Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

Wild bootstrapping in finite populations with auxiliary information

R.Helmers and M.H.Wegkamp

Department of Operations Reasearch, Statistics, and System Theory

BS-R9520 1995

Report BS-R9520
ISSN 0924-0659

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Wild Bootstrapping in Finite Populations with Auxiliary Information

R. Helmers
CWI
P.O.Box 94079, 1090 GB Amsterdam, The Netherlands

M.H. Wegkamp
Department of Mathematics and Computer Science,
University of Leiden
P.O.Box 9512, 2300 RA Leiden, The Netherlands

Abstract

Consider a finite population u, which can be viewed as a realization of a superpopulation model. A simple ratio model (linear regression, without intercept) with heteroscedastic errors is supposed to have generated u. A random sample is drawn without replacement from u. In this setup a two stage wild bootstrap resampling scheme as well as several other useful forms of bootstrapping in finite populations will be considered. Some asymptotic results for various bootstrap approximations for normalized and Studentized versions of the well-known ratio and regression estimator are given. Bootstrap based confidence intervals for the population total and for the regression parameter of the underlying ratio model are also discussed.

AMS Subject Classification (1991): 62G09, 62E20, 62D05.

Keywords & Phrases: wild bootstrapping, finite populations, ratio estimator, regression estimator, auxiliary information, superpopulation model, ratio model, heteroscedastic errors, bootstrap based confidence intervals.

1 Introduction

Resampling methods for finite populations is an important topic of current research. We refer to [2] for a survey. In the present paper the situation is considered where the finite population is viewed as a realization of a certain superpopulation model.

This enables us to incorporate auxiliary information (past experience) in the statistical analysis. The authors first came across this problem in a 1994 statistical consultation project at CWI with the Netherlands postal services PTT Post. In this setup a new resampling scheme called 'two stage wild bootstrapping' is proposed and studied.

Suppose that the finite population $y_{1,N}, \ldots, y_{N,N}$ is a realization of the following superpopulation model ξ:

$$Y_{i,N} = \beta x_{i,N} + \varepsilon_{i,N} \tag{1.1}$$

where $\varepsilon_{i,N}$ are independent random variables with

$$E_\xi \varepsilon_{i,N} = 0, \;\; E_\xi \varepsilon_{i,N}^2 = \sigma_{i,N}^2, \;\; i = 1, \ldots, N, \quad N = 1, 2, \ldots$$

I.e. a simple ratio model (linear regression, without intercept) with heteroscedastic errors is imposed: $E_\xi Y_{i,N} = \beta x_{i,N}, \; \sigma_\xi^2(Y_{i,N}) = \sigma_{i,N}^2, \; i = 1, \ldots, N$. Usually the $x_{i,N}$'s $(1 \leq i \leq N)$ are all known positive real numbers, but the regression parameter β and the variances $\sigma_{i,N}^2$ $(1 \leq i \leq N)$ are unknown and are to be estimated from data. In the sequel, we shall assume that the auxiliary quantities $x_{i,N}$ are indeed known and also strictly positive. Auxiliary information of the simple form (1.1) is of course not always available. The superpopulation model may have a more complicated structure, e.g. instead of (1.1) one may employ a general linear regression model. Another possibility, recently explored in [4], is to use non-parametric regression.

Given a realization $y_{1,N}, \ldots, y_{N,N}$ we draw a random sample without replacement. Denote by $s = \{i_1, \ldots, i_n\} \subset u = \{1, \ldots, N\}$, the coordinates of the (necessarily distinct) drawn observations. Let P_ξ denote the probability measure generated by the superpopulation model and let π be the probability induced by random sampling without replacement from a finite population. Our asymptotic analysis will only require $n \to \infty$ and $N - n \to \infty$. No further restrictions on the sample fraction $f = n/N$ are needed.

It should be noted that the heteroscedastic regression superpopulation model (1.1) involves $N + 1$ parameters, while the number $n = |s|$ of elements drawn without replacement from the finite population $\{y_{1,1}, \ldots, y_{N,N}\}$ is only equal to fN; i.e., the sample size n is usually considerably less than the number of parameters involved. In such cases consistent estimation of the underlying high-dimensional model (1.1) is clearly impossible, but - essentially because of a CLT type argument similar to the one given by R. Beran in his contribution to a discussion of a paper by C.F.J. Wu ([18])- some appropriate form of wild bootstrapping may still work. In fact, only certain weighted averages of the $\sigma_{i,N}^2$ $(i = 1, \ldots, N)$, like $\sum_{i=1}^N \sigma_{i,N}^2$ or $\sum_{i=1}^N x_{i,N}^2 \sigma_{i,N}^2$, will show up in the asymptotics, rather than all the $\sigma_{i,N}^2$'s $(1 \leq i \leq N)$ separately. Our main results (c.f. section 3) can be viewed as an extension of the already existing theory of wild bootstrapping for heteroscedastic regression models (c.f., e.g. [10], [11], [18]) to the situation considered in the present paper, where such models serve as an underlying superpopulation structure for a finite population $\{y_{1,1}, \ldots, y_{N,N}\}$ at hand. In a way the only thing we do is prove that some suitable forms of wild bootstrapping indeed provide consistent estimates for the distribution of various statistics of interest in a finite population context, such as the population total. In particular, we will propose and study a resampling scheme called *two stage wild bootstrapping*, which not only imitates the underlying ξ-model (1.1), but also properly reflects the random sampling scheme without replacement from a finite population in the 'wild bootstrap world'. Some other useful forms of bootstrapping in finite populations will also be considered.

In this paper we consider the ratio estimator:

$$\hat{\beta}_{RA} = \sum_{i\in s} y_{i,N} / \sum_{i\in s} x_{i,N} \tag{1.2}$$

and the regression estimator:

$$\hat{\beta}_{RE} = \sum_{i\in s} y_{i,N} x_{i,N} / \sum_{i\in s} x_{i,N}^2. \tag{1.3}$$

Note that $\hat{\beta}_{RA}$ and $\hat{\beta}_{RE}$ can be viewed as the solution of a least squares problem: minimize $\sum_{i\in s}(Y_{i,N} - \beta x_{i,N})^2 x_{i,N}^{-1}$, respectively $\sum_{i\in s}(Y_{i,N} - \beta x_{i,N})^2$, as a function of β. More generally, we may as well consider the class of estimators given by:

$$\hat{\beta}_{BLU} = \sum_{i\in s} x_{i,N} y_{i,N} v_{i,N}^{-1} / \sum_{i\in s} x_{i,N}^2 v_{i,N}^{-1} \tag{1.4}$$

If we take $v_{i,N} = 1$ for all i and N then $\hat{\beta}_{BLU} = \hat{\beta}_{RE}$ and if we take $v_{i,N} = x_{i,N}$ for all i and N then $\hat{\beta}_{BLU} = \hat{\beta}_{RA}$. Note that $\hat{\beta}_{BLU}$ is the solution of $\min_\beta \sum_{i\in s}(Y_{i,N} - \beta x_{i,N})^2 v_{i,N}^{-1}$. If the variances $\sigma_{i,N}$ were known, one should certainly take $v_{i,N} = \sigma_{i,N}^2$. Clearly $\hat{\beta}_{RE}$ is the least squares estimate in the homoscedastic model, while $\hat{\beta}_{RA}$ is the least squares estimate when $\sigma_{i,N}^2 = x_{i,N}$.

Our aim is two fold: In the first place we want to validate bootstrap based inference about unknown parameters of the actual finite population $\{y_{1,N}, \ldots, y_{N,N}\}$ at hand, e.g. the parameter $\theta_N = \sum_{i=1}^N y_{i,N}$, the population total. Secondly we focus on the regression parameter of the underlying superpopulation model β, which in a way describes how the finite population is supposed to be generated.

Estimators of θ_N and β based on $\hat{\beta}_{RA}$ and $\hat{\beta}_{RE}$ are discussed in section 2. This section also contains some preliminary results of CLT-type. Some asymptotic theory for various bootstrapped versions of these estimators will be developed in section 3. The proofs of the main results are given in the appendix.

