Wild Bootstrapping in Finite Populations with Auxiliary Information

ROELOF HELMERS

CWI

MARTEN WEGKAMP

Yale University

ABSTRACT. Consider a finite population u, which can be viewed as a realization of a super-population model. A simple ratio model (linear regression, without intercept) with hetero-scedastic errors is supposed to have generated u. A random sample is drawn without replacement from u. In this set-up 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.

Key words: auxiliary information, bootstrap based confidence intervals, finite populations, heteroscedastic errors, ratio estimator, ratio model, regression estimator, super-population model, wild bootstrapping.

1. Introduction

Resampling methods for finite populations is an important topic of current research. We refer to Booth *et al.* (1994) for a survey. In the present paper the situation is considered where the finite population is viewed as a realization of a certain super-population 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 set-up 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 super-population ξ:

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

where $\varepsilon_{i,N}$ are independent random variables, all defined on the same probability space, with

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

That is, 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 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 super-population 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 Dorfman (1994), 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 super-population 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 super-population 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 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 (1986)—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 ($1 \le i \le N$) separately. Our main results (cf. section 3) can be viewed as an extension the already existing theory of wild bootstrapping for heteroscedastic regression models (cf., e.g. Liu, 1988; Liu & Singh, 1992; Wu, 1986) to the situation considered in the present paper, where such models serve as an underlying super-population 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 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

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

and the regression estimator

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

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

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

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 super-population 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 boot-strapped versions of these estimators will be developed in section 3. The proofs of the main results are given in the appendix.

2. Preliminaries

Define $\bar{y}_N = N^{-1}\sum_{i\in u}y_{i,N}$, the population mean, $\bar{y}_n = n^{-1}\sum_{i\in s}y_{i,N}$, the sample mean. The quantities $\bar{x}_N$, $\bar{x}_n$, $\bar{\varepsilon}_N$, $\bar{\varepsilon}_n$ are defined in the similar way. In addition we define $\overline{x_N y_N} = N^{-1}\sum_{i\in u}x_{i,N}y_{i,N}$ and $\overline{x_N^2} = N^{-1}\sum_{i\in u}x_{i,N}^2$. To begin with, we rewrite $\hat{\beta}_{RA}$ and $\hat{\beta}_{RE}$. Because $x/y = x_0/y_0 + [x - y(x_0/y_0)]/y$ is valid for all real x, x_0 and non-zero y, y_0, one easily checks

$$\hat{\beta}_{RA} = \frac{\bar{y}_N}{\bar{x}_N} + \frac{1}{n\bar{x}_n}\sum_{i\in s}\left(y_{i,N} - \frac{\bar{y}_N}{\bar{x}_N}x_{i,N}\right), \tag{2.1}$$

and

$$\hat{\beta}_{RE} = \frac{\overline{y_N x_N}}{\overline{x_n^2}} + \frac{1}{n\bar{x}_n^2}\sum_{i\in s}\left(y_{i,N}x_{i,N} - \frac{\overline{y_N x_N}}{\overline{x_N^2}}x_{i,N}^2\right). \tag{2.2}$$

For notational convenience we also define

$$B_N = \bar{y}_N/\bar{x}_N,$$

$$a_{i,N} = y_{i,N} - B_N x_{i,N} = (\beta - B_N)x_{i,N} + \varepsilon_{i,N},$$

$$\tilde{B}_N = \overline{y_N x_N}/\overline{x_N^2}, \tag{2.3}$$

$$\tilde{a}_{i,N} = y_{i,N}x_{i,N} - \tilde{B}_N x_{i,N}^2 = (\beta - \tilde{B}_N)x_{i,N}^2 + \varepsilon_{i,N}x_{i,N}.$$

So we may as well write

$$\hat{\beta}_{RA} = B_N + \frac{1}{n\bar{x}_n}\sum_{i\in s}a_{i,N}$$

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

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 $D_N^2 = N^{-1}\sum_{i\in u}a_{i,N}^2$ and $\tilde{D}_N^2 = N^{-1}\sum_{i\in u}\tilde{a}_{i,N}^2$. Note that $\sum_{i\in u}a_{i,N} = \sum_{i\in u}\tilde{a}_{i,N} = 0$.

