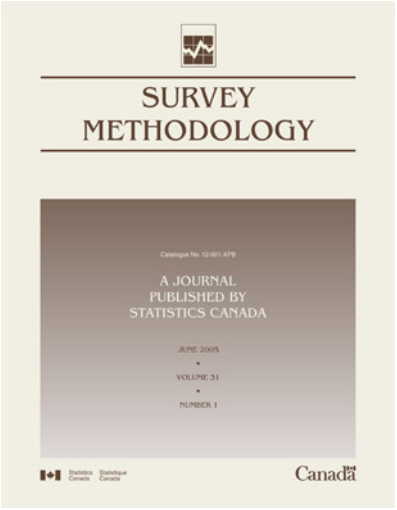
Catalogue no. 12-001-XIE

Survey Methodology

June 2005



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Business Survey Methods Division

Survey Methodology

June 2005

Published by authority of the Minister responsible for Statistics Canada

May 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 31, No. 1, pp. 95-99
Statistics Canada, Catalogue No. 12-001

An Optimal Calibration Distance Leading to the Optimal Regression Estimator

Per Gösta Andersson and Daniel Thorburn [1]

Abstract

When there is auxiliary information in survey sampling, the design based "optimal (regression) estimator" of a finite population total/mean is known to be (at least asymptotically) more efficient than the corresponding GREG estimator. We will illustrate this by some simulations with stratified sampling from skewed populations. The GREG estimator was originally constructed using an assisting linear superpopulation model. It may also be seen as a calibration estimator; *i.e.*, as a weighted linear estimator, where the weights obey the calibration equation and, with that restriction, are as close as possible to the original "Horvitz-Thompson weights" (according to a suitable distance). We show that the optimal estimator can also be seen as a calibration estimator in this respect, with a quadratic distance measure closely related to the one generating the GREG estimator. Simple examples will also be given, revealing that this new measure is not always easily obtained.

Key Words: Horvitz-Thompson estimator; Regression estimator; Survey sampling theory.

1. Notation and Basics

Consider a finite population U consisting of N objects labelled 1, ..., N with associated study values y_1, ..., y_N and $J-$dimensional auxiliary (column) vectors $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$. We want to estimate the population total $t_y = \sum_{i \in U} y_i$ by drawing a random sample s of size n (fixed or random) from U, with first and second order inclusion probabilities $\pi_i = P(i \in s)$, $\pi_{ij} = P(i, j \in s)$, $i, j = 1, ..., N$. The study values and the auxiliary vectors are recorded for the sampled objects and before the sample is drawn we assume that at least $\boldsymbol{t}_x = \sum_{i \in U} \boldsymbol{x}_i$ is known.

This is the standard setup for a regression estimator. In section 2 we discuss different regression estimators: the common GREG estimator (Särndal, Swensson and Wretman 1992), the optimal estimator (Montanari 1987, Andersson, Nerman and Westhall 1995) and calibration estimators (Deville and Särndal 1992). It is well known that the GREG estimator can be obtained as a calibration estimator. In section 3 it is shown that this holds also for the optimal estimator, but with a more complicated distance measure. In the last two sections this and the optimal estimator are illustrated, first by theoretical examples and then by simulations.

Finally some comments about matrix notation in this paper: Generally, the transpose of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^T$ and if $\boldsymbol{A}$ is square, the inverse (generalised inverse) is written $\boldsymbol{A}^{-1}(\boldsymbol{A}^-)$. We further let the column vectors $\boldsymbol{y} = (y_i)_{i \in s}$ and $\mathbf{w}_0 = (1/\pi_i)_{i \in s}$, $\boldsymbol{X}$ be the $J \times n$ "design" matrix of the auxiliary information from s and finally $\boldsymbol{I}_n$ means a unit diagonal matrix of size n.