2 Preliminaries

Define $\overline{y}_N = N^{-1}\sum_{i=1}^N y_{i,N}$, the population mean, $\overline{y}_n = n^{-1}\sum_{i\in s} y_{i,N}$, the sample mean. The quantities $\overline{x}_N$, $\overline{x}_n$, $\overline{\varepsilon}_N$, $\overline{\varepsilon}_n$ are defined in the similar way. In addition we define

$\overline{x_N y_N} = N^{-1}\sum_{i=1}^N x_{i,N}y_{i,N}$ and $\overline{x_N^2} = N^{-1}\sum_{i=1}^N x_{i,N}^2$. To begin with, let us investigate the asymptotic behavior of $\hat\beta_{RA}$ and $\hat\beta_{RE}$. If we expand $\hat\beta_{RA}$ and $\hat\beta_{RE}$ in Taylor series, we obtain :

$$\hat\beta_{RA} = \frac{\overline{y}_N}{\overline{x}_N} + \frac{1}{n\overline{x}_N}\sum_{i\in s}(y_{i,N} - \frac{\overline{y}_N}{\overline{x}_N}x_{i,N}) + R_{N,n} \tag{2.1}$$

where $\sqrt{n/(1-f)}R_{N,n} \stackrel{\pi}{\longrightarrow} 0$ in P_ξ probability and

$$\hat\beta_{RE} = \frac{\overline{y_N x_N}}{\overline{x_N^2}} + \frac{1}{n\overline{x_N^2}}\sum_{i\in s}(y_{i,N}x_{i,N} - \frac{\overline{y_N x_N}}{\overline{x_N^2}}x_{i,N}^2) + \tilde R_{N,n} \tag{2.2}$$

where $\sqrt{n/(1-f)}\tilde R_{N,n} \stackrel{\pi}{\longrightarrow} 0$ in P_ξ probability (c.f. the proof of Theorem 2.1). For notational convenience we also define

$$\begin{aligned}
&B_N = \overline{y}_N/\overline{x}_N,\\
&a_{i,N} = y_{i,N} - B_N x_{i,N} = (\beta - B_N)x_{i,N} + \varepsilon_{i,N},\\
&\tilde B_N = \overline{y_N x_N}/\overline{x_N^2},\\
&\tilde a_{i,N} = y_{i,N}x_{i,N} - \tilde B_N x_{i,N}^2 = (\beta - \tilde B_N)x_{i,N}^2 + \varepsilon_{i,N}x_{i,N}.
\end{aligned} \tag{2.3}$$

So we may as well write

$$\hat\beta_{RA} = B_N + \frac{1}{n\overline{x}_N}\sum_{i\in s} a_{i,N} + R_{N,n}$$

$$\hat\beta_{RE} = \tilde B_N + \frac{1}{n\overline{x_N^2}}\sum_{i\in s} \tilde a_{i,N} + \tilde R_{N,n}.$$

Finally, we define the population variances of $a_{1,N},\ldots,a_{N,N}$ resp. $\tilde a_{1,N},\ldots,\tilde a_{N,N}$ by $D_N^2 = N^{-1}\sum_{i=1}^N a_{i,N}^2$ and $\tilde D_N^2 = N^{-1}\sum_{i=1}^N \tilde a_{i,N}^2$. Note that $\sum_{i=1}^N a_{i,N} = \sum_{i=1}^N \tilde a_{i,N} = 0$.

A useful probabilistic tool for our asymptotic analysis is given in the following lemma.

Lemma 2.1 *Set $S_{n,N} = \sum_{i\in s} a_{i,N}$ and $\tilde S_{n,N} = \sum_{i\in s}\tilde a_{i,N}$. Under the following conditions*

$$(A)\quad \max_{1\le i\le N} E_\xi|\varepsilon_{i,N}|^{2+\delta} = \mathcal{O}(1),\ \text{as } N\to\infty \text{ for some } 0<\delta<1;$$

$$(B)\quad \max_{1\le i\le N}|x_{i,N}| = \mathcal{O}(1),\ \text{as } N\to\infty;$$

$$(C)\quad \sum_{i=1}^N x_{i,N}^2 \sim c_1 N,\ \text{as } N\to\infty,\ \text{for some } 0<c_1<\infty;$$

and either

$$(D)\quad \sum_{i=1}^N \sigma_{i,N}^2 \sim c_2 N,\ \text{as } N\to\infty,\ \text{for some } 0<c_2<\infty;$$

or

$$(\tilde D)\quad \sum_{i=1}^N x_{i,N}^2\sigma_{i,N}^2 \sim c_2 N,\ \text{as } N\to\infty,\ \text{for some } 0<c_2<\infty;$$

we have,in P_ξ-probability , as $n \to \infty,\ N-n \to \infty$

$$\frac{S_{n,N}}{\sqrt{f(1-f)}\sqrt{\sum_{i=1}^N a_{i,N}^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \tag{2.4}$$

$$\frac{\tilde{S}_{n,N}}{\sqrt{f(1-f)}\sqrt{\sum_{i=1}^N \tilde{a}_{i,N}^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.5}$$

In addition, we may replace $\sum_{i=1}^N a_{i,N}^2$ by $\sum_{i=1}^N \sigma_{i,N}^2$ in formula formula (2.4) and $\sum_{i=1}^N \tilde{a}_{i,N}^2$ by $\sum_{i=1}^N (x_{i,N}\sigma_{i,N})^2$ in (2.5).

For any estimator $\hat{\beta}$ the 'estimated residuals' are given by

$$\hat{\varepsilon}_{i,N} = Y_{i,N} - \hat{\beta} x_{i,N},\ \ i \in s. \tag{2.6}$$

With the aid of Lemma 2.1, we easily obtain a CLT for normalized ((2.7), (2.9)) and Studentized ((2.8), (2.10)) versions of $\hat{\beta}_{RA}$ and $\hat{\beta}_{RE}$.

Theorem 2.1 *Under the same conditions as in Lemma 2.1, we have in P_ξ-probability, as $n \to \infty,\ N-n \to \infty$,*

$$\frac{\sqrt{n}\,\overline{x}_N}{\sqrt{1-f}\sqrt{N^{-1}\sum_{i\in u} a_{i,N}^2}}(\hat{\beta}_{RA} - B_N) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.7}$$

$$\frac{\sqrt{n}\,\overline{x}_N}{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s} \hat{\varepsilon}_{i,N}^2}}(\hat{\beta}_{RA} - B_N) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.8}$$

Similarly,

$$\frac{\sqrt{n}\,\overline{x_N^2}}{\sqrt{1-f}\sqrt{N^{-1}\sum_{i\in u} \tilde{a}_{i,N}^2}}(\hat{\beta}_{RE} - \tilde{B}_N) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.9}$$

$$\frac{\sqrt{n}\,\overline{x_N^2}}{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s} (\hat{\varepsilon}_{i,N} x_{i,N})^2}}(\hat{\beta}_{RE} - \tilde{B}_N) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.10}$$

Results somewhat related to Theorem 2.1 can be found in [6] and [15].

Remark 2.1 An important issue in practice is estimation of the population total $\theta_N = \sum_{i=1}^N y_{i,N}$ of an actual finite population at hand. We can estimate this quantity simply by $\hat{\beta}_{RA} \cdot x_N$ (set $x_N = \sum_u x_{i,N}$). Proving asymptotic normality of $\hat{\beta}_{RA} \cdot x_N$ becomes straightforward in view of Theorem 2.1. Let $\hat{\theta}_{RA} = \hat{\beta}_{RA} \cdot x_N$ and note that

$$\begin{aligned} \hat{\theta}_{RA} - \theta_N &= x_N \left(\hat{\beta}_{RA} - B_N\right) = \frac{x_N}{n\,\overline{x}_N} \sum_{i \in s} a_{i,N} + x_N \cdot R_{n,N} \\ &= \frac{N}{n} \sum_{i \in s} a_{i,N} + x_N \cdot R_{n,N} \end{aligned}$$

and hence, the counterpart of (2.7) becomes: in P_ξ-probability we have, as $n \to \infty$, $N - n \to \infty$,

$$\frac{\sqrt{n}}{N\sqrt{1-f}\sqrt{N^{-1}\sum_{i \in u} a_{i,N}^2}} \left(\hat{\theta}_{RA} - \theta_N\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.11}$$

Similarly, setting $\tilde{\theta}_N = x_N \cdot \tilde{B}_N$ and $\hat{\theta}_{RE} = \hat{\beta}_{RE} \cdot x_N$, we have in P_ξ-probability, as $n \to \infty$ and $N - n \to \infty$,

$$\frac{\sqrt{n}\,\overline{x_N^2}}{x_N\sqrt{1-f}\sqrt{N^{-1}\sum_{i \in u} \tilde{a}_{i,N}^2}} \left(\hat{\theta}_{RE} - \tilde{\theta}_N\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.12}$$

However $\tilde{\theta}_N$, in contrast with θ_N, is not a very interesting quantity. Instead of $\hat{\theta}_{RE}$, consider:

$$\hat{\theta}_{RE,c} = N\overline{y}_n + \hat{\theta}_{RE}\left(1 - \frac{\overline{x}_n}{\overline{x}_N}\right)$$

as our estimator based on $\hat{\theta}_{RE}$. The estimator $\hat{\theta}_{RE,c}$ for the population total θ_N also appears in [17] as the 'Combined regression through the origin' estimator. We note in passing that Wright ([17]; see also [14]) views $\hat{\theta}_{RA}$ and $\hat{\theta}_{RE}$ as 'predictors' of the random variable θ_N, because the randomness induced by the ξ-model is taken into account. In contrast, we condition on the finite population at hand, but use the auxiliary information to motivate the estimators $\hat{\theta}_{RA}$ and $\hat{\theta}_{RE,c}$, for the population total.

