Variance of Calibrated Estimator with Sampling Constraints and Their Effects

C. Ceccarelli, G.M. Giorgi and A. Guandalini[1]

ABSTRACT

The use of auxiliary information considerably improves the estimates and allows their production for unplanned domains, as well.

Different kinds of auxiliary information can be used to improve estimates. Demographic information from administrative sources is frequently used, but sampling information can also be used.

In this context we present an extension of calibrated estimator (CAL) of a total, originally proposed by Deville and Särndal in 1992. Our proposal of CAL estimator is able to take into account both demographic and sampling information. It also agrees with the needs of the Italian Labour Force Survey.

The joint use of these kinds of auxiliary information in CAL estimators raises some questions. How their sampling error can affect the efficiency of estimates and whether introducing exogenous elements of sampling error is justified by the increased efficiency or at least improvement in estimates?

The aim of this article is: (i) to derive the expression for the estimate of variance of the CAL estimator with sampling information, (ii) to present a new tool, based on sampling error, quick and able to support the introduction of additional auxiliary variables in the constraints system of the CAL estimator.

---

[1] Claudio Ceccarelli, Italian National Institute of Statistics, Department for Environmental and Social Statistics, Rome, Italy. E-mail: claudio.ceccarelli@istat.it; Giovanni M. Giorgi, "La Sapienza" University, Department of Statistical Sciences, Rome, Italy. E-mail: giovanni.giorgi@uniroma1.it; Alessio Guandalini, "La Sapienza" University, Department of Statistical Sciences, Rome, Italy. E-mail: alessio.guandalini@uniroma1.it.

Some interesting results are achieved both analytically and by applying the proposal methodology to the monthly estimates of the Italian Labour Force Survey.

KEY WORDS: calibrated estimator; composite estimator; sample constraints; exogenous variance; constraints effect; constraints selection.

1. INTRODUCTION

The use of demographic variables from administrative sources, with no sampling errors, is already consolidated both for the strategy of sampling units and for the determination of estimators.

Moreover, several National Institutes of Statistics use information from previous periods to improve estimates and make them less volatile.

Besides the efficiency gain, they also have the need to build a coherent system of surveys and the possibility to produce also estimates for unplanned domains.

The Italian Labour Force Survey ($It\text{-}LFS$), following the experience of Canadian Labour Force Survey, uses longitudinal information to increase the efficiency of estimates. However, the use of regression composite estimator by Singh, Kennedy and Wu (2001) and Fuller and Rao (2001) can have some trouble when applied to the $It\text{-}LFS$. The impact of imputation for a high fraction of previous period missing values for the present month's respondent can be serious (Singh, Kennedy and Wu, 2001, p. 34). Moreover looking at the transition among working status, for the Italian labour market, the probability to stay in the same status of a previous periods is considered quite close to one, both for employed and inactive workers. Therefore, we can consider them as straight constraints for our estimates.

Our proposal of Calibrated estimator (CAL) allows to exploit auxiliary information from previous periods to improve efficiency of estimates (Wolter, 1979, p. 604). Moreover, in the perspective of National Institute of Statistics, it allows to provide estimates consisting with those of simultaneous surveys for building a coherent system of estimates.

The CAL estimator, developed by Deville and Särndal (1992), uses the total of auxiliary variables as constraints to produce unbiased, efficient and consistent estimates.

In general in CAL estimator the totals of auxiliary variables are exactly-known from administrative sources (administrative constraints), without sampling error. They are structural variables, such as sex, age and location.

However, constraints in CAL estimator do not necessarily have to be totals exactly-known. They can be also derived from different sample surveys, with sampling error (Deville 1999, p. 207).

The introduction of this kind of constraints ensures that our estimates are consistent with total providing from simultaneous surveys or with the same survey of previous periods.

Therefore, in variance estimate, in addition to the sampling error, we must considered the error related to sampling estimates we put as constraints.

Our paper suggests an analytical expression of variance estimate of CAL estimator of total with sampling constraints. Furthermore, we also propose a prior evaluation of efficiency increase that we can expect adding further constraints in CAL estimator (similarly to "variable selection" we can call this "constraints selection").