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

Lemma 2.1

Consider the super-population model (1.1), *and 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) $\max_{1\leqslant i\leqslant N} E_\xi|\varepsilon_{i,N}|^{2+\delta} = \mathcal{O}(1)$, *as* $N \to \infty$ *for some* $0 < \delta$;

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

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

and either

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

or

$$(\tilde{D}) \quad \sum_{i=1}^{N} x_{i,N}^2 \sigma_{i,N}^2 \sim \tilde{c}_2 N, \quad as \ N \to \infty, \quad for \ some \ 0 < \tilde{c}_2 < \infty;$$

we have, with P_ξ-probability 1, as $n \to \infty$, $N - n \to \infty$,

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

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

In addition, we may replace $\sum_{i=1}^{N} a_{i,N}^2$ by $\sum_{i=1}^{N} \sigma_{i,N}^2$ in (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

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

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 for almost every realization $Y_{1,N}, \ldots, Y_{N,N}$, as $n \to \infty$, $N - n \to \infty$,

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

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

Similarly,

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

$$\frac{\sqrt{n\bar{x}_N^2}}{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in u}(\hat{\varepsilon}_{i,N}x_{i,N})^2}}(\hat{\beta}_{RE} - \tilde{B}_N) \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1),$$ (2.10)

Results somewhat related to theorem 2.1 can be found in Fuller (1975) and Scott & Wu (1981).

Remark 2.1. An important issue in practice is estimation of the population total $\theta_N = \sum_{i\in u}y_{i,N}$ of an actual finite population at hand. We can estimate this quantity simply by $\hat{\beta}_{RA}\cdot x_N$ (set $x_N = \sum_{i\in u}x_{i,N}$). 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

$$\hat{\theta}_{RA} - \theta_N = x_N(\hat{\beta}_{RA} - B_N) = \frac{N}{n}\cdot\frac{\bar{x}_N}{\bar{x}_n}\cdot\sum_{i\in s}a_{i,N},$$

and hence, the counterpart of (2.7) becomes: with P_ξ-probability 1 we have, as $n \to \infty$, $N - n \to \infty$,

$$\frac{\sqrt{n}}{N\sqrt{1-f}\sqrt{N^{-1}\sum_{i\in u}a_{i,N}^2}}(\hat{\theta}_{RA} - \theta_N) \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1).$$ (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 P_ξ-a.s., as $n \to \infty$ and $N - n \to \infty$,

$$\frac{\sqrt{n\bar{x}_N^2}}{x_N\sqrt{1-f}\sqrt{N^{-1}\sum_{i\in u}\tilde{a}_{i,N}^2}}(\hat{\theta}_{RE} - \tilde{\theta}_N) \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1).$$ (2.12)

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

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

as our estimator based on $\hat{\theta}_{RE}$. The estimator $\hat{\theta}_{RE,c}$ for the population total θ_N also appears in Wright (1983) as the "combined regression through the origin" estimator. We note in passing that Wright (1983); see also Särndal *et al.* (1992) 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:

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

where $a'_{i,N}$, $i \in s$ are centred random variables $a'_{i,N} = [y_{i,N} - \tilde{B}_N x_{i,N}] - [\bar{y}_N - \tilde{B}_N \bar{x}_N]$. Note that there is only a minor difference between $a_{i,N}$ and $a'_{i,N}$ (cf. (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): with P_ξ-probability 1, as $n \to \infty$ and $N - n \to \infty$,

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

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

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

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 $N^{-1}\sum_{i \in u}a^2_{i,N}$ and $N^{-1}\sum_{i \in u}(a'_{i,N})^2$ appearing in (2.11) and (2.15) by estimates, i.e. by $n^{-1}\sum_{i \in s}\hat{\varepsilon}^2_{i,N}$ and $n^{-1}\sum_{i \in s}\hat{\varepsilon}^2_{i,N} - (n^{-1}\sum_{i \in s}\hat{\varepsilon}_{i,N})^2$ 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 (cf. 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

$$\frac{\sum\limits_{i \in u} x_{i,N}}{\sqrt{\sum\limits_{i \in u}\sigma^2_{i,N}}} \cdot (B_N - \beta) \xrightarrow{\mathscr{D}} \mathscr{N}(0, 1).$$

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

$$\frac{N}{n}\sum\limits_{i \in s}\hat{\varepsilon}^2_{i,N} \Big/ \left(\sum\limits_{i \in u}x_{i,N}\right)^2$$

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:

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

The first term at the right 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 $\sigma^2_\xi(B_N)$ is

consistently estimated by $(N/n)\sum_{i \in s}\hat{\varepsilon}_{i,N}^2/(\sum_{i \in u}x_{i,N})^2$. 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