In view of the results obtained in Lemma 2.1 and Theorem 2.1, proving asymptotic normality of $\hat{\theta}_{RE,c}$ is an easy task. Observe that $\hat{\theta}_{RE,c}$ is an asymptotically unbiased estimator of θ_N and that we have the following decomposition:

$$\begin{aligned} \hat{\theta}_{RE,c} - \theta_N &= N\left\{(\overline{y}_n - \overline{y}_N) - \hat{\beta}_{RE}(\overline{x}_n - \overline{x}_N)\right\} \\ &= N\left(\overline{y}_n - \overline{y}_N\right) - N\left(\hat{\beta}_{RE} - \tilde{B}_N\right)\left(\overline{x}_n - \overline{x}_N\right) - N\tilde{B}_N(\overline{x}_n - \overline{x}_N) \\ &= N\left\{(\overline{y}_n - \tilde{B}_N\overline{x}_n) - (\overline{y}_N - \tilde{B}_N\overline{x}_N)\right\} - N\left(\hat{\beta}_{RE} - \tilde{B}_N\right)\left(\overline{x}_n - \overline{x}_N\right) \\ &= \frac{N}{n}\sum_{i \in s} a'_{i,N} - N\left(\hat{\beta}_{RE} - \tilde{B}_N\right)\left(\overline{x}_n - \overline{x}_N\right), \qquad (2.13) \end{aligned}$$

where $a'_{i,N}$, $i \in s$ are centered random variables $a'_{i,N} = [y_{i,N} - \tilde{B}_N x_{i,N}] - [\overline{y}_N - \tilde{B}_N \overline{x}_N]$. Note that there is only a minor difference between $a_{i,N}$ and $a'_{i,N}$ (c.f. (2.3)). As a consequence we have indeed under the same conditions as in Lemma 2.1 asymptotic normality of the first term in (2.13): in P_ξ-probability, as $n \to \infty$ and $N - n \to \infty$,

$$\frac{\sum_{i \in s} a'_{i,N}}{\sqrt{f(1-f)}\sqrt{\sum_{i \in u}(a'_{i,N})^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.14}$$

By Theorem 2.1 we have that $\sqrt{n/(1-f)}\left(\hat{\beta}_{RE} - \tilde{B}_N\right)$ is asymptotically normal and that the difference $|\overline{x_n} - \overline{x_N}|$ tends in π-probability to zero. This result and the identity (2.13) and the asymptotic normality (2.14) entail, in P_ξ-probability, as $n \to \infty$ and $N - n \to \infty$,

$$\frac{\sqrt{n}}{N\sqrt{1-f}\sqrt{N^{-1}\sum_{i \in u}(a'_{i,N})^2}}\left(\hat{\theta}_{RE,c} - \theta_N\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.15}$$

With the aid of (2.11) and (2.15) we can validate normal based confidence intervals for the population total θ_N. Clearly we have to replace the quantities $N^{-1}\sum_{i \in u} a_{i,N}^2$ and $N^{-1}\sum_{i \in u}(a'_{i,N})^2$ appearing in (2.11) and (2.15) by estimates, i.e. by $n^{-1}\sum_{i \in s}\hat{\varepsilon}_{i,N}^2$ and $n^{-1}\sum_{i \in s}\hat{\varepsilon}_{i,N}^2 - (n^{-1}\sum_{i \in s}\hat{\varepsilon}_{i,N})^2$ respectively. Note that we are concerned here with a problem in conditional inference; i.e. the resulting confidence intervals for θ_N are valid asymptotically for a fixed sequence of finite populations $\{y_{1,N}, \ldots, y_{N,N}\}$, $N = 1, 2, \ldots$, for which the results of Theorem 2.1 hold true. In the next section we introduce bootstrap based confidence intervals for θ_N (c.f. Remark 3.2).

Remark 2.2 In the special case that $f \to 0$ then B_N, respectively $\tilde{B}_N$, can with impunity be replaced by β in (2.7) and (2.8), respectively (2.9) and (2.10). In general, however, when the sample size n may be of the same order as the population size N, the bias $B_N - \beta$ of $\hat{\beta}_{RA}$ in estimating β is not negligible, but it can be estimated consistently if we introduce some additional randomness. Notice that

$$E_\xi B_N = \beta, \quad \sigma_\xi^2(B_N) = \sum_{i=1}^N \sigma_{i,N}^2 / (\sum_{i=1}^N x_{i,N})^2, \quad \frac{B_N - \beta}{\sigma_\xi(B_N)} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Let $Z_{N,n}$ be centered normally distributed random variable with variance

$$N\, n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,N}^2 / (\sum_{i=1}^N x_{i,N})^2$$

and define the bias corrected estimate $\hat{\beta}_{RA,b}$ by $\hat{\beta}_{RA} - Z_{N,n}$. The claim is now that $\hat{\beta}_{RA}$ can be replaced by $\hat{\beta}_{RA,b}$ and B_N by β in (2.7) and (2.8). For this purpose we employ the following identity:

$$\sqrt{n}\left(\hat{\beta}_{RA,b} - \beta\right) = \sqrt{n}\left(\hat{\beta}_{RA} - B_N\right) + \sqrt{f}\sqrt{N}\left((B_N - \beta) - Z_{N,n}\right). \tag{2.16}$$

The first term in the right-hand side of (2.16) tends to a (non degenerate) normal distribution by Theorem 2.1, while the second term $\sqrt{N}\,((B_N-\beta)-Z_{N,n})$ tends in P_ξ-probability to zero. This is a direct consequence of our definition of $Z_{N,n}$ since the variance $\sigma_\xi^2(B_N)$ is consistently estimated by $N\,n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,N}^2/(\sum_{i=1}^N x_{i,N})^2$. A similar analysis can easily be carried out, using $\hat{\beta}_{RE}$ instead of $\hat{\beta}_{RA}$.

In practice, one may be interested in confidence intervals for β. A normal based confidence interval for β can now be based on $\hat{\beta}_{RA}-Z_{N,n}$, using (2.8). This interval estimate for β is given by

$$-u_{1-\frac{\alpha}{2}}\frac{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}\hat{\varepsilon}_{i,N}^2}}{\sqrt{n}\overline{x}_N}-Z_{n,N}+\hat{\beta}_{RA}\leq\beta\leq$$
$$\hat{\beta}_{RA}-Z_{n,N}-u_{\frac{\alpha}{2}}\frac{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}\hat{\varepsilon}_{i,N}^2}}{\sqrt{n}\overline{x}_N}, \tag{2.17}$$

where $u_{\frac{\alpha}{2}}$ and $u_{1-\frac{\alpha}{2}}$ denote the $(\alpha/2)$th and $(1-\alpha/2)$th quantile of the standard normal distribution. The confidence interval based on $\hat{\beta}_{RE}$ is obtained from (2.17) by replacing $\hat{\beta}_{RA}-Z_{n,N}$ by a bias corrected version of $\hat{\beta}_{RE}$ (c.f. (2.10)). In the next section we introduce bootstrap based confidence intervals for β (c.f. Remark 3.3).

3 Bootstrapping

Much is known about different forms of bootstrapping in a variety of regression models, such as the *residual method*, the *paired bootstrap*, and the *wild bootstrap*. We refer to [10], [11], [13] and [18]. Bootstrapping in finite population models also received a lot of attention. For instance, the asymptotic behavior of the bootstrap for stratified sampling without replacement from a finite population has recently been explored in [3] (see also the references given in their paper). These authors proposed a two stage resampling procedure in order to mirror the original sampling scheme: simple random sampling without replacement in each stratum and show that the resulting bootstrap is second order efficient. Our situation as described in the introduction, is somewhat intermediate between these two models. We work conditionally given a realization of the superpopulation model (1.1) (i.e. conditionally given the finite population at hand) and employ the auxiliary information provided by the regression model (1.1) to motivate the use of statistics like $\hat{\beta}_{RA}$, $\hat{\beta}_{RE}$, $\hat{\theta}_{RA}$, $\hat{\theta}_{RE}$ and $\hat{\theta}_{RE,c}$ in our study.

In this section we propose and study three different bootstrap resampling schemes for estimating the distributions of normalized and Studentized versions of $\hat{\beta}_{RA}$ and $\hat{\beta}_{RE}$. As an application various bootstrap confidence intervals for the population total $\theta_N=\sum_{i=1}^N y_{i,N}$ and the parameter β of the superpopulation model are given in the Remarks 3.2 and 3.3.

Another approach in general regression problems for estimating β is the so called 'residual method'. This resampling strategy generally fails for the heteroscedastic case since the variances of $\hat{\beta}_{RE}-\beta$ and $\hat{\beta}^*_{RE}-\hat{\beta}_{RE}$ are typically different (see, e.g. [10]). Also in the finite population context as considered in this paper, apparently this resampling method does not work.

