Survey Methodology, December 1996
Vol. 22, No. 2, pp. 117-125
Statistics Canada



# Variance Estimation for Calibration Estimators: A Comparison of Jackknifing Versus Taylor Linearization

DIANA M. STUKEL, MICHAEL A. HIDIROGLOU and CARL-ERIK SÄRNDAL[1]

ABSTRACT

The use of auxiliary information in estimation procedures in complex surveys, such as Statistics Canada's Labour Force Survey, is becoming increasingly sophisticated. In the past, regression and raking ratio estimation were the commonly used procedures for incorporating auxiliary data into the estimation process. However, the weights associated with these estimators could be negative or highly positive. Recent theoretical developments by Deville and Särndal (1992) in the construction of "restricted" weights, which can be forced to be positive and upwardly bounded, has led us to study the properties of the resulting estimators. In this paper, we investigate the properties of a number of such weight generating procedures, as well as their corresponding estimated variances. In particular, two variance estimation procedures are investigated via a Monte Carlo simulation study based on Labour Force Survey data; they are Jackknifing and Taylor Linearization. The conclusion is that the bias of both the point estimators and the variance estimators is minimal, even under severe "restricting" of the final weights.

KEY WORDS: Auxiliary information; Raking ratio estimators; Regression estimators; Restricted weighting.

## 1. INTRODUCTION

Auxiliary information has many uses in survey sampling. One typical use is its incorporation at the estimation stage through the use of regression estimators or raking ratio estimators. For these estimators, a unit's sampling weight is multiplied by an adjustment factor to produce the final weight. A well-known shortcoming associated with the regression estimator is that some of the adjustment factors may be negative, resulting in negative final weights. On the other hand, for the raking ratio estimator, some adjustment factors may be very large and positive, resulting in unduly large final weights. These shortcomings can be overcome by considering a family of estimators, known as "calibration estimators". Developed by Deville and Särndal (1992), the estimators in this family incorporate auxiliary information, and in certain cases, non-negative weights can be ensured by prespecifying lower and upper bounds on the weights. These "calibration" weights are obtained by minimizing functions which measure the distances between original sampling weights and final calibrated weights, while respecting a set of benchmarking constraints. Huang and Fuller (1978) and Singh and Mohl (1996) have developed similar estimators which maintain the above properties. Ordinarily, there are very small differences between the point estimates corresponding to the various distance functions.

Historically, Statistics Canada's Labour Force Survey (LFS) has used, at different points in time, both the Taylor and Jackknife variance estimation techniques in tandem with regression and raking ratio estimators. Recently, the LFS has also allowed for the option of using other calibration estimators in addition to the previously available regression estimator, to eliminate the problem of potential negative weights. It is therefore of interest to investigate the behaviour of these point estimators and their corresponding Taylor and Jackknife variance estimators, particularly for those estimators that allow bounding on the weights. Therein lies the main focus of this paper. Now, both the Taylor and the Jackknife have their advantages. The Taylor method is computationally much less intensive than the Jackknife method, but requires working out new expressions for each different parameter that is considered; this is particularly a burden in multipurpose surveys where many different parameters may be of interest. On the other hand, for the Jackknife method, cumbersome variance expressions need not be derived for each new parameter; only the functional form of the point estimator itself is required.

The paper is structured as follows: section 2 provides the theoretical underpinnings of calibration estimation and introduces a family of related distance functions. In section 3, variances for calibration estimators are discussed. Section 4 provides the results of a Monte Carlo simulation study, in which the bias of both the point estimators and their corresponding Taylor and Jackknife variance estimators (relative to a "true" variance) is tracked, for a variety of distance functions from calibration theory. In section 5, some concluding remarks are made.

## 2. DISTANCE FUNCTIONS AND CALIBRATION ESTIMATORS

We begin by introducing the basic idea behind calibration estimation. Let $U = \{1, ..., k, ..., N\}$ denote the index set for

[1] Diana M. Stukel, Household Survey Methods Division, and Michael A. Hidiroglou, Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6; Carl-Erik Särndal, Département de Mathématiques et de Statistique, Université de Montréal, C.P. 6128, Succursale A, Montréal, P.Q., H3C 3J7.

the $N$ units of a finite population of units. In survey sampling, one is often interested in estimating parameters of a finite population such as totals, means and ratios. For the sake of simplicity, we will focus on totals, although the ideas presented in this paper may easily be extended to include other parameters. Thus, suppose the objective is to estimate the population total $Y = \sum_{k\in U} y_k$, where $y_k$ is the value of $y$, the variable of interest for the $k$-th population unit.

A probability sample $s$ is drawn from $U$ by a given sampling design which induces the inclusion probabilities $\pi_k = P(k\in s)$. These are assumed known and positive. Let $a_k = 1/\pi_k$ be the sampling weight associated with the $k$-th unit. Finally, let the auxiliary information be specified in the form of known population totals of one or more auxiliary variables.

An elementary estimator of $Y$ is the Horvitz-Thompson (HT) estimator:

$$\hat{Y}_a = \sum_{k\in s} a_k y_k.$$

The HT estimator possibly but not necessarily (depending on the sampling design) incorporates auxiliary information at the design stage only; what is sought is an improved estimator which incorporates the auxiliary information at the estimation stage, as well. The incorporation of auxiliary information can be reflected in the creation of new weights, denoted by $w_k$; $k\in s$. The new estimator is then of the form:

$$\hat{Y}_w = \sum_{k\in s} w_k y_k. \tag{2.1}$$

The approach of Deville and Särndal (1992) and Deville, Särndal and Sautory (1993) involves determining these new weights $\{w_k; k\in s\}$ by making them as close as possible to the original sampling weights $\{a_k; k\in s\}$ according to a specified distance function. Constraints placed on the new weights are such that, when applied to each of the auxiliary variables, the known population total $\boldsymbol{X}$ is reproduced. That is,

$$\sum_{k\in s} w_k \boldsymbol{x}_k = \boldsymbol{X} \tag{2.2}$$

is required to hold, leading to a problem in constrained minimization. Here $\boldsymbol{x}_k' = (x_{1k}, x_{2k}, ..., x_{pk})$ is a vector of length $p$ containing the values of the auxiliary variables for the $k$-th individual, and the auxiliary information available from an external source is summarized by the known vector total $\boldsymbol{X} = \sum_{k\in U} \boldsymbol{x}_k$.