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

$$
-u_{\alpha/2} \frac{\sqrt{1-f}\sqrt{n^{-1}\sum_{i \in s}\hat{\varepsilon}_{i,N}^2}}{\sqrt{n}\bar{x}_N}, \tag{2.17}
$$

where $u_{\alpha/2}$ and $u_{1-\alpha/2}$ 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 $\hat{\beta}_{RE}$ (cf. (2.10)). In the next section we introduce bootstrap based confidence intervals for β (cf. 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 Liu (1988), Liu & Singh (1992), Mammen (1992) and Wu (1986). Bootstrapping in finite population models also received a lot of attention. For instance, the asymptotic behaviour of the bootstrap for stratified sampling without replacement from a finite population has recently been explored in Chen & Sitter (1993) (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 super-population 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 $\theta_N = \sum_{i \in u}y_{i,N}$ and the parameter β of the super-population model are given in 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. Liu, 1988). 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 $\hat{\beta}$:

(1) calculate "estimated residuals" $\hat{\varepsilon}_{i,N} = y_{i,N} - \hat{\beta}x_{i,N}$, $i \in s$ (cf. (2.6));
(2) 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 $\varepsilon_{i,N}^* = \hat{\varepsilon}_{i,N}Z_i$, $i \in s$;
(3) two-stage resampling procedure put $n' = ([nf] + 1) \wedge n$ and $k = [n/n']$.

Stage 1. Draw without replacement from s a bootstrap sample $s_1^* = \{i_{j1}, \ldots, i_{j_{n'}}\}$ with size n' (the indices $i_{j1}, \ldots, i_{j_{n'}}$, being necessarily distinct);

Stage 2. Repeat this (stage 1) independently k times, replacing the resamples each time, and obtain a bootstrap sample $s^* = s_1^* \cup \cdots \cup s_k^*$ 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 $\hat{\beta}_{RE}^* = \sum x_{i,N} y_{i,N}^* / \sum x_{i,N}^2$ 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 Chen & Sitter (1993), 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'. We mirror—using their terminology—the original sampling scheme since the sampling fractions $f' = n'/n$ and $f = n/N$ agree asymptotically. The second stage is needed to *match* the skewness. In a way we imitate the ξ-model (cf. (1.1)) by a wild bootstrap version ξ^*:

$$Y_{i,N}^* = \hat{\beta} x_{i,N} + \varepsilon_{i,N}^*, \quad i \in s \tag{3.1}$$

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

$$E_{\xi^*} \varepsilon_{i,N}^* = 0, \quad E_{\xi^*}[\varepsilon_{i,N}^*]^2 = \hat{\varepsilon}_{i,N}^2$$

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

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

$$\frac{\sqrt{n^*}\bar{x}_n}{\sqrt{1-f'}\sqrt{1/n^* \sum_{i \in s^*} (\varepsilon_{i,N}^*)^2}}(\hat{\beta}_{RA}^* - \hat{\beta}_{RA}) \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1), \tag{3.3}$$

Similarly,

$$\frac{\sqrt{n^*}\,\overline{x_n^2}}{\sqrt{1-f'}\sqrt{1/n \sum_{i \in s} (\hat{\varepsilon}_{i,N} x_{i,N})^2}}(\hat{\beta}_{RE}^* - \hat{\beta}_{RE}) \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1), \tag{3.4}$$

$$\frac{\sqrt{n^*}\,\overline{x_n^2}}{\sqrt{1-f'}\sqrt{1/n^* \sum_{i \in s^*} (x_{i,N} \varepsilon_{i,N}^*)^2}}(\hat{\beta}_{RE}^* - \hat{\beta}_{RE}) \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1), \tag{3.5}$$

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 at the left 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 Serfling (1974). 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 did not pursue this point here.

Here is our second result. In a way we adapt the familiar "paired bootstrap" resampling scheme (cf. also Liu & Singh, (1992) to our finite population set-up. 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:

(1) put $n' = ([nf] + 1) \wedge n$ and $k = [n/n']$;

(2) 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 $s^* = s_1^* \cup \cdots \cup s_k^*$ of size $n^* = n'k$;

(3) compute $\hat{\beta}_{RA}^{*P} = \sum_{s^*} y_{i,N} / \sum_{s^*} x_{i,N}$ and $\hat{\beta}_{RE}^{*P} = \sum_{s^*} x_{i,N} y_{i,N} / \sum_{s^*} x_{i,N}^2$.

