic imputation procedure, and we apply hot-deck-random imputation.<sup>10</sup> First, we construct imputation classes over the cross classification of variable BAS and CIT. Imputation classes that are too small are assigned to larger classes. Afterwards, for each missing value in an imputation class, a donor is drawn randomly from the observed value of the same imputation class. As a result of the stochastic imputation procedure, we apply the adjustment of imputed values as described in Shao (2002) instead of the reimputation. Due to the small overall sampling fraction, we do not consider any of the corrections described in Mashreghi et al. (2014) in this scenario.

<sup>10</sup> We apply an unweighted hot-deck-random imputation since in our case the resulting point estimator was nearly Monte Carlo unbiased. However, with respect to Haziza (2009) in other applications it may be necessary to apply a weighted hot-deck-random imputation, particularly, when the sampling weight is highly correlated with the variable of interest  $y$ .

**Table 2** Resampling methods considered in the simulation study

Variance estimator	Description
Resc_obs	Rescaling bootstrap that treats the imputed values as <i>actually observed</i> without reimputation
Resc_whole	Rescaling bootstrap that conducts the reimputation of the imputed values on the basis of the whole sample, without distinction between the elements drawn and not drawn for the subsample
Resc_only_Sub	Rescaling bootstrap that does not reimpute the imputed values in the categories C1, C3, C5, and C7
Resc_stage1	Rescaling bootstrap in which the reimputation is conducted separately for the imputed values whose corresponding PSU is drawn for the subsample and for the imputed values whose corresponding PSU is not drawn for the subsample
Resc_stage2	Rescaling bootstrap in which the reimputation is conducted separately for the imputed values whose corresponding SSU is drawn for the subsample and for the imputed values whose corresponding SSU is not drawn for the subsample
Resc_stage3	Rescaling bootstrap in which the reimputation is done separately with respect to the imputed values drawn at the third stage and the imputed values not drawn at the third stage
Resc_Mod1	Rescaling bootstrap in which the reimputation of imputed values is done separately in all categories from C1 to C8
Resc_Mod2	Rescaling bootstrap in which categories with the same weight adjustments are combined, and reimputation is done in the remaining categories

### 3.3 Included estimator and benchmark

As shown in Table 2, we consider the following modifications of the rescaling bootstrap in the simulation study.

The modifications of the rescaling bootstrap `Resc_Mod1` and `Resc_Mod2` are the proposed methods of this paper. These are the only modifications that consider the three requirements to modify the rescaling bootstrap in the multistage design with respect to imputation that are described in Sect. 2.2. The other modifications of the rescaling bootstrap that are included in the simulation study do neglect at least one of the requirements. The aim of including these modifications of the rescaling bootstraps is to show the effect of important components not being considered or applied adequately, and that it is important to take all three requirements into account when applying the rescaling bootstrap in a multistage sampling design.

As a result, we include the estimator `Resc_whole` that applies the reimputation on the basis of the whole sample without distinction between the elements drawn and not drawn for the subsample and thus does not consider, for example, the second requirement. In case of a deterministic imputation method, this estimator should lead to the same results as the estimator that treats imputed values as actually observed. This issue is examined by including the estimator `Resc_obs`. Both estimators do not consider the imputation process adequately.

The estimators `Resc_stage1`, `Resc_stage2`, and `Resc_stage3` do not consider certain stages of the sampling design and the subsampling process of the rescaling bootstrap in the reimputation process and thus the third requirement. The reimputation of these estimators only considers one stage of the sampling design. For example, the reimputation in the estimator `Resc_stage1` is done separately only for the imputed values whose corresponding PSU is drawn for the subsample ( $\delta_d = 1$ ) and for the imputed values whose corresponding PSU is not drawn for the subsample ( $\delta_d = 0$ ). Stages 2 and 3 of the sampling design are neglected. The reimputation for the estimator `Resc_stage2` is done separately only for the imputed values whose corresponding SSU is drawn for the subsample ( $\delta_{dq} = 1$ ) and for the imputed values whose corresponding SSU is not drawn for subsample ( $\delta_{dq} = 0$ ). Thus, stages 1 and 3 of the sampling design are neglected. The estimator `Resc_stage3` does not consider the stages 1 and 2, and the reimputation is done separately only with respect to the imputed values drawn at the third stage ( $\delta_{dqi} = 1$ ) and the imputed values not drawn at the third stage ( $\delta_{dqi} = 0$ ). Instead of the eight categories in Table 1, we only have two categories in which the reimputation is applied within each estimator.

Similar to study of Bruch (2019) with respect to a single stage design, we also examine the need to reimpute the imputed values in a multistage design to avoid the unintentional variance component dependent on the sampling fractions, and the influence of not reimputed imputed missing values. This issue is related to the first requirement in Sect. 2.2 and is examined by including the estimator `Resc_only_Sub`. In contrast to the single stage design, statements regarding the need to reimpute imputed values dependent on the sampling fractions are not straightforward. That is due to more parameters having an impact on the weight adjustment, while the classification of elements being drawn (or not drawn) concerns the different stages.

With respect to the first point, we can only consider a limited number of parameter constellations in the simulation study, particularly, a scenario with small and a scenario with large sampling fractions. Thus, the statements focus mainly on these constellations. Regarding the second point, the classification of elements, we examine the effect of a missing reimputation dependent on the sampling fractions with respect

to certain categories of Table 1 rather than with respect to a classification of elements drawn and not drawn for the subsample. Thus, for the estimator `Resc_only_Sub`, we do not reimpute the imputed values in the categories C1, C3, C5, and C7. We chose these categories because their elements receive same weight adjustments as described in Sect. 2.2. Thus, the effect of smaller and larger sampling fractions should be the same with respect to the influence of the elements of the categories C1, C3, C5, and C7 on the variance estimation. Particularly, the influence of the elements of these categories is very small when sampling fractions are small and larger when sampling fractions are large. This is shown in the Appendix A.1 and in Table 4, in which we compute the relative influence of USUs of a certain category  $c$ . The relative influence of categories C1, C3, C5, and C7 is altogether 4% in the scenario with small sampling fractions and 20% in the scenario with large sampling fractions. As a consequence, we expect a larger unintentional variance component and thus overestimation of the variance in the scenario with larger sampling fractions with estimator `Resc_only_Sub`.