2. Regression and Calibration Estimators

An unbiased simple estimator of t_y is the Horvitz-Thompson estimator $\hat{t}_y = \sum_{i \in s} y_i / \pi_i = \boldsymbol{y}^T \boldsymbol{w}_0$. However, more efficient estimators may be obtained utilising the auxiliary information, *e.g.*, the well-known model assisted GREG estimator, see Särndal *et al.* (1992). For example, constructed from the assumption of a homoscedastic linear regression superpopulation model the GREG estimator is

$$\hat{t}_{yr} = \boldsymbol{y}^T \boldsymbol{w}_0 + (\boldsymbol{y}^T \boldsymbol{R}_r \boldsymbol{X}^T)(\boldsymbol{X} \boldsymbol{R}_r \boldsymbol{X}^T)^{-1}(\boldsymbol{t}_x - \hat{\boldsymbol{t}}_x) \tag{1}$$

$$= \boldsymbol{y}^T \boldsymbol{g}, \tag{2}$$

where $\boldsymbol{R}_r = \boldsymbol{w}_0 \boldsymbol{I}_n$, $\hat{\boldsymbol{t}}_x = \sum_{i \in s} \boldsymbol{x}_i / \pi_i$ and

$$\boldsymbol{g} = \left(\frac{1}{\pi_i} (1 + \boldsymbol{x}_i^T (\boldsymbol{X} \boldsymbol{R}_r \boldsymbol{X}^T)^{-1} (\boldsymbol{t}_x - \hat{\boldsymbol{t}}_x)) \right)_{i \in s}.$$

Now, the expression (2) for the GREG estimator is interesting since we also have that

$$\boldsymbol{x}^T \boldsymbol{g} = \boldsymbol{t}_x, \tag{3}$$

which is called the *calibration equation*. This brings us to an alternative possible derivation of the GREG estimator according to Deville and Särndal (1992). Suppose that we seek an estimator $\boldsymbol{y}^T \boldsymbol{w}$ of t_y with a vector $\boldsymbol{w}$ of sample-dependent weights $(w_i)_{i \in s}$, which respects the corresponding calibration equation, while also minimising the distance between $\boldsymbol{w}$ and $\boldsymbol{w}_0$ according to the quadratic distance measure

1. Per Gösta Andersson, Mathematical Statistics, Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden; Daniel Thorburn, Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden.

$$(\boldsymbol{w}-\boldsymbol{w}_0)^T\,\boldsymbol{R}\,(\boldsymbol{w}-\boldsymbol{w}_0),$$

where $\mathbf{R}=(\boldsymbol{w}_0\ \boldsymbol{I}_n)^{-1}$.

This results in

$$\boldsymbol{w}=\boldsymbol{w}_0+\boldsymbol{R}^{-1}\boldsymbol{x}^T\,(\boldsymbol{X}\ \boldsymbol{R}^{-1}\boldsymbol{X}^T)^{-1}(\boldsymbol{t}_x-\hat{\boldsymbol{t}}_x), \tag{4}$$

which means that $\boldsymbol{w}=\boldsymbol{g}$, since here $\boldsymbol{R}=\boldsymbol{R}_r^{-1}$.

Turning to the optimal estimator, consider first the vector $(\hat{t}_y,\ \hat{\boldsymbol{t}}_{\boldsymbol{x}}^{\boldsymbol{T}})$ and let $\sum_{y,x}$ be the covariance (row) vector of $\hat{t}_y$ and $\hat{\boldsymbol{t}}_{\boldsymbol{x}}$ and $\sum_{x,x}$ the covariance matrix of $\hat{\boldsymbol{t}}_{\boldsymbol{x}}$. Now, the minimum-variance, see Montanari (1987), unbiased linear estimator (in $\hat{t}_y$ and $\hat{\boldsymbol{t}}_{\boldsymbol{x}}$) of t_y is the difference estimator

$$\hat{t}_y+\sum_{y,x}\sum_{x,x}^{-1}(\boldsymbol{t}_x-\hat{\boldsymbol{t}}_{\boldsymbol{x}}). \tag{5}$$