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

$$\frac{\sqrt{n^* \bar{x}_n}}{\sqrt{1 - f'} \sqrt{1/n \sum_{i \in s} \hat{\varepsilon}_{i,N}^2}} (\hat{\beta}_{RA}^{*P} - \hat{\beta}_{RA}) \xrightarrow{\mathcal{L}} \mathcal{I}(0, 1), \tag{3.6}$$

$$\frac{\sqrt{n^*} \, \overline{x_n^2}}{\sqrt{1 - f'} \sqrt{1/n \sum_{i \in s} (x_{i,N}^2 \hat{\varepsilon}_{i,N}^2)^2}} (\hat{\beta}_{RE}^{*P} - \hat{\beta}_{RE}) \xrightarrow{\mathcal{L}} \mathcal{I}(0, 1), \tag{3.7}$$

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_is ("wild bootstrap component") properly to reflect sampling from a finite population as well.

Theorem 3.3

Consider the following resampling scheme:

(1) generate n independent copies of a random variable Z with $EZ = 0$, $EZ^2 = 1 - f$ and $E|Z|^{2+\eta} < \infty$ for some $\eta > 0$;

(2) 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 $\hat{\beta} = \hat{\beta}_{RE}$;

(3) compute $\hat{\beta}_{RA}^{*W} = \sum_s y_{i,N}^* / \sum_{i \in s} x_{i,N}$ and $\hat{\beta}_{RE}^{*W} = \sum_s x_{i,N} y_{i,N}^* / \sum_{i \in s} x_{i,N}^2$.

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

$$\frac{\sqrt{n}\bar{x}_n}{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}\hat{\varepsilon}_{i,N}^2}}(\hat{\beta}_{RA}^{*W} - \hat{\beta}_{RA}) \xrightarrow{\mathcal{L}}. I\,{}'(0, 1),$$ (3.8)

$$\frac{\sqrt{n}\bar{x}_n^2}{\sqrt{1-f}\sqrt{n^{-1}\sum_{i\in s}x_{i,N}^2\hat{\varepsilon}_{i,N}^2}}(\hat{\beta}_{RE}^{*W} - \hat{\beta}_{RE}) \xrightarrow{\mathcal{L}}. I\,{}'(0, 1),$$ (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 $\theta_N = \sum_{i\in u}y_{i,N}$ 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

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

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

$$\frac{\sqrt{n^*}}{n\sqrt{1-f'}\sqrt{1/n^*\sum_{i\in s^*}(\varepsilon_{i,N}^*)^2}}(\hat{\theta}_{RA}^* - \hat{\theta}_{RA}).$$ (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 $1 - \alpha + \sigma(1)$, 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 $1 - \alpha + \sigma(n^{-1/2})$, provided the distribution of the Zs 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 $c_{\alpha/2}^*$ and $c_{1-\alpha/2}^*$ 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 super-population model ξ (cf. (1.1)). Replace in (2.17) the normal quantiles by the bootstrap quantiles of the distribution of

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

the left-hand side of (3.3). A difference from 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. That is, we have shown that the bootstrap approximations given in section 3 are asymptotically consistent. However, the question remains: how well do these bootstrap approximations (cf. theorems 3.1, 3.2 and 3.3) compare with the more traditional normal approximations (cf. 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. Although our estimators differ from the ones considered by Chen and Sitter (linear combinations of stratum means), they can be expressed as an average as well (cf. (2.1), (2.2) and (5.10) and (5.13) ahead). The finite population aspect is reflected by the two-stage procedure as in Chen & Sitter (1993) whereas the hetero-scedastic nature of the underlying regression model (1.1) is captured by the wild bootstrap component. At this point we may add an assumption on EZ^3 (cf. also Liu, 1988) and a more careful choice of the design parameters n' and k of the resampling scheme will be needed. We do not 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.

Second one may be interested in a slight extension of our set-up, namely the case that the super-population model ξ (cf.(1.1)) is still valid, but stratified sampling, instead 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 (cf. Wright, 1983; see also Särndal et al., 1992). However, at this point, we shall 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 $n_l \rightarrow \infty$ and $N_l - n_l \rightarrow \infty$ 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 super-population 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.

Acknowledgements

References

Von Bahr, B. & Esseen, C. G. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leqslant r \leqslant 2$. *Ann. Math. Statist.* **36**, 299–303.

Booth, J. G., Butler, R. W. & Hall, P. (1994). Bootstrap methods for finite populations. *J. Amer. Statist. Assoc.* **89**, 1282–1289.

Chen, J. & Sitter, R. R. (1993). Edgeworth expansion and the bootstrap for stratified sampling without replacement from a finite population. *Canad. J. Statist.* 21, 347–357.

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

Erdős, P. & Rényi, A. (1959). On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hung. Acad. Sci.* 4, 49–61.

Freedman, D. (1977). A remark on the difference between sampling with and without replacement. *J. Amer. Statist. Assoc.* 72, 681.