In the first part of the simulation study, all estimators presented in Table 2 are considered. In this part of the simulation study, those estimators should perform well that fulfil the basic requirements for a variance estimation under imputation in a multistage sampling design, particularly, the three requirements defined in Sect. 2.2. As described in Sect. 3.1.1 using a multivariate normal distribution to create the population ensures, that all stages of the sampling design may have to be considered and that other influences (such as the skewed variable of interest) that confound the effects that should be analysed are avoided. As a result, this scenario allows to focus on the analysis of the basic requirements of the variance estimation under imputation and is rather simple with respect to other influences. Thus, estimators that do not perform well in this scenario do not fulfil the requirements (as used in this study) to obtain a valid variance estimation under imputation in a multistage sampling design. For that reason, these estimators are not considered any further in the second part of the simulation study.

As a benchmark, we use the Monte Carlo variance, which is computed on the basis of the total value point estimators in (9) of 50,000 samples. This large number of samples for the benchmark is necessary, since the simulation study applied by Bruch (2016) shows that the Monte Carlo variance can be very unstable when the number of samples is not large. However, the variance estimation that applies resampling methods may converge much faster if the number of subsamples is large enough (Bruch 2019). Thus, and as a result of computation reasons, in the scenarios, that are based on the multinormal distributions, we computed the resampling methods on the basis of 125 subsamples and only for 1000 samples. In the scenario in which we use the AMELIA dataset the convergence was much slower due to the highly skewed variable of interest. For that reason, we increase the number of samples to 20,000.

## 4 Results

We present the results of the simulation study by using the relative bias. The results are visualized by boxplots that are produced by the **R** package `lattice` (Sarkar 2008). The black point in each boxplot indicates the mean value of the estimations of the

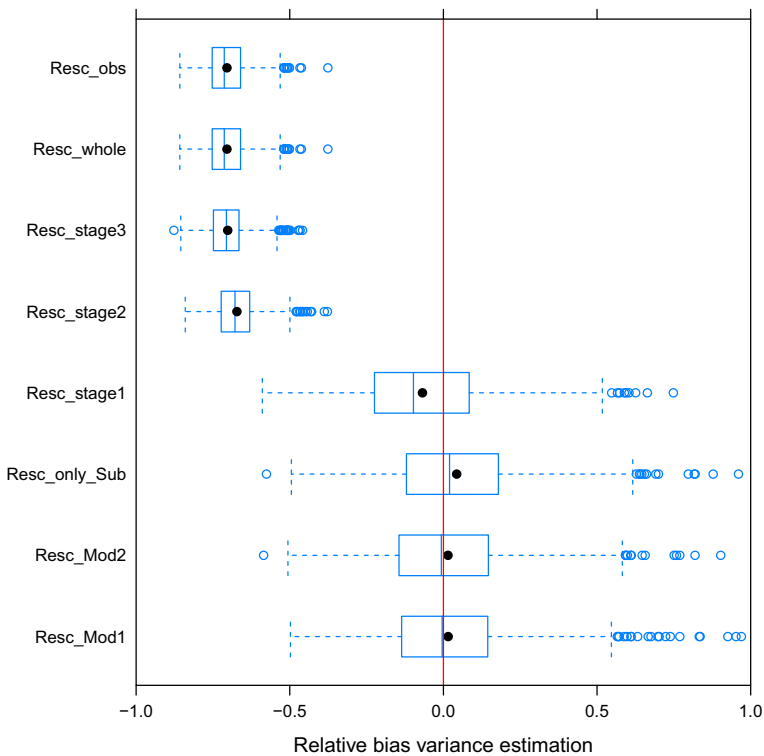


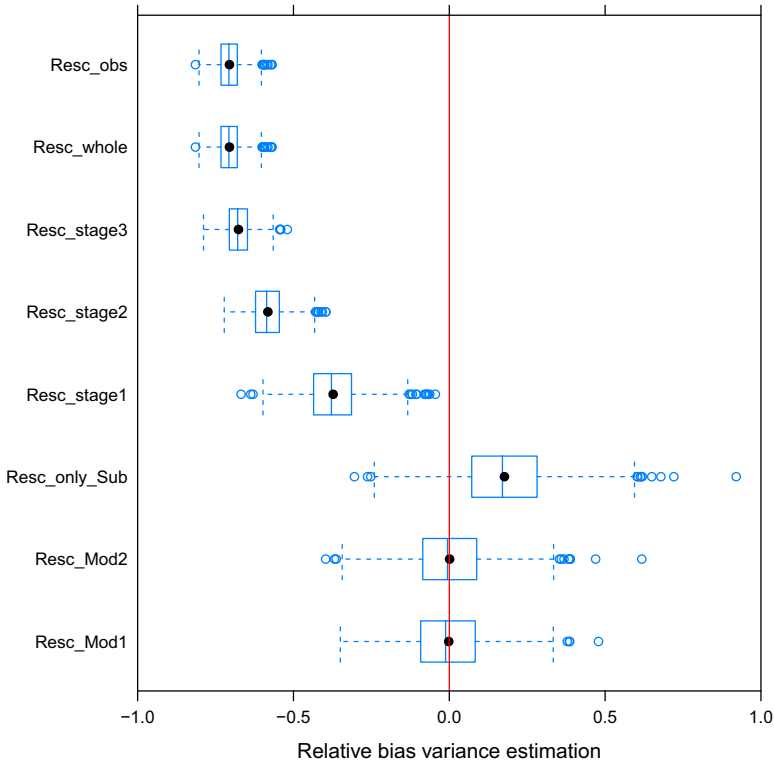
Fig. 1 Results of the scenario with small sampling fractions and a fixed response vector

different samples for a certain estimator. The red line across all boxplots indicates the Monte Carlo variance, whose relative bias is zero due to it being the benchmark.