Our first and perhaps most promising bootstrap resampling scheme, which we call *two stage wild bootstrapping*, is as follows: given a sample $s = \{i_1, \ldots, i_n\}$ from population u and any estimator $\hat{\beta}$

- calculate 'estimated residuals' $\hat{\varepsilon}_{i,N} = y_{i,N} - \hat{\beta} x_{i,N}$, $i \in s$ (c.f. (2.6));
- *wild bootstrap component.* generate n independent copies $Z_1, \ldots, Z_n$ of a random variable Z with $EZ = 0$ and $EZ^2 = 1$ and set $y^*_{i,N} = \hat{\beta} x_{i,N} + \varepsilon^*_{i,N}$ with $\varepsilon^*_{i,N} = \hat{\varepsilon}_{i,N} Z_i$, $i \in s$;
- *two stage resampling procedure.* put $n' = ([nf]+1) \wedge n$ and $k = [n/n']$.
  stage 1. draw without replacement from s a bootstrap sample $s^*_1 = \{i_{j_1}, \ldots, i_{j_{n'}}\}$ with size n' (the indices $i_{j_1}, \ldots, i_{j_{n'}}$ being necessarily distinct);
  stage 2. repeat this (stage 1) independently k times, replacing the resamples each time, and obtain a bootstrap sample $s^* = s^*_1 \cup \ldots \cup s^*_k$ of size $n^* = n'k$.

The bootstrap version of $\hat{\beta}_{RA}$ becomes $\hat{\beta}^*_{RA} = \sum y^*_{i,N} / \sum x_{i,N}$ and we obtain $\hat{\beta}^*_{RE} = \sum x_{i,N} y^*_{i,N} / \sum x^2_{i,N}$ for $\hat{\beta}_{RE}$ where the summation is now taken over s^*. The integers n' and k can be considered as the design parameters of our resampling procedure. Similarly as in [3], our resampling scheme can be viewed as a stratified resample without replacement, with k identical strata of size n and within-stratum resample size n'. In a way we imitate the ξ-model (c.f. (1.1)) by a wild bootstrap version ξ^*:

$$Y^*_{i,N} = \hat{\beta} x_{i,N} + \varepsilon^*_{i,N}, \ i \in s \tag{3.1}$$

where $\varepsilon^*_{i,N}$ are independent random variables with

$$E_{\xi^*} \varepsilon^*_{i,N} = 0, \ E_{\xi^*} \{\varepsilon^*_{i,N}\}^2 = \hat{\varepsilon}^2_{i,N}.$$

and resample from the finite population $\{Y^*_{i,N}, \ i \in s\}$ our two stage resampling procedure.

We are now ready to state the first main result of this section:

Theorem 3.1 *Set $f' = n'/n$. Under the same conditions as in Lemma 2.1*

$$\frac{\sqrt{n^*}\,\overline{x}_n}{\sqrt{1-f'}\sqrt{\frac{1}{n}\sum_{i\in s}\hat{\varepsilon}^2_{i,N}}} (\hat{\beta}^*_{RA} - \hat{\beta}_{RA}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.2}$$

$$\frac{\sqrt{n^*}\,\overline{x}_n}{\sqrt{1-f'}\sqrt{\frac{1}{n^*}\sum_{i\in s^*}(\varepsilon^*_{i,N})^2}} (\hat{\beta}^*_{RA} - \hat{\beta}_{RA}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.3}$$

Similarly

$$\text{(3.4)} \qquad \frac{\sqrt{n^*}\,\overline{x_n^2}}{\sqrt{1-f'}\sqrt{\frac{1}{n}\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2}}(\hat{\beta}^*_{RE}-\hat{\beta}_{RE}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

$$\text{(3.5)} \qquad \frac{\sqrt{n^*}\,\overline{x_n^2}}{\sqrt{1-f'}\sqrt{\frac{1}{n^*}\sum_{i\in s^*}(x_{i,N}\varepsilon^*_{i,N})^2}}(\hat{\beta}^*_{RE}-\hat{\beta}_{RE}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

in π-probability.

Combination of (3.4) with (2.9) and the fact that $n^* \sim n$ directly yields that the bootstrap approximation given by the left hand side of (3.4) is indeed asymptotically consistent in estimating the distribution of the left hand side of (2.9). A similar remark applies of course to (3.5) and (2.10). The corresponding assertions concerning the normalized and Studentized versions of $\hat{\beta}_{RA}$ are now obvious and therefore omitted.

Remark 3.1 In principle, it appears possible to strengthen the assertions of Theorem 3.1 slightly, by showing that the weak convergences in (3.2)-(3.5) are valid in a stronger almost sure sense, rather than in π-probability. To show this one would certainly need some of the results in [16]. Because Theorem 3.1 (as well as the other theorems in this section) in its present form, appears to be sufficient for practical applications we didn't pursued this point here.

Here is our second result. In a way we adapt the familiar 'paired bootstrap' resampling scheme (c.f. also [11]) to our finite population setup. Similarly as in Theorem 3.1 we apply a two stage procedure to obtain consistent bootstrap approximations.

Theorem 3.2 *Consider the following resampling scheme:*

- *put $n' = ([nf]+1) \wedge n$ and $k = [n/n']$;*
- *stage 1. draw without replacement from s a bootstrap sample s_1^* with size n'; stage 2. repeat the previous step (stage 1) k times independently, replacing the resamples each time, and obtain a bootstrap sample $s^* = s_1^* \cup \ldots \cup s_k^*$ of size $n^* = n'k$;*
- *compute $\hat{\beta}^{*P}_{RA} = \sum_{s^*} y_{i,N} / \sum_{s^*} x_{i,N}$ and $\hat{\beta}^{*P}_{RE} = \sum_{s^*} x_{i,N}y_{i,N} / \sum_{s^*} x^2_{i,N}$.*

Then, under the conditions of Lemma 2.1, we have in π-probability

$$\text{(3.6)} \qquad \frac{\sqrt{n^*}\,\overline{x_n}}{\sqrt{1-f'}\sqrt{\frac{1}{n}\sum_{i\in s}\hat{\varepsilon}^2_{i,N}}}\left(\hat{\beta}^{*P}_{RA}-\hat{\beta}_{RA}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

$$\text{(3.7)} \qquad \frac{\sqrt{n^*}\,\overline{x_n^2}}{\sqrt{1-f'}\sqrt{\frac{1}{n}\sum_{i\in s}x^2_{i,N}\hat{\varepsilon}^2_{i,N}}}\left(\hat{\beta}^{*P}_{RE}-\hat{\beta}_{RE}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

with $f' = n'/n$.

Finally we propose an extremely simple *wild* bootstrap variant. Instead of employing a two stage resampling procedure to mimic random sampling without replacement in the bootstrap world, we chose the Z_i's ('wild bootstrap component') properly to reflect sampling from a finite population as well.

Theorem 3.3 *Consider the following resampling scheme:*

- *generate* n *independent copies of a random variable* Z *with* $\mathbf{E}Z = 0$, $\mathbf{E}Z^2 = 1 - f$ *and* $\mathbf{E}|Z|^{2+\eta} < \infty$ *for some* $\eta > 0$.
- *compute* $y^*_{i,N} = \hat{\beta} x_{i,N} + \hat{\varepsilon}_{i,N} Z_i$, $i \in s$ *for both* $\hat{\beta} = \hat{\beta}_{RA}$ *and* $\hat{\beta} = \hat{\beta}_{RE}$;
- *compute* $\hat{\beta}^{*W}_{RA} = \sum_s y^*_{i,N} / \sum_s x_{i,N}$ *and* $\hat{\beta}^{*W}_{RE} = \sum_s x_{i,N} y^*_{i,N} / \sum_s x^2_{i,N}$.

Then, under the conditions of Lemma 2.1, we have in π*-probability*

$$\frac{\sqrt{n}\,\overline{x_n}}{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}\hat{\varepsilon}^2_{i,N}}}\left(\hat{\beta}^{*W}_{RA} - \hat{\beta}_{RA}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.8}$$

$$\frac{\sqrt{n}\,\overline{x^2_n}}{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}x^2_{i,N}\hat{\varepsilon}^2_{i,N}}}\left(\hat{\beta}^{*W}_{RE} - \hat{\beta}_{RE}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.9}$$

Studentized versions of (3.6), (3.7) and (3.8), (3.9) are certainly also possible. In view of (3.3) and (3.4), this appears to be an easy matter and is therefore omitted.