Since $\sum_{y,x}$ and $\sum_{x,x}$ in practice are unknown, we let the optimal estimator be

$$\begin{aligned}\hat{t}_{y\mathrm{opt}}&=\boldsymbol{y}^T\ \boldsymbol{w}_0+\hat{\sum}_{y,x}\,\hat{\sum}_{x,x}^{-1}(\boldsymbol{t}_x-\hat{\boldsymbol{t}}_x)\\&=\hat{t}_y+(\boldsymbol{y}^T\ \boldsymbol{R}_{\mathrm{opt}}\ \boldsymbol{X}^T)\,(\boldsymbol{X}\ \boldsymbol{R}_{\mathrm{opt}}\ \boldsymbol{X}^T)^{-1}(\boldsymbol{t}_x-\hat{\boldsymbol{t}}_x),\end{aligned} \tag{6}$$

where $\boldsymbol{R}_{\mathrm{opt}}=((\pi_{ij}-\pi_i\ \pi_j)/(\pi_{ij}\ \pi_i\ \pi_j))_{i,\ j\in s}$.

In an asymptotic context, where $n\to\infty$ and $N\to\infty$, $\hat{\Sigma}_{x,\,y}$ and $\hat{\Sigma}_{x,\,x}$ may be viewed as components of the asymptotic covariance matrix of $(\hat{t}_y,\ \hat{\boldsymbol{t}}_x^T)$. Under the assumption of consistency of $\hat{\Sigma}_{x,\,y}$ and $\hat{\Sigma}_{x,\,x}$, which holds under very mild conditions, see Andersson *et al.* (1995), the optimal estimator has the same asymptotic variance as the difference estimator (5). In particular it follows that the optimal estimator is asymptotically better than the usual GREG estimator, see Rao (1994), Montanari (2000) and Andersson (2001), *i.e.*, its asymptotic variance is never larger and usually smaller. In section 5 we actually present some simple simulations showing that the optimal estimator can be much more efficient than GREG. However, one does not know anything about the efficiency for finite samples, since the covariance estimator may converge slowly. The rate of convergence is illustrated in section 5. Note also that in some cases there exist asymptotically even better estimators which are not linear.

Now, the fact that the GREG estimator is also a calibration estimator using

$$(\boldsymbol{w}-\boldsymbol{w}_0)^T\,\boldsymbol{R}_r^{-1}\,(\boldsymbol{w}-\boldsymbol{w}_0) \tag{7}$$

as the distance measure and comparing (1) with (6), leads one to believe that replacing $\boldsymbol{R}_r$ by $\boldsymbol{R}_{\mathrm{opt}}$ in (7) should imply that we instead derive the optimal regression estimator as a calibration estimator. That this actually holds is shown below.

3. The Main Result

In order to show existence of a distance measure corresponding to the optimal estimator, we will first state and prove a result in the general case.

Lemma: With $\boldsymbol{R}$ denoting an arbitrary positive definite $n\times n$ matrix,

$$(\boldsymbol{w}-\boldsymbol{w}_0)^T\,\boldsymbol{R}\,(\boldsymbol{w}-\boldsymbol{w}_0) \tag{8}$$

subject to the constraint $\boldsymbol{X}\,\boldsymbol{w}=\boldsymbol{t}_x$, is minimised by

$$\boldsymbol{w}=\boldsymbol{w}_0+\boldsymbol{R}^{-1}\ \boldsymbol{X}^T\,(\boldsymbol{X}\ \boldsymbol{R}^{-1}\ \boldsymbol{X}^T)^{-1}(\boldsymbol{t}_x-\hat{\boldsymbol{t}}_x).$$

