An empirical comparison of BRR and linearization variance estimators

Discussion paper 03006

Harm Jan Boonstra and Nico Nieuwenbroek

Statistics Netherlands

Voorburg/Heerlen, August 2003

Publisher
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg
The Netherlands

Printed by
Statistics Netherlands - Facility Services

Cover design
WAT ontwerpers, Utrecht

Information
E-mail: infoservice@cbs.nl

Where to order
E-mail: verkoop@cbs.nl

Internet
http://www.cbs.nl

ISSN: 1572-0314
Key figure: X-10
Production code: 6008303005

Statistics Netherlands

AN EMPIRICAL COMPARISON OF BRR AND LINEARIZATION VARIANCE ESTIMATORS

A simulation study is performed to compare linearization and balanced repeated replication (BRR) variance estimates in various situations. The simulations are based on small samples according to a simple two-per-stratum design drawn from a population of businesses, and larger samples according to a two-stage design drawn from a population of persons. Variance estimates for linear and non-linear statistics, including regression estimators, are considered. Several variants of the BRR method are used: BRR with different values of the Fay factor, grouped BRR and BRR with artificial strata.

Keywords: Resampling, Taylor linearization, General regression estimator, Simulation

1 Introduction

The quality of sample-based estimators of finite population quantities can often be improved by using sensibly chosen auxiliary information about the population. For the simulation studies described in this paper we use the general regression estimator for this purpose. The resulting estimates for population totals can be conveniently presented as weighted sums of sample observations. If no auxiliary information is used, the general regression estimator reduces to the Horvitz-Thompson estimator with weights equal to the design weights, i.e. the reciprocal values of the inclusion probabilities. For variance estimation in the design-based framework, the combination of sampling design and estimation method must be taken into account. For complex combinations this may not always be possible and we often have to rely on one or another approximation method.

In this report, we carry out a simulation study of two qualitatively different variance estimation methods. The first method is based on an analytical derivation of the design variance, after Taylor linearization, and is referred to as the Taylor method. The second method is balanced repeated replication (BRR), which is an example of a resampling method.

Two populations are used for the simulations. In the first study, small samples of two elements per stratum are drawn from a population of businesses. For the two-per-stratum design it is well-known that the BRR estimator equals the Taylor variance estimator for linear statistics. So the main purpose here is to compare BRR and Taylor variance estimates of non-linear statistics. In the second study, larger samples are drawn from a population of persons based on the Dutch Labour Force Survey. The design used in this case is stratified two-stage simple random sampling without replacement at both stages. Here the main purpose is to compare Taylor and BRR variance estimates for this particular complex sampling design. Another essential difference with the first study, affecting in particular the BRR method, is that more than two primary sampling units per stratum are selected.

The variance estimators are applied to Horvitz-Thompson (HT) estimators and general regression estimators for population totals and ratios of population totals. The true parameters and design variances of the (linearized) estimators can be computed from the populations. For non-linear statistics it is generally not possible to compute exact design-based variances, so we simulate these variances by means of a large number of samples.

The weighting package Bascula is used to compute Taylor and BRR variance estimates. In the next two sections the Taylor linearization and BRR methods as they are implemented in Bascula are described, see also Nieuwenbroek and Boonstra (2002). Sections 4 and 5 report about the first and second simulation studies, respectively. The final section contains conclusions.

2 Taylor variance

In this section, we give the expression for the Taylor variance estimators used in the simulations. The formulas are written out for a general regression estimator in combination with a stratified two-stage cluster design with simple random sampling without replacement at both stages. The formula for single-stage element or cluster sampling designs follows easily from it. For details about the regression estimator and its variance we refer to Särndal et al. (1992).

Let H denote the number of strata, N_h the number of primary sampling units (PSUs) in stratum h from which n_h are sampled and $f_{1h} = n_h/N_h$ the PSU fraction. In the second stage we have M_{hi} secondary sampling units (SSUs) within the ith PSU of stratum h from which m_{hi} are in the sample and $f_{2hi} = m_{hi}/M_{hi}$ is the corresponding SSU fraction. As the second simulation study below involves a second-stage sample of complete households, we consider here the general case that each SSU consists of a cluster of elements all of which are observed. This can be viewed as a special case of three-stage element sampling, but without any variance contribution from the third stage. The design weight for the kth respondent of cluster j in the ith PSU of stratum h is denoted by d_{hijk}. Cluster sizes are denoted by m_{hij}. It holds that $d_{hijk} = 1/(f_{1h}f_{2hi})$.