Fuller, W. A. (1975). Regression analysis for sample survey, *Sankhyā Ser. C* 37, 117–132.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population, *Period Math.* 5, 361–374.

Hájek, J. (1981). *Sampling from a finite population*, Statistics: textbooks and monographs, vol. 37, Dekker.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* 16, 1696–1708.

Liu, R. Y. & Singh, K. (1992). Efficiency and robustness in resampling. *Ann. Statist.* 20, 370–385.

Loève, M. (1963). *Probability theory*, Third edn, D. van Nostrand.

Mammen, E. (1992). *When does bootstrap work?*, Lecture notes in statistics, Vol. 77, Springer-Verlag, New York.

Petrov, V. V. (1975). *Sums of independent random variables.* Ergebnisse der Mathematik und ihrer Grenzgebiete 82, Springer-Verlag.

Särndal, C., Swensson, B. & Wretman, J. (1992). *Model assisted survey sampling*, Springer-Verlag, Berlin.

Scott, A. & Wu, C. F. (1981). On the asymptotic distribution of ratio and regression estimators. *J. Amer. Statist. Assoc.* 76, 98–102.

Serfling, R. J. (1974). Probability inequalities for the sum in sampling without replacement. *Ann. Statist.* 2, 39–48.

Wright, R. L. (1983). Finite population sampling with multivariate auxiliary information, *J. Amer. Statist. Assoc.* 78, 879–883.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* 14, 1261–1350.

Marten Wegkamp. Department of Statistics, Yale University, Box 208290 Yale Station, New Haven, CT 06520-8290, USA.

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 Erdős & Rényi (1959) and Hájek (1960), we have to show that the Hájek–Lindeberg condition

$$\frac{1}{\sum_{i \in u} \tilde{a}_{i,N}^2} \sum_{i \in u} \tilde{a}_{i,N}^2 \mathbb{1}\left\{ |\tilde{a}_{i,N}| > \eta \sqrt{f(1-f) \sum_{i \in u} \tilde{a}_{i,N}^2} \right\} \to 0 \quad \text{for all } \eta > 0 \tag{A.1}$$

as $N - n \to \infty$, $n \to \infty$ holds with P_ξ-probability one. We shall use a strong law of large numbers (SLLN) for independent zero mean random variables (cf. th. 12, p. 272 in Petrov (1975)) to prove that

$$\left| \frac{1}{N} \sum_{i \in u} \tilde{a}_{i,N}^2 - \frac{1}{N} \sum_{i \in u} (x_{i,N} \sigma_{i,N})^2 \right| \to 0 \quad \text{as } N \to \infty \tag{A.2}$$

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

$$\tilde{D}_N^2 = \frac{1}{N} \sum_{i \in u} (\beta - \tilde{B}_N)^2 x_{i,N}^4 + \frac{1}{N} \sum_{i \in u} x_{i,N}^2 \varepsilon_{i,N}^2 + \frac{2}{N} \sum_{i \in u} (\beta - \tilde{B}_N) x_{i,N}^3 \varepsilon_{i,N}. \tag{A.3}$$

Note that with P_ξ-probability one,

$$\tilde{B}_N - \beta = \sum_{i \in u} x_{i,N} \varepsilon_{i,N} \Big/ \sum_{i \in u} x_{i,N}^2 \to 0, \tag{A.4}$$

where we need assumptions (A), (B) and (C) in order to apply the mentioned SLLN for independent r.v.s. As a result of (A.4) condition (B) we have $(\beta - \tilde{B}_N)^2 N^{-1} \sum_{i \in u} x_{i,N}^4 = \wp(1)$ P_ξ-a.s. The last term in (A.3) vanishes too, because by the Cauchy–Schwarz inequality

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

It remains to consider the middle term in (A.3) and this term will in fact turn out to be the dominant one. Again by means of the SLLN and condition (A), we can see that (A.2) is true. In view of (A.2), it suffices to check whether

$$\frac{1}{N\tilde{\sigma}_N^2} \sum \tilde{a}_{i,N}^2 \mathbb{1}\{|\tilde{a}_{i,N}| > \eta \sqrt{f(1-f)N}\tilde{\sigma}_N\} \to 0 \quad \text{for all } \eta > 0 \tag{A.5}$$

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

$$\tilde{\sigma}_N^2 = N^{-1} \sum_{i \in u} (x_{i,N}\sigma_{i,N})^2. \tag{A.6}$$