Remark 3.2 We establish bootstrap confidence intervals for the population total $\theta_N = \sum_{i=1}^N y_{i,N}$ of the actual finite population at hand. For instance, a $(1-\alpha)$ confidence interval for θ_N, obtained by the two stage wild bootstrapping procedure, is given by

(3.10)
$$-c^*_{1-\frac{\alpha}{2}}\frac{N\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}\hat{\varepsilon}^2_{i,N}}}{\sqrt{n}} + \hat{\theta}_{RA} \le \theta_N \le \hat{\theta}_{RA} - c^*_{\frac{\alpha}{2}}\frac{N\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}\hat{\varepsilon}^2_{i,N}}}{\sqrt{n}}$$

where $c^*_{\frac{\alpha}{2}}$ and $c^*_{1-\frac{\alpha}{2}}$ denote the $(\frac{\alpha}{2})$th and $(1-\frac{\alpha}{2})$th quantile of the (bootstrap) distribution of

$$\frac{\sqrt{n^*}}{n\sqrt{1-f'}\sqrt{\frac{1}{n^*}\sum_{i\in s^*}(\varepsilon^*_{i,N})^2}}(\hat{\theta}^*_{RA} - \hat{\theta}_{RA}). \tag{3.11}$$

This is a simple consequence of (3.3), similarly as (2.11) is easily implied by (2.7). The two stage wild bootstrap confidence interval (3.10) for θ_N has coverage probability $1 - \alpha + o(1)$, as $n \to \infty$, $N - n \to \infty$. Because Studentization was employed in the construction of (3.10) and the two stage resampling procedure mimics sampling without replacement from a finite population in the bootstrap world one may expect that in fact the interval is second order efficient, i.e. the coverage probability is equal to $1-\alpha+o(\frac{1}{\sqrt{n}})$, provided the distribution of the Z's is chosen such that the skewness of (2.11) (with $N^{-1}\sum_{i\in u} a_{i,N}^2$ replaced by $n^{-1}\sum_{i\in s}\hat{\varepsilon}_{i,N}^2$) matches the skewness of (3.11) (see also section 4). Similarly, a $(1-\alpha)$ bootstrap confidence interval for θ_N based on $\hat{\theta}_{RE,c}$ is given by (3.10) with $\hat{\theta}_{RA}$ replaced by $\hat{\theta}_{RE,c}$ and $n^{-1}\sum_{i\in s}\hat{\varepsilon}_{i,N}^2$ by $n^{-1}\sum_{i\in s}\hat{\varepsilon}_{i,N}^2 - (n^{-1}\sum_{i\in s}\hat{\varepsilon}_{i,N})^2$, while the bootstrap quantiles $c^*_{\frac{\alpha}{2}}$ and $c^*_{1-\frac{\alpha}{2}}$ are now determined by (3.11), with $\sum_{i\in s^*}(\varepsilon_{i,N}^*)^2$ replaced by $\sum_{i\in s^*}(x_{i,N}\varepsilon_{i,N}^*)^2$ and $\hat{\theta}_{RE}(\hat{\theta}^*_{RE})$ instead of $\hat{\theta}_{RA}(\hat{\theta}^*_{RA})$.

Remark 3.3 Wild bootstrapping can now also be employed for the construction of confidence intervals for the parameter β of the superpopulation model ξ (c.f. (1.1)). Replace in (2.17) the normal quantiles by the bootstrap quantiles of the distribution of

$$\frac{\sqrt{n^*}\,\overline{x_n}}{\sqrt{1-f'}\sqrt{\frac{1}{n^*}\sum_{i\in s^*}(\varepsilon_{i,N}^*)^2}}(\hat{\beta}^*_{RA}-\hat{\beta}_{RA}),$$

the left hand side of (3.3). A difference with Remark 2.2 is that the bootstrap version for $\hat{\beta}$ is unbiased, whereas for $f \not\to 0$, $\hat{\beta}-\beta$ is biased.

4 Possible extensions

In this section we discuss very briefly some possible extensions of our results. To begin with let us note that so far we have only dealt with first order asymptotics. I.e. we have shown that the bootstrap approximations given in section 3 are asymptotically consistent. However, the question remains: how well do these bootstrap approximations (c.f. Theorems 3.1, 3.2 and 3.3) compare with the more traditional normal approximations (c.f. Theorem 2.1)? A second order analysis (involving Edgeworth expansions) is necessary to confirm our conjecture that two stage wild bootstrapping of Studentized statistics like (3.3) and (3.4) is second order correct. At this point we may add an assumption on EZ^3 (c.f. also [10]) and a more careful choice of the design parameters n' and k of the resampling scheme will be needed. We don't expect however, that the other two bootstrap methods we discussed in section 3 will be second order efficient. An investigation along these lines appears feasible, but outside the scope of the present paper. The authors hope to report on these matters elsewhere.

Secondly one may be interested in a slight extension of our set up, namely the case that the superpopulation model ξ (c.f.(1.1)) is still valid, but stratified sampling, in stead of simple random sampling, from the finite population is employed. I.e., we now assume that the population $\Pi = \{y_{1,N},\ldots,y_{N,N}\}$ at hand, is divided into L disjoint strata $\Pi_1,\ldots,\Pi_L$, where $\Pi=\bigcup_{l=1}^{L}\Pi_l$. We note in passing that one may try to use the super population model (1.1) to obtain an efficient stratification of Π (c.f. [17]; see also [14]). However, at this point, we will consider the situation that the stratification of Π

in L subpopulations (strata) is already known a priori, as appears to be frequently the case in sample surveys. In each stratum (of size N_l) a simple random sample (of size n_l) is drawn. I.e., if we take a sample (without replacement) of size n_l from stratum Π_l and require $n_l \to \infty$ and $N_l - n_l \to \infty$ in each stratum Π_l $(l = 1, \ldots, L)$, while the number of strata L is kept fixed, we may extend the results of this paper in a fairly straightforward manner. We omit further details.

In the third place one may want to extend our results to a more general class of statistics, e.g. the one described in (1.4). This appears to be a straightforward matter and is therefore omitted. Finally it may be of some interest - as already alluded to in the introduction - to consider superpopulation models of a more general type, such as the general linear regression model. Such an extension is certainly possible, but will not be pursued here.

5 Appendix

This section contains the proofs. We only prove the statements in the case of the regression estimator but for the ratio estimator similar arguments hold.

Proof of Lemma 2.1. According to [5] and [8], we have to show that the Hájek-Lindeberg condition

$$(5.1) \qquad \frac{1}{\sum_u \tilde{a}_{i,N}^2} \sum_u \tilde{a}_{i,N}^2 \mathbf{1}\left\{ |\tilde{a}_{i,N}| > \eta \sqrt{f(1-f) \sum_u \tilde{a}_{i,N}^2} \right\} \to 0 \text{ for all } \eta > 0$$

as $N - n \to \infty$, $n \to \infty$ holds, in P_ξ-probability. We shall use the Von Bahr-Esseen inequality (c.f. [1]) to prove that

$$(5.2) \qquad \left| \frac{1}{N} \sum_{i=1}^N \tilde{a}_{i,N}^2 - \frac{1}{N} \sum_{i=1}^N (x_{i,N}\sigma_{i,N})^2 \right| \xrightarrow{P_\xi} 0, \; N \to \infty$$

holds. Rewrite first $\tilde{D}_N^2 = N^{-1} \sum_{i=1}^N \tilde{a}_{i,N}^2$ into

$$(5.3) \quad \tilde{D}_N^2 = \frac{1}{N} \sum_{i=1}^N (\beta - \tilde{B}_N)^2 x_{i,N}^4 + \frac{1}{N} \sum_{i=1}^N x_{i,N}^2 \varepsilon_{i,N}^2 + \frac{2}{N} \sum_{i=1}^N (\beta - B_N) x_{i,N}^3 \varepsilon_{i,N}.$$

Note that

$$E_\xi \tilde{B}_N = \beta \;\text{ and }\; \sigma_\xi^2(\tilde{B}_N) = (\sum_{i=1}^N x_{i,N}^2)^{-2} \sum_{i=1}^N x_{i,N}^2 \sigma_{i,N}^2.$$

Conditions (C) and ($\tilde{\mathrm{D}}$) ensure that $\tilde{B}_N - \beta = \mathcal{O}_{P_\xi}(N^{-1/2})$ by Chebyshev's inequality and therefore together with condition (B) we have $(\beta - \tilde{B}_N)^2 N^{-1} \sum_{i=1}^N x_{i,N}^4 =$

$\mathcal{O}_{P_\xi}(N^{-1})$. The last term in equation (5.3) vanishes too, (order $O_{P_\xi}(N^{-1/2})$) because by the Cauchy-Schwarz inequality

$$|\frac{1}{N}\sum_{i=1}^{N} x_{i,N}^3 \varepsilon_{i,N}| \leq \left(\frac{1}{N}\sum_{i=1}^{N} x_{i,N}^6\right)^{1/2} \left(\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,N}^2\right)^{1/2}.$$

