The effect of measurement errors on the calibrated weights

Daniela Marella

Dipartimento di Statistica, Probabilità e Statistiche Applicate

Università di Roma "La Sapienza"

Abstract. The aim of the paper is to evaluate the effect on both the calibrated weights and the corresponding calibration estimator of measurement errors affecting the sample auxiliary information.

Keywords. Calibrated weights, Generalized regression estimator, Measurement errors.

1 Introduction

The first assessment of calibration estimation was carried out by Deville et al. (1992). Assume that one or more auxiliary variables are available. The auxiliary information can be used at the estimation stage to construct estimators that are approximately unbiased with a variance smaller than that of the Horvitz-Thompson estimator.

The basic idea is to define a new weights system which are as close as possible, according to a given distance measure, to the original sampling design weights while respecting a set of constraints (calibration equations). These state that the new weights, called calibrated weights, must provide perfect estimates when applied to each auxiliary variables. Clearly, different distance measures lead to a different sets of weights, all calibrated to the same information, and thereby to new estimators.

An important result by Deville et al. (1992) is that, within the classical framework and under some regularity conditions, the calibration estimators with different choice of objective function are asymptotically equivalent (*i.e.* with the same asymptotic variance) to the generalized regression estimator, obtained under a chi-square distance measure.

This paper explores the effect both on the calibrated weights and the resulting calibration estimator of measurement errors affecting the sample auxiliary information. In the sequel we assume that the auxiliary information is one-dimensional and global, $i.e$ only the total is known. Besides, we restrict our discussion to the generalized regression estimator (GREG).

The paper is organized as follows. In Section 2 we introduce the general framework for the calibration estimation with regard to a single-stage cluster sampling. In Section 3 the effect of measurement errors on the calibrated weights is evaluated by introducing a model describing their generating mechanism. In Section 4 we compute the bias and the variance of the calibration estimator whose weights are affected by measurement errors. In Section 5 a simulation study is performed. Finally, in Section 6 the conclusions of the paper are given.

2 General Framework

Consider a finite population of N elements $U = \{1, .., N\}$ grouped into c disjoint subpopulations, called clusters and denoted by $\mathbf{U}_c = \{U_1, .., U_c\}$. The number of population elements in the ith cluster is denoted by N_i. In single-stage cluster sampling a probability sample of clusters s_c of size n_c is drawn according to the design $p(\cdot)$ and every population element in the clusters is observed. The first and second order cluster inclusion probabilities induced by the design $p(\cdot)$ are π_i and π_{ij} respectively. Let $d_i = 1/\pi_i$ denote the original sampling weights, fixed by design.

As known, single stage-cluster sampling is used in many large scale surveys where direct element sampling is not used because the corresponding frame could not exist as well as to reduce the cost of the field work related to travel expenses (for instance if the population units are scattered over a wide area). Consider as an example a survey conducted in a large city with city blocks as primary sampling units and buildings as elements. The study variable Y measures some aspect of the kth building, such as habitable floor space.

The basic goal is to estimate the (unknown) population total t_y. Let X^* be an auxiliary variable with total t_{x^*}, which is assumed to be accurately known from an outside source. It can be obtained from census data or administrative data files. For instance, for each city block we could have as auxiliary information the number or/and the type of buildings, the number of inhabitants and so on. Formally, we assume the following auxiliary information to be available:

1. The population total t_{x^*} is known;

2. The cluster totals $(t_{y_i}, t_{x_i^*})$ of Y and X^* respectively are known for every cluster $i \in s_c$.

In order to estimate the unknown total t_y, one possibility is the simple unbiased Horvitz-Thompson estimator

$$\hat{t}_{y,ht} = \sum_{i \in s_c} d_i t_{y_i}$$

where d_i is the inverse of cluster inclusion probability. An estimator of t_y that uses explicitly the information t_{x^*}, through a more efficient weighting of the observed totals t_{y_i}, is the calibration estimator. This term was coined by Deville et al. (1992) to describe an estimator of the form

$$\hat{t}_y^* = \sum_{i \in s_c} w_i^* t_{y_i} \tag{1}$$

where the weights w_i^* satisfy the calibration equation

$$\sum_{i \in s_c} w_i^* t_{x_i^*} = t_{x^*} \tag{2}$$

The constraint (2) states that if we apply the weighting system w_i^* to the totals $t_{x_i^*}$ (for $i \in s_c$) and sum over s_c, the estimator (1) agrees exactly with the total known value t_{x^*}. The set of weights, which is chosen to minimize the average distance from the basic design weights, is called calibrated to the information t_{x^*}. Calibration is a highly desirable property for survey weights, since the control totals are disseminated as benchmark values then reproducing them from the sample is reassuring to the user.

As previously stressed, the calibration technique provides an alternative derivation of the generalized regression estimator (GREG), obtained under a chi-square distance measure. Formally, the GREG estimator can be expressed as (1), where the weight w_i^* associated to the ith cluster total is given by

$$w_i^* = d_i + \frac{d_i t_{x_i^*}(t_{x^*} - \sum_{i \in s_c} d_i t_{x_i^*})}{\sum_{i \in s_c} d_i t_{x_i^*}^2} = d_i + R_{i,s_c} \tag{3}$$

Clearly, w_i^* depend on both ith cluster and the whole sample s_c containing i. The original sampling weights d_i are suitably modified to reflect the information in the auxiliary variable. As a consequence the generalized regression estimator can be written as

$$\hat{t}_{y,greg}^* = \hat{t}_{y,ht} + \hat{B}^*(t_{x^*} - \hat{t}_{x^*,ht}) \tag{4}$$

where $\hat{B}^*$ is the estimated regression coefficient obtained regressing the cluster total t_{y_i} on $t_{x_i^*}$ for the clusters belonging to the sample s_c. The estimator (4) is explained and illustrated in several textbooks, for instance in Chapters 6 and 7 of Särndal et al. (1992). It is important to stress that t_{x^*} should be an accurate value of the population total. If an erroneous external total is used, then the GREG estimator could be severely biased, with a bias depending on the distance between the erroneous and the correct total of the auxiliary information. In the sequel we assume that the total t_{x^*} is considered reliable. The aim of the paper is to evaluate the effect on both the calibrated weights and the corresponding calibration estimator of measurement errors affecting the sample auxiliary information.