Let Y and Z be two study variables, X a vector of auxiliary variables, and y_{hijk}, z_{hijk}, x_{hijk} the observed individual values. We consider two types of quantities to be estimated, totals and ratios of totals, with corresponding estimators

$$\hat{\theta}_{\text{tot}}(y) = \sum_{h,i,j,k} w_{hijk} y_{hijk} \,, \qquad \hat{\theta}_{\text{ratio}} = \frac{\hat{\theta}_{\text{tot}}(y)}{\hat{\theta}_{\text{tot}}(z)} \,. \tag{1}$$

The weights w_{hijk} are regression-based weights, depending on the auxiliary variables. The same formulas can be used for Horvitz-Thompson estimators provided that the regression weights are replaced by design weights. This is the special case that no auxiliary information is incorporated. For convenience we shall call estimators that are functions of regression estimators, such as $\hat{\theta}_{\text{ratio}}$, simply regression estimators, and estimators based on HT estimators simply HT estimators.

Under the assumption that all n_h and m_{hi} are at least of size 2, a variance estimator of the regression estimator $\hat{\theta}_{\text{tot}}$ for a total, can now be written as

$$v_{\text{Taylor}}(\hat{\theta}_{\text{tot}}) = \sum_{h=1}^{H} \left\{ n_h(1-\lambda f_{1h})\tilde{s}_{1h}^2 + \lambda f_{1h} \sum_{i=1}^{n_h} m_{hi}(1-f_{2hi})\tilde{s}_{2hi}^2 \right\}, \quad (2)$$

where

$$\begin{aligned} \tilde{s}_{1h}^2 &= \frac{1}{n_h-1}\left\{ \sum_{i=1}^{n_h} (\tilde{t}_{hi})^2 - \frac{1}{n_h}\left(\sum_{i=1}^{n_h} \tilde{t}_{hi}\right)^2 \right\}, \\ \tilde{s}_{2hi}^2 &= \frac{1}{m_{hi}-1}\left\{ \sum_{j=1}^{m_{hi}} (\tilde{t}_{hij})^2 - \frac{1}{m_{hi}}\left(\sum_{j=1}^{m_{hi}} \tilde{t}_{hij}\right)^2 \right\}, \\ \tilde{t}_{hi} &= \sum_{j=1}^{m_{hi}} \tilde{t}_{hij}, \quad \text{and} \quad \tilde{t}_{hij} = \sum_{k=1}^{m_{hij}} d_{hijk} g_{hijk} e_{hijk}^{(y)}. \end{aligned} \quad (3)$$

The last expression is a weighted sum of sample fit residuals $e_{hijk}^{(y)} = y_{hijk} - \hat{y}_{hijk}$, where $\hat{y}_{hijk} = \hat{B}^t x_{hijk}$ with coefficients $\hat{B}$ estimated from the sample. The weights used here are the regression weights $w_{hijk} = g_{hijk} d_{hijk}$, in which g_{hijk} can be interpreted as adjustment factors.

The variance estimator written here is a specialization to the two-stage cluster design described above of the "g-weight" variance estimator, which was motivated in Särndal et al. (1989). The basic linearization variance estimator involves not the calibrated weights w_{hijk} but only the design weights d_{hijk} and so is obtained by setting the g-weights to 1 in (3). The value of λ in (2) must be 1 or 0, depending on whether the PSUs are selected without or with replacement, respectively. In the latter case, the formula is applicable to general multi-stage designs (with replacement in the first stage), even if PSUs are drawn with unequal probabilities, e.g. proportional to size.

A variance estimator for the ratio $\hat{\theta}_{\text{ratio}}$, where numerator and denominator totals are calibrated with respect to the same population totals, is obtained from (2) after the following adjustments: divide by an overall factor $(\hat{\theta}_{\text{tot}}(z))^2$ and change the residuals $e_{hijk}^{(y)}$ into $e_{hijk}^{(y)} - \hat{\theta}_{\text{ratio}} e_{hijk}^{(z)}$, where $e_{hijk}^{(z)} = z_{hijk} - \hat{z}_{hijk}$. Note that in the case that no auxiliary information is used, i.e. in the case of HT estimators, the residuals simply have to be replaced by y_{hijk} or $y_{hijk} - \hat{\theta}_{\text{ratio}} z_{hijk}$ for totals or ratios, respectively.