## 4.1 Multivariate normal distribution

### 4.1.1 Scenario 1: Small sampling fractions, fixed response vector

Figure 1 presents the results for the small sampling fractions at all stages of the sampling design and a fixed response vector. As expected, the rescaling bootstrap that conducts the reimputation of the imputed values on the basis of the whole sample without distinction between the elements drawn and not drawn for the subsample (Resc\_whole) leads to the same results as the rescaling bootstrap that treats the imputed value as actually observed (Resc\_obs). Applying the reimputation of the imputed values on the basis of the whole sample without distinction between the elements drawn and not drawn for the subsample leads to the same reimputed value as the imputed value in the original sample when the imputation procedure is deterministic. In this case, this approach matches the rescaling bootstrap that treats the imputed value of the original sample as actually observed. Both estimators lead to a severe underestimation of the



**Fig. 2** Results of the scenario with large sampling fractions and a fixed response vector

variance because they do not consider the effect of missing values and imputation. Particularly, it shows the need to meet requirement 2 in Sect. 2.2.

Furthermore, the figure shows that the three modifications of the rescaling bootstrap that neglect certain stages of the sampling design in the reimputation process (Resc\_stage1, Resc\_stage2 and Resc\_stage3) underestimate the variance. These results strengthen the importance to meet requirement 3 in Sect. 2.2. However, the underestimation of the rescaling bootstrap that only considers the first stage in the reimputation process is not as large as the underestimation of the other two estimators.

As expected, the estimator that does not reimpute the imputed values in the categories C1, C3, C5, and C7 (Resc\_only\_Sub) leads only to a small overestimation of the variance. Elements of these categories have a minimal influence if the sampling fractions at the several stages are small as seen in Table 4 in the Appendix A.1. Thus, the unintentional variance component is small that leads to a small overestimation of the variance.

However, our two proposed modifications of the rescaling bootstrap—the one that makes more use of the weight adjustment process of the rescaling bootstrap (Resc\_Mod2) and the other that separately does the reimputation in all the categories (Resc\_Mod1)—lead to very similar results. On average, the estimations of both methods are close to the benchmark.



### 4.1.2 Scenario 2: Large sampling fractions, fixed response vector

These results of the previous scenario become more obvious when increasing the sampling fractions at the several stages. The results for large sampling fractions are presented in Fig. 2. Again, the estimators `Resc_whole` and `Resc_obs` lead to a huge underestimation of the variance. These results further confirm the need to meet requirement 2 in Sect. 2.2.

The modifications of the rescaling bootstrap that neglect certain stages of the sampling design (`Resc_stage1`, `Resc_stage2` and `Resc_stage3`) strongly underestimate the variance, which is also the case for the rescaling bootstrap that only considers the first stage in the reimputation process. All estimations are below the benchmark for all three modifications. These results further show the need to meet requirement 3 in Sect. 2.2.

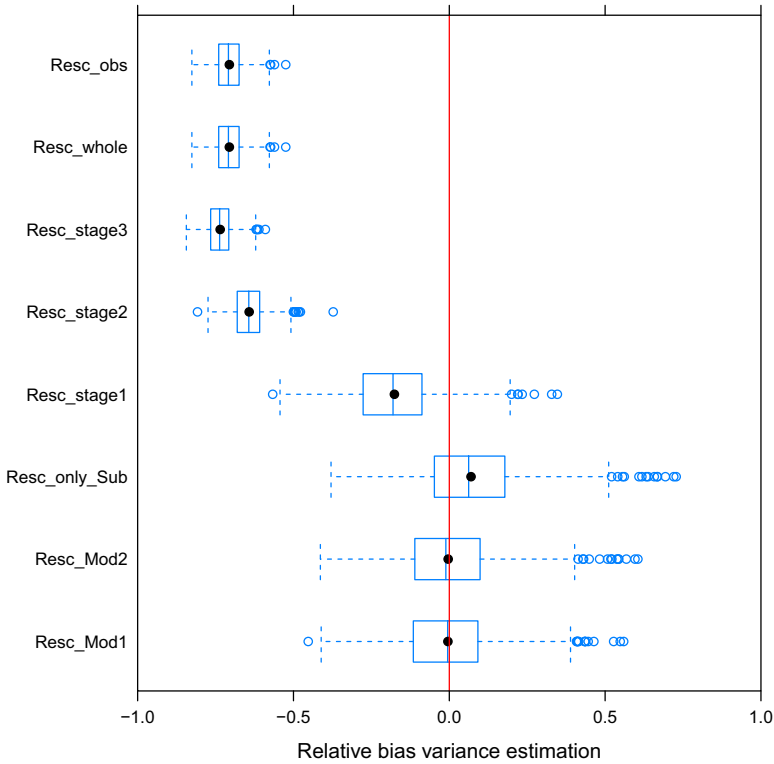
As expected, the estimator that does not reimpute the imputed values in the categories C1, C3, C5, and C7 (`Resc_only_Sub`) leads to a larger overestimation in this scenario than in the scenario with the small sampling fractions. This larger overestimation is a result of the increasing influence of elements of categories C1, C3, C5, and C7 with larger sampling fractions at the several stages as shown in Table 4 in the Appendix A.1. Thus, the unintentional variance component also increases and the overestimation is much larger. The results emphasise the need to meet requirement 1 in Sect. 2.2 and show the need to reimpute all imputed values of all (combined) categories when elements of the category have an influence on the variance estimation. Otherwise, the unintentional variance component can also appear in a multistage sampling design.