It remains to consider the middle term in (5.3) and this term will in fact turn out to be the dominant one. By the Von Bahr-Esseen inequality, we have for $p \in (0,1)$ a constant $C_p > 0$ such that

$$P_\xi(|\frac{1}{N}\sum_{i=1}^{N}\{x_{i,N}^2(\varepsilon_{i,N}^2 - \sigma_{i,N}^2)\}| > \eta) \leq \frac{C_p \sum_{i=1}^N E_\xi |x_{i,N}^2(\varepsilon_{i,N}^2 - \sigma_{i,N}^2)|^{1+p}}{(N\eta)^{1+p}} \to 0,$$

where also condition (A) has been invoked. This implies that (5.2) is true. In view of (5.2), it suffices to check whether

$$\frac{1}{N\tilde{\sigma}_N^2}\sum \tilde{a}_{i,N}^2 \mathbf{1}\left\{|\tilde{a}_{i,N}| > \eta\sqrt{f(1-f)N}\tilde{\sigma}_N\right\} \to 0 \text{ for all } \eta > 0 \tag{5.4}$$

if $n \to \infty$, $N - n \to \infty$ holds, in P_ξ-probability, where

$$\tilde{\sigma}_N^2 = N^{-1}\sum_{i=1}^{N}(x_{i,N}\sigma_{i,N})^2. \tag{5.5}$$

Since $\tilde{a}_{i,N} = (\beta - \tilde{B}_N)x_{i,N}^2 + x_{i,N}\varepsilon_{i,N}$ (c.f. (2.3)), we can bound the left-hand side of (5.4) above by

$$\begin{aligned}&\frac{1}{N\tilde{\sigma}_N^2}\sum \tilde{a}_{i,N}^2 \mathbf{1}\left\{|\beta - \tilde{B}_N|x_{i,N}^2 > \frac{\eta}{2}\sqrt{f(1-f)N}\tilde{\sigma}_N\right\} + \\ &\frac{1}{N\tilde{\sigma}_N^2}\sum \tilde{a}_{i,N}^2 \mathbf{1}\left\{|x_{i,N}\varepsilon_{i,N}| > \frac{\eta}{2}\sqrt{f(1-f)N}\tilde{\sigma}_N\right\} \\ &:= d_{n,N}^{(1)}(\frac{\eta}{2}) + d_{n,N}^{(2)}(\frac{\eta}{2}).\end{aligned}$$

Assumption (B) and the fact that $\tilde{B}_N - \beta = \mathcal{O}_{P_\xi}(N^{-1/2})$ entail that for every $\eta > 0$, $\lim_{n\to\infty,\, N-n\to\infty} d_{n,N}^{(1)}(\eta/2) = 0$ in P_ξ-probability.
Application of Hölder's inequality yields

$$E_\xi d_{N,n}^{(2)}(\frac{\eta}{2}) \leq \frac{1}{N\tilde{\sigma}_N^2}\sum_{i=1}^{N}(E_\xi \tilde{a}_{i,N}^{2(1+\delta)})^{\frac{1}{1+\delta}} \left(P\left\{|\varepsilon_{i,N}| > \frac{\eta}{2}\sqrt{\frac{n}{N}\cdot\frac{(N-n)}{N}}\sqrt{N}\tilde{\sigma}_N\right\}\right)^{\frac{\delta}{1+\delta}} \to 0$$

as $\max_{1\leq i\leq N} E_\xi \tilde{a}_{i,N}^{2(1+\delta)} < \infty$ and

$$\max_{1\leq i\leq N} P_\xi\left\{|\varepsilon_{i,N}| > \frac{\eta}{2}\sqrt{\frac{n}{N}\cdot\frac{(N-n)}{N}}\sqrt{N}\tilde{\sigma}_N\right\} \leq \max_{1\leq i\leq N}\frac{E_\xi|\varepsilon_{i,N}|^{2+\delta}}{(\tilde{\sigma}_N^2\frac{\eta}{2})^{2+\delta}}\cdot\frac{1}{(f(N-n))^{\frac{2+\delta}{2}}} \to 0$$

for some $0 < \delta < 1$, $n \to \infty, N - n \to \infty$. Hence $d^{(2)}_{n,N}(\eta/2) \xrightarrow{P_\xi} 0$ for $n \to \infty, N - n \to \infty$ $\forall \eta > 0$.

This proves the lemma.□

Proof of Theorem 2.1. According to the first order Taylor expansion (2.2) we may write (c.f the display after (2.3)):

$$\hat{\beta}_{RE} - \tilde{B}_N = \frac{1}{n\overline{x^2_N}} \sum_{i \in s} \tilde{a}_{i,N} + \tilde{R}_{N,n}$$

$$\tilde{R}_{N,n} = 2(\overline{x^2_n} - \overline{x^2_N})^2 \frac{\xi(xy)}{(\xi(x))^3} - (\overline{x^2_n} - \overline{x^2_N})(\overline{x_n y_n} - \overline{x_N y_N}) \frac{1}{(\xi(x))^2}$$

with $\xi(xy)$ between $\overline{x_n y_n}$ and $\overline{x_N y_N}$ and $\xi(x)$ between $\overline{x^2_n}$ and $\overline{x^2_N}$.

First we show that

$$\frac{\sqrt{n}\overline{x^2_N}}{\sqrt{(1-f)N^{-1}\sum_u \tilde{a}^2_{i,N}}} \tilde{R}_{N,n} \xrightarrow{\pi} 0. \tag{5.6}$$

Define the inclusion indicators

$$\delta_{i,N} = \begin{cases} 1 & \text{if } i \in s; \\ 0 & \text{otherwise.} \end{cases}$$

It is well-known that for $i = 1, \ldots, N$, $j = 1, \ldots, N$ with $i \neq j$, (c.f. [9])

$$E\delta_{i,N} = f, \quad \sigma^2(\delta_{i,N}) = f(1-f), \quad \mathrm{Cov}(\delta_{i,N}, \delta_{j,N}) = -\frac{n}{N^2} \frac{N-n}{N-1}$$

Easy calculations yield

$$E_\pi(\overline{x^2_n}) = \overline{x^2_N} = \mathcal{O}(1), \quad \sigma^2_\pi(\overline{x^2_n}) = \mathcal{O}(\frac{1-f}{n}); \; E_\xi \overline{x_N y_N} = \frac{1}{N} \sum_u \beta x^2_{i,N} = \mathcal{O}(1),$$

$$\sigma^2_\xi(\overline{x_N y_N}) = \mathcal{O}(N^{-1}); \; E_\pi \overline{x_n y_n} = \overline{x_N y_N} = \mathcal{O}_{P_\xi}(1), \quad \sigma^2_\pi(\overline{x_n y_n}) = \mathcal{O}_{P_\xi}(\frac{1-f}{n}).$$

Thus both $\overline{x^2_n} - \overline{x^2_N}$ and $\overline{x_n y_n} - \overline{x_N y_N}$ are of order $\mathcal{O}_\pi\left(\sqrt{(1-f)/n}\right)$ by Chebyshev's inequality. Note that condition (C) implies that $\xi(x) \asymp c_1$. It follows now directly that $\tilde{R}_{N,n} = \mathcal{O}_\pi((1-f)/n)$ in P_ξ probability.

The scaling factor in (5.6) is of order $\mathcal{O}(\sqrt{n/(1-f)})$ so indeed the influence of the remainder term is asymptotically negligible, i.e. (5.6) is true.

Note that Lemma 2.1 implies

$$\frac{1}{\sqrt{f(1-f)\sum_u \tilde{a}^2_{i,N}}} \sum_s \tilde{a}_{i,N} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{5.7}$$

Hence (2.9) is proved. It remains to check the validity of (2.10). It suffices clearly to show that

$$\frac{1}{n}\sum_s(\hat{\varepsilon}_{i,N}x_{i,N})^2/\frac{1}{N}\sum_u(\hat{\varepsilon}_{i,N}x_{i,N})^2 \xrightarrow{\pi} 1 \tag{5.8}$$

and

$$\left|\frac{1}{N}\sum_u(\hat{\varepsilon}_{i,N}x_{i,N})^2 - \frac{1}{N}\sum_u \tilde{a}^2_{i,N}\right| \xrightarrow{P_\xi} 0, \tag{5.9}$$

where $\hat{\varepsilon}_{i,N} = Y_{i,N} - \hat{\beta}_{RE}x_{i,N}$ for all $i \in u$. But this is an easy matter in view of condition $\tilde{D}$ and the last assertions of Lemma 2.1.□

Next we first prove Theorem 3.2 ('paired bootstrap') and then proceed by proving Theorem 3.1 ('two stage wild bootstrapping'). In part, we will exploit the argument given in our proof of Theorem 3.2 again in the somewhat more complicated setting we consider in the proof of Theorem 3.1. Finally, we conclude the appendix with a fairly easy proof of Theorem 3.3.