We denote the distance from $w_k$ to $a_k$ by $F^*(w_k, a_k)$. Deville and Särndal (1992) limit their discussions to distance functions of the form $F^*(w_k, a_k) = a_k c_k F(w_k/a_k)$ where $w_k/a_k = g_k$, the ratio of the final calibrated weight to original sampling weight, is called the "$g$-factor". Here $c_k$ is a known positive weight unrelated to $a_k$; the uniform weighting $c_k = 1$ is often used in applications. Note that equation (2.1) can alternatively be written as:

$$\hat{Y}_w = \sum_{k\in s} a_k g_k y_k.$$

It is assumed that $F$ is non-negative and convex, and that $F(1) = 0$, implying that when $w_k = a_k$ the distance between the weights is zero. Moreover, it is required that $F'$ is continuous, one-to-one, and that $F'(1) = 0$ and $F''(1) > 0$ which makes $w_k = a_k$ a local minimum. (See Deville, Särndal and Sautory 1993.) The total distance, $\sum_{k\in s} a_k c_k F(w_k/a_k)$, is minimized subject to the constraint (2.2). That is,

$$\sum_{k\in s} a_k c_k F(w_k/a_k) - \lambda'\left(\sum_{k\in s} w_k \boldsymbol{x}_k - \boldsymbol{X}\right)$$

is minimized with respect to the $w_k$, where $\lambda$ is a $p$-vector of Lagrange multipliers. Differentiating with respect to $w_k$, equating to zero, and solving for $w_k$ leads to the calibrated weights $w_k = a_k g_k = a_k g(\lambda' \boldsymbol{x}_k/c_k)$ where $g$ is the inverse function of $f$ and $f(z) = dF(z)/dz$. To compute $w_k$, one must first obtain $\lambda$ as the solution of the calibration equation implied by (2.2), namely,

$$\sum_{k\in s} a_k g(\lambda' \boldsymbol{x}_k/c_k)\boldsymbol{x}_k = \boldsymbol{X}. \tag{2.3}$$

The solution of this (possibly) nonlinear system of $p$ equations in $p$ unknowns may require the use of some iterative procedure, such as the Newton-Raphson method.

A number of distance functions are considered by Deville and Särndal (1992), Huang and Fuller (1978) and Singh and Mohl (1996). Two important distance functions which we first discuss are the Generalized Least Squares (GLS) distance function and the Raking Ratio (RR) distance function, both given in Deville and Särndal(1992).

The GLS distance function is defined by:

$$F^*(w_k,a_k) = F^*_{\mathrm{GLS}}(w_k,a_k)$$
$$= c_k(w_k - a_k)^2/a_k = a_k c_k (w_k/a_k - 1)^2. \tag{2.4}$$

It generates the well-known generalized regression estimator (GREG), which encompasses as special cases the ratio estimator, the simple regression estimator, and the simple post-stratified estimator, among others. It follows from (2.3) that the calibrated weights corresponding to the GLS distance function are:

$$w_k = a_k g_k = a_k[1 + (\boldsymbol{X} - \hat{\boldsymbol{X}}_a)'\left(\sum_{j\in s} a_j \boldsymbol{x}_j \boldsymbol{x}_j'/c_j\right)^{-1} \boldsymbol{x}_k/c_k]$$

where $\hat{\boldsymbol{X}}_a = \sum_{k\in s} a_k \boldsymbol{x}_k$ is the HT estimator of $\boldsymbol{X}$. The corresponding estimator of $Y$ can be written in the usual regression estimator form as

$$\hat{Y}_{w(\mathrm{GREG})} = \hat{Y}_a + (\boldsymbol{X} - \hat{\boldsymbol{X}}_a)'\hat{\beta} \tag{2.5}$$

where

$$\hat{\beta} = \left(\sum_{k\in s} a_k \boldsymbol{x}_k \boldsymbol{x}_k'/c_k\right)^{-1} \sum_{k\in s} a_k \boldsymbol{x}_k y_k/c_k. \tag{2.6}$$

Thus, the regression estimator can be thought of as the HT estimator plus an adjustment term. A drawback of the GLS distance function is that it may give rise to negative weights, particularly if the system is overconstrained. In practice, negative weights are rare; however, it is desirable to eliminate them entirely since it may be difficult to give them any meaningful interpretation.

The Raking Ratio (RR) distance function is defined by:

$$F^*(w_k,a_k) = F^*_{RR}(w_k,a_k)$$
$$= c_k[w_k \log(w_k/a_k) - w_k + a_k] \qquad (2.7)$$
$$= a_k c_k[(w_k/a_k)\log(w_k/a_k) - (w_k/a_k) + 1].$$

Solving for $g$-factors using the RR distance function and the constraint defined by equation (2.3) can be shown to be equivalent to using the Iterative Proportional Fitting (IPF) algorithm of Deming and Stephan (1940) when calibrating on known marginals of frequency tables of dimension two or higher. Unlike the GLS distance function, which has a closed form solution, the calibration equations for the RR distance function can only be solved iteratively. Computer software exists for this purpose; for example, the CALMAR software (see Deville, Särndal and Sautory 1993) solves the calibration equations for the RR distance function using the Newton-Raphson method, rather than the IPF algorithm originally proposed by Deming and Stephan. The RR distance function always ensures positive weights; however, it also has the undesirable property that some of the resulting calibration weights can be excessively large.