Since $\tilde{a}_{i,N} = (\beta - \tilde{B}_N)x_{i,N}^2 + x_{i,N}\varepsilon_{i,N}$ (cf. (2.3)), we can bound the left-hand side of (A.5) above by $d_{N,n}^{(1)}(\eta/2) + d_{N,n}^{(2)}(\eta/2)$, defined by

$$d_{N,n}^{(1)}(\eta) = \frac{1}{N\tilde{\sigma}_N^2} \sum_{i \in u} \tilde{a}_{i,N}^2 \mathbb{1}\{|\beta - \tilde{B}_N|x_{i,N}^2 > \eta \sqrt{f(1-f)N}\tilde{\sigma}_N\},$$

$$d_{N,n}^{(2)}(\eta) = \frac{1}{N\tilde{\sigma}_N^2} \sum_{i \in u} \tilde{a}_{i,N}^2 \mathbb{1}\{|x_{i,N}\varepsilon_{i,N}| > \eta \sqrt{f(1-f)N}\tilde{\sigma}_N\}.$$

Assumption (B) and the fact that $\tilde{B}_N - \beta = \wp(1)$ P_ξ-a.s. entail that for every $\eta > 0$, $\lim_{n \to \infty, N-n \to \infty} d_{n,N}^{(1)}(\eta/2) = 0$ with P_ξ-probability one. Note that

$$\frac{1}{N} \sum_{i \in u} \mathbb{1}\{|x_{i,N}\varepsilon_{i,N}| > \eta \sqrt{f(N-n)}\tilde{\sigma}_N\} \to 0 \quad \text{as } N - n \to \infty, \, n \to \infty$$

as $N - n \to \infty$, $n \to \infty$, P_ξ-a.s. Application of Hölder's inequality yields $d_{n,N}^{(2)}(\eta/2) \to 0$ P_ξ-a.s. for $n \to \infty$, $N - n \to \infty$ for every $\eta > 0$. This proves the lemma.

Proof of theorem 2.1. Recall that (cf. (2.2))

$$\hat{\beta}_{RE} - \tilde{B}_N = \frac{1}{n\overline{x_n^2}} \sum_{i \in s} \tilde{a}_{i,N} = \frac{1}{n\overline{x_n^2}} \cdot \frac{\overline{x_N^2}}{\overline{x_n^2}} \sum_{i \in s} \tilde{a}_{i,N}.$$

Note that lemma 2.1 implies

$$\frac{1}{\sqrt{f(1-f) \sum_{i \in u} \tilde{a}_{i,N}^2}} \sum_{i \in s} \tilde{a}_{i,N} \xrightarrow{\mathscr{L}} \mathscr{N}(0, 1). \tag{A.7}$$

Since $\overline{x_N^2}/\overline{x_n^2} \xrightarrow{\pi} 1$, (2.9) follows from Slutsky's lemma. It remains to check the validity of (2.10). It suffices clearly to show that

$$\frac{1}{n}\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2 \Big/ \frac{1}{N}\sum_{i\in u}(\hat{\varepsilon}_{i,N}x_{i,N})^2 \xrightarrow{\pi} 1 \tag{A.8}$$

and with P_ξ-probability one

$$\left|\frac{1}{N}\sum_{i\in u}(\hat{\varepsilon}_{i,N}x_{i,N})^2 - \frac{1}{N}\sum_{i\in u}\tilde{a}_{i,N}^2\right| \to 0, \tag{A.9}$$

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 representation:

$$\hat{\beta}_{RE}^{*P} - \hat{\beta}_{RE} = \frac{1}{\sum_{i\in s^*}x_{i,N}^2}\sum_{i\in s^*}\hat{\varepsilon}_{i,N}x_{i,N} = \frac{1}{\sum_{i\in s^*}x_{i,N}^2}\sum_{j=1}^{k}\sum_{i\in s_j^*}\hat{\varepsilon}_{i,N}x_{i,N}. \tag{A.10}$$

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 (A.8) holds: for every sequence of Borel sets $C_1, \ldots, C_N$,

$$\left|\frac{1}{n}\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2 1_{C_i} - \frac{1}{N}\sum_{i\in u}(\hat{\varepsilon}_{i,N}x_{i,N})^2 1_{C_i}\right| \xrightarrow{\pi} 0, \quad N\to\infty, n\to\infty.$$

This implies that the Hájek–Lindeberg condition

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

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

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

for all $\eta > 0$ where $f' = n'/n$. Recall that $x_{i,N}\hat{\varepsilon}_{i,N} = (\beta - \hat{\beta}_{RE})x_{i,N}^2 + \varepsilon_{i,N}x_{i,N}$. By the triangle inequality, we have

$$|\beta - \hat{\beta}_{RE}| \leq |\beta - \tilde{B}_N| + |\tilde{B}_N - \hat{\beta}_{RE}|,$$

where the first term at the right is $\sigma(1)$ P_ξ-a.s. and the second term is of order $\sigma_\pi([1-f]/n)^{1/2}$. Therefore we have P_ξ-a.s.