Proof of Theorem 3.2. We start with the following expansion:

$$\begin{aligned}\hat{\beta}^{*P}_{RE} - \hat{\beta}_{RE} &= \frac{1}{n^*\,\overline{x^2_n}}\sum_{i\in s^*}\hat{\varepsilon}_{i,N}x_{i,N} + R^* \\ &= \frac{1}{n^*\overline{x^2_n}}\sum_{j=1}^{k}\sum_{i\in s^*_j}\hat{\varepsilon}_{i,N}x_{i,N} + R^*,\end{aligned} \tag{5.10}$$

where the remainder R^* is given by

$$R^* = 2(\overline{x^2_{n^*}} - \overline{x^2_n})^2\frac{\xi(xy)}{\{\xi(x)\}^3} - (\overline{x^2_{n^*}} - \overline{x^2_n})(\overline{x_{n^*}y_{n^*}} - \overline{x_ny_n})\frac{1}{\{\xi(x)\}^2}$$

with $\xi(x)$ between $\overline{x^2_{n^*}}$ and $\overline{x^2_n}$ and $\xi(xy)$ between $\overline{x_{n^*}y_{n^*}}$ and $\overline{x_ny_n}$. We suppress the dependence of R^* on n', k and n.

Next we prove a CLT for $\sum_{i\in s^*_1}\hat{\varepsilon}_{i,N}x_{i,N}$ where s^*_1 is the first bootstrap sample obtained by random sampling without replacement from s. For this purpose we first observe that $\sum_{i\in s}\hat{\varepsilon}_{i,N}x_{i,N} = 0$. Note that a more general statement than (5.8) holds: for every sequence of Borel sets $C_1, \ldots, C_N$,

$$\left|\frac{1}{n}\sum_s(\hat{\varepsilon}_{i,N}x_{i,N})^2\mathbf{1}_{C_i} - \frac{1}{N}\sum_u(\hat{\varepsilon}_{i,N}x_{i,N})^2\mathbf{1}_{C_i}\right| \xrightarrow{\pi} 0,\ N\to\infty, n\to\infty.$$

This implies that the Hájek-Lindeberg condition

$$\frac{1}{\sum_s(x_{i,N}\hat{\varepsilon}_{i,N})^2}\sum_s(x_{i,N}\hat{\varepsilon}_{i,N})^2\mathbf{1}\left\{|x_{i,N}\hat{\varepsilon}_{i,N}| > \eta\sqrt{f'(1-f')\sum_s(x_{i,N}\hat{\varepsilon}_{i,N})^2}\right\} \to 0,$$

for all $\eta > 0$ may with impunity be replaced by

$$\frac{1}{\sum_u (x_{i,N}\hat{\varepsilon}_{i,N})^2} \sum_u (x_{i,N}\hat{\varepsilon}_{i,N})^2 \mathbf{1}\left\{ |x_{i,N}\hat{\varepsilon}_{i,N}| > \eta \sqrt{f'(1-f')f \sum_u (x_{i,N}\hat{\varepsilon}_{i,N})^2} \right\} \to 0,$$

for all $\eta > 0$ where $f' = n'/n$. Recall that $x_{i,N}\hat{\varepsilon}_{i,N} = (\beta - \hat{\beta}_{RE})x_i^2 + \varepsilon_i x_i$ and deduce that

$$\begin{aligned} |\beta - \hat{\beta}_{RE}| &\leq |\beta - \tilde{B}_N| + |\tilde{B}_N - \hat{\beta}_{RE}| \\ &= \mathcal{O}_{P_\xi}\left(N^{-1/2}\right) + \mathcal{O}_\pi \left(\frac{1-f}{n}\right)^{1/2}. \end{aligned}$$

Another application of the Von Bahr-Esseen inequality shows that

$$\left| \frac{1}{N} \sum_u (\hat{\varepsilon}_{i,N} x_{i,N})^2 - \frac{1}{N} \sum_u (\sigma_i x_i)^2 \right| \xrightarrow{P_\xi} 0.$$

This reduces the Hájek-Lindeberg condition to

$$\begin{aligned} &\frac{1}{N\tilde{\sigma}_N^2} \sum_u (x_{i,N}\hat{\varepsilon}_{i,N})^2 \mathbf{1}\left\{ |\beta - \hat{\beta}_{RE}| x_i^2 > \frac{\eta}{2}\sqrt{f'(1-f')n}\tilde{\sigma}_N \right\} + \\ &\frac{1}{N\tilde{\sigma}_N^2} \sum_u (x_{i,N}\hat{\varepsilon}_{i,N})^2 \mathbf{1}\left\{ |x_{i,N}\varepsilon_{i,N}| > \frac{\eta}{2}\sqrt{f'(1-f')n}\tilde{\sigma}_N \right\} \to 0 \end{aligned}$$

for all $\eta > 0$, in P_ξ-probability, whenever $n' \to \infty$ and $n - n' \to \infty$. A similar reasoning as in the proof of Lemma 2.1 entails

$$\frac{1}{\sqrt{f'(1-f')}\sqrt{\sum_{i\in s}(\hat{\varepsilon}_{i,N} x_{i,N})^2}} \sum_{s_1^*} \hat{\varepsilon}_{i,N} x_{i,N} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{5.11}$$

whenever $n' \to \infty$ and $n - n' \to \infty$. Since the second stage resampling is performed independently, we have then

$$\frac{1}{\sqrt{f'(1-f')}\sqrt{\sum_{i\in s}(\hat{\varepsilon}_{i,N} x_{i,N})^2}} \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \sum_{s_j^*} \hat{\varepsilon}_{i,N} x_{i,N} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{5.12}$$

This completes the proof for the case that n' and $n - n'$ both get large. From the definition of the design-parameters n' and k of the two stage wild bootstrap resampling scheme it follows that $n - n' \sim n - (n^2/N) = n(N-n)/N \to \infty$. If n' remains bounded (note that this happens only when the original sample fraction $f \to 0$ at the rate $\mathcal{O}(N^{-\frac{1}{2}})$) however, a slightly different argument is needed. In this case we may simply replace with impunity sampling *without* replacement by sampling *with* replacement. This is an easy consequence of a result on the difference in the total variation distance between sampling with and without replacement by D.Freedman ([7]). Relation (5.12) now follows directly from Lindeberg's CLT for triangular arrays. Therefore (c.f. relation (5.10))

$$\frac{n^* \overline{x_n^2}}{\sqrt{k}\sqrt{f'(1-f')\ \sum \hat{\varepsilon}_i x_i^2}} \left(\hat{\beta}_{RE}^{*P} - \hat{\beta}_{RE}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

because, by similar reasoning as in the proof of Theorem 2.1, we have

$$\frac{n^* \overline{x_n^2}}{\sqrt{k}\sqrt{f'(1-f')}\sqrt{\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2}} R^* \xrightarrow{\pi^*} 0,$$

where π^* denotes the probability induced by random sampling without replacement from sample s. □

Proof of Theorem 3.1. We first prove (3.4). The bootstrap sample can be written as a union of k independent samples: $s^* = s_1^* \cup \ldots \cup s_k^*$. Consider the stochastic expansion of $\hat{\beta}_{RE}^* - \hat{\beta}_{RE}$:

$$\begin{aligned}\hat{\beta}_{RE}^* - \hat{\beta}_{RE} &= \frac{1}{n^* \overline{x_n^2}} \sum_{i\in s^*} \hat{\varepsilon}_{i,N} x_{i,N} Z_i + R^* \\ &= \frac{1}{n^* \overline{x_n^2}} \sum_{j=1}^{k} \sum_{i\in s_j^*} \hat{\varepsilon}_{i,N} x_{i,N} Z_i + R^*. \qquad (5.13)\end{aligned}$$

Observe that generally $\sum_s \hat{\varepsilon}_{i,N} x_{i,N} z_i \neq 0$ although it has expectation zero under P_Z, the probability induced by $Z_1, \ldots, Z_n$. Define

$$M(Z) = \frac{1}{n}\sum_s \hat{\varepsilon}_{i,N} x_{i,N} Z_i, \quad \Delta^2(Z) = \frac{1}{n}\sum_s \left(\hat{\varepsilon}_{i,N} x_{i,N} Z_i - M(Z)\right)^2.$$

In the sequel we consider the following statistic:

$$T_n^* = \frac{n^* \overline{x_n^2}}{\sqrt{k}\sqrt{f'(1-f')}\sqrt{n}} \left(\hat{\beta}_{RE}^* - \hat{\beta}_{RE}\right) =$$
$$\frac{1}{\sqrt{k}} \sum_{j=1}^{k} \frac{1}{\sqrt{f'(1-f')}\sqrt{n}} \sum_{i\in s_j^*} \left[\hat{\varepsilon}_{i,N} x_{i,N} Z_i - M(Z)\right] + \sqrt{\frac{k}{1-f'}} M(Z) + \frac{\sqrt{n^*}\,\overline{x_n^2}}{\sqrt{1-f'}} R^*.$$

The left hand side of (3.4) is precisely equal to $T_n^* / \sqrt{n^{-1}\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2}$. We can not apply the Erdös-Rényi CLT for samples drawn without replacement from a finite population (see [5], [8]) at once, we first have to condition on $Z = (Z_1, \ldots, Z_n)$. Similarly as in the proof of Theorem 3.2 (c.f. the argument leading to (5.12) can be repeated, with