Neither the possibility of negative weights produced by the GLS distance function nor the possibility of large positive weights produced by the RR distance function are desirable. One can define restricted distance functions whereby the range of the resulting weights $w_k$ are limited. This is achieved by imposing restrictions on the distance function $F(w_k/a_k)$ in such a way that the $g$-factors $g_k = w_k/a_k$ are bounded within a prespecified interval. To this end, one can specify a lower bound $L$ and an upper bound $U$, such that $L < 1 < U$. To guarantee positive weights, one would choose $L > 0$. Now, Deville and Särndal (1992) define restricted versions of the two distance functions given above; they are: the Restricted GLS (RGLS) distance function and the Restricted Raking Ratio (RRR) or Logit distance function. Two other methods of restricting final weights are proposed by Huang and Fuller (1978) and Singh and Mohl (1996). All four restricted distance functions are considered in this paper; they are also discussed in detail in Singh and Mohl (1996), but from a different perspective.

The Restricted GLS distance function is defined by:

$$F^*(w_k,a_k) =$$
$$F^*_{RGLS}(w_k,a_k) = \begin{cases} c_k(w_k - a_k)^2/a_k & \text{if} \quad L < w_k/a_k < U \\ \infty \quad \text{otherwise.} \end{cases} \qquad (2.8)$$

The Restricted RR (or Logit) distance function is defined by:

$$F^*(w_k,a_k) = F^*_{RRR}(w_k,a_k) =$$
$$\begin{cases} A^{-1}c_k[(w_k/a_k - L)\log[(w_k/a_k - L)/(1 - L)] \\ \quad + (U - w_k/a_k)\log[(U - w_k/a_k)/(U - 1)]] \\ \qquad\qquad\qquad\qquad \text{if} \quad L < w_k/a_k < U \\ \infty \quad \text{otherwise} \end{cases} \qquad (2.9)$$

where $A = (U - L)/\{(1 - L)(U - 1)\}$. The specification $L = 0$, $U = \infty$ gives the RR distance function. It is easy to show that the Restricted GLS and Restricted RR distance functions share the property that the corresponding weights $w_k$ satisfy $L < w_k/a_k < U$.

Now, Huang and Fuller (1978) propose a method for adjusting regression weights such that the calibration constraints given by equation (2.2) are satisfied and such that the $g$-factors are restricted to lie close to one. Singh and Mohl (1996) show that their method can be written in terms of minimizing a distance function which changes from iteration to iteration. Singh and Mohl also modify the original method to allow for arbitrary restrictions on the $g$-factors, similar to the restricted distance functions above, and show that the estimator resulting from the modified distance function is asymptotically equivalent to the regression estimator. The Modified Huang-Fuller (MHF) distance function is given by:

$$F^*(w_k^{(\nu-1)}, a_k) = F^{(\nu)*}_{MHF}(w_k^{(\nu-1)}, a_k)$$
$$= (w_k^{(\nu-1)} - a_k)^2/a_k q_k^{(\nu-1)*}; \quad \nu = 1,2,\ldots \qquad (2.10)$$

where $q_k^{(\nu-1)*} = q_k^{(\nu-1)} \ldots q_k^{(1)} q_k^{(0)}$ with $q_k^{(0)} = 1$ and where $\nu$ is the iteration number. Here,

$$q_k^{(\nu-1)} = \begin{cases} 1 & \text{if} \quad \xi_k^{(\nu-1)} < .5 \\ 1 - \delta(\xi_k^{(\nu-1)} - .5)^2 & \text{if} \quad .5 \le \xi_k^{(\nu-1)} < 1 \\ (1 - \delta/4)/\xi_k^{(\nu-1)} & \text{if} \quad \xi_k^{(\nu-1)} \ge 1 \end{cases}$$

for $\delta$ arbitrarily chosen such that $0 < \delta < 1$. Also

$$\xi_k^{(\nu-1)} = \begin{cases} (g_k^{(\nu-1)} - 1)/(L' - 1) & \text{if} \quad g_k^{(\nu-1)} \le 1 \\ (g_k^{(\nu-1)} - 1)/(U' - 1) & \text{otherwise} \end{cases}$$

where $L' = \alpha L + 1 - \alpha$ and $U' = \alpha U + 1 - \alpha$ for $\alpha$ arbitrarily chosen such that $0 < \alpha < 1$ and $L$ and $U$ are as in earlier restricted distance functions. The parameters $\alpha$ and $\delta$ serve to speed up the convergence of the iterative algorithm used to provide a solution. Singh and Mohl (1996) empirically test a variety of values for these parameters using large data sets, and suggest that $\alpha = .67$ and $\delta = .8$ work well in practice. Finally, the $g$-factor at each iteration is

$$g_k^{(\nu-1)} = 1 + (\boldsymbol{X} - \hat{\boldsymbol{X}}_w^{(\nu-2)})' \left( \sum_{j \in s} a_j q_j^{(\nu-2)*} \boldsymbol{x}_j \boldsymbol{x}_j' \right)^{-1} \boldsymbol{x}_k; \quad \nu = 2, 3, \ldots$$

where $\hat{\boldsymbol{X}}_w^{(\nu-2)} = \sum_{k \in s} w_k^{(\nu-2)} \boldsymbol{x}_k$; $\nu = 2, 3, \ldots$ and where $w_k^{(\nu-2)} = a_k g_k^{(\nu-2)}$; $\nu = 2, 3, \ldots$. Starting values are given by $g_k^{(0)} = 1$ and $w_k^{(0)} = a_k$.