The only estimators that lead to estimations close to benchmark are the two proposed modifications of the rescaling bootstrap `Resc_Mod1` and `Resc_Mod2`. These are the only estimators that meet all requirements in Sect. 2.2. The differences between these two modifications are again rather small.

### 4.1.3 Scenario 3: Varying response vector

This section presents the results of the simulation study when generating a new response vector in each simulation run. The results for sampling fractions  $f_1 = 0.25$ ,  $f_2 = 0.25$ , and  $f_3 = 0.25$  are displayed by Fig. 3. In this case, the total variance in (7) may also include the variance component  $V_2$ . For a large sampling fraction  $f$ , the proportion of  $V_2$  may be relevant. Under such conditions, the modifications described in this paper may lead to an underestimation of the overall variance, since they do not cover this variance component.

However, the results for the sampling fractions  $f_1 = 0.25$ ,  $f_2 = 0.25$ , and  $f_3 = 0.25$  are very similar to the results of the previous section. The estimators `Resc_whole`, `Resc_obs`, `Resc_stage1`, `Resc_stage2`, and `Resc_stage3` result in an underestimation of the variance, and the estimator `Resc_only_Sub` overestimates the variance. On average, the estimations of the modifications `Resc_Mod1` and `Resc_Mod2` are close to the benchmark. It seems that the variance component  $V_2$  only makes a small contribution to the total variance, even if the sampling fractions are rather large compared to many



**Fig. 3** Results generating a new response vector in each simulation run

**Table 3** Relative bias (RB) for the two proposed modifications, depending on the sampling fractions when generating a new response vector in each simulation run

	Sampling fractions at different stages			
Variance estimator	$f_1 = 0.25,$ $f_2 = 0.25,$ $f_3 = 0.25,$ $f \approx 0.016$	$f_1 = 0.3,$ $f_2 = 0.3$ $f_3 = 0.3,$ $f_3 = 0.027,$	$f_1 = 0.4,$ $f_2 = 0.4,$ $f_3 = 0.4,$ $f = 0.064$	$f_1 = 0.6,$ $f_2 = 0.5,$ $f_3 = 0.5,$ $f = 0.15$
Resc_Mod1	RB = -0.004	RB = -0.009	RB = -0.035	RB = -0.069
Resc_Mod2	RB = -0.004	RB = -0.008	RB = -0.039	RB = -0.073

practical applications. The overall sampling fraction with  $f \approx 0.016$  may be too small for the component  $V_2$  to have an influence on the overall variance.

Thus, to examine this point in more detail, we further increase the sampling fractions in our simulation study. Table 3 presents the results for estimators Resc\_Mod1 and Resc\_Mod2. As can be seen, only from sampling fractions  $f_1 = 0.4, f_2 = 0.4, f_3 = 0.4$  and  $f = 0.064$  onwards, some underestimations arise due to the non-consideration of the variance component  $V_2$ . The underestimations for  $f_1 = 0.4, f_2 =$

0.4,  $f_3 = 0.4$  and overall sampling fraction  $f = 0.064$  are rather small, with relative biases of  $-0.035$  and  $-0.039$ . Even for the sampling fractions  $f_1 = 0.6$ ,  $f_2 = 0.5$ , and  $f_3 = 0.5$ , with an overall sampling fraction  $f$  of 15%, the underestimations are not much larger with relative biases of  $-0.069$  and  $-0.073$ . Thus, with respect to the conditions of our simulation study, it seems that the overall sampling fraction must be considerably larger for the variance component  $V_2$  to have an influence on the overall variance, and to be significant. However, as seen at the parameter constellations, such large sampling fractions may not often achieve in practical application. Thus, the modifications of the rescaling bootstrap described in Sect. 2.2 may ensure a valid variance estimation for the overall variance in many situations. Nevertheless, if the variance component  $V_2$  has an influence on the variance, the approaches described in Mashreghi et al. (2014) may be used as described in Sect. 2.3.

## 4.2 Simulation study using the AMELIA dataset

The results of the variance estimation using both of the proposed modifications of the rescaling bootstrap for the AMELIA dataset are presented by Fig. 4. Basically, the proposed modifications perform well under these challenging circumstances. The estimator `Resc_Mod1` leads to a relative bias of  $-0.0246$  on average and the estimator `Resc_Mod2` results in a relative bias of  $-0.02352$  on average, which is small when considering that the variable of interest `INC` is highly skewed. The overall sampling fraction is very small with 0.3%. However, as can be seen, due to the highly skewed variable of interest, the distributions of the variance estimations of both modifications are highly skewed as well. Particularly, larger outliers make the variance estimation less robust, a problem many variance estimations methods will be confronted with, when applying on such skewed data.

## 4.3 Computational efficiency

To evaluate the computational efficiency of the proposed modifications and the other included estimator, we evaluate the elapsed time of the different procedures for a certain sample in Scenario 1 (Sect. 4.1.1) with  $n = 4000$  using the `R`-command `system.time` of the base-package (R Core Team 2021). The computational times can only be evaluated with some stronger limitations, since it depends also on parameters that are not related to the actual different estimators, such as the computer power or the efficiency of the written `R`-Code of the different procedures.