Proof: Introducing the $J\times 1$ vector $\boldsymbol{\lambda}$ of Lagrange multipliers, we get after differentiation the equation system

$$2\boldsymbol{R}(\boldsymbol{w}-\boldsymbol{w}_0)+\boldsymbol{X}^T\,\boldsymbol{\lambda}\ =\ 0 \tag{9}$$

$$\boldsymbol{X}W-t_x=0 \tag{10}$$

Multiplying (9) by $\boldsymbol{X}\boldsymbol{R}^{-1}$, using (10) and solving for $\boldsymbol{\lambda}$, yields with $\boldsymbol{X}\,\boldsymbol{w}_0=\hat{\boldsymbol{t}}_x$:

$$\boldsymbol{\lambda}=2(\boldsymbol{X}\,\boldsymbol{R}^{-1}\boldsymbol{X}^T)^{-1}(\hat{\boldsymbol{t}}_x-\boldsymbol{t}_x). \tag{11}$$

Putting this into (9) and solving for $\boldsymbol{w}$ finally leads to

$$\boldsymbol{w}=\boldsymbol{w}_0+\boldsymbol{R}^{-1}\ \boldsymbol{X}^T\,(\boldsymbol{X}\ \boldsymbol{R}^{-1}\ \boldsymbol{X}^T)^{-1}(\boldsymbol{t}_x-\hat{\boldsymbol{t}}_x).$$

From the lemma we thus have the following main result:

Theorem: With $\boldsymbol{R}_{\mathrm{opt}}$ being positive (semi –) definite and using the optimal calibration distance-measure, which we get by letting $\boldsymbol{R}=\boldsymbol{R}_{\mathrm{opt}}^{-1}$ ($\boldsymbol{R}_{\mathrm{opt}}^{-}$) in (8), the calibration estimator will become the optimal regression estimator.

Remark: $\boldsymbol{R}_{\mathrm{opt}}$ may in some cases be indefinite (see below). The only thing we know is that it is an unbiased estimator of a covariance matrix. If it is not positive semi-definite there also exist $x-$values such that $\boldsymbol{X}\,\boldsymbol{R}_{\mathrm{opt}}\ \boldsymbol{X}^T$ is not positive semi-definite, but the probability of such $x-$values goes to zero as the population and sample sizes increase (and if $\sum_{x,x}$ is positive definite). A strict minimisation of a distance with "a negative component" would lead to infinitely large corrections. This problem of the optimal estimator has, to our knowledge, not been pointed out previously.

The simplest way to find a distance which gives the optimal estimator as a calibration estimator is to find a matrix $\boldsymbol{R}_{\mathrm{dist}}$ which has the same eigenvectors as $\boldsymbol{R}_{\mathrm{opt}}$ but where the eigenvalues are replaced by their absolute values. (This result can be shown along the same lines as the proof of the lemma above. The distance can be seen as the sum of

the products of the eigenvalues and the squared eigenvectors. Putting the derivatives equal to zero means that in the proposition we found the extremes *i.e.*, the minima for positive eigenvalues and the maxima for negative eigenvalues. By changing all negative signs the extremes will all be minima).

4. Examples

Positive definite $\boldsymbol{R}_{\text{opt}}$: Suppose that the objects in U are independently drawn with inclusion probabilities $\pi_1, ..., \pi_N$ (Poisson sampling); thus implying a random sample size n, where $E[n] = \sum_{i \in U} \pi_i$. Due to the independence of drawings, $\boldsymbol{R}_{\text{opt}}$ is diagonal and specifically

$$\boldsymbol{R}_{\text{opt}}^{-1} = \boldsymbol{I}_n \left(\frac{\pi_i^2}{1-\pi_i} \right)_{i \in s}.$$

Positive semi-definite $\boldsymbol{R}_{\text{opt}}$: Suppose n objects are drawn according to simple random sampling, *i.e.*, each object has inclusion probability $\pi_i = n/N$. The elements of $\boldsymbol{R}_{\text{opt}}$ are

$$i = j: \left(\frac{N}{n} \right)^2 \frac{N-n}{N}$$

$$i \neq j: \left(\frac{N}{n} \right)^2 \frac{n-N}{N(n-1)}.$$

This means that $\boldsymbol{R}_{\text{opt}}$ is singular with rank $n-1$.