3 Balanced Repeated Replication

BRR was originally developed for stratified multi-stage designs where in each stratum two primary sampling units (PSUs) or clusters are drawn with replacement in the first stage, see McCarthy (1969). A half-sample is a subset consisting of one of the two units of each of the H strata. The basic idea is to select a set of, say R, half-samples from the set of all 2^H half-samples and recalculate the estimator $\hat{\theta}$ for each of these half-samples. From these values an estimate of the variance of $\hat{\theta}$ can be computed. It turns out that certain balanced sets of only a moderate number of half-samples give the same result as would be obtained from the full set of half-samples, at least when $\hat{\theta}$ is a linear statistic.

BRR has been extended to stratified multi-stage sampling where from some or all strata more than two PSUs are selected with replacement using orthogonal arrays, see Wu (1991) and Sitter (1993). However, for large sample sizes these elegant extensions of BRR are not easy to apply, because of the large number of replicates required.

Therefore, as relatively simple alternatives, the grouped BRR and BRR with artificial strata are eligible. These alternative BRR methods have in common that the original stratified multi-stage design is approximated by a design for which the basic two-per-stratum procedure can be applied. In the grouped case this is done by randomly forming two groups of PSUs per stratum. In the case with artificial strata, the design strata are randomly subdivided into artificial strata each with two (groups of) PSUs. The price to pay for grouping is that the variances are estimated less precise compared to the more advanced procedures.

Grouped BRR and BRR with artificial strata can be applied straightforwardly to stratified multi-stage designs with replacement in the first stage. By means of a simple modification for the finite population, these techniques are also applicable to stratified single-stage designs with simple random sampling without replacement in each stratum, see e.g. Wolter (1985). Inspired by Rao and Wu (1988) who suggested bootstrap replications for two-stage designs in which simple random sampling without replacement is used at both stages, Renssen et al. (1997) modified BRR to be applied to such stratified two-stage designs. We now describe this method, using the same notation as in Section 2 and again assuming that all n_h and m_{hi} are at least of size 2.

If n_h is even then the PSUs in the hth stratum are randomly divided into two groups of sizes $n_h/2$. For strata with n_h odd, first one randomly selected PSU is deleted. Subsequently, the PSUs are randomly divided into two groups of sizes $(n_h - 1)/2$. For notational convenience we denote by

$$n_h^* = \begin{cases} n_h & \text{if } n_h \text{ is even} \\ n_h - 1 & \text{if } n_h \text{ is odd,} \end{cases}$$

and we introduce two dummy variables Δ_{hi}^1 and Δ_{hi}^2 indicating to which group the (h, i)th PSU belongs. If the (h, i)th PSU belongs to the first group then Δ_{hi}^1 equals 1 and 0 otherwise, and similarly for Δ_{hi}^2. Analogously to the procedure for PSUs, each SSU can be randomly divided into two groups of size $m_{hi}/2$ or

$(m_{hi}-1)/2$ depending on whether m_{hi} is even or odd. Let us denote by

$$m_{hi}^* = \begin{cases} m_{hi} & \text{if } m_{hi} \text{ is even} \\ m_{hi}-1 & \text{if } m_{hi} \text{ is odd,} \end{cases}$$

and introduce dummies Δ_{hij}^1 and Δ_{hij}^2. Here Δ_{hij}^1 equals 1 if the jth SSU in PSU (h,i) belongs to the first group, otherwise it is 0, and similarly for Δ_{hij}^2.

Next, a set of R balanced half-samples is selected. This set may be defined by an $R \times H$ matrix with elements δ_{rh} either $+1$ or -1 according to whether the rth half-sample of the hth stratum contains the PSUs of the first or the second group, $r = 1 \ldots R$, $h = 1 \ldots H$. A minimum set of balanced half-samples can be constructed using Hadamard matrices. A set of R half-samples is said to be balanced if

$$\sum_{r=1}^{R} \delta_{rh}\delta_{rh'} = 0 \tag{4}$$

for all pairs of strata $h \neq h'$. This requires that R must be equal to the smallest quadruple greater than or equal to the number of strata. Now we have the ingredients for carrying out the BRR method. The procedure consists of two steps.