Estimator `Resc_Mod1` requires approximately 2.8 s per sample, `Resc_Mod2` about 2.1 seconds and `Resc_only_Sub` 1.9 s. The estimators `Resc_stage1`, `Resc_stage2` and `Resc_stage3` have an elapsed time of about 1.6 seconds. With 1.4 seconds and 1.3 s, the estimators `Resc_whole` and `Resc_obs` are the fastest. The computational times show that the more complex the reimputation process is applied in the different estimator, the longer the computational time. `Resc_Mod1` conducts the reimputation by considering the most number of categories. `Resc_obs` does not apply the reimputation. Hence, this may be one reason for the differences in computational time. The computational time

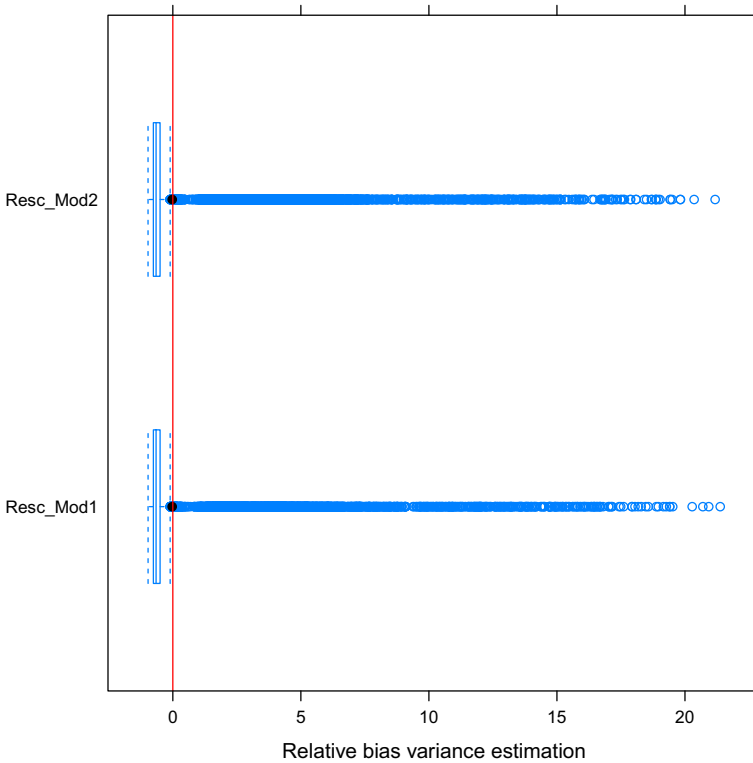


Fig. 4 Results for the AMELIA dataset

can increase strongly with sample size in Scenario 2 (Sect. 4.1.2,  $n = 1,500,000$ ). For example, Resc\_Mod1 needs about 562 s and Resc\_obs around 423 s.

### 5 Discussion and conclusion

In practice, frequently, there will be a trade off between the aim to achieve good variance estimates and the increase in complexity when using variance estimation procedures that include the sampling design and the imputation such as it was done in the proposed modifications. The theoretical part of Sects. 1 and 2 show the complexity to include the sampling design, particularly, the different stages and the imputation process that additionally concerns the implementation. As a result, the data analyst may have to adjust these procedures for their particular application, since they may be too specific for standard statistical software implementations. Moreover, the computational time with the proposed modifications increases since more components of the sampling design, particularly, more stages are taken into account as well as the reimputation has to be conducted again for each subsample.

However, from a theoretical view, the variance of an estimator is derived with respect to a certain sampling design that thus has to be considered in the variance

estimation process. The imputation process also has an influence on the variance. Thus, the sampling design as well as the imputation process have to be considered adequately, otherwise, highly biased variance estimates may occur. This is confirmed in our simulation study. Our results show the need to consider the imputation process correctly in the variance estimation since our proposed modifications that include the sampling design and the imputation process adequately outperforms the other included estimators. Particularly, the subsampling process of the rescaling bootstrap and thus the different stages of the sampling design has to be taken into account in the reimputation process. As shown in the simulation study, the variance is heavily underestimated when the imputation process is not included in the variance estimation or the sampling design is not considered in the reimputation correctly. In case certain imputed values are not reimputed, the variance is overestimated. Thus, the results show that the consequences in the practical application may be severe since variance estimates are used, for example, in hypothesis testing and confidence intervals. For the estimators of the simulation study that do not consider the imputation process or the sampling design correctly in the reimputation, hypothesis tests with respect to a certain significance level may lead to falsely significant results since the variance estimate is too low. Confidence intervals and thus the coverage probability may be too small [see also the explanations in Haziza (2009)]. When using these variance estimates in subsequent surveys, the sample size may be chosen too small.

Using the variance estimations of the estimator that do not reimpute certain imputed values leads to opposite problems: hypothesis tests may be wrongly insignificant with respect to a certain significance level, confidence intervals are too large and too large sample sizes may be chosen in subsequent surveys.

The large extent of the bias produced by most of the considered estimators in comparison to the proposed modifications shows the need to include adequately the imputation and sampling design in the variance estimation and may overrule the disadvantage of a complex implementation and higher computational times. The extent of the Monte Carlo bias of most of the variance estimation procedures that do not consider the components correctly in our simulation study was so large that they cannot be applied in practice.

Thus, it is important to receive a more valid variance estimation when applying a complex sampling design, such as a multistage design, and imputation is applied. This can be ensured by using the proposed modifications of the rescaling bootstrap for multistage sampling designs of Preston (2009) that lead in our simulation study to estimations that are on average close to the benchmark when the variance component  $V_2$  is negligible.

Our simulation study did not detect any larger differences between the two proposed modifications. The decision for one of the two proposed modifications may be done based on the simplicity of implementation in a particular application.

However, our proposed modifications are limited to some degree, since they are combinations of different procedure that themselves have some limitations.