3 The effect of measurement errors on the calibrated weights

3.1 Measurement Errors Model

By the term measurement error we denote the difference between the recorded value of the variable X^* for the kth element and its "true value". This kind of error occurs during the data collection stage for a variety of reasons. Sources of measurement errors can be categorized into four principal factors: the questionnaire, the respondents, the interviewer and the data collection mode which determines the types of measurement errors (face to face interviewing, telephone interviewing, self-administered methods) . For instance, the respondent may intentionally or unintentionally give incorrect answers, in a personal interview the interviewer may record the responses incorrectly and may influence the respondent answers, some items in the questionnaire may be poorly formulated and so on. An overview on the measurement errors sources is given in Biemer et al. (1991). This kind of nonsampling error cannot be completely avoided in practice, but it is possible to keep it under control through an accurate survey planning: the use of qualified interviewers or the interviewers training, the supervision phase, an accurate questionnaire wording. In the sequel we assume that the population total t_{x^*} comes from an external source considered reliable, but that the observed values of X^* for the sample elements are subject to measurement errors. More specifically, with regard to the measurement procedure we assume that the data are collected through a personal interview. As previously stressed, the interviewer represents a possible source of measurement errors introducing bias, variance and correlation into the responses.

In order to evaluate the effect of measurement errors on the calibrated weights, we introduce a fairly general statistical model describing the mechanism that generates the observed values. Basic contributions to the methodology of measurement errors models in survey sampling were given by Mahalanobis (1946), Hansen et al. (1951, 1961, 1964). In the sequel, the finite population observed values are considered as a realization from an infinite population and we describe our uncertainty about what particular values will appear through a probabilistic model (superpopulation approach). Clearly, the model should be adapted to the specific conditions of the particular survey at hand, and should formalize our prior knowledge about the population. In fact, if a realistic superpopulation model can give powerful inferences, on the other side invalid inferences would result if the assumed model were invalid.

As in Särndal et al. (1992) (pag. 618) we assume to have c interviewers, each is linked to a cluster U_i in a deterministic, nonrandom manner. The preassigned interviewer carries out all interviews from his own cluster. For instance, the population can be geographically divided into districts with one interviewer permanently stationed in each district who carries out all interviewers there. Formally, the measurement model, denoted by m, specifies that the observed value for the population unit k in cluster i is composed of a constant true value x^*_{ki} plus an unobserved error term ϵ_{ki}

$$x_{ki} = x^*_{ki} + \epsilon_{ki} \tag{5}$$

with the following stochastic structure

$$\begin{cases} E_m(\epsilon_{ki}) = b_i + \delta \quad \forall k \in U_i \\ var_m(\epsilon_{ki}) = \nu_i \quad \forall k \in U_i \\ cov_m(\epsilon_{ki}, \epsilon_{lj}) = 0 \quad \forall k \in U_i, \quad \forall l \in U_j, \quad \forall i \neq j \\ cov_m(\epsilon_{ki}, \epsilon_{li}) = \rho_i \nu_i \quad \forall k, l \in U_i, \quad \forall k \neq l \end{cases} \tag{6}$$

where $b_i, \nu_i, \rho_i, \delta$ are unknown model parameters, and $|\rho_i| < 1$. The observed population values of X^* are considered to be the realized values of random variables, whose distribution is described by (6). As a matter of fact, a superpopulation model is invoked but the inference still concerns the finite population parameter t_y.

Model (6) implies that the measurements made by the same interviewer are correlated and affected by the same constant effect b_i (interviewer effect). An extensive literature exists on the influence that the interviewer demographic and socioeconomic characteristics (such as sex, race and age) can have on the responses, see, for instance, Groves (1989). However, the correlation between responses is not uniquely due to interviewers, since there are other possible sources for correlated errors. Factors such as coders and supervisors may introduce correlation between responses for the units they are associated with. Note that b_i represents a interviewer fixed-effect, and that the parameter δ represents a systematic error that consistently affects the measurement process, no matter to which cluster (interviewer) the element belongs.

Under a single-stage cluster sampling with no missing data, the measurement model (5) can be expressed as

$$t_{x_i} = t_{x^*_i} + \sum_{k=1}^{N_i} \epsilon_{ki} = t_{x^*_i} + \eta_i$$

under the stochastic structure

$$\begin{cases} E_m(\eta_i) = N_i(b_i + \delta) = \mu_i \quad \forall i \\ var_m(\eta_i) = N_i \nu_i + N_i(N_i - 1)\rho_i \nu_i \quad \forall i \\ cov_m(\eta_i, \eta_j) = -N_i N_j (b_i + \delta)(b_j + \delta) \quad \forall i \neq j \end{cases}$$

where $t_{x_i}, t_{x_i^*}$ represent the observed and the true total for the ith cluster respectively. When the sample observations are affected by measurement errors, the calibrated weight assigned to the total t_{yi} becomes

$$w_i = d_i + \frac{d_i t_{x_i}(t_{x^*} - \sum_{j\in s_c} d_j t_{x_j})}{\sum_{j\in s_c} d_j t_{x_j}^2} = d_i + \widehat{R}_{i,s_c} \qquad (7)$$

which is different from the "true" calibrated weight (3). Given the sample s_c, w_i is a function of t_{x_j}-values which are random variables under the measurement model m. As a consequence, the model based evaluation of the observed calibrated weight w_i consists in computing the expectation $E_m(w_i|s_c)$. In order to accomplish this, let $\widehat{V}$ be the ratio

$$\widehat{V} = \frac{\sum_{j\in s_c} d_j t_{x_j}^2}{\sum_{j\in s_c} d_j t_{x_j^*}^2}$$

and let $cov_m(\widehat{R}_{i,s_c}, \widehat{V}|s_c)$ denote the covariance between $\widehat{R}_{i,s_c}$ and $\widehat{V}$, given by

$$\begin{aligned} cov_m(\widehat{R}_{i,s_c}, \widehat{V}|s_c) &= E_m(\widehat{R}_{i,s_c}\widehat{V}|s_c) - E_m(\widehat{R}_{i,s_c}|s_c)E_m(\widehat{V}|s_c) \\ &= \frac{1}{\gamma}\left[\frac{E_m[d_i t_{x_i}(t_{x^*} - \sum_{j\in s_c} d_j t_{x_j})]}{E_m\left[\sum_{j\in s_c} d_j t_{x_j}^2\right]} - E_m(\widehat{R}_{i,s_c}|s_c)\right] \end{aligned}$$

where

$$\gamma = \frac{\sum_{j\in s_c} d_j t_{x_j^*}^2}{E_m[\sum_{j\in s_c} d_j t_{x_j}^2]}$$

Hence

$$\begin{aligned} E_m(\widehat{R}_{i,s_c}|s_c) &= \frac{E_m[d_i t_{x_i}(t_{x^*} - \sum_{j\in s_c} d_j t_{x_j})]}{E_m\left[\sum_{j\in s_c} d_j t_{x_j}^2\right]} - \gamma cov_m(\widehat{R}_{i,s_c}, \widehat{V}|s_c) \\ &= \gamma[R_{i,s_c} + C - cov_m(\widehat{R}_{i,s_c}, \widehat{V}|s_c)] \end{aligned}$$

where

$$C = \frac{d_i t_{x^*}\mu_i - \sum_{j\in s_c} d_i d_j[cov_m(t_{x_i}, t_{x_j}) + \mu_j t_{x_i^*} + \mu_i(t_{x_j^*} + \mu_j)]}{\sum_{j\in s_c} d_j t_{x_j^*}^2}$$