Singh and Mohl (1996) also propose a new distance function which changes from iteration to iteration called the Shrinkage-Minimization (SM) distance function, and show that the estimator resulting from this distance function is also asymptotically equivalent to the regression estimator. It is given by:

$$F^*(w_k^{(\nu-1)}, a_k) = F_{\mathrm{SM}}^{(\nu)*}(w_k^{(\nu-1)}, a_k) = (w_k^{(\nu-1)} - a_k^{(\nu-1)*})^2 / a_k^{(\nu-1)*}; \qquad \nu = 1, 2, \ldots \quad (2.11)$$

where

$$a_k^{(\nu-1)*} = \begin{cases} L' a_k & \text{if } w_k^{(\nu-1)} < L'' a_k \\ U' a_k & \text{if } w_k^{(\nu-1)} > U'' a_k \\ w_k^{(\nu-1)} & \text{otherwise.} \end{cases} \qquad \nu = 2, 3, \ldots$$

Terms in the above equations are defined as follows: $L' = \alpha L + (1 - \alpha)$, $U' = \alpha U + (1 - \alpha)$, $L'' = \eta L + (1 - \eta)$ and $U'' = \eta U + (1 - \eta)$ for $\alpha$ and $\eta$ arbitrarily chosen such that $0 < \alpha < \eta \leq 1$. As before, the parameters $\alpha$ and $\eta$ serve to speed up the convergence of the iterative algorithm used to provide a solution; Singh and Mohl (1996) suggest that $\alpha = .67$ and $\eta = .9$ work well in practice. Finally, $w_k^{(\nu-1)} = a_k g_k^{(\nu-1)}$; $\nu = 2, 3, \ldots$ where

$$g_k^{(\nu-1)} = \frac{a_k^{(\nu-2)*}}{a_k} \left[ 1 + (\boldsymbol{X} - \hat{\boldsymbol{X}}_w^{(\nu-2)})' \left( \sum_{j \in s} a_j^{(\nu-2)*} \boldsymbol{x}_j \boldsymbol{x}_j' \right)^{-1} \boldsymbol{x}_k \right]; \qquad \nu = 2, 3, \ldots$$

and where $\hat{\boldsymbol{X}}_w^{(\nu-2)}$ is as before. Starting values are given by $a_k^{(0)*} = a_k$ and $w_k^{(0)} = a_k$.

A property of the Modified Huang-Fuller and Shrinkage-Minimization distance functions is that the calibration constraints (equation (2.2)) are met at every iteration whereas the range restrictions on the $g$-factors are met only upon convergence. For the Restricted GLS and Restricted Raking Ratio distance functions, the range restrictions on the $g$-factors are met at every iteration whereas the calibration constraints are only met upon convergence. Now, it is often useful to specify an upper bound on the number of iterations to convergence; this feature may be programmed into the iterative algorithm for operational expediency. If this upper bound is exceeded due to slow convergence, the iterative algorithm may be terminated prematurely. Regardless, for the Modified Huang-Fuller and Shrinkage-Minimization distance functions, the calibration constraints will be met. Likewise, for the Restricted GLS and Restricted Raking Ratio distance functions, the range restrictions will be met.

Now, the behaviour of the $g$-factors from some of the distance functions has been studied extensively; see, for example, Deville, Särndal and Sautory (1993). Stukel and Boyer (1992) empirically show that the GLS and RR distance functions, as well as their restricted counterparts having loose bounds imposed on them, give $g$-factors whose distributions over a given data set adhere to normality rather closely. However, as the bounds on the restricted distance functions are squeezed together more closely, the distributions exhibit a "pile-up" of $g$-factors at the lower and upper bounds. Regardless, even under extreme squeezing, the restricted distance functions seem to give point estimates that are close to their unrestricted counterparts, as the results of our empirical study will verify. However, the biases of both the point and variance estimators under extreme squeezing on the restricted distance functions have not been investigated. This investigation is of interest to surveys such as the LFS, where an augmentation to the current estimation system has been implemented, which now allows users the option of choosing from amongst the Restricted GLS distance function and the Shrinkage-Minimization distance function, in addition to the previously available GLS distance function.

## 3. VARIANCE ESTIMATION FOR CALIBRATION ESTIMATORS

The exact variance of the calibration estimator $\hat{Y}_w$ is intractable since the point estimator itself is nonlinear. In addition, there is no explicit unbiased method of variance estimation. Therefore, approximately unbiased methods, such as the Taylor and the Jackknife, are often used in practice.

Now, for stratified multistage designs, "with replacement" sampling is not often used in practice since the possibility of drawing the same unit more than once is unappealing. Therefore, the preponderance of surveys use "without replacement" sampling, at least at the first stage of sampling. Even so, if the first stage sampling fraction is small (say, less than 10 percent as a rule of thumb), it may be reasonable to use a simplified variance formula that assumes "with replacement" sampling at the first stage of sampling. For the generalized regression estimator (GLS distance function) under a stratified multi-stage design this simplification of the variance estimator yields:

$$\hat{V}_T^*(\hat{Y}_{w(\mathrm{GREG})}) = \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \sum_{k \in s_{hi}} a_{hik} e_{hik} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{k \in s_{hi}} a_{hik} e_{hik} \right]^2 \quad (3.1)$$

where $s_{hi}$ is the sample of individuals in the $i$-th primary sampling unit (PSU) and the $h$-th stratum, $a_{hik}$ is the original sampling weight under the stratified multi-stage design for

sampled individual $k$ in PSU $i$ and stratum $h$, and $n_h$ is the number of sampled PSUs in stratum $h$. Also $e_{hik} = y_{hik} - \boldsymbol{x}'_{hik}\hat{\beta}$ is the estimated residual associated with the regression estimator where $\hat{\beta} = (\sum_{hik\in s} a_{hik}\boldsymbol{x}_{hik}\boldsymbol{x}'_{hik}/c_{hik})^{-1} \sum_{hik\in s} a_{hik}\boldsymbol{x}_{hik}y_{hik}/c_{hik}$. For many designs, the "with replacement" formula given by (3.1) overestimates the true variance (see Särndal, Swensson and Wretman 1992, section 4.6). Note that although, technically speaking, this simplified variance estimator is *not* the Taylor variance estimator, it is often referred to as such for historical reasons and so will it be in this paper.