In the paper, we assumed that an imputation procedure is applied that allows for an inclusion of the imputation process in the rescaling bootstrap via the reimputation or the adjustment of imputed values. This is especially a consequence of the derivation of the proposed modifications to be an estimator for component  $V_s(\hat{\theta}_I | \mathbf{z})$ . However,

there may be imputation procedures for which both procedures may not be applied and thus the rescaling bootstrap may not be applied for these imputation procedures, at least in the current version. The reimputation may lead to an overestimation of the variance when the imputation procedure includes random components and the subsample size is not of the original sample size. This circumstance is also the case for the rescaling bootstrap. The adjustment of imputed values may be used instead but that would require metric variables to be imputed. Thus, other adjustments may be necessary for other imputation procedures. Such adjustments may also be included in the rescaling bootstrap but an evaluation is necessary in order to examine how to include these adjustments.

Additionally, for stochastic imputation procedures, for large sampling fractions, the joint bootstrap distribution of the sampling and random component may be a poor approximation of the corresponding joint distribution in the original sampling as result of the rescaling [see Mashreghi et al. (2014) for further explanations]. Thus, further corrections may be necessary in such a case. However, this limitation may be again rather irrelevant in survey practice with generally small sampling fractions.

When the component  $V_2$  makes a significant contribution to the overall variance, the proposed methods may be modified with respect to the approaches described in Mashreghi et al. (2014). However, a larger contribution of the variance component  $V_2$  to the overall variance was detected only in our simulation study for high sampling fractions, which are rather unrealistic in practice. As already described, a large overall sampling fraction  $f$ , which is the main determinant of the variance component  $V_2$ , may be difficult to achieve in multistage sampling designs. Thus, in many situations, the proposed modifications may be sufficient to estimate the variance of an imputed estimator when multistage sampling designs are applied. However, due to the large number of possible parameter constellations, our simulation study could only be conducted for a few parameter constellations. In particular, the influence of the variance component  $V_2$  should be examined in more detail in further simulation studies under other conditions.

Furthermore, the simulation study based on the AMELIA dataset led to large outliers with respect to the variance estimation that can be explained by the highly skewed variable of interest. Thus, our modifications show potentials for improvements with respect to the robustness.

Finally, in our simulation study, we can control all influences, particularly the response mechanism, by setting the corresponding parameters and by considering the variables that are responsible for the nonresponse. This requirement is important to evaluate the proposed modifications. In practice, it is often not possible to receive a lot of information with respect to the nonresponse mechanism, and a biased point estimation cannot be avoided. This biased estimation also may affect the variance estimation. Future research also may examine our proposed modification under such circumstances in more detail. However, this goes beyond of the scope of this paper in which the proposed methods were evaluated if they are basically an appropriate instrument to estimate the variance of an imputed estimator in a multistage sampling design.

**Acknowledgements** The author acknowledges Ralf Münnich and Siegfried Gabler for valuable advice and suggestions on this research. Furthermore, the author thanks Matthias Sand, Wayne Egers, Hannah Laumann and the anonymous reviewers for the helpful comments. Finally, the author thanks the Mannheim Centre for European Social Research (MZES), where preliminary work to this research was conducted.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Code availability** The R-Code of the proposed modifications are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The author declares no conflicts of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Appendix

### A.1 Relative influence

In this paper, we also aim to examine the effect of a missing reimpuation with respect to imputed values of categories C1, C3, C5, and C7 of Table 1 dependent on the sampling fractions and their influence on the variance estimation. To evaluate the influence of certain categories, we compute the relative influence of the USUs of a certain category  $C$  by:

$$Rel.Inf_C = \frac{\sum_{i \in C} w_{dqi}^*}{\sum_{i \in S} w_{dqi}^*}. \quad (13)$$

In detail, this measure shows the influence of the USUs of a certain category on the rescaling point estimator and thus on the variance estimation. Table 4 shows the relative influence of USUs for the categories C1, . . . , C8, according to the two scenarios in Sects. 4.1.1 and 4.1.2 in which a fixed response vector is used.

With respect to small sampling fractions, the relative influence of USUs of category C8 is very high. This category includes the USUs that are drawn for the subsample as well as their PSU und SSU. Also, the USUs of the categories C2, C4, and C6 have a high relative influence. These categories include the USUs for which at least their PSU is drawn for the subsample but not their SSU or not the USUs themselves or not both. The relative influence of the USUs in categories C1, C3, C5 and C7 is very small. Together, they have a relative influence of only 4%. In contrast, the relative influence of the elements of these categories is much larger when the sampling fractions are large. The

relative influence of each category C1, C3, C5, and C7 is 0.05, and together they have a relative influence of 20%. Thus, the influence of the elements of these categories on the variance estimation is much larger in the scenario with large sampling fractions than in the scenario with small sampling fractions. This circumstance is particularly important when analysing the effects of estimator Resc\_only\_Sub. As a result, with estimator Resc\_only\_Sub, we can examine the effect of a missing reimpuation dependent on the variation of the relative influence of certain categories and dependent on the sampling fraction. Particularly, we expect that in the scenario with larger sampling fractions and a large influence of the not reimputed imputed value, a larger unintentional variance component appears that leads to an overestimation of the variance.