Suppose instead (as in the following simulation study) that U is partitioned into L strata of sizes $N_1, ..., N_L$, from which we draw independent simple random samples of sizes $n_1, ..., n_L$. The elements of $\boldsymbol{R}_{\text{opt}}$ then are

$$i = j: \left(\frac{N_h}{n_h} \right)^2 \frac{N_h - n_h}{N_h}$$

$$i \neq j: \left(\frac{N_h}{n_h} \right)^2 \frac{n_h - N_h}{N_h (n_h - 1)},$$

when in the latter case i and j both belong to stratum h, $h = 1, ..., L$ and 0 otherwise. Therefore $\boldsymbol{R}_{\text{opt}}$ has rank $N-h$.

Non positive semi-definite $\boldsymbol{R}_{\text{opt}}$: Let U consist of four elements and s of two elements. Suppose that a systematic sample is taken with probability 0.94 and a simple random sample with probability 0.06, *i.e.*, $\pi_{13} = \pi_{24} = 0.48$ and $\pi_{12} = \pi_{14} = \pi_{23} = \pi_{34} = 0.01$. In that case

$$\boldsymbol{R}_{\text{opt}} = \begin{pmatrix} 2 & 23/12 \\ 23/12 & 2 \end{pmatrix} \tag{12}$$

with probability 0.96 and

$$\boldsymbol{R}_{\text{opt}} = \begin{pmatrix} 2 & -96 \\ -96 & 2 \end{pmatrix} \tag{13}$$

with probability 0.04. The second matrix has a negative eigenvalue.

The problem does not necessarily disappear if N is large. Consider instead a population consisting of $N/4$ strata with four elements each. Suppose that the above sampling procedure is used independently in each stratum. In that case $\boldsymbol{R}_{\text{opt}}$ will consist of a matrix with the above 2×2 – matrices along the diagonal and zeroes elsewhere.

5. A Simulation Study

5.1 Notation and Outline

In order to make empirical comparisons between the optimal estimator (OPT) and the GREG estimator (GREG) and also compare these estimators with the Horvitz-Thompson estimator (HT), we have conducted a small simulation study. In the previous sections we mentioned that OPT is Best Linear Asymptotic Efficient and a calibration estimator. Even though it has many nice properties it may for reasonable sample sizes be inefficient. Here we will in some simulated situations show that the optimal estimator can be a substantial improvement compared to GREG also for moderate sample sizes when the population is (deliberately) chosen to be unfavourable for GREG. A simple but non-trivial situation for which OPT is not equal to GREG arises for stratified simple random sampling, in particular, when the slopes differ between the different strata and the unstratified population. Consider therefore a population of size N, which is partitioned into L strata of sizes $N_1, ..., N_L$. From each stratum h a simple random sample s_h of size n_h is drawn, where $s_1 + ... s_L = s$ and $n_1 + ... + n_L = n$. For simplicity we further assume that the auxiliary information is one-dimensional and global, *i.e.*, only t_x is known beforehand. For GREG we have chosen the homoscedastic simple linear regression model, see Särndal *et al.* (1992).

The resulting expressions for HT, OPT and GREG respectively are

$$\hat{t}_y = N \bar{y}_{st}$$
$$\hat{t}_{y\,\text{opt}} = N(\bar{y}_{st} + \hat{B}_{\text{opt}} (\bar{x} - \bar{x}_{st}))$$
$$\hat{t}_{yr} = N(\bar{y}_{st} + \hat{B}_r (\bar{x} - \bar{x}_{st})),$$

where $\bar{x} = (1/N) \sum_{i=1}^{N} x_i$, $\bar{y}_{st} = (1/N) \sum_{h=1}^{L} N_h \bar{y}_{s_h}$, ($\bar{x}_{st}$ analogous) and