An improvement to equation (3.1), which includes the $g$-factor in the variance formula (recall that $w_{hik} = a_{hik}g_{hik}$), is suggested by Hidiroglou, Fuller and Hickman (1980). It is given by:

$$\hat{V}_T(\hat{Y}_{w(\mathrm{GREG})}) = \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \sum_{k\in s_{hi}} w_{hik}e_{hik} - \frac{1}{n_h}\sum_{i=1}^{n_h}\sum_{k\in s_{hi}} w_{hik}e_{hik} \right]^2. \quad (3.2)$$

An analogue of equation (3.2) is also suggested by Särndal (1982) in the context of two-stage sampling, but for Yates-Grundy type variance estimators. Now Deville and Särndal (1992) show that any distance function which obeys a set of general conditions will produce an estimator that is asymptotically equivalent to the one produced by the GLS distance function, that is, $\hat{Y}_{w(\mathrm{GREG})}$ given by (2.5). Singh and Mohl (1996) extend this result to include the Modified Huang-Fuller and Shrinkage-Minimization distance functions. As a result, the asymptotic variance of the calibration estimator $\hat{Y}_w$ can be considered to be roughly equal to that of $\hat{Y}_{w(\mathrm{GREG})}$. This observation leads to a method for estimating the Taylor variance which is common to all calibration estimators, namely, to estimate the variance of $\hat{Y}_w$ using a modification of the Taylor variance estimator employed for $\hat{Y}_{w(\mathrm{GREG})}$, rather then rederiving the Taylor formula for each of the distance functions separately. Thus, whenever a variance estimator associated with a distance function different from the GLS is required, equation (3.2) is used, replacing the final weights $\{w_{hik}\}$ from the GLS distance function with those from the distance function in question.

It is straightforward to apply the Jackknife procedure to obtain a variance estimator for $\hat{Y}_w$, regardless of the distance function used to obtain the final calibrated weights. An expression for the variance formula under a stratified multi-stage design using with replacement sampling at the first stage is given by:

$$\hat{V}_J(\hat{Y}_w) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} (\hat{Y}_w(hi) - \hat{Y}_w)^2 \quad (3.3)$$

where $\hat{Y}_w(hi)$ is often referred to as the "replicate estimator"; "replicates" are formed by taking what remains of the sample after removing PSU $i$ from stratum $h$. Thus, $\hat{Y}_w(hi)$ is calculated by recomputing $\hat{Y}_w$ after removing the $i$-th PSU from the $h$-th stratum, $h = 1, ..., L$; $i = 1, ..., n_h$, *i.e.*, with the original sampling weights altered to reflect the PSU removal and the $g$-factors recalculated based on the reduced sample or replicate. Finally, the Jackknife estimator is constructed by repeatedly removing PSUs one at a time, calculating the corresponding replicate estimator, and then assembling the final estimator using (3.3). The Jackknife variance estimator given by (3.3) is the most conservative among the four variations suggested in the extensive discussion on the subject by Wolter (1985).

It is interesting to note that, for the GREG estimator, Yung and Rao (1996) obtain (3.2) as an approximation to the Jackknife variance estimator given by (3.3); they call (3.2) the "Jackknife Linearization Variance Estimator". Their simulation study shows that biases (both conditional and unconditional) of the Taylor variance estimator (equation (3.1)), the Jackknife Linearization variance estimator (equation (3.2)) and the Jackknife variance estimator (equation (3.3)) behave similarly. While their simulation focuses on variance estimators for the unrestricted GREG estimator, our simulation study, which we discuss next, focuses on variance estimators for the GREG as well as for estimators based on other restricted and unrestricted distance functions.

## 4. MONTE CARLO SIMULATION STUDY

### 4.1 Design of the Study

In order to compare the performance of the calibration estimators and their corresponding Taylor and Jackknife variance estimators, we undertook a Monte Carlo simulation study, in which we investigated their finite sample design-based frequentist properties.

December 1990 Labour Force Survey (LFS) sample data for the province of Newfoundland was used to simulate a finite population, from which repeated samples were drawn. The LFS is the largest ongoing household sample survey conducted by Statistics Canada. Monthly data relating to the labour market is collected using a complex multi-stage sampling design with several levels of stratification. The details of the design of the survey prior to the 1991 redesign can be found in Singh, Drew, Gambino and Mayda (1990). In general, provinces are stratified into "economic regions", which are large areas of similar economic structure; Newfoundland has four such economic regions. The economic regions are further substratified into "self-representing units" (SRUs) and "non self-representing units" (NSRUs), which are, in turn, further substratified into lower level substrata. SRUs are cities whose population exceeds 15,000, such as St. John's and Cornerbrook, in the case of Newfoundland. Now, the lowest level of stratification in Newfoundland yielded 45 strata, each of which contained less than 6 primary sampling units (PSUs), which was an insufficient number from which to sample, for the purposes of the simulation. Thus, the 45 strata were collapsed down to 18, each containing between 6 and 18 PSUs. In collapsing the strata,

economic regions were kept intact, as were the Census Metropolitan Areas (CMAs) of St. John's and Cornerbrook.

For the Monte Carlo study, $R = 4{,}000$ samples, each of size approximately 1,000, were drawn from the Newfoundland "population" (which was of size 9,152), according to a two-stage design. For collapsed strata belonging to NSRUs, two PSUs were selected at the first stage using Probability Proportional to Size (PPS) with replacement (WR) sampling, where the size measure used was the number of dwellings in the PSU. At the second stage, one in five dwellings were selected from the sampled PSUs using Simple Random Sampling (SRS) without replacement (WOR). For collapsed strata belonging to SRUs, three PSUs were selected at the first stage using PPS WR sampling. At the second stage, all the dwellings in the sampled PSUs were selected, reducing this part of the design to one-stage take-all cluster sampling. This feature was necessary since there were not enough dwellings per PSU to subsample in SRUs. The selection of two PSUs in NSRU strata versus three in SRU strata was driven by the fact that, in general, NSRU strata had fewer population PSUs from which to sample than did SRU strata. In all, there were 47 sampled PSUs. In either case (NSRUs or SRUs), all dwelling members were included in the sample. Although this design is a hybrid between a one and two-stage design, we shall refer to it as a two-stage design, for convenience.