Then the bias of w_i is given by

$$\begin{aligned} E_m(w_i|s_c) - w_i^* &= E_m(\widehat{R}_{i,s_c}|s_c) - R_{i,s_c} \\ &= (\gamma - 1)R_{i,s_c} + \gamma[C - cov_m(\widehat{R}_{i,s_c}, \widehat{V}|s_c)] \qquad (8) \end{aligned}$$

For instance, if $\mu_j = 0$ for each sample cluster, that is if the expected measurements values reproduce the true values, then

$$E_m(w_i|s_c) - w_i^* = (\gamma - 1)R_{i,s_c} - \gamma cov_m(\widehat{R}_{i,s_c}, \widehat{V}|s_c)$$

where $\gamma < 1$ and $C = 0$. Expression (8) does not give us any useful information about the sign and the magnitude of the bias since depends on the unknown quantity $cov_m(\widehat{R}_{i,s_c}, \widehat{V}|s_c)$. However, since w_i is a nonlinear function of the n_c sample clusters totals $(t_{x_1}, ..., t_{x_{n_c}})$, by the first-order Taylor approximation evaluated at the point $P_0 = [t_{x_1^*}, .., t_{x_{n_c}^*}]$ we obtain

$$w_i \simeq d_i + \frac{d_i t_{x_i^*}[t_{x^*} - \sum_{i \in s_c} d_i t_{x_i^*}]}{\sum_{i \in s_c} d_i t^2_{x_i^*}} + \sum_{j \in s_c} \frac{\partial w_i}{\partial t_{x_j}}\bigg|_{P_0} (t_{x_j} - t_{x_j^*})$$

so that the weight w_i can be expressed by a main term, which is linear in $(t_{x_1}, ..., t_{x_{n_c}})$ and a remainder term that is assumed negligible if compared to the main term. Taking the expectation with respect to the measurement model m, we have

$$\begin{aligned} E_m(w_i|s_c) &\simeq w_i^* + \sum_{j \in s_c} \frac{\partial w_i}{\partial t_{x_j}}\bigg|_{P_0} E_m(t_{x_j} - t_{x_j^*}) \\ &= w_i^* + \sum_{j \in s_c} \frac{\partial w_i}{\partial t_{x_j}}\bigg|_{P_0} N_j(b_j + \delta) \\ &= w_i^* + \sum_{j \in s_c} \frac{\partial w_i}{\partial t_{x_j}}\bigg|_{P_0} \mu_j \end{aligned}$$

where

$$\begin{cases} \frac{\partial w_i}{\partial t_{x_i}}\bigg|_{P_0} = \frac{d_i\left[(t_{x^*} - \sum_{j \in s_c} d_j t_{x_j})(\sum_{j \in s_c} d_j t^2_{x_j} - 2 d_i t^2_{x_i}) - d_i t_{x_i} \sum_{j \in s_c} d_j t^2_{x_j}\right]}{(\sum_{j \in s_c} d_j t^2_{x_j})^2}\bigg|_{P_0} \\ \frac{\partial w_i}{\partial t_{x_j}}\bigg|_{P_0} = -\frac{d_i d_j t_{x_i}[\sum_{j \in s_c} d_j t^2_{x_j} + 2 t_{x_j}(t_{x^*} - \sum_{j \in s_c} d_j t_{x_j})]}{(\sum_{j \in s_c} d_j t^2_{x_j})^2}\bigg|_{P_0} \quad j \neq i \end{cases} \tag{9}$$

Given the sample s_c, the bias in the calibrated weight w_i is given by

$$E_m(w_i|s_c) - w_i^* \simeq \sum_{j \in s_c} \frac{\partial w_i}{\partial t_{x_j}}\bigg|_{P_0} \mu_j = \sum_{j \in s_c} \frac{\partial w_i}{\partial t_{x_j}}\bigg|_{P_0} N_j(b_j + \delta) \tag{10}$$

If $E_m(t_{x_j}) = t_{x_j^*}$ for each $j \in s_c$ then w_i is approximately unbiased for w_i^*. In the next section we proceed to analyze the bias (10), which depends on the clusters measurement errors, clusters sizes as well as the sign and the magnitude of derivatives (9).

3.2 The measurement bias in the calibrated weights

In order to analyze the bias (10), in the sequel we assume that the auxiliary variable $X^* > 0$. Given the clusters sample s_c:

1. If $(t_{x^*} - \hat{t}_{x^*,ht}) = 0$ then $w_i^* = d_i$. The calibrated weight reproduces the original sampling design weight. In presence of measurement errors affecting the sample values of the auxiliary variable X^*, let $\hat{t}_{x,ht}$ denote the Horvitz-Thompson estimator based on the observed totals t_{x_j}. If $\mu_j > 0$ for each $j \in s_c$ then the term $\widehat{R}_{i,s_c}$ in (7), coming from the measurement bias of $\hat{t}_{x,ht}$ as estimator of t_{x^*}, is negative. As a consequence, the calibrated weight w_i underestimates w_i^* ($w_i < w_i^* = d_i$). On the other hand, if $\mu_j < 0$ for each $j \in s_c$ the weight w_i overestimates w_i^* ($w_i > w_i^* = d_i$).

   The same results can be obtained analyzing the bias expression (10) together with derivatives (9). Both derivatives are negative under the initial condition $(t_{x^*} - \hat{t}_{x^*,ht}) = 0$.

2. If $(t_{x^*} - \hat{t}_{x^*,ht}) > 0$ then $w_i^* > d_i$. With regard to the sign of derivatives (9), the second one is always negative while the first one could be positive or negative. In more detail, we have

$$\left.\frac{\partial w_i}{\partial t_{x_i}}\right|_{P_0} = \frac{d_i\left[(t_{x^*} - \sum_{j\in s_c} d_j t_{x_j^*})(\sum_{j\in s_c} d_j t^2_{x_j^*} - 2d_i t^2_{x_i^*}) - d_i t_{x_i^*}\sum_{j\in s_c} d_j t^2_{x_j^*}\right]}{(\sum_{j\in s_c} d_j t^2_{x_j^*})^2}$$

$$= \frac{d_i}{\sum_{j\in s_c} d_j t^2_{x_j^*}}[(t_{x^*} - \sum_{j\in s_c} d_j t_{x_j^*})(1-2\lambda_i) - d_i t_{x_i^*}] \quad (11)$$

   where $\lambda_i = d_i t^2_{x_i^*}/(\sum_{j\in s_c} d_j t^2_{x_j^*})$ represents the weight of ith cluster in the selected sample s_c, with $0 < \lambda_i < 1$. The derivative (11) is negative if

$$\overline{\lambda}_i = \frac{1}{2}\left[\frac{t_{x^*} - \sum_{j\in s_c} d_j t_{x_j^*} - d_i t_{x_i^*}}{t_{x^*} - \sum_{j\in s_c} d_j t_{x_j^*}}\right] < \lambda_i < 1 \quad (12)$$

   where $\overline{\lambda}_i < \frac{1}{2}$.