Step 1: Calculate modified design weights for each of the half-samples. In the original BRR setting with two units per stratum sampled with replacement, the modified design weights are 0 or twice the original design weights. However, possibly odd stratum and PSU sizes as well as finite population correction factors imply different modifications of the design weights. Besides, following an idea of Fay (1989), milder perturbations of the design weights can be accomplished through the introduction of a so-called Fay factor ϕ, with $0 < \phi \leq 1$. Altogether, this leads to the following modified design weights for resample r:

$$d_{hijk}^r = d_{hijk}\left(1 + \phi\delta_{rh}\left\{(\Delta_{hi}^1 - \Delta_{hi}^2)\sqrt{(1-\lambda f_{1h})\frac{n_h}{n_h^*}} + \lambda(\Delta_{hij}^1 - \Delta_{hij}^2)\sqrt{(1-f_{2hi})\frac{m_{hi}}{m_{hi}^*}}\right\}\right). \tag{5}$$

It can be shown that for $\phi < \sqrt{1/3} \approx 0.577$ it is guaranteed that these weights are always non-negative.

Step 2: Perform a calibration of each half-sample to the population totals just as it is done for the parent sample. This results into regression weights w_{hijk}^r for each resample, which can be calculated similarly as the regression weights for the parent sample, using d_{hijk}^r instead of d_{hijk} as starting weights. Replicate estimates corresponding to (1) are then computed as

$$\hat{\theta}_{\text{tot}}^r(y) = \sum_{h,i,j,k} w_{hijk}^r y_{hijk}\,, \quad \hat{\theta}_{\text{ratio}}^r = \frac{\hat{\theta}_{\text{tot}}^r(y)}{\hat{\theta}_{\text{tot}}^r(z)}\,. \tag{6}$$

The BRR variance estimate is

$$v_{\text{BRR}}(\hat{\theta}) = \frac{1}{\phi^2 R}\sum_{r=1}^{R}(\hat{\theta}^r - \hat{\theta})^2\,. \tag{7}$$

The whole procedure needs $R+1$ sets of (regression) weights, i.e. one for the parent sample and one for each of the half-samples. Once these weights have been computed, the BRR variance estimates can be computed straightforwardly for an arbitrary set of statistics. This way, variances for all kinds of quantities can be estimated using the same simple expression. More details on the method described here can be found in Renssen et al. (1997). There it is also shown that if the grouped method is applied to the case of a linear statistic, the variance estimator (7), using the design weights (5), is unbiased for (2).

4 Two-per-stratum sampling

In the first study we compare variance estimates for several estimators of population parameters under a sampling design well-suited to BRR: simple random sampling of two units per stratum (without replacement). We use a population of businesses stratified according to activity code and size class in 84 strata of sizes between 10 and 20. This population is based on an existing sample.

A total of 2000 samples of size 168 (= 84 × 2) were drawn. Point estimates were computed for three parameters in each sample: total turnover for small businesses, θ_1, total turnover for large businesses, θ_2, and a ratio θ_r of two totals, turnover and value added. These parameters were estimated using HT estimators and regression estimators. The regression model used in the regression estimator was $SizeClass_2 + SizeClass_2 \times TurnoverTax$ where $SizeClass_2$ is a categorical variable with two size categories and $TurnoverTax$ is a quantitative variable available from the tax office.

Along with the point estimates, several variance estimates for these were computed, namely

1. Taylor: the standard linearization variance estimator
2. Taylor(g): the same as Taylor but using regression weights instead of design weights in the variance estimator
3. BRR: standard BRR, i.e. with Fay factor equal to 1
4. BRR(.57): BRR with non-standard Fay factor 0.57
5. BRR(.01): BRR with Fay factor equal to 0.01.

For BRR estimation, a balanced set of 84 resamples was used.

We first discuss the results for the case without auxiliary information, i.e. variance estimates for the HT estimators. Taylor and Taylor(g) estimates are the same in this case. For linear estimators, here the HT estimators $\hat{\theta}_1^{HT}$ for θ_1 and $\hat{\theta}_2^{HT}$ for θ_2, BRR variance estimates are independent of the Fay factor. Besides, it is well-known that BRR estimates agree with direct (Taylor) variance estimates for linear statistics in the two-per-stratum case. This was verified here in case of $\hat{\theta}_1^{HT}$ and $\hat{\theta}_2^{HT}$. The ratio θ_r was estimated by the ratio $\hat{\theta}_r^{HT}$

of two HT estimators. This is a non-linear estimator and the direct (Taylor) method relies on linearization. For $\hat{\theta}_1^{HT}$, $\hat{\theta}_2^{HT}$ and $\hat{\theta}_r^{HT}$, only Taylor and BRR(1) estimates were computed and their mean performances in terms of relative bias and relative root mean square error (RMSE) are summarized in Table 1.