We took $Y$, the total number of unemployed, to be the parameter of interest. This was calculated from the finite population by: $Y = \sum_{k \in U} y_k = \sum_{k=1}^{9152} y_k$ where $y_k = 1$ if individual $k$ was unemployed; 0 otherwise. For each of the $R = 4{,}000$ samples, we calculated $\hat{Y}_w$, the estimated total number of unemployed as $\hat{Y}_w = \sum_{k \in s} w_k y_k$. The $\{w_k : k \epsilon s\}$ were determined by the following six distance functions discussed earlier:

(1) the Generalized Least Squares (GLS) Distance Function (equation (2.4)),
(2) the Raking Ratio (RR) Distance Function (equation (2.7)),
(3) the Restricted GLS (RGLS) Distance Function (equation (2.8)),
(4) the Restricted RR (RRR) or Logit Distance Function (equation (2.9)),
(5) the Modified Huang-Fuller (MHF) Distance Function ($\alpha = .67$, $\delta = .8$) (equation (2.10)), and
(6) the Shrinkage-Minimization (SM) Distance Function ($\alpha = .67$, $\eta = .9$) (equation (2.11)).

For the latter four distance functions, the following four sets of bounds were imposed on each to restrict the minimization: (i) $L = 0$, $U = 4$, (ii) $L = .4$, $U = 2$, (iii) $L = .68$, $U = 1.6$ and (iv) $L = .8$, $U = 1.3$. This yielded a total of eighteen point estimators. For each of the eighteen point estimators, the calibration used auxiliary information based on Census projections at the province level for 10 mutually exclusive and exhaustive age/sex categories (age categories: < = 14, 15-24, 25-44, 45-64, > = 65 crossed with the two sexes) and the four economic regions of Newfoundland. Thus, the auxiliary information for each individual was a vector of length fourteen having exactly two ones and twelve zeros. However, for computational purposes, the dimensionality of the vector had to be reduced to thirteen when using the Newton-Raphson procedure to solve equation (2.3). For the first four distance functions, we set $c_k = 1$.

For each of the $R = 4{,}000$ samples and each of the eighteen point estimators, we calculated the Jackknife variance estimator given by equation (3.3). We also calculated the Taylor variance estimator given by equation (3.2), and the modification suggested in section 3 was used for distance functions other than the GLS. Note that since PPSWR, rather than PPSWOR, was used at the first stage of sampling, the use of the variance estimator given by equation (3.2) was entirely appropriate for our simulation. Finally, for the GLS distance function only, the formula (3.1) was calculated to observe the impact of omitting $g$-factors from the variance estimator.

For each of the six distance functions given above, a number of frequentist properties were investigated. These are given below.

(A) The Percent Relative Bias of the Estimated Number of Unemployed (with respect to the population value) is estimated by:

$$\frac{E_M(\hat{Y}_w) - Y}{Y} * 100 \qquad (4.1)$$

where

$$E_M(\hat{Y}_w) = \frac{1}{R} \sum_{r=1}^{R} \hat{Y}_{w_r}$$

is the Monte Carlo expectation of the point estimator $\hat{Y}_w$ taken over the $R$ samples, and $\hat{Y}_{w_r}$ is the value of $\hat{Y}_w$ for sample $r$.

(B) The Percent Relative Bias of the Taylor/Jackknife Variance Estimator (with respect to the true variance) is estimated by:

$$\frac{(E_M(\hat{V}(\hat{Y}_w)) - V_{\text{true}})}{V_{\text{true}}} * 100 \qquad (4.2)$$

where

$$E_M(\hat{V}(\hat{Y}_w)) = \frac{1}{R} \sum_{r=1}^{R} \hat{V}_r(\hat{Y}_w)$$

and

$$V_{\text{true}} = \frac{1}{R} \sum_{r=1}^{R} (\hat{Y}_{w_r} - E_M(\hat{Y}_w))^2$$

and $\hat{V}_r(\hat{Y}_w)$ is the value of $\hat{V}(\hat{Y}_w)$ (Taylor or Jackknife) for sample $r$.

(C) The Percent Coefficient of Variation of the Taylor/Jackknife Variance Estimator (with respect to the true variance) is estimated by:

$$\frac{\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{V}_r(\hat{Y}_w) - V_{\text{true}})^2}}{V_{\text{true}}} * 100 \qquad (4.3)$$

*i.e.*, the root mean squared error of the variance estimator divided by the true variance, expressed as a percentage. Although most studies focus on the *bias* of the variance estimators, it is also of secondary interest to look at the *coefficient of variation* of the variance estimators to see how variable the variance estimates themselves are.

Note that in equations (4.2) and (4.3), it may have been more appropriate to make comparisons relative to a "true mean squared error" rather than a "true variance". However, for our simulation, the relative biases were so small that the differences between the two types of comparisons are virtually negligible.

Finally, in order to assess the appropriateness of the choice of number of repeated samples, we calculated Monte Carlo errors, using as a measure the Percent Coefficient of Variation of $E_M(\hat{V}(\hat{Y}_w))$, given by:

$$\frac{\sqrt{\frac{1}{R^2}\sum_{r=1}^{R}[\hat{V}_r(\hat{Y}_w) - E_M(\hat{V}(\hat{Y}_w))]^2}}{E_M(\hat{V}(\hat{Y}_w))} * 100. \qquad (4.4)$$

The Monte Carlo errors were found to be consistently low (between .99% and 3.60%) for both the Jackknife and Taylor using $R = 4{,}000$, indicating stable results.

### 4.2 Results of the Study

Table 1 gives the Percent Relative Bias of the Point Estimators (equation (4.1)) as well as the Percent Relative Bias of the Taylor and Jackknife Variance Estimators (equation (4.2)) and the Percent CVs of the Taylor and Jackknife Variance Estimators (equation (4.3)). The percent relative bias for all the point estimates (column two) is negligible, ranging in value from 0.10% to 0.52%, but much less than 1% in all cases. The fact that all point estimates have a similar bias seems reasonable, given the asymptotic equivalence of all calibration estimators to the regression estimator.