3. If $(t_{x^*} - \hat{t}_{x^*,ht}) < 0$ then $w_i^* < d_i$. Under such circumstances the first derivative is always negative, the second derivative

$$\left.\frac{\partial w_i}{\partial t_{x_j}}\right|_{P_0} = -\frac{d_i d_j t_{x_i^*}[\sum_{j\in s_c} d_j t^2_{x_j^*} + 2t_{x_j^*}(t_{x^*} - \sum_{j\in s_c} d_j t_{x_j^*})]}{(\sum_{j\in s_c} d_j t^2_{x_j^*})^2}$$

$$= -\frac{d_i t_{x_i^*}}{t_{x_j^*}(\sum_{j\in s_c} d_j t^2_{x_j^*})}[d_i t_{x_j^*} - 2\lambda_j(\sum_{j\in s_c} d_j t_{x_j^*} - t_{x^*})] \quad (13)$$

is negative if

$$0 < \lambda_j < \frac{d_j t_{x_j^*}}{2(\sum_{j\in s_c} d_j t_{x_j^*} - t_{x^*})} = \overline{\lambda}_j \tag{14}$$

where $\lambda_j = d_j t^2_{x_j^*}/(\sum_{j\in s_c} d_j t^2_{x_j^*})$, with $0 < \lambda_j < 1$ for $j \neq i$.

For instance, given the sample s_c, suppose that $(t_{x^*} - \hat{t}_{x^*,ht}) > 0$. Then

(a) If $\overline{\lambda}_i < \lambda_i < 1$, both derivatives (9) are negative then the calibrated weight w_i underestimates w_i^* if $\mu_j > 0$ for each $j \in s_c$. Viceversa if $\mu_j < 0$.

(b) If $0 < \lambda_i < \overline{\lambda}_i$ the derivative (11) is positive. As a consequence, the final effect depends on how the parameters appearing in (10) interact each other. For instance, if $\mu_j > 0$ for each $j \in s_c$ the positive term in (10) due to ith cluster could compensate for the negative term coming from the remaining $n_c - 1$ clusters. In such circumstances, the calibrated weight w_i overestimates w_i^*.

As a matter of fact, note that the values of $\overline{\lambda}_i$ and $\overline{\lambda}_j$ given by (12) and (14) respectively depend on the composition of the selected sample s_c. However, for $X^* > 0$ and for sample sizes large enough the derivative (11) tends to be negative since the interval $(\overline{\lambda}_i < \lambda_i < 1)$ tends to the interval $(0 < \lambda_i < 1)$, being $\overline{\lambda}_i < 0$. The same consideration holds for (13), with regard to the interval $(0 < \lambda_j < \overline{\lambda}_j)$ being $\overline{\lambda}_j > 1$. This means that, as a consequence of the Horvitz-Thompson estimator consistency, both derivatives tends to be negative for sample size n_c large enough.

Then, the bias in the calibrated weight w_i depends on: (i) the composition of clusters sample; (ii) the sample clusters sizes; (iii) the clusters measurements errors. The first two factors influence the bias (10) through the quantities (λ_j, N_j) respectively, the third one through the superpopulation model parameters. How these factors combine each other is fundamental in determining the sign and the magnitude of the bias in the calibrated weights. As a consequence, the calibrated weight could be approximately unbiased in spite of the condition $E_m(t_{x_j}) = t_{x_j^*}$ for each $j \in s_c$, is not satisfied. However, in practice this happens very rarely. The bias in the calibrated weight w_i will be analyzed via simulation in Section 5.

4 The effect of measurement errors on the calibration estimator

Let us now consider the GREG estimator and the effect that the measurement errors in the X^* sample values have on its accuracy. By replacing the expression

(7) in the calibrated estimator (1), we obtain

$$\begin{aligned}
\hat{t}_y &= \sum_{i\in s_c} w_i t_{y_i} = \sum_{i\in s_c}\left[d_i + \frac{d_i t_{x_i}(t_{x^*} - \sum_{i\in s_c} d_i t_{x_i})}{\sum_{i\in s_c} d_i t_{x_i}^2}\right] t_{y_i} \\
&= \hat{t}_{y,ht} + \frac{\sum_{i\in s_c} d_i t_{x_i} t_{y_i}}{\sum_{i\in s_c} d_i t_{x_i}^2}(t_{x^*} - \sum_{i\in s_c} d_i t_{x_i}) \\
&= \hat{t}_{y,ht} + \frac{\sum_{i\in s_c} d_i t_{x_i} t_{y_i}}{\sum_{i\in s_c} d_i t_{x_i}^2}(t_x - \sum_{i\in s_c} d_i t_{x_i}) + \frac{\sum_{i\in s_c} d_i t_{x_i} t_{y_i}}{\sum_{i\in s_c} d_i t_{x_i}^2}(t_{x^*} - t_x) \\
&= \hat{t}_{y,greg} + \frac{\sum_{i\in s_c} d_i t_{x_i} t_{y_i}}{\sum_{i\in s_c} d_i t_{x_i}^2}(t_{x^*} - t_x) \\
&= \hat{t}_{y,greg} + \hat{B}(t_{x^*} - t_x) \qquad (15)
\end{aligned}$$

where $\hat{t}_{y,greg}$ is the generalized regression estimator under the assumption that the external total has been generated from the same measurement model. The evaluation of the estimator (15) must take into account both the measurement model m and the probability mechanism used to select the sample $p(s_c)$. More specifically, the first stage contributing to the randomness arises from the variation in the measurement on X^* produced by an hypothetical observational process regarding the all population, the second one arises from the sampling design.