After introducing notation and methodological bases (section 2), in section 3 we derive the amount of error imported from the survey from which the total estimate, used as constraint, is taken. We call it exogenous variance. A tool for deciding the constraints addition, based on evaluation of estimator efficiency, is introduced in section 4. In section 5 the whole methodology illustrated in the previous sections is applied to monthly estimates of the Italian Labour Force Survey. Finally, in section 6, there are conclusions and remarks.

2. NOTATION AND METHODOLOGICAL BASES

In sample surveys, for each sample unit, the values of several auxiliary variables, be noted by the vector $\boldsymbol{x} = \left(x_{1k}, \dots, x_{jk}, \dots, x_{Jk}\right)'$, are observed in addition to the variable of interest Y. Hence, for each sample unit we have a vector $(y_k, \boldsymbol{x}_k)$ with $J+1$ elements, being J the number of auxiliary variables.

We assume complete responses for variable Y, so the value y_k is observed on each k units of sample s.

The vector of auxiliary variables, as well as the vector of their known totals $\boldsymbol{X} = \sum_{k \in U} \boldsymbol{x}_k = \left(\sum_{k \in U} \boldsymbol{x}_{1k}, \dots, \sum_{k \in U} \boldsymbol{x}_{jk}, \dots, \sum_{k \in U} \boldsymbol{x}_{Jk}\right)$, can be used to improve the estimate of Y (Bethlehem, Keller, 1987, p. 143 and Wright, 1983, p. 879).

The CAL estimator of a total, $\hat{Y}_{CAL} = \sum_{k \in s} y_k w_k$, estimates the total amount of interest variable using a system of weights that "*perform well for the auxiliary variable*" (Deville & Särndal, 1992, p. 367). Weights w_k are computed by solving the constrained optimization problem:

$$\begin{cases} min\left\{\sum_{k \in s} G(w_k, d_k)\right\} \\ \sum_{k \in s} w_k y_k = \boldsymbol{X} \end{cases}, \qquad (2.1)$$

where G is a quadratic distance function (Deville & Särndal, 1992, pp. 377-379; Singh & Mohl, 1996, pp. 108-114).

For the sake of simplicity, we use the notation w_k and d_k instead w_{k_s} and d_{k_s} even if these depends on both the unit k and the sample s.

The constrained optimization problem assures that all individuals of the same family have the same final weights w_k (Lemaître & Dufour, 1987, p. 199). The final weights are, for a given distance function, on average, as close as possible to d_k and respected a constraints system with the vector of totals $\boldsymbol{X}$. The d_k (equal to the inverse of inclusion probabilities, π_k) are previously adjusted for nonresponse (Deville, Särndal & Sautory, 1993, p. 1013).

For large sample size CAL estimator is asymptotically equivalent to the $GREG$ estimator, because distance functions, generally used to measure distance between weights w_k and d_k, are linear in $\boldsymbol{x}$ (Deville & Särndal, 1992, pp.377-380 and Singh & Mohl, 1996, pp. 108-114). The variance of CAL estimator is asymptotically equal to that of $GREG$ estimator and its unbiased estimate is:

$$var(\hat{Y}_{CAL}) = \sum_{k \in s} \sum_{l \neq k} \frac{\Delta_{kl}}{\pi_{kl}} \left(\gamma_k \frac{\hat{e}_k}{\pi_{k_s}} \right) \left(\gamma_l \frac{\hat{e}_l}{\pi_{l_s}} \right), \quad (2.2)$$

where $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$. Moreover $\hat{e}_{\cdot}$ and $\gamma_{\cdot}$ are respectively the residuals estimated by the regression model and the corrector of basic design weights (Deville & Särndal, 1992, pp. 379-380 and Estevao, Hidiroglou & Särndal, 1995, p. 184).

3. CALIBRATED ESTIMATOR WITH SAMPLE CONSTRAINTS

The CAL estimator assures consistence with the external source (administrative or census) without sampling error, because its estimates are constrain by $\boldsymbol{X}$ vector of exactly-known totals. However, the estimates can be also constrained to a vector of sampling-known totals, or to totals estimated by another sample survey. In this case the vector of sampling-known totals is $\hat{\hat{\boldsymbol{X}}}$ (Deville, 1999, p. 207) and the constrained optimization problem (2.1) becomes:

$$\begin{cases} min\left\{\sum_{k \in s} G(w_k, d_k)\right\} \\ \sum_{k \in s} w_k \boldsymbol{x}_k = \left(\boldsymbol{X}, \hat{\hat{\boldsymbol{X}}}\right) \end{cases} \tag{3.1}$$

where $\left(\boldsymbol{X}, \hat{\hat{\boldsymbol{X}}}\right)' = \left(X_1, \dots, X_j, \dots, X_{J-r-1}, \hat{\hat{X}}_{J-r}, \dots, \hat{\hat{X}}_J\right)$ and the first J elements are exactly-known totals, $\boldsymbol{X}$, while the last r elements are sampling-known totals, $\hat{\hat{\boldsymbol{X}}}$, which are estimates with sampling error.