Relative bias and relative RMSE are defined as

$$\text{rel. bias} = \frac{\overline{v_M(\hat{\theta}) - \nu(\hat{\theta})}}{\nu(\hat{\theta})}, \quad \text{rel. RMSE} = \frac{\sqrt{\overline{\left(v_M(\hat{\theta}) - \nu(\hat{\theta})\right)^2}}}{\nu(\hat{\theta})}, \tag{8}$$

where a bar denotes the average over the simulation runs, $v_M(\hat{\theta})$ is a variance estimate for $\hat{\theta}$ obtained by method M (e.g. Taylor), and $\nu(\hat{\theta})$ is the true design variance. We also give the coverage, i.e. the percentage of simulated intervals $\left[\hat{\theta} - z_{1-\alpha/2}\sqrt{v_M(\hat{\theta})}\,,\ \hat{\theta} + z_{1-\alpha/2}\sqrt{v_M(\hat{\theta})}\right]$ containing the true parameter θ, where $z_{1-\alpha/2}$ is a standard normal quantile, which we take at $\alpha = 0.05$.

For $\hat{\theta}_1^{HT}$ and $\hat{\theta}_2^{HT}$ the true design variance was computed directly from the population, whereas for the non-linear statistic $\hat{\theta}_r^{HT}$ it was simulated using 40000 runs.

Note that Table 1 shows simulated biases, although it is known that the variance estimators for the linear statistics $\hat{\theta}_1^{HT}$ and $\hat{\theta}_2^{HT}$ are unbiased. Not much can be said about these results; for linear statistics Taylor and BRR methods agree as they should, and for the non-linear statistic $\hat{\theta}_r^{HT}$, the differences between Taylor and BRR variance estimates seem unimportant.

Table 1. *Mean performance over 2000 simulation runs of Taylor and BRR variance estimates for HT estimators concerning a population of businesses.*

Statistic	Method	rel. bias (%)	rel. RMSE (%)	coverage (%)
$\hat{\theta}_1^{HT}$	Taylor	-0.973	37.4	92.2
	BRR	-0.973	37.4	92.2
$\hat{\theta}_2^{HT}$	Taylor	-1.61	40.2	91.4
	BRR	-1.61	40.2	91.4
$\hat{\theta}_r^{HT}$	Taylor	-0.666	56.4	91.5
	BRR(1)	0.735	57.2	91.9

Next, we discuss regression estimators and their variance estimates. For this case Taylor and Taylor(g) estimates are different because of the adjustment of weights. BRR estimates were computed for three different values of the Fay factor. Simulation results are listed in Table 2.

We noticed that Taylor(g) and BRR(.01) gave almost exactly the same variance estimates with relative differences less than a tenth of a percent in all samples. This is in line with a result of Rao and Shao (1999) who prove that as the Fay factor ϕ approaches 0, the BRR(ϕ) variance estimate approaches the linearization variance estimate Taylor. In the case of regression-weighted estimates we observe here that the BRR(ϕ) variance estimate approaches the Taylor(g) estimate.

Table 2. *Mean performance over 2000 simulation runs of Taylor and BRR variance estimates for regression estimators concerning a population of businesses.*

Statistic	Method	rel. bias (%)	rel. RMSE (%)	coverage(%)
$\hat{\theta}_1^R$	Taylor	-11.8	34.0	90.8
	Taylor(g)	-9.07	33.0	92.1
	BRR(.01)	-9.07	33.0	92.1
	BRR(.57)	-5.66	33.9	92.6
	BRR	4.34	39.3	93.6
$\hat{\theta}_2^R$	Taylor	-13.3	36.1	90.6
	Taylor(g)	-13.3	35.8	91.7
	BRR(.01)	-13.3	35.8	91.7
	BRR(.57)	-8.77	37.6	92.2
	BRR	4.94	48.9	93.1
$\hat{\theta}_r^R$	Taylor	-13.1	34.0	89.7
	Taylor(g)	-13.3	35.9	90.9
	BRR(.01)	-13.3	35.9	90.9
	BRR(.57)	-9.03	37.3	91.4
	BRR	-13.3	45.4	92.6

All methods, and especially the Taylor methods, tend to underestimate the variance of the regression estimator for a total, except the standard BRR method (Fay factor = 1), which instead shows a tendency to slightly overestimate the variance. These results are in line with results in Canty and Davison (2002).