Denoting by $E_m(\cdot)$ and $E_p(\cdot)$ the expectations with respect to the superpopulation model m and the sampling design $p(\cdot)$ respectively, the expected value of $\hat{t}_y$ is given by

$$\begin{aligned}
E(\hat{t}_y) &= E_m E_p(\hat{t}_y) \\
&= E_m E_p(\hat{t}_{y,greg} + \hat{B}(t_{x^*} - t_x)) \\
&= E_m E_p(\hat{t}_{y,greg}) + E_m E_p[\hat{B}(t_{x^*} - t_x)] \\
&\simeq E_m(t_y) + E_m\left[B(t_{x^*} - t_x)\right] \\
&= t_y + E_m\left[B(t_{x^*} - t_x)\right] \qquad (16)
\end{aligned}$$

where $E_p(\hat{B}) \simeq B$ is the regression coefficient under a hypothetical complete enumeration of the population, where we observe t_{y_i} and t_{x_i} for each cluster. It follows that the bias of $\hat{t}_y$, given by

$$Bias(\hat{t}_y) = E_m E_p(\hat{t}_y) - t_y \simeq E_m\left[B(t_{x^*} - t_x)\right] \qquad (17)$$

is due to measurement errors. In order to obtain an explicit expression for the measurement bias (17), note that it is a function of c population clusters totals. By first-order Taylor approximation technique evaluated at the point $P_1 = [E_m(t_{x_1}), ..., E_m(t_{x_c})]$, the approximation

$$B(t_{x^*} - t_x) \simeq B(t_{x^*} - t_x)\Big|_{P_1} + \sum_{i \in U_c} \left(\frac{\partial}{\partial t_{x_i}} B(t_{x^*} - t_x) \right) \Big|_{P_1} [t_{x_i} - E_m(t_{x_i})] \tag{18}$$

is obtained. It will be useful in computing the variance of the estimator. Taking the expected value of (18) with respect to the measurement model m we have

$$E_m \left[B(t_{x^*} - t_x) \right] \simeq -\frac{\sum_{i \in \mathbf{U}_c} (t_{x_i^*} + \mu_i) t_{yi}}{\sum_{i \in U_c} (t_{x_i^*} + \mu_i)^2} \sum_{i \in \mathbf{U}_c} \mu_i$$

As stressed in Section 3, if $E_m(t_{x_j}) = t_{x_j^*}$ the calibrated weight (7) will be approximately unbiased for (3). As a consequence, the estimator (15) will be approximately unbiased for the total of Y. Note that the bias in the calibration estimator mainly comes from the difference between the observed and the true total of X^* or equivalently from the measurement bias affecting $\hat{t}_{x,ht}$ as estimator of t_{x^*}. In fact, if the same measurement model generates both the sample values and the external total (t_x), then the measurement errors will not tend to introduce bias into the calibration estimator but only lead to a loss of efficiency. Formally, it is easy to show that $E_p(\hat{t}_{y,greg}) \simeq t_y$.

The variance of the calibration estimator is shown to be composed of two components

$$var(\hat{t}_y) = var_m E_p(\hat{t}_y) + E_m var_p(\hat{t}_y) \tag{19}$$

for details see Appendix. With regard to the first component, and using the approximation (18), we have

$$var_m E_p(\hat{t}_y) = \sum_{i \in \mathbf{U}_c} a_i^2 var_m(t_{x_i}) + \sum_{i \in \mathbf{U}_c} \sum_{j \neq i \in \mathbf{U}_c} a_i a_j cov_m(t_{x_i}, t_{x_j}) \tag{20}$$

where

$$a_i = \frac{\partial}{\partial t_{x_i}} \left[B(t_{x^*} - t_x) \right] \Big|_{P_1}$$

Note that (20), called measurement variance, represents the increase of variance due to the presence of measurement errors. More specifically, the former component comes from the variability of measurements (simple measurement variance), the latter component from their correlation due to the presence of a common interviewer (correlated measurement variance). The variability induced by the sampling design is represented by

$$
\begin{aligned}
E_m var_p(\hat{t}_y) &\simeq \sum_{i\in\mathbf{U}_c}\sum_{j\in\mathbf{U}_c}\left(\frac{d_i d_j}{d_{ij}}-1\right)E_m(E_iE_j)\\
&+ \sum_{i\in\mathbf{U}_c}\sum_{j\in\mathbf{U}_c}\left(\frac{d_i d_j}{d_{ij}}-1\right)E_m\left[\frac{(t_{x^*}-t_x)^2 t_{x_i}t_{x_j}E_iE_j}{\sum_{i\in\mathbf{U}_c}t_{x_i}^2}\right]\\
&+ 2E_m\left[(t_{x^*}-t_x)cov_p(\hat{t}_{y,greg},\hat{B})\right]
\end{aligned}
$$

where the first two components are expressed in terms of residuals $E_i = t_{yi} - Bt_{xi}$ for the whole population. In conclusion, the Mean Square Error of the calibration estimator $\hat{t}_y$ can be written as the sum of variance (19) and squared bias (17).

The results obtained for a single-stage cluster sampling can be easily extended to a two stage sampling or multistage sampling. Various practical situations are covered by this general information statement. For instance city blocks can be sampled at first stage and buildings within blocks at the second. In a two stage-cluster sampling, the auxiliary information can be available both for units and for clusters. Estevao & Särndal (2006) show various alternatives to compute the calibrated weights using a single step or a two steps approach.

5 A Simulation Study

In the previous sections we have formally evaluated the effect on both the calibrated weights and the resulting calibration estimator of measurement errors affecting the sample auxiliary information. In this section we perform a simulation experiment. In detail, a finite population of size $N = 2000$ units was generated from $log(y) = 1 + x + e$, where $x \sim Gamma(5, 1)$ and $e \sim Normal(0, 1)$.

Without loss of generality we assume that the population is partitioned in 10 clusters, each having the same size $N_i = 200$. The tendency of the units in the same cluster to resemble each other with regard to the Y variable is expressed by the homogeneity coefficient $\tau = 0.52$, see Särndal et al. (1992) (pag. 130).

From the population a simple random sampling without replacement of size $n = 3$ was taken. For each cluster $i \in \mathbf{U}_c = \{U_1, .., U_c\}$, we have randomly generated the measurement errors from a normal distribution with mean and variance equal to m_i and ν_i respectively. In order to evaluate the expectation $E_m(w_i|s_c)$, the hypothetical observational process regarding the all population has been repeated $B = 300$ times.

Given the clusters sample $s_c = (1, 2, 3)$, we begin by evaluating the effect of measurement errors on the calibrated weight of *1*th cluster. Let us suppose that the measurement model parameters are the same for each sample cluster, that is $m_1 = m_2 = m_3 = m$ and $\nu_1 = \nu_2 = \nu_3 = \nu = 1$. Different values of m have been used

$$m = (-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1, 2, 3, 4, 5)$$

The result is reported in Figure 1, where the horizontal line represents the true calibrated weight $w_1^* = 3.27$ obtained in absence of measurement errors. With regard to the calibrated weight curve characterized by $m_1 = m_2 = m_3$, if the sample clusters expected measurement values reproduce the true values (*i.e.* $m = 0$) then w_1 is approximately unbiased for w_1^*. If $m < 0$ then w_1 overestimates w_1^*. Viceversa if $m > 0$.