We distinguish between constraints from overlapped samples (OS) and constraints from non-overlapped samples (NOS).

We have OS constraints from a repeated survey with overlapping samples. Therefore we use, as constraints, totals derived from previous period of a periodic survey with a rotating design. In this case we constrain the estimate of our parameter to a total obtained on a fraction of the same units.

When the overlapping does not occur we are in NOS case. The totals used as constraints are taken from a different survey in the same period, with the same definitions of variables.

The use of sampling-known totals, both NOS and OS, does not change the expression of the estimator. Indeed the sampling-known totals are considered as simple constraints, and the way of determining the system of weights remains essentially unchanged.

However, the presence of sampling constraints modifies the expression of the variance of the CAL estimator, because of the use of sampling-known totals.

Therefore, we introduce a portion of error from the other sample where the totals are derived.

A first suggestion for variance estimate of CAL estimator with these kinds of constraints, for a simple random sampling and for NOS constraints only, is provided by Ballin, Falorsi and Russo (2000, p. 49).

Readjusting sample dependent weights and introducing a discount rate depending on the proportion of sample units for which the sampling constraints act, f_j, we have a variance estimator useful for complex design surveys.

We define $\widehat{CAL}$ the CAL estimator with sampling constraints. An asymptotically unbiased estimate of variance $\widehat{CAL}$, if at least a NOS constraint is considered, is:

$$var\left(\hat{Y}_{\widehat{CAL}^{NOS}}\right) =$$

$$= \sum_{k\in s}\sum_{l\neq k}\frac{\Delta_{kl}}{\pi_{kl}}\left(\gamma_{ks}\frac{\hat{e}_k}{\pi_k}\right)\left(\gamma_{ls}\frac{\hat{e}_l}{\pi_l}\right) + \sum_{j=1}^{J} f_j\left(B_j^2 var\left(\hat{\hat{X}}_j\right) + \sum_{j'\neq j} B_j B_{j'} cov\left(\hat{\hat{X}}_j, \hat{\hat{X}}_{j'}\right)\right), \qquad (3.2)$$

In the above expression the first element is the sampling error of Y variable for the survey in issue and it is equal to the variance (2.2). The second element is the additional error we introduce when sampling-known constraints are considered. Later we will define it as “exogenous variance”.

When we use at least an OS constraint, we ought to introduce a component part of joint variability that, under the hypothesis of constancy of regression coefficients and of unbiased estimates of sampling-known totals used as constraints, is equal to $2\hat{Y}$ (see Appendix, Proof 1.). Hence, the variance estimate of CAL estimator with OS constraints is:

$var(\hat{Y}_{\overline{CAL}}os) =$

$$= \sum_{k\in s}\sum_{l\neq k}\frac{\Delta_{kl}}{\pi_{kl}}\left(\gamma_{ks}\frac{\hat{e}_k}{\pi_k}\right)\left(\gamma_{ls}\frac{\hat{e}_l}{\pi_l}\right) + \sum_{j=1}^{J} f_j\left(B_j^2 var\left(\hat{\hat{X}}_j\right) + \sum_{j'\neq j} B_j B_{j'} cov\left(\hat{\hat{X}}_j, \hat{\hat{X}}_{j'}\right) + 2\hat{Y}\right). \qquad (3.3)$$