$$\hat{B}_{\text{opt}} = \frac{\sum_{h=1}^{L} \frac{N_h^2}{n_h - 1}\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\sum_{i \in s_h}(x_i - \bar{x}_{s_h})(y_i - \bar{y}_{s_h})}{\sum_{h=1}^{L} \frac{N_h^2}{n_h - 1}\left(\frac{1}{n_h} - \frac{1}{N_h}\right)\sum_{i \in s_h}(x_i - \bar{x}_{s_h})^2}$$

$$\hat{B}_r = \frac{\sum_{h=1}^{L} \frac{N_h}{n_h}\sum_{i \in s_h}(x_i - \bar{x}_{st})(y_i - \bar{y}_{st})}{\sum_{h=1}^{L} \frac{N_h}{n_h}\sum_{i \in s_h}(x_i - \bar{x}_{st})^2}.$$

It is easily seen from these formulae that the optimal regression coefficient is the mean of the within stratum slopes and that the GREG regression coefficient is the global slope. When there is a large difference between these slopes the GREG correction becomes bad. We are here particularly interested in comparing the qualities of these estimators when the assisting (linear) model for GREG fails. We have thus generated $x-$ and $y-$values from correlated lognormally distributed random variables X and Y, where $\ln X$ is normally distributed with expectation 0 and variance σ_1^2 ($N(0, \sigma_1)$) and $\ln Y$ is $N(0, \sigma_2)$. The variances σ_1^2 and σ_2^2 and the correlation between $\ln X$ and $\ln Y$ can then be chosen to obtain prespecified values of the variances σ_x^2 of X and σ_y^2 of Y and their correlation $\rho(X, Y)$. Values generated from bivariate normal distributions were obtained by MATLAB (version 6.0). Twelve populations have in this manner been created, each of size $N = 10{,}000$, including four combinations of variances σ_x^2 and σ_y^2 (10 and 100) and three values of the correlation $\rho(X, Y)$ (0.5, 0.7 and 0.9). For these populations a variance of 10 implies a skewness of 9.37 and the variance 100 leads to skewness 38.59.

Now, before stratification, the objects of each population are ordered with respect to ascending $y-$values. The number of strata is $L = 5$ throughout with sizes $N_1 = 4{,}000$, $N_2 = 2{,}500$, $N_3 = 2{,}000$, $N_4 = 1{,}000$ and $N_5 = 500$. These strata are constructed in such a way that objects with the smallest $y-$values constitute stratum 1, and so forth. From each stratified population we have drawn samples of sizes $n = 250$, 1,000 and 2,500, where for each sample $n_1 = ... = n_5$. This means that we have created an approximate πps (probability proportional to size) design, with for example, objects in stratum 5 having the largest inclusion probability (n_5 / N_5). The number of simulated samples was $K = 25{,}000$ for each of the $12 \times 3 = 36$ cases and HT, OPT and GREG were then computed for each sample.

In general, common measures of quality for an estimator $\hat{t}$ of a total t from a sequence $\hat{t}_1$, ..., $\hat{t}_L$ are the estimated relative bias

$$\frac{\bar{\hat{t}} - t}{t}$$

and the estimated variance

$$S^2 = \frac{1}{K-1}\sum_{i=1}^{K}(\hat{t}_i - \bar{\hat{t}})^2,$$

where $\bar{\hat{t}} = (1/K)\sum_{i=1}^{K}\hat{t}_i$.

Since we are mainly concerned with comparisons of OPT and GREG, we will only display results of the relative measures of variance (or equivalently standard deviation)

$$\frac{S_{y\,\text{opt}}^2}{S_{y\,\text{HT}}^2} \text{ and } \frac{S_{yr}^2}{S_{y\,\text{HT}}^2},$$

from which we can compare the estimated variances of OPT and GREG with HT and also determine which of OPT and GREG have the lowest estimated variance.