Suppose now that the interviewer effect corresponding to the first cluster increases, while the interviewer effect of the remaining sample clusters does not change. More specifically, we assume that $m_1 = 2, m_2 = m_3$. As Figure 1 shows the curve shifts down, then the calibrated weight w_i underestimates w_1^* for $m > -1$. Besides, when $m_1 = 2, m_2 = m_3 = -1$ the weight w_1 is approximately unbiased for w_1^* since in the approximation

$$E_m(w_i|s_c) \simeq w_i^* + \sum_{j\in s_c} \left.\frac{\partial w_i}{\partial t_{x_j}}\right|_{P_0} \mu_j \quad for \quad i = 1 \tag{21}$$

the second term on the right side is near to zero. Then, the bias in the calibrated weight w_i could be zero in spite of the expected measurements values on elements do not agree with the true values. In order to investigate the dependence of (21) from the sample clusters sizes, suppose to increase the size of *1*th cluster setting $N_1 = 300, N_2 = N_3 = 150$. The calibrated weight curve shifts down again, since the *1*th cluster interviewer effect is intensified by the larger sample size N_1.

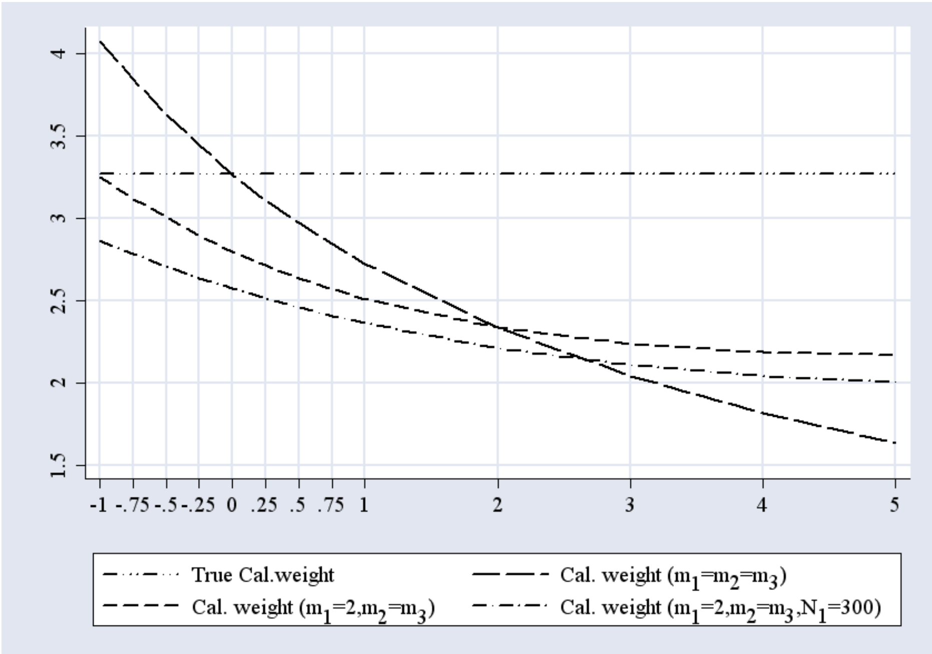
The above simulation has been repeated for different values of the sample clusters measurement errors variance. The results are reported in Table 1 as ν increases for different values of m. As Table 1 shows, the effect of the measurement errors variability on the calibrated weight w_1 is negligible.

Table 1: *Calibrated weight for* 1*th cluster in the sample* $s_c = (1, 2, 3)$*, as the sample clusters measurement errors variability* ν *increases and for different values of* m*.*

m	$\nu = 0.5$	$\nu = 1$	$\nu = 2$	$\nu = 5$	$\nu = 10$
-1	4.09	4.09	4.08	4.13	4.08
0	3.27	3.27	3.26	3.27	3.32
1	2.72	2.73	2.73	2.73	2.73
4	1.81	1.81	1.81	1.80	1.82

Next, we proceed to evaluate the Bias and the Mean Square Error of the calibration estimator (15). More specifically, assuming that the measurement model parameters are the same for each population cluster ($m_i = m, \nu_i = \nu = 1$), the

Figure 1: *Calibrated weight for* 1*th cluster in the sample* $s_c = (1, 2, 3)$.



relative bias of $\hat{t}_y$ is reported in Figure 2. Its expression is given by

$$RB = \frac{Bias(\hat{t}_y)}{t_y} 100$$

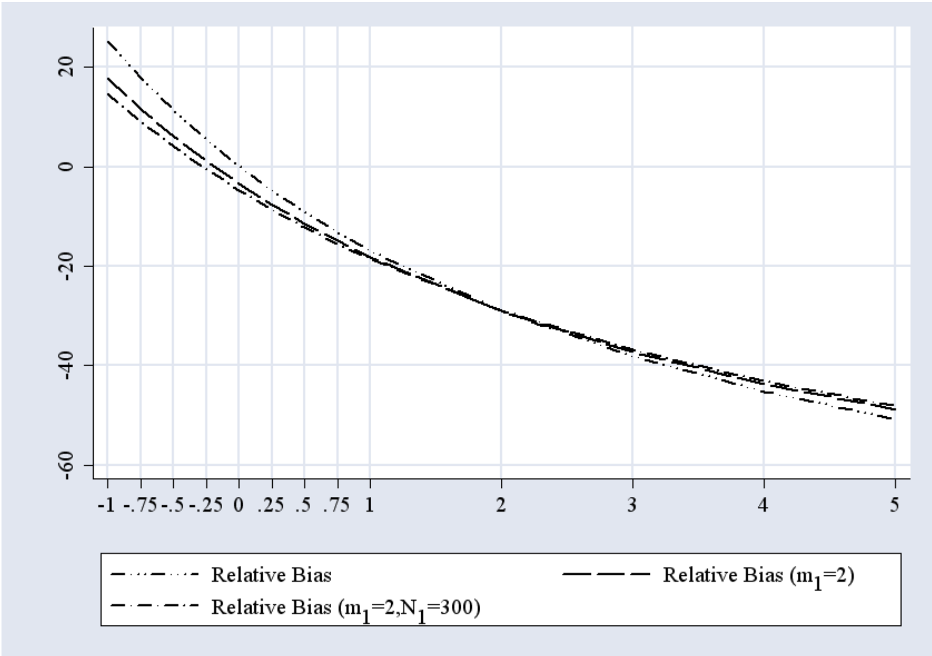
In Figure 3, a measure of relative efficiency given by

$$RE = \sqrt{\frac{MSE(\hat{t}_y)}{MSE(\hat{t}_{y,ht})}} \tag{22}$$

is shown, where $MSE(\hat{t}_{y,ht})$ represents the Mean Square Error of Horvitz-Thompson estimator.