## A.2 The appearance of the unintentional variance component in a single stage design in case of a large sampling fraction

In Sect. 1.3.2, we explained that it is necessary to reimpute also the imputed values not drawn for the subsample in the rescaling bootstrap in a single stage design. This is a results of the studies of Bruch (2016, 2019) that show that the reimpuation done only for the imputed values drawn for the subsample leads to an unintentional variance component, which results in the overestimation of the variance. This unintentional variance component is a result of the two different kinds of imputed values in the rescaling point estimator when not reimputing the imputed values not drawn for the subsample. In this case, if the missing value is drawn for the subsample, the reimputed imputed value is used, whereas the imputed value of the original sample is used if the missing value is not drawn for the subsample. Thus, the unintentional variance component occurs when computing the variance over all rescaling bootstrap point estimators and when—for the same missing value—some point estimators of the rescaling bootstrap include the reimputed imputed value, whereas other point estimators include the imputed value of the original sample. The variance component is called *unintentional* because it is not related to the original imputation process that does not include two different kinds of imputed values. The unintentional variance only occurs for increasing sampling fractions in a single-stage sampling design, since elements not drawn for the subsample only have a small impact when sampling fractions are small. Thus, in

**Table 4** Relative influence of the USUs of the different categories in Scenario 1 and 2

Category	Small sampling fraction Sect. 4.1.1	Large sampling fractions Sect. 4.1.2
C1	0.01	0.05
C2	0.19	0.11
C3	0.01	0.05
C4	0.27	0.21
C5	0.01	0.05
C6	0.19	0.11
C7	0.01	0.05
C8	0.32	0.40



the case of large sampling fractions, the imputed values not drawn for the subsample must be reimputed separately from the imputed values drawn for the subsample.

### A.3 Derivation and explanations with respect to the proposed modifications to be a valid estimator

The derivation of the proposed modifications to be a valid estimator for variance component  $V_s(\hat{\theta}_I|\mathbf{z})$  follows from the explanations and derivations of Mashreghi et al. (2014) in conjunction with Shao and Sitter (1996) and Preston (2009).  $V_s(\hat{\theta}_I|\mathbf{z})$  is the sampling variance of the imputed estimator, given the response vector  $\mathbf{z}$ .

In the case without missing values, the rescaling bootstrap for a three stage design of Preston (2009) is reduced to the standard variance estimator in a three stage sampling design when the statistic of interest is linear. This is shown in the Appendix of Preston (2009). Thus, the multistage bootstrap of Preston (2009) covers the three stage sampling variance by the corresponding subsampling at each stage and corresponding weight adjusting. However, in case of the imputation being applied the sampling variance do not pertain to the estimator that is based on complete estimations but the imputed estimator  $\hat{\theta}_I$  and thus the variance component  $V_s(\hat{\theta}_I|\mathbf{z})$ . With respect to Mashreghi et al. (2014), this component can be estimated by using the so called Shao and Sitter procedure (Shao and Sitter 1996). This means paired subsamples from the elements  $\mathbf{y}_I$  and their corresponding response status  $\mathbf{z}$  of the original sample are taken, applying any complete bootstrap method with respect to the sample design (Mashreghi et al. (2014) show this for a one stage stratified design). Afterwards, in case of a deterministic imputation procedure, the reimputation as described in Sect. 1.3.1 is applied to the bootstrap sample (Mashreghi et al. 2014). To transfer this procedure to a multistage sampling design and to derive our modifications, we simply apply the multistage rescaling bootstrap of Preston (2009) that describes a complete bootstrap method with respect to the Shao and Sitter procedure. However, the reimputation has to be conducted with respect to the subsampling process of the rescaling bootstrap that leads to the division in elements drawn for the subsample and not drawn for the subsample, as mentioned in Bruch (2019). Transferring this division to the rescaling bootstrap in the multistage sampling (particularly, to the several stages) leads to the both modifications described in Sect. 2.2. Thus, our modifications represent the Shao Sitter principal with respect to the requirements of the rescaling bootstrap in a multistage design. As a result, it estimates  $V_s(\hat{\theta}_I|\mathbf{z})$  and thus  $V_1$ . As a further consequence, the procedure is not able to estimate  $V_2$  in (7). This is a limitation of the procedure. However, this component is only significant in case of a large  $f$ . In our paper we explain that it is often difficult in a multistage sampling design in practice, to achieve such a large that overall sampling fraction that  $V_2$  makes a strong contribution. Thus, for many situations in survey practice, our proposed modifications should be a valid variance estimator.

For a stochastic imputation procedure, the reimputation may lead to an overestimation of the variance when the imputation procedure includes random components and the subsample size is not of the original sample size. This condition applies also for the rescaling bootstrap. The adjustment of imputed values can be used instead when

metric variables are imputed (Shao 2002). Furthermore, a further variance component appears as a result of the variability of the random component of the imputation procedure. The Shao Sitter procedure and thus our proposed modifications also involves this component (Mashreghi et al. 2014). However, with respect to Mashreghi et al. (2014) for large sampling fractions, the joint bootstrap distribution of the sampling and random component may be a poor approximation of the corresponding joint distribution in the original sampling as result of the rescaling. Thus, further corrections may be necessary in such a case. However, in practice, as said before, sampling fractions may be rather small for multistage designs in practice.