The standard BRR method has largest mean square error, but then it also has slightly better coverage than the others. Lowest coverage is displayed by method Taylor (without g). The estimates corresponding to this method were seen to be negatively correlated with the actual squared error of the point estimates. Generally, the two-per-stratum simulation study described in this section shows only modest differences between Taylor and BRR methods for the statistics studied.

5 Grouped BRR and BRR with artificial strata

In the second study we compare variances of Taylor and BRR methods in the case that more than two elements per stratum are sampled according to a two-stage cluster design. The population used in this study was derived from the Dutch population as it was simulated for the Dacseis project using frequency tables derived from Labour Force Survey data. Municipalities were used as strata and groups of households were formed to represent primary sampling units (PSUs). The population consisted of 188216 persons divided over 70 municipalities (strata) each containing between ten and a few hundred PSUs.

Each PSU contained between 15 and about 40 households.

A total of 500 samples were drawn according to a two-stage design with simple random sampling without replacement in both stages. Complete households were sampled with a first-stage sampling fraction of 1/5 and a second-stage sampling fraction of 1/5 as well. Thus, approximately 4% of the population was sampled, yielding sample sizes around 7500.

For each sample, point estimates were computed for the number of persons falling into the three categories of the variable $Employment_3$: (1) number of employed labour force θ_1, (2) number of unemployed labour force θ_2, and (3) number of people not belonging to the labour force θ_3. These parameters were estimated using HT estimators and regression estimators. The weight used in the HT estimators is simply the design weight, i.e. $1/(5 \times 5) = 25$ for each person. The available auxiliary information for the regression estimator was purely categorical and the regression model used was $Sex_2 \times Age_6 + Sex_2 \times MaritalStatus_3 + Sex_2 \times Ethnicity_3$, where the subscripts denote the number of categories of each auxiliary variable.

The following variance estimators were compared:

1. Taylor: the standard linearization variance estimator
2. Taylor(g): the same as Taylor but using regression weights instead of design weights in the variance estimator
3. BRR: grouped BRR
4. BRRA: BRR with artificial strata.

For the grouped BRR method a balanced set of 72 resamples was used. For method BRRA the 70 existing strata were further subdivided such that a total of 120 artificial strata were obtained and the same number of resamples was used. Note that these 120 artificial strata are still far from a two-per-stratum situation, so that grouping was still needed. All BRR estimates were computed using Fay factor $\phi = 0.57$, which is the default used by Bascula. The variance estimates were computed for each of the three point estimates in all 500 samples. For the linear HT estimates there is no distinction between Taylor and Taylor(g).

True values for the variances of the HT and regression estimators were simulated using 10000 samples. For the HT estimators the results in terms of relative bias and relative root mean square error, defined in (8), are listed in Table 3. Table 4 holds the results for the regression estimators.

Table 3. *Mean performance over 500 simulation runs of Taylor and BRR variance estimates for HT estimators concerning a population of persons.*

Statistic	Method	rel. bias (%)	rel. RMSE (%)	coverage (%)
$\hat{\theta}_1^{HT}$	Taylor	2.21	15.4	95.2
	BRR	-0.079	29.2	94.6
	BRRA	2.09	22.5	95.0
$\hat{\theta}_2^{HT}$	Taylor	2.10	9.20	94.0
	BRR	1.55	25.5	92.2
	BRRA	2.19	16.6	93.6
$\hat{\theta}_3^{HT}$	Taylor	0.59	12.6	94.2
	BRR	1.45	29.5	93.0
	BRRA	1.17	20.5	93.6

Table 4. *Mean performance over 500 simulation runs of Taylor and BRR variance estimates for regression estimators concerning a population of persons.*