$$\left|\frac{1}{N}\sum_{i\in u}(\hat{\varepsilon}_{i,N}x_{i,N})^2 - \frac{1}{N}\sum_{i\in u}(\sigma_{i,N}x_{i,N})^2\right| \xrightarrow{\pi} 0.$$

This reduces the Hájek–Lindeberg condition to

$$\frac{1}{N\tilde{\sigma}_N^2}\sum_{i\in u}(x_{i,N}\hat{\varepsilon}_{i,N})^2 1\left\{|\beta - \hat{\beta}_{RE}|x_{i,N}^2 > \frac{\eta}{2}\sqrt{f'(1-f')}n\tilde{\sigma}_N\right\}$$

$$+\frac{1}{N\tilde{\sigma}_N^2}\sum_{i\in u}(x_{i,N}\hat{\varepsilon}_{i,N})^2 1\left\{|x_{i,N}\varepsilon_{i,N}| > \frac{\eta}{2}\sqrt{f'(1-f')}n\tilde{\sigma}_N\right\} \to 0$$

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

$$\frac{1}{\sqrt{f'(1-f')}\sqrt{\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2}}\sum_{i\in s_1^*}\hat{\varepsilon}_{i,N}x_{i,N} \xrightarrow{\mathscr{L}} \Gamma(0, 1), \tag{A.11}$$

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

$$\frac{1}{\sqrt{f'(1-f')}\sqrt{\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2}}\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\sum_{i\in s_j^*}\hat{\varepsilon}_{i,N}x_{i,N} \xrightarrow{\mathscr{L}} \Gamma(0, 1). \tag{A.12}$$

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 (this happens only when the original sample fraction $f \to 0$ at the rate $\Gamma(N^{-1/2})$), 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 Freedman (1977). Relation (A.12) now follows directly from Lindeberg's CLT for triangular arrays. Therefore (cf. (A.10) and Slutsky's lemma)

$$\frac{n^*\overline{x_n^2}}{\sqrt{k}\sqrt{f'(1-f')\sum_{i\in s}\hat{\varepsilon}_i x_i^2}}(\hat{\beta}_{RE}^{*P} - \hat{\beta}_{RE}) \xrightarrow{\mathscr{L}} \Gamma(0, 1),$$

which completes the proof.

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

$$\hat{\beta}_{RE}^* - \hat{\beta}_{RE} = \frac{1}{\sum_{i\in s^*}x_{i,N}^2}\sum_{i\in s^*}\hat{\varepsilon}_{i,N}x_{i,N}Z_i = \frac{1}{\sum_{i\in s^*}x_{i,N}^2}\sum_{j=1}^{k}\sum_{i\in s_j^*}\hat{\varepsilon}_{i,N}x_{i,N}Z_i. \tag{A.13}$$

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

$$M(Z) = \frac{1}{n} \sum_{i \in s} \hat{\varepsilon}_{i,N} x_{i,N} Z_i,$$

$$\Delta^2(Z) = \frac{1}{n} \sum_{i \in s} (\hat{\varepsilon}_{i,N} x_{i,N} Z_i - M(Z))^2,$$

$$T_n^* = \frac{\sum_{i \in s^*} x_{i,N}^2}{\sqrt{k}\sqrt{f'(1 - f')}\sqrt{n}} (\hat{\beta}_{RE}^* - \hat{\beta}_{RE}).$$

Then we can rewrite T_n^* as

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

The left-hand side of (3.4) is precisely equal to

$$\frac{\frac{1}{n} \sum_{i \in s} x_{i,N}^2}{\frac{1}{n^*} \sum_{i \in s^*} x_{i,N}^2} \cdot T_n^* \bigg/ \sqrt{\frac{1}{n} \sum_{i \in s} (\hat{\varepsilon}_{i,N} x_{i,N})^2}.$$

Note that $M(Z) \xrightarrow{P_Z} 0$ since $EM(Z) = 0$ and $EM^2(Z) = \ell (n^{-1})$. If we define

$$\Delta^2 = E_Z \Delta^2(Z) = n^{-1} \sum_{i \in s} (x_{i,N} \hat{\varepsilon}_{i,N})^2 - n^{-2} \left(\sum_{i \in s} x_{i,N} \hat{\varepsilon}_{i,N}\right)^2 \sim n^{-1} \sum_{i \in s} (x_{i,N} \hat{\varepsilon}_{i,N})^2,$$

the convergence $\Delta(Z) - \Delta \xrightarrow{P_Z} 0$ is easily verified by the Von Bahr–Esseen inequality (cf. Von Bahr & Esseen, 1965).