Figure 2 shows that, if $m_i = m = 0$ for each population cluster then $\hat{t}_y$ will be approximately unbiased for the Y variable total. Otherwise, for $m_i = m < 0$ since the calibrated weights overestimate w_i^* the calibration estimator $\hat{t}_y$ will have a positive bias. Viceversa for $m_i = m > 0$. If we assume that $m_1 = 2, m_i = m$ for each $i \neq 1$ then the bias decreases for $m < -0.25$ and $m > 2$ while increases for $-0.25 < m < 2$. In the former case ($m < -0.25$) a positive interviewer effect for the first cluster partially balances the negative interviewer effects of the remaining clusters. Clearly, as shown in Figure 2, this effect will be more enhanced if the *1*th cluster size is larger than the others. We show it setting $N_1 = 300, N_2 = N_3 = 150, N_i = 200$ for each $i \neq (1, 2, 3)$.

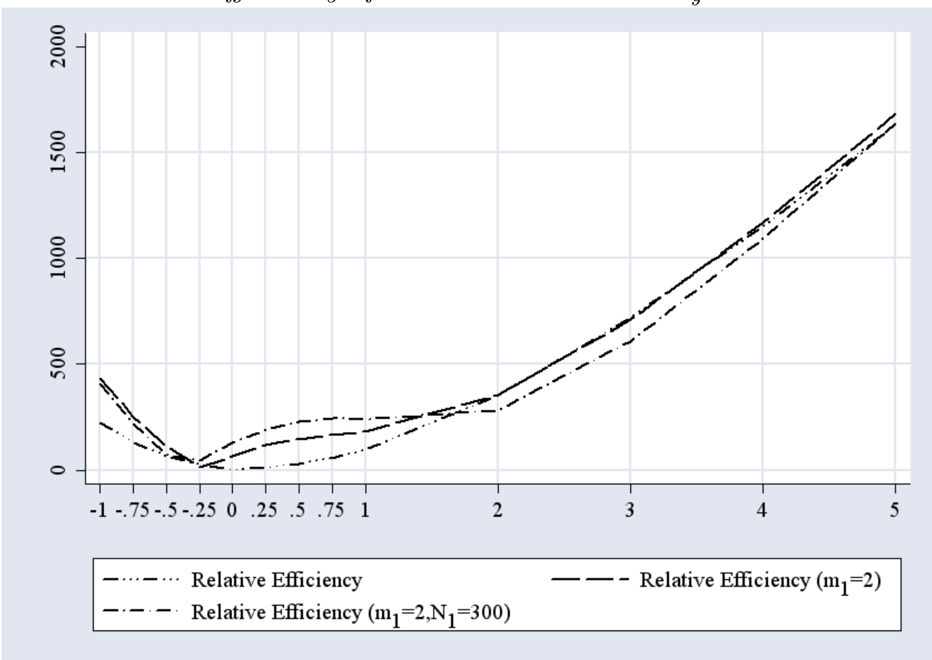
Figure 2: *Relative Bias of calibration estimator* $\hat{t}_y$.



The performance of the calibration estimator (15), evaluated in terms of Mean Square Error, is reported in Figure 3. If $m_i = m, \nu_i = \nu = 1$ for each population cluster, the relative efficiency measure exhibits a parabolic behavior whose minimum equal to 1.4 corresponds to the measurement model parameter $m = 0$. Under such a condition, since the estimator (15) is approximately unbiased the measure (22) becomes a ratio between two variances. If the interviewer effect of first cluster increases ($m_1 = 2$) then the Mean Square Error shifts up for $m < 2$ because of the increase in the estimator bias. Moreover, if we increase also the cluster size setting $N_1 = 300, N_2 = N_3 = 150, N_i = 200$ for each $i \neq (1, 2, 3)$, the relative efficiency curve shifts down below the other curves for $m \geqslant 2$. Under such a circumstance, while the measurement model parameters corresponding to the population clusters $i \in \mathbf{U}_c$ for each $i \neq 1$ increase ($m = 3, 4, 5$), the *1*th cluster model parameter remains constant ($m_1 = 2$). Moreover, its effect on the estimator accuracy will be more enhanced because of the larger cluster size $N_1 = 300$. Such a circumstance implies a gain of efficiency for $m \geqslant 2$.

Note that, in absence of measurement errors and assuming that the clusters have the same size $N_i = 200$, the relative efficiency of the calibration estimator is less than one and equal to 0.08. As previously stressed, in presence of measurement errors (*i.e.* $m_i = m = 0$) the relative efficiency increases to 1.4. Since the estimator (15) is approximately unbiased if the expected measurements values reproduce the true values, the loss of efficiency comes from the variability and the correlation between measurements obtained from the same interviewer. As a consequence, the calibration estimator is not more efficient than the Horvitz-Thompson estimator that ignores such a information.

Figure 3: *Relative Efficiency of calibration estimator* $\hat{t}_y$.



6 Conclusions

The basic idea of the calibration approach is to modify the original sampling design weights d_i using the available auxiliary information. In this paper, the total t_{x^*} is assumed to be accurately known while the X^* sample values are affected by measurement errors. As shown in Section 3, the bias in the calibrated weight w_i depends essentially on the following factors: (i) the composition of clusters sample; (ii) the sample clusters sizes; (iii) the clusters measurement errors. The first two factors influence the bias (10) through the quantities (λ_j, N_j) respectively, the third one through the superpopulation model parameters. How these factors combine each other determines the sign and the magnitude of the bias in the calibrated weight.

As a matter of fact, the presence of measurement errors in X^* affects the accuracy of $\hat{t}_y$ through the weights w_i. In particular, as shown from (17) the bias in $\hat{t}_y$ mainly comes from the difference between the observed and the true total of X^*, that is from the measurement bias of $\hat{t}_{x,ht}$ as estimator of the known external total t_{x^*}. In fact, as stressed in Section 4 if the same measurement model generates both the sample values and the external total, then the measurement errors will not tend to introduce bias into the calibration estimator but only lead to a loss of efficiency. The variance of $\hat{t}_y$ is composed by two components. The first one depends on sampling design, the second one comes from the variability and the correlation between measurements obtained from the same interviewer.