Statistic	Method	rel. bias (%)	rel. RMSE (%)	coverage (%)
$\hat{\theta}_1^{R}$	Taylor	0.097	4.83	94.8
	Taylor(g)	-0.41	4.77	94.2
	BRR	-0.86	25.0	93.8
	BRRA	0.60	13.7	93.4
$\hat{\theta}_2^{R}$	Taylor	2.43	8.93	94.6
	Taylor(g)	1.99	8.71	94.4
	BRR	2.16	25.8	92.2
	BRRA	2.38	16.3	94.2
$\hat{\theta}_3^{R}$	Taylor	-0.91	4.81	94.0
	Taylor(g)	-1.43	4.87	94.0
	BRR	-1.46	24.8	92.6
	BRRA	-0.20	13.3	93.6

It comes as no surprise that the grouped BRR method is significantly less efficient than is the Taylor method. The differences are more pronounced for the regression estimators. It is also clear that using a finer (artificial) stratification improves the efficiency of the BRR estimator. For the regression estimator this improvement is quite significant. We further observe that the Taylor and Taylor(g) variance estimates behave in nearly the same way. This is also as expected, since the sample size is reasonably large and, moreover, the auxiliary information used is only of categorical character.

6 Conclusions

For two-per-stratum designs the differences we found between Taylor and BRR estimates are not very large. For smaller samples and/or less smooth statistics than studied here the differences will presumably be greater, and in such cases the BRR estimator (using a not too small Fay factor) may be preferable. For smooth statistics there is a direct connection between BRR variance estimates and Taylor linearization estimates including g-weights: they approach each other as the Fay factor approaches zero.

In practice, designs are mostly not two-per-stratum. In the grouped setting BRR variance estimates are not as efficient as Taylor estimates. The differences can be rather large if strata sample sizes are large. In principle, one could define a much finer stratification and use a two-per-stratum design to obtain good BRR estimates of the design variance. Alternatively, one might construct artificial strata, which in the case study of Section 5 indeed improves the BRR estimates. However, for large sample sizes this can be done only to a limited extent, since otherwise the number of resamples becomes impractically large. Another approach to improve grouped BRR estimates is to do the BRR computations repeatedly with different random groupings. This approach has not been used in our simulation study.

BRR, as other resampling methods, has the advantage that no separate implementations are required for different types of statistics. Besides, if there are many population parameters to be estimated, the computational disadvantage of BRR as compared with Taylor linearization may disappear due to the costly computations of regression residuals required for the latter. In conclusion, there are certain situations (small samples perhaps, large number of non-linear statistics of various types) in which BRR has its advantages over the linearization approach. However, for larger samples this usually comes at the cost of lower precision.

References

1. Canty, A.J. and A.C. Davison (2002), *Resampling-based variance estimation for labour force surveys*, Dacseis research paper.
2. Fay, R.E. (1989), *Theory and Application of Replicate Weighting for Variance Calculations*, Proceedings of the Social Statistics Session, American Statistical Association, 212–217.
3. McCarthy, P.J. (1969), *Pseudo-replication: half samples*, Review of the International Statistical Institute **37**, 239–264.
4. Nieuwenbroek, N.J. and H.J. Boonstra (2002), *Bascula 4.0 Reference Manual*, Statistics Netherlands.
5. Rao, J.N.K. and J. Shao (1999), *Modified balanced repeated replication for complex survey data*, Biometrika **86**, 403–415.
6. Rao, J.N.K. and C.F.J. Wu (1988), *Resampling Inference with Complex Survey Data*, Journal of the American Statistical Association **83**, 231–241.
7. Renssen, R.H., N.J. Nieuwenbroek, and G.T. Slootbeek (1997), *Variance Module in Bascula 3.0: Theoretical Background*, Research paper no. 9712, Statistics Netherlands.

8. Särndal, C.E., B. Swensson, and J.H. Wretman (1989), *The weighted residual technique for estimating the variance of the general regression estimator of the finite population total*, Biometrika **76**, 527–537.
9. Särndal, C.E., B. Swensson, and J.H. Wretman, *Model Assisted Survey Sampling*, Springer-Verlag, New York.
10. Sitter, R.R. (1993), *Balanced Repeated Replications based on Orthogonal Multi-Arrays*, Biometrika **80**, 211–221.
11. Wolter, K.M. (1985), *Introduction to Variance Estimation*, Springer-Verlag, New York.
12. Wu, C.F.J. (1991), *Balanced Repeated Replications based on Mixed Orthogonal Arrays*, Biometrika **78**, 181–188.