We can not apply the Erdös–Rényi CLT for samples drawn without replacement from a finite population (see Erdös & Rényi, 1959; Hájek, 1960) at once, we first have to condition on $Z = (Z_1, \ldots, Z_n)$.

Similarly as in the proof of theorem 3.2 (cf. the argument leading to (A.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 $z = (z_1, \ldots, z_n)$ in a set A_n with arbitrarily large probability as $n \to \infty$, we have

$$\frac{1}{\sqrt{k}\sqrt{f'(1 - f')}n\Delta(z)} \sum_{i \in s^*} (x_{i,N} \hat{\varepsilon}_{i,N} z_i - M(z)) \xrightarrow{\mathscr{L}} \mathscr{I}(0, 1). \tag{A.14}$$

Such a sequence A_n exists due to the convergence of $M(Z)$ and $\Delta(Z)$ stated above. Thus we have proved that, for every $z \in A_n$,

$$\left| P^*(T_n^* \leq y | Z = z \in A_n) - \Phi\left(\frac{y - M(z)}{\Delta(z)}\right) \right| \to 0, \quad \text{as } n \to \infty, \tag{A.15}$$

where Φ denotes the standard normal distribution. If we define $r_n(Z)$ as the difference between these two probabilities at the left of (A.15), we have $r_n(Z) \xrightarrow{P_Z} 0$, $|r_n(\cdot)| \leq 2$ and hence, by dominated convergence, $E_Z |r_n(Z)| \to 0$ as $n \to \infty$. Consequently, we may integrate the Z out. Now, we employ the following Taylor expansion

$$\Phi\left(\frac{y - M(Z)}{\Delta(Z)}\right) = \Phi\left(\frac{y}{\Delta}\right) + \left\{\frac{y - M(Z)}{\Delta(Z)} - \frac{y}{\Delta}\right\} \phi(\xi) \tag{A.16}$$

where ϕ is the standard normal density 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. Also note that $|[y - M(Z)]/\Delta(Z) - y/\Delta|\varphi(\xi) \leqslant 2$ because of (A.16) and the triangle inequality. As a result we obtain by dominated convergence that $E_Z\Phi([y - M(Z)]/\Delta(Z)) = \Phi(y/\Delta) + \mathit{r}(1)$. 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

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

$$\leqslant \left|\frac{1}{n^*}\sum_{i\in s^*}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2 - \frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2\right| + \left|\frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N})^2(Z_i^2 - 1)\right|. \tag{A.17}$$

As before, conditionally on $Z_1, \ldots, Z_n$, the first part of (A.17) 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 $\hat{\beta}_{RE}^{*W}$ in

$$\hat{\beta}_{RE}^* = \frac{\sum_{i\in s} x_{i,N}y_{i,N}^*}{\sum_{i\in s}x_{i,N}^2} = \hat{\beta}_{RE} + \frac{\sum_{i\in s}x_{i,N}\hat{\varepsilon}_{i,N}Z_i}{\sum_{i\in s}x_{i,N}^2}. \tag{A.18}$$

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

$$\left|\frac{1}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N}Z_i)^2 - \frac{1-f}{n}\sum_{i\in s}(x_{i,N}\hat{\varepsilon}_{i,N})^2\right| \xrightarrow{P_Z} 0. \tag{A.19}$$

This entails (see Loève, 1963, p. 331, prob. 5) that conditionally given s,

$$\frac{\sum_{i\in s}\hat{\varepsilon}_{i,N}x_{i,N}Z_i}{\sqrt{1-f}\sqrt{\sum_{i\in s}(\hat{\varepsilon}_{i,N}x_{i,N})^2}} \xrightarrow{\mathscr{D}} \mathit{l}\,'(0, 1) \tag{A.20}$$

because

$$E\left(\sum_{i\in s}\hat{\varepsilon}_{i,N}x_{i,N}Z_i\right) = 0 \quad \text{and} \quad \sigma^2\left(\sum_{i\in s}\hat{\varepsilon}_{i,N}x_{i,N}Z_i\right) = (1 - f)\sum_{i\in s}\hat{\varepsilon}_{i,N}^2 x_{i,N}^2. \tag{A.21}$$

Combination of (A.19) and (A.21) gives the desired result (3.9).