The third column gives the percent relative bias of the Taylor variance estimator. Here, the true variance is always underestimated, but never by more than 6.2%. In the case of the regression estimator, it appears to make little difference whether or not the $g$-factor is included in the variance formula (equation (3.1) versus (3.2)); the bias improves only slightly for the case of the $g$-factor included (-5.82% versus -6.01%). The Jackknife variance estimator (column four), on the other hand, outperforms the Taylor variance estimator uniformly. The Jackknife almost always underestimates the true variance, but by less than 2% in all cases.

To produce a solution, all distance functions but the GLS required an iterative algorithm. This being the case, some of the 4,000 samples experienced convergence problems, particularly in the case of extreme bounding on the $g$-factors. Those samples for which the algorithm did not converge were discarded. Thus, they did not contribute to the various Monte Carlo measures. The number of such discarded samples is

**Table 1**
Percent Relative Bias of the Point Estimators, and Percent Relative Bias and Percent CV of the Taylor and Jackknife Variance Estimators (Sample Size About 1000)

| Distance Function | | Percent Relative Bias Point Estimator | Percent Relative Bias Taylor Variance | Percent Relative Bias Jackknife Variance | Percent CV Taylor Variance | Percent CV Jackknife Variance | Number of Discarded Samples (From 4000) |
|---|---|---|---|---|---|---|---|
| GLS (Regression) | | .11 | -6.01 (eq 3.1)<br>-5.82 (eq 3.2) | -1.73 | 60.79 (eq 3.1)<br>59.60 (eq 3.2) | 62.86 | 0 |
| Restricted GLS | ($L = 0$, $U = 4$) | .11 | -5.82 | -1.73 | 59.60 | 62.86 | 0 |
| | ($L = .4$, $U = 2$) | .10 | -5.36 | -1.27 | 59.93 | 63.21 | 32 |
| Raking Ratio | | .52 | -6.20 | 0.84 | 59.45 | 63.35 | 0 |
| Restricted RR | ($L = 0$, $U = 4$) | .50 | -6.09 | -0.31 | 59.48 | 63.47 | 0 |
| | ($L = .4$, $U = 2$) | .46 | -5.69 | -0.39 | 59.81 | 64.21 | 32 |
| Modified Huang-Fuller | ($L = 0$, $U = 4$) | .11 | -5.82 | -1.73 | 59.60 | 62.86 | 0 |
| | ($L = .4$, $U = 2$) | .10 | -5.36 | -1.20 | 59.94 | 63.27 | 32 |
| Shrinkage-Minimization | ($L = 0$, $U = 4$) | .11 | -5.82 | -1.73 | 59.60 | 62.86 | 0 |
| | ($L = .4$, $U = 2$) | .10 | -5.36 | -1.27 | 59.94 | 63.25 | 32 |

indicated in the last column of Table 1. In the case of extreme bounds ($L$ = .68, $U$ = 1.6 and $L$ = .8, $U$ = 1.3), so many samples were discarded (between 231 and 234 for the cases $L$ = .68, $U$ = 1.6 and between 1,562 and 1,602 for the cases $L$ = .8, $U$ = 1.3) that the results were not considered reliable, and so are not reported here. However, these tighter bounds were of interest, so the simulation was rerun using approximately double the sample size (increase from roughly 1,000 to 2,000). Note that Deville and Särndal (1992) show that convergence is achieved for all distance functions with probability one as the sample size increases.

Columns five and six of Table 1 give the Percent CVs of the Taylor and Jackknife Variance Estimators. The coefficients of variation are similar for all distance functions, ranging in value from 59.45% to 64.21%. However, the CVs corresponding to the Jackknife are always slightly larger than that of Taylor. Coefficients of variation of this magnitude, although large, have been encountered in other simulation studies relating to variances. See, for example, Kovačević, Yung and Pandher (1995). However, we were interested in seeing if the key results relating to the bias of the variance estimators would still hold if the CVs were lowered. Therefore, at the suggestion of a referee, we reran the simulation, increasing the number of PSUs drawn from 47 to 83, since CVs of variance estimators are known to be approximately inversely related to the number of PSUs drawn. The PSUs were increased in such a way that the overall design was made self-weighting; this approach appeared to have the greatest effect on lowering the CVs. The second stage of sampling remained the same as before. Rerunning the simulation had the secondary benefit of roughly doubling the sample size, and thus, solving the convergence problems referred to in the last paragraph.

The results from the second run of the simulation are reported in Table 2. The last column in Table 2 shows the reduced number of discarded samples due to convergence problems. The fifth and sixth column of this table show that the CVs are significantly reduced to between 22.70% and 24.2% with the Jackknife consistently exhibiting slightly higher values. Now, as before, the percent relative bias in the point estimator is negligible, always being well under 1%. In the previous run, the percent relative biases for the Taylor estimator were always roughly -6%; here, they are always about -3%, again implying underestimation of the true variance. Once more, in the case of the GLS distance function, there is very little difference in the bias that results from using equation (3.1) versus (3.2). The percent relative bias in the Jackknife estimator (always roughly -1.5%) is consistently

**Table 2**
Percent Relative Bias of the Point Estimators, and Percent Relative Bias and Percent CV of the Taylor and Jackknife Variance Estimators (Sample Size About 2000)