5.2 Results

Firstly, as reference, the absolute value of the estimated relative bias of the unbiased HT did not in any case exceed $4 \cdot 10^{-4}$. The corresponding maximum values for OPT and GREG were $6 \cdot 10^{-3}$, which means that we may concentrate on the ratios of estimated variances in order to evaluate relative efficiencies of HT, OPT and GREG.

As seen from Table 1, OPT is superior to both HT and GREG (with one exception: $\rho(X, Y) = 0.9$, $\sigma_x^2 = 10$, $\sigma_y^2 = 100$ and $n = 250$, where GREG has slighly less estimated variance). For the lowest correlation though, the decrease in estimated variance for OPT compared with HT is not substantial. GREG on the other hand does not compete well with the others and this anomaly is particularly accentuated for the largest sample size $n = 2{,}500$. Changing $\rho(X, Y)$ to 0.7 means improvement for both OPT and GREG, but GREG is also now for most cases inferior to HT. Finally, for $\rho(X, Y) = 0.9$ GREG still displays poor behavior compared with HT for $n = 2{,}500$ (with the exception of $\sigma_x^2 = 100$ and $\sigma_y^2 = 10$). In general GREG is closing in on OPT for increasing values of $\rho(X, Y)$ (the assisting linear model becoming less misspecified), while OPT, on the other hand, is increasing its superiority over GREG for increasing sample sizes, which should come as no surprise since OPT is asymptotically well motivated.

Table 1
Relative Estimated Efficiencies (Given as Percentages) of OPT $(S^2_{y\,\mathrm{opt}} / S^2_{y\,\mathrm{HT}})$ and GREG $(S^2_{y\,r} / S^2_{y\,\mathrm{HT}})$ to HT, Based on 25,000 Simulated Samples for Each Sample Size

	$\sigma_x^2 = 10$		$\sigma_x^2 = 10$		$\sigma_x^2 = 100$		$\sigma_x^2 = 100$	
	$\sigma_y^2 = 10$		$\sigma_y^2 = 100$		$\sigma_y^2 = 10$		$\sigma_y^2 = 100$	
	OPT	GREG	OPT	GREG	OPT	GREG	OPT	GREG
$\rho(X, Y) = 0.5$								
$n = 250$	99.1	232.8	97.4	176.8	93.9	179.4	91.4	122.3
$n = 1{,}000$	98.3	247.1	98.0	193.7	97.5	183.5	99.9	141.9
$n = 2{,}500$	96.8	756.7	96.8	1,455.0	97.8	534.7	96.8	1,625.5
$\rho(X, Y) = 0.7$								
$n = 250$	89.7	197.6	83.8	101.2	73.6	120.4	64.3	72.9
$n = 1{,}000$	91.0	227.5	89.8	117.2	81.2	120.5	71.7	84.0
$n = 2{,}500$	93.8	648.2	91.5	1,308.6	93.1	218.6	93.1	673.5
$\rho(X, Y) = 0.9$								
$n = 250$	56.5	76.1	41.2	38.8	27.2	43.4	40.4	41.4
$n = 1{,}000$	61.8	87.3	44.1	44.2	27.6	44.1	41.5	45.4
$n = 2{,}500$	77.0	237.4	59.8	335.4	63.6	66.0	74.6	259.8

References

Andersson, P.G. (2001). Improving estimation quality in large sample surveys. Ph. D. Thesis, Department of Mathematics, Chalmers University of Technology and Göteborg University.

Andersson, P.G., Nerman, O. and Westhall J. (1995). Auxiliary information in survey sampling. *Technical Report NO 1995:3*, Department of Mathematics, Chalmers University of Technology and Göteborg University.

Deville, J.C., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Montanari, G.E. (1987). Post-sampling efficient QR-prediction in large-sample surveys. *International Statistical Review*, 55, 191-202.

Montanari, G.E. (2000). Conditioning on auxiliary variable means in finite population inference. *Australian & New Zealand Journal of Statistics*, 42, 407-421.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.