We can write (3.2) and (3.3) in the form:

$$var(\hat{Y}_{\overline{CAL}}) \cong var(\hat{Y}_{GREG}) + \mathbf{1}'(\boldsymbol{L})\,\mathbf{1} + 2\hat{Y} diag(\boldsymbol{F})'\boldsymbol{I}[j], \qquad (3.4)$$

where $\mathbf{1}$ is a J-dimension vector of one, $\boldsymbol{L} = (\boldsymbol{F} \otimes \boldsymbol{B} \otimes \boldsymbol{S})$ is the Hadamard product of diagonal matrix $\boldsymbol{F}$ in which the element f_j is the proportion of sample units for which the j-th sampling constraints acts, and $\boldsymbol{B}$, $\boldsymbol{S}$ are two symmetric J-matrices defined as:

$$\boldsymbol{B} = \begin{bmatrix} B_1^2 & \dots & B_j B_1 & \dots & B_{J-r-1}B_1 & B_{J-r}B_1 & \dots & B_J B_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ B_1 B_j & \dots & B_j^2 & \dots & B_{J-r-1}B_j & B_{J-r}B_j & \dots & B_J B_j \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ B_1 B_{J-r-1} & \dots & B_j B_{J-r-1} & \dots & B_{J-r-1}^2 & B_{J-r}B_{J-r-1} & \dots & B_J B_{J-r-1} \\ B_1 B_{J-r} & \dots & B_j B_{J-r} & \dots & B_{J-r-1}B_{J-r} & B_{J-r}^2 & \dots & B_J B_{J-r} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ B_1 B_J & \dots & B_j B_J & \dots & B_{J-r-1}B_J & B_{J-r-1}B_j & \dots & B_J^2 \end{bmatrix}$$

$\boldsymbol{S}$=

$$\begin{bmatrix} var(X_1) & \dots & cov(X_j, X_1) & \dots & cov(X_{J-r-1}, X_1) & cov(X_{J-r}, X_1) & \dots & cov(X_J, X_1) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(X_1, X_j) & \dots & var(X_j) & \dots & cov(X_{J-r-1}, X_j) & cov(X_{J-r}, X_j) & \dots & cov(X_J, X_j) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(X_1, X_{J-r-1}) & \dots & cov(X_j, X_{J-r-1}) & \dots & var(X_{J-r-1}) & cov(X_{J-r}, X_{J-r-1}) & \dots & cov(X_J, X_{J-r-1}) \\ cov(X_1, X_{J-r}) & \dots & cov(X_j, X_{J-r}) & \dots & cov(X_{J-r-1}, X_{J-r}) & var(X_{J-r}) & \dots & cov(X_J, X_{J-r}) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov(X_1, X_J) & \dots & cov(X_j, X_J) & \dots & cov(X_{J-r-1}, X_J) & cov(X_{J-r}, X_J) & \dots & var(X_J) \end{bmatrix}$$

where B_j is the regression coefficient of j-th auxiliary variable and $\boldsymbol{S}$ is the sampling variance-covariance matrix of auxiliary variables. Note that $diag(\boldsymbol{F})' = (f_1, \dots, f_{J-r-1}, f_{J-r} \dots, f_J)$ and $\boldsymbol{I}[j]$ is a column-vector where each element is equal to 1 if j is the OS constraint and 0 otherwise.

We stress that actually matrix $\boldsymbol{L}$ possesses the structure:

$$
\boldsymbol{L} = \begin{bmatrix} 0 & 0 \\ 0 & \begin{matrix} B_{J-r}^2 var(X_{J-r}) & \cdots & B_{J-r} B_J cov(X_J, X_{J-r}) \\ \vdots & \ddots & \vdots \\ B_J B_{J-r} cov(X_{J-r}, X_J) & \cdots & B_J^2 var(X_J) \end{matrix} \end{bmatrix}
$$

because $var(X_j) = 0$ for $j = 1, \ldots, J - r - 1$ and $cov(X_j, X_{j'}) = \rho_{j,j'} \sqrt{var(X_j) var(X_{j'})} = 0$ for $j = 1, \ldots, J - r - 1$ e $j' = 1, \ldots, J - r - 1$ with $j \neq j'$.

If all totals of constraints are exactly-known ($r = 0$), variance estimate (3.4) becomes simply (2.2), because both $\mathbf{1}'(\boldsymbol{L})\,\mathbf{1}$ and $2\hat{Y}\, diag(\boldsymbol{F})'\boldsymbol{I}[j]$ are equal to 0. When there is at least a sampling constraint, if it is a NOS constraint, $var(\hat{Y}_{\widehat{PV}}) \cong var(\hat{Y}_{GREG}) + \mathbf{1}'(\boldsymbol{L})\mathbf{1}$, whilst if it is an OS constraint (3.4) holds. Moreover, if we define $var