In conclusion, as shown both formally and via simulation, the presence of measurement errors can eliminate the major efficiency of the calibration estimator respect to the Horvitz-Thompson estimator that ignores the available auxiliary information. Formally, the use of calibration is justified if the relative efficiency measure (22) is less than one. Such a circumstance happens if the bias and the variability increase in the estimator due to measurement errors does not eliminate completely the variance reduction due to calibration. Clearly, the final effect depends on the characteristics of the finite population regarding the relationship between X^* and Y, as well as the measurement errors generating mechanism. In fact, the larger is the correlation between X^* and Y the larger the benefits in terms of Mean Square Error of the calibration, and then the larger the tolerance level for the measurement errors. On the other side, if the population is not well described by a linear regression model, the improvement in terms of variance reduction on the Horvitz-Thompson estimator is modest and it can be easily compensated by the loss of efficiency due to measurement errors. Then, if the X^* sample values are of poor quality the calibration approach could not bring to any gain of efficiency. In such circumstances, estimators that ignore such a information are preferable.

ACKNOWLEDGEMENT

The author would like to thank Prof. Pier Luigi Conti for his helpful comments.

Appendix

In this appendix we derive the variance of the calibration estimator $\hat{t}_y$. Such a variance is shown to be composed of two components

$$var(\hat{t}_y) = var_m E_p(\hat{t}_y) + E_m var_p(\hat{t}_y)$$

Consider each of these separately. With regard to the first component, we have

$$\begin{aligned}
var_m E_p(\hat{t}_y) &= var_m\left[t_y + B(t_{x^*} - t_x)\right] \\
&= var_m\left[B(t_{x^*} - t_x)\right] \\
&\simeq var_m\left[\sum_{i\in\mathbf{U}_c} a_i[t_{x_i} - (t_{x_i^*} + N_i(b_i + \delta))]\right] \\
&= \sum_{i\in\mathbf{U}_c}\sum_{j\in\mathbf{U}_c} a_i a_j cov_m(t_{x_i}, t_{x_j}) \\
&= \sum_{i\in\mathbf{U}_c} a_i^2 var_m(t_{x_i}) + \sum_{i\in\mathbf{U}_c}\sum_{j\neq i\in\mathbf{U}_c} a_i a_j cov_m(t_{x_i}, t_{x_j})
\end{aligned}$$

where using (18) we have set $a_i = \frac{\partial}{\partial t_{x_i}}\left[B(t_{x^*} - t_x)\right]\Big|_{P_1}$.

The variability induced by the sampling design is represented by

$$\begin{aligned}
E_m var_p(\hat{t}_y) &= E_m var_p\left[\hat{t}_{y,greg} + \hat{B}(t_{x^*} - t_x)\right] \\
&= E_m\left[var_p(\hat{t}_{y,greg})\right] + E_m\left[(t_{x^*} - t_x)^2 var_p(\hat{B})\right] \\
&\quad +2E_m\left[(t_{x^*} - t_x)cov_p(\hat{t}_{y,greg}, \hat{B})\right] \\
&= A_1 + A_2 + A_3 \qquad (23)
\end{aligned}$$

where

$$\begin{aligned}
A_1 &\simeq E_m\left[\sum_{i\in\mathbf{U}_c}\sum_{j\in\mathbf{U}_c}\left(\frac{d_i d_j}{d_{ij}} - 1\right)E_i E_j\right] \\
&= \sum_{i\in\mathbf{U}_c}\sum_{j\in\mathbf{U}_c}\left(\frac{d_i d_j}{d_{ij}} - 1\right)E_m(E_i E_j)
\end{aligned}$$

is expressed by means of residuals $E_i = t_{y_i} - Bt_{x_i}$ and

$$\begin{aligned}
A_2 &= E_m\left[\frac{(t_{x^*} - t_x)^2}{\sum_{i\in\mathbf{U}_c} t_{xi}^2}\sum_{i\in\mathbf{U}_c}\sum_{j\in\mathbf{U}_c}\left(\frac{d_i d_j}{d_{ij}} - 1\right)(t_{x_i}E_i)(t_{x_j}E_j)\right] \\
&= \sum_{i\in\mathbf{U}_c}\sum_{j\in\mathbf{U}_c}\left(\frac{d_i d_j}{d_{ij}} - 1\right)E_m\left[\frac{(t_{x^*} - t_x)^2 t_{x_i}E_i t_{x_j}E_j}{\sum_{i\in\mathbf{U}_c} t_{x_i}^2}\right]
\end{aligned}$$

With regard to the term A_3 in (23), note that

$$
\begin{aligned}
cov_p(\hat{t}_{y,greg}, \hat{B}) &= E_p[(\hat{t}_{y,greg} - E_p(\hat{t}_{y,greg}))(\hat{B} - E_p(\hat{B}))] \\
&\simeq E_p[(\hat{t}_{y,greg} - t_y)(\hat{B} - B)]
\end{aligned}
$$

since $\hat{t}_{y,greg}$ and $\hat{B}$ are approximately unbiased for t_y and B respectively. For sample size large enough, the $cov_p(\hat{t}_{y,greg}, \hat{B})$ can be approximated by the covariance between their linear approximations at the point $P_1 = (t_y, t_x, t_{xy}, t_{x^2})$. Then

$$
\begin{aligned}
cov_p(\hat{t}_{y,greg}, \hat{B}) &\simeq \frac{cov_p(\hat{t}_{y,ht}, \hat{t}_{x^2,ht})}{t_{x^2}} - \frac{BCov_p(\hat{t}_{y,ht}, \hat{t}_{xy,ht})}{t_{x^2}} + \\
&- \frac{BCov_p(\hat{t}_{x,ht}, \hat{t}_{x^2,ht})}{t_{x^2}} + \frac{B^2 cov_p(\hat{t}_{x,ht}, \hat{t}_{xy,ht})}{t_{x^2}}
\end{aligned}
$$

References

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A.,Sudman,S., (1991). Measurement Errors in Surveys. New York : Wiley

Deville, J.C., and Särndal, C.E., (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376-382.

Estevao, V.M., Särndal, C.E., (2006). Survey estimates by calibration on complex auxiliary information. International Statistical Review, 74, 127-147.

Groves, R.M.,(1989). Survey Errors and Survey Costs. New York : Wiley.

Hansen, M.H., Hurwitz, W.N., Bershad,M.A., (1961). Measurement errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38, 359-374.

Hansen, M.H., Hurwitz, W.N., Marks, E.S., Mauldin, W.P., (1951). Response Errors in Surveys. *Journal of the American Statistical Association*, 46,147-190.

Hansen, M.H., Hurwitz, W.N., Pritzker, L., (1964). The estimation and interpretation of gross differences and the Simple Response Variance. In Contributions to Statistics (presented to P.C. Mahalanobis on the occasion of his 70th birthday).

Mahalanobis, P.C., (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, 109, 325-370.

Särndal, C.E., Swensson, B., Wretman, J., (1992). Model assisted survey sampling. New York : Springer-Verlag.