| Distance Function | | Percent Relative Bias Point Estimator | Percent Relative Bias Taylor Variance | Percent Relative Bias Jackknife Variance | Percent CV Taylor Variance | Percent CV Jackknife Variance | Number of Discarded Samples (From 4000) |
|---|---|---|---|---|---|---|---|
| GLS (Regression) | | .02 | -2.71 (eq 3.1)<br>-2.61 (eq 3.2) | -1.43 | 23.03 (eq 3.1)<br>22.84 (eq 3.2) | 23.29 | 0 |
| Restricted GLS | ($L$ = 0, $U$ = 4) | .02 | -2.61 | -1.43 | 22.84 | 23.29 | 0 |
| | ($L$ = .4, $U$ = 2) | .02 | -2.61 | -1.43 | 22.84 | 23.29 | 0 |
| | ($L$ = .68, $U$ = 1.6) | .02 | -2.61 | -1.44 | 22.84 | 23.29 | 0 |
| | ($L$ = .8, $U$ = 1.3) | .02 | -2.75 | -1.56 | 22.70 | 23.15 | 118 |
| Raking Ratio | | .25 | -2.75 | -1.15 | 22.84 | 23.43 | 0 |
| Restricted RR | ($L$ = 0, $U$ = 4) | .17 | -2.67 | -1.36 | 22.84 | 23.30 | 0 |
| | ($L$ = .4, $U$ = 2) | .16 | -2.70 | -1.42 | 22.84 | 23.29 | 0 |
| | ($L$ = .68, $U$ = 1.6) | .31 | -2.77 | -0.49 | 22.83 | 24.20 | 0 |
| | ($L$ = .8, $U$ = 1.3) | .27 | -2.91 | * | 22.70 | * | 118 |
| Modified Huang-Fuller | ($L$ = 0, $U$ = 4) | .02 | -2.61 | -1.43 | 22.84 | 23.29 | 0 |
| | ($L$ = .4, $U$ = 2) | .02 | -2.61 | -1.43 | 22.84 | 23.29 | 0 |
| | ($L$ = .68, $U$ = 1.6) | .02 | -2.61 | -1.44 | 22.84 | 23.29 | 0 |
| | ($L$ = .8, U = 1.3) | .02 | -2.58 | -1.36 | 22.73 | 23.18 | 116 |
| Shrinkage-Minimization | ($L$ = 0, $U$ = 4) | .02 | -2.61 | -1.43 | 22.84 | 23.29 | 0 |
| | ($L$ = .4, $U$ = 2) | .02 | -2.61 | -1.43 | 22.84 | 23.29 | 0 |
| | ($L$ = .68, $U$ = 1.6) | .02 | -2.61 | -1.44 | 22.84 | 23.29 | 0 |
| | ($L$ = .8, $U$ = 1.3) | .02 | -2.61 | -1.24 | 22.73 | 23.63 | 118 |

smaller in absolute value than that of Taylor. For the Jackknife estimator, there is one case (Restricted RR ($L$ = .8, $U$ = 1.3)) where there were convergence problems; those results are omitted, indicated by a "*". Surprisingly, for both the Taylor and Jackknife, there is virtually no change in bias for the restricted distance functions as the bounds are made successively more tight. In fact, there seems to be very little difference in the percent relative bias across all of the distance functions, for both the Taylor and the Jackknife. Note that for the rerun of the simulation, the Monte Carlo errors ranged between .37% and 2.13%.

## 5. CONCLUSIONS

This paper focused on exploring the behaviour of point estimators and their corresponding Taylor and Jackknife variance estimators for a number of different distance functions available through calibration theory. Particular emphasis was given to those distance functions which allowed range restrictions to be imposed on the $g$-factors, eliminating the possibility of negative and high positive final weights. All of the point estimators which were investigated exhibited a negligible bias.

Both the Jackknife and Taylor variance estimators exhibited small underestimation of the true variance, although the Jackknife consistently had smaller biases (in absolute value) than the Taylor. The most striking result was that, for both Taylor and Jackknife, the biases remained roughly the same in the cases of extreme bounding on the $g$-factors as in the cases of less restrictive bounding. In general, however, caution should be exercised in the use of extreme bounds, due to the convergence problems that may be experienced, particularly when Jackknifing is used for variance estimation and the point estimators must be recalculated repeatedly. If the main objective of using the restricted distance functions is to eliminate the possibility of negative or high positive weights, then modest bounds on the $g$-factors should suffice.

As a final remark, it is interesting to note that roughly 97% of the computing time was spent Jackknifing while the remaining 3% was spent on Taylor linearization. This rather extreme difference in computation time may give the Taylor method an advantageous edge if measures of precision are required for a large number of domains. However, given recent developments in the computational efficiency of the Jackknife variance estimator (for example, the program WESVARPC (1995)), it may be possible to offset this imbalance. Even so, it should be noted that, at this time, WESVARPC has improved the computational efficiency for designs having only two PSUs per stratum, and poststratified estimators having only one dimension.

In conclusion, since our study does not conclusively show either variance estimator to be clearly superior and shows both to behave reasonably well for all distance functions, it is up to the user to decide which variance/ distance function combination best fits the system requirements.

## ACKNOWLEDGEMENTS

The authors would like to thank Chris Mohl for providing us with some of the computer code which we used in the simulation study. We would also like to thank an Associate Editor and two referees for useful comments on the earlier version of this paper.

## REFERENCES

DEMING, W.E., and STEPHAN, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.

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

DEVILLE, J.-C., SÄRNDAL, C.-E., and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

HIDIROGLOU, M.A., FULLER, W.A., and HICKMAN, R.D. (1980). SUPERCARP, Department of Statistics, Iowa State University, Ames, Iowa.

HUANG, E.T., and FULLER, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Social Statistics Section, American Statistical Association*, 300-305.

KOVAČEVIĆ, M.S., YUNG, W., and PANDHER, G.S. (1995). Estimating the sampling variances of measures of income inequality and polarization – An empirical study. Methodology Branch Working Paper, HSMD-95-007E, Statistics Canada.

SÄRNDAL, C.-E. (1982). Implications of survey design for generalized regression estimation of linear functions. *Journal of Statistical Planning and Inference*, 7, 155-170.

SÄRNDAL, C.-E., SWENSSON, B., and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SINGH, A.C., and MOHL, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, 107-115.

SINGH, M.P., DREW, J.D., GAMBINO, J.G., and MAYDA, F. (1990). *Methodology of the Canadian Labour Force Survey: 1984-1990*. Catalogue No. 71-526, Statistics Canada.

STUKEL, D.M., and BOYER, R. (1992). Calibration estimation: An application to the Canadian Labour Force Survey. Methodology Branch Working Paper, SSMD-92-009E, Statistics Canada.

WESVARPC (1995). Westat Inc., Rockville, Maryland.

WOLTER, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

YUNG, W., and RAO, J.N.K. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*, 22, 23-31.