## References

- Alfons A, Filzmoser P, Hulliger B, Kolb J-P, Kraft S, Münnich R, Templ M (2011) Synthetic data generation of silc data, Technical report, AMELI deliverable D6.2. <https://www.uni-trier.de/index.php?id=24676&L=2>
- Bruch C (2016) Varianzschätzung unter Imputation und bei komplexen Stichprobendesigns, PhD thesis, Trier University. <https://ubt.opus.hbz-nrw.de/frontdoor/index/year/2016/docId/734>
- Bruch C (2019) Applying the rescaling bootstrap under imputation: a simulation study. *J Stat Comput Simul* 89(4):641–659. <https://doi.org/10.1080/00949655.2018.1563898>
- Bruch C, Münnich R, Zins S (2011) Variance estimation for complex surveys, Technical report, AMELI deliverable D3.1. <https://www.uni-trier.de/index.php?id=24676&L=2>
- Burgard JP, Ertz F, Merkle H, Münnich R (2020) AMELIA—Data description v0.2.3.1. <http://amelia.uni-trier.de/?p=287>
- Burgard JP, Kolb J-P, Münnich R, Merkle H (2017) Synthetic data for open and reproducible methodological research in social sciences and official statistics. *ASTA Wirtschafts- und Sozialstatistisches Archiv* 11(3–4):233–244
- Chipperfield J, Preston J (2007) Efficient bootstrap for business surveys. *Surv Methodol* 33(2):167–172
- Davison AC, Sardy S (2004) Resampling methods for variance estimation, Technical report, Dacseis Project. Workpackage 5. Deliverable 5.2. <https://www.uni-trier.de/index.php?id=29730>
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7(1):1–26
- Efron B (1994) Missing data, imputation, and the bootstrap. *J Am Stat Assoc* 89(426):463–475. <https://doi.org/10.1080/01621459.1994.10476768>
- Enderle T, Münnich R, Bruch C (2013) On the impact of response patterns on survey estimates from access panels. *Surv Res Methods* 7:91–101. <https://ojs.ub.uni-konstanz.de/srm/article/view/5036>
- European Social Survey (2018) European social survey round 9 sampling guidelines: principles and implementation, Technical report, The ESS Sampling and Weighting Expert Panel. [https://www.europeansocialsurvey.org/docs/round9/methods/ESS9\\_sampling\\_guidelines.pdf](https://www.europeansocialsurvey.org/docs/round9/methods/ESS9_sampling_guidelines.pdf)
- Fay RE (1991) A design-based perspective on missing data variance. In: Proceedings of the 1991 annual research conference, U.S. Bureau of the Census, pp 429–440
- Funaoka F, Saigo H, Sitter RR, Toida T (2006) Bernoulli bootstrap for stratified multistage sampling. *Surv Methodol* 32(2):151–156
- Genz A, Bretz F (2009) Computation of multivariate normal and T probabilities, 1st edn. Springer, Incorporated
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2014). mvtnorm: multivariate normal and t distributions. R package version 1.0-2. <http://CRAN.R-project.org/package=mvtnorm>
- Häder S (2004) Stichproben in der Praxis, Technical report, Mannheim, GESIS-Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). <https://www.gesis.org/gesis-survey-guidelines/statistics/stichproben-in-der-praxis-1-1>
- Haziza D (2009) Chapter 10-Imputation and inference in the presence of missing data. In: Rao C (ed) *Handbook of statistics*, vol 29. Elsevier, Amsterdam, pp 215–246
- Haziza D (2010) Resampling methods for variance estimation in the presence of missing survey data. In: Proceedings of the annual conference of the Italian statistical society

- Kolb J-P (2012) Methoden zur Erzeugung synthetischer Simulationsgesamtheiten, PhD thesis, Trier University. <https://ubt.opus.hbz-nrw.de/frontdoor/index/index/year/2013/docId/590>
- Lohr S (1999) Sampling: design and analysis. Duxbury Press, Pacific Grove
- Lynn P, Häder S, Gabler S, Laaksonen S (2007) Methods for achieving equivalence of samples in cross-national surveys: the European social survey experience. *J Off Stat* 23(1):107–124
- Mashreghi Z, Haziza D, Léger C (2016) A survey of bootstrap methods in finite population sampling. *Stat Surv* 10:1–52. <https://doi.org/10.1214/16-SS113>
- Mashreghi Z, Léger C, Haziza D (2014) Bootstrap methods for imputed data from regression, ratio and hot-deck imputation. *Can J Stat* 42(1):142–167
- Moya PJ, Munoz JF, Álvarez Verdejo E, Blanco-Encomienda FJ (2020) Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest. *Commun Stat Simul Comput* 2020:1–30. <https://doi.org/10.1080/03610918.2020.1859539>
- Preston J (2009) Rescaled bootstrap for stratified multistage sampling. *Surv Methodol* 35(2):227–234
- R Core Team (2021) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rao JNK, Wu CFJ (1988) Resampling inference with complex survey data. *J Am Stat Assoc* 83(401):231–241. <https://doi.org/10.1080/01621459.1988.10478591>
- Rao J, Shao J (1992) Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* 79(4):811–822
- Rao J, Wu C, Yue K (1992) Some recent work on resampling methods for complex surveys. *Surv Methodol* 18(2):209–217
- Saigo H (2010) Comparing four bootstrap methods for stratified three-stage sampling. *J Off Stat* 26(2):193–207
- Saigo H, Shao J, Sitter RR (2001) A repeated half-sample and balanced repeated replications. *Surv Methodol* 27(2):189–196
- Sarkar D (2008) Lattice: multivariate data visualization with R. Springer, New York
- Särndal C-E, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer, New York
- Shao J (2002) Replication methods for variance estimation in complex surveys with imputed data. In: Groves RM, Dillman DA, Eltinge JL, Little RJA (eds) Survey nonresponse. Wiley, New-York
- Shao J, Sitter R (1996) Bootstrap for imputed survey data. *J Am Stat Assoc* 91(435):1278–1288
- Shao J, Steel P (1999) Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *J Am Stat Assoc* 94(445):254–265
- Shao J, Tu D (1995) The jackknife and bootstrap. Springer, New York
- Sitter RR (1992a) Comparing three bootstrap methods for survey data. *Can J Stat* 20(2):135–154. <https://doi.org/10.2307/3315464>
- Sitter RR (1992b) A resampling procedure for complex survey data. *J Am Stat Assoc* 87(419):755–765. <https://doi.org/10.1080/01621459.1992.10475277>
- Wolter K (2007) Introduction to variance estimation. Springer Series in Statistics. Springer, New York. <https://books.google.de/books?id=79pfvdk2zjUC>
- Zhou H, Elliott MR, Raghunathan TE (2016) Synthetic multiple-imputation procedure for multistage complex samples. *J Off Stat* 32(1):231–256

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.