$x_{i,N}\hat{\varepsilon}_{i,N}$ replaced by $x_{i,N}\hat{\varepsilon}_{i,N}z_i$), one can show that, for every realization $z = (z_1, \ldots, z_n)$, we have

$$\frac{1}{\sqrt{k}\sqrt{f'(1-f')n}\Delta(z)} \sum_{i \in s^*} (x_{i,N}\hat{\varepsilon}_{i,N}z_i - M(z)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{5.14}$$

and

$$\frac{\sqrt{n^*}\,\overline{x_n^2}}{\sqrt{k}\sqrt{f'(1-f')n}\Delta(z)} R^* \xrightarrow{\pi^*} 0. \tag{5.15}$$

Thus we have proved that, for every value of the vector Z,

$$\left| P^*\left(T_n^* \le y \mid Z = z\right) - \Phi\left(\frac{y - M(z)}{\Delta(z)}\right)\right| \to 0, \text{ as } n \to \infty, \tag{5.16}$$

where Φ denotes the standard normal distribution. Next we take the expected value w.r.t. Z:

$$P^*\left(T_n^* \le y\right) = E_Z P^*\left(T_n^* \le y \mid Z\right) = E_Z\Phi\left(\frac{y - M(Z)}{\Delta(Z)} \mid Z\right) + E_Z\, r_n(Z).$$

As $r_n(z) \to 0$, pointwise in z, and $|r_n(z)| \le 2$, we have by dominated convergence $E_Z|r_n(Z)| \to 0$ as $n \to \infty$.

Now, we employ the following Taylor expansion:

$$\Phi\left(\frac{y - M(z)}{\Delta(z)}\right) = \Phi\left(\frac{y}{\Delta}\right) + \left\{\left|\frac{y - M(Z)}{\Delta(Z)} - \frac{y}{\Delta}\right|\right\} \varphi(\xi) \tag{5.17}$$

where $\Delta^2 = E_Z\Delta^2(Z) = n^{-1}\sum_s (x_{i,N}\hat{\varepsilon}_{i,N})^2 - n^{-2}(\sum_s x_{i,N}\hat{\varepsilon}_{i,N})^2 \sim n^{-1}\sum_s (x_{i,N}\hat{\varepsilon}_{i,N})^2$, φ is the standard normal density and and ξ is a point between $(y - M(z))/\Delta(z)$ and y/Δ. Note that, for every fixed value of y, the difference $|[y - M(Z)]/\Delta(Z) - y/\Delta|$ goes in P_Z-probability to zero because $EM(Z) = 0$, $EM^2(Z) = \mathcal{O}(n^{-1})$ uniform in Z, $E\Delta^2(Z) \sim n^{-1}\sum_s (\hat{\varepsilon}_{i,N}x_{i,N})^2$ and $E|\Delta^2(Z) - \Delta^2|^r = o(1)$ for some $r > 0$ by an application of the Von Bahr-Esseen inequality. Also note that $|[y - M(Z)]/\Delta(Z) - y/\Delta|\,\varphi(\xi) \le 2$ because of (5.17) and the triangle inequality. As a result we obtain by dominated convergence that $E_Z\Phi\left([y - M(Z)]/\Delta(Z)\right) = \Phi(y/\Delta) + o(1)$. This completes the proof of (3.4).

Finally we prove the weak convergence of the Studentized version (3.5). It clearly suffices to show that we may replace $n^{-1}\sum_{i \in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2$ by $\Delta^2(Z)$ in (3.5). Recall the definition of $s^* = \cup_{j=1}^k s_j^*$ so that

$$\frac{1}{n^*}\sum_{i \in s^*}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2 = \frac{1}{k}\sum_{j=1}^k \frac{1}{n'}\sum_{i \in s_j^*}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2.$$

By the triangle inequality we have

$$(5.18)\qquad \left|\frac{1}{n^*}\sum_{i\in s^*}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2-\frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N})^2\right|\le$$

$$\left|\frac{1}{n^*}\sum_{i\in s^*}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2-\frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2\right|+\left|\frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2-\frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N})^2\right|.$$

As before, conditionally on $Z_1,\ldots,Z_n$, the first part of (5.18) tends to zero in π^*-probability, whereas the second part goes to zero in P_Z-probability. Notice that at this point we also invoke moments inequalities and the Von Bahr-Esseen inequality. □

Proof of Theorem 3.3. Rewrite $\hat{\beta}^{*W}_{RE}$ in

$$\hat{\beta}^{*W}_{RE}=\frac{\sum_{i\in s}x_{i,N}y^*_{i,N}}{\sum_{i\in s}x^2_{i,N}}=\frac{\sum_{i\in s}\{\hat{\beta}_{RE}x^2_{i,N}+x_{i,N}\hat{\varepsilon}_{i,N}Z_i\}}{\sum_{i\in s}x^2_{i,N}}$$

$$(5.19)\qquad =\hat{\beta}_{RE}+\frac{\sum_{i\in s}x_{i,N}\hat{\varepsilon}_{i,N}Z_i}{\sum_{i\in s}x^2_{i,N}}.$$

As $EZ^2=1-f$ and $E|Z|^{2+\zeta}<\infty$ for some $\zeta>0$, we have by an application of the Von Bahr-Esseen inequality

$$(5.20)\qquad \left|\frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2-\frac{1-f}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N})^2\right|\xrightarrow{P^*_Z}0.$$

This entails (see [12], p331, problem 5) that conditionally given s,

$$(5.21)\qquad \frac{\sum_{i\in s}\hat{\varepsilon}_{i,N}x_{i,N}Z_i}{\sqrt{1-f}\sqrt{\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)$$

because

$$(5.22)\quad E\left(\sum_{i\in s}\hat{\varepsilon}_{i,N}x_{i,N}Z_i\right)=0\text{ and }\sigma^2\left(\sum_{i\in s}\hat{\varepsilon}_{i,N}x_{i,N}Z_i\right)=(1-f)\sum_{i\in s}\hat{\varepsilon}^2_{i,N}x^2_{i,N}.$$

Combination of formulas (5.19) and (5.21) gives the desired result (3.9). □

Acknowledgements. The authors are grateful to R.Beran and Z.Prášková for their useful comments and to A.Dorfman for sending the interesting preprint [4].

References

[1] B. Von Bahr & C.G. Esseen (1965), Inequalities for the rth absolute moment of a sum of random variables, $1\le r\le 2$, Ann. Math. Statist. 36, no.1, pp.299-303.

[2] J.G. Booth, R.W. Butler, P. Hall (1994), Bootstrap methods for finite populations, Jor. Amer. Stat. Ass., vol. 89, 428, pp. 1282-1289.

[3] J. Chen & R.R. Sitter (1993), Edgeworth expansion and the bootstrap for stratified sampling without replacement from a finite population, Canadian Jour. Stat. 21, pp.347-357.

[4] A.H. Dorfman (1994), Non-parametric regression for estimating totals in finite populations, Preprint.

[5] P. Erdös & A. Rényi (1959), On the central limit theorem for samples from a finite population, Publ. Math. Inst. Hung. Acad. Sci., 4, pp.49-61

[6] W.A. Fuller (1975), Regression analysis for sample survey, Sankhyā, 37, series C, pp.117-132.

[7] D. Freedman (1977), A remark on the difference between sampling with and without replacement , Jour. Amer. Stat. Ass., 72, pp.681.

[8] J. Hájek (1960), Limiting distributions in simple random sampling from a finite population, Periodica Mathematica, 5, pp.361-374.

[9] J. Hájek (1981), *Sampling from a finite population*, Statistics: textbooks and monographs, vol. 37, Dekker, inc./New York.

[10] R.Y. Liu (1988), Bootstrap procedures under some non-i.i.d. models, Ann.Statist. 16, pp.1696-1708.

[11] R.Y. Liu & K. Singh (1992), Efficiency and robustness in resampling, Ann.Statist. 20, pp.370-385.

[12] M. Loève (1963), *Probability Theory*, Third Edition, D. van Nostrand Company, Inc.

[13] E. Mammen (1992), *When does bootstrap work?,* Lecture notes in Statistics, Vol 77, Springer-Verlag

[14] C. Särndal, B. Swensson, J. Wretman (1992), *Model assisted survey sampling*, Springer-Verlag

[15] A. Scott & C.F. Wu (1981), On the asymptotic distribution of ratio and regression estimators, Jour. Amer. Stat. Ass., 76, pp.98-102.

[16] R.J.Serfling (1974), Probability inequalities for the sum in sampling without replacement, Ann. Statist. 2, pp.39-48.

[17] R.L. Wright (1983), Finite population sampling with multivariate auxiliary information, Jour. Amer. Stat. Ass., vol. 78, 384, pp.879-883.

[18] C.F.J. Wu (1986), Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). Ann. Statist. 14, pp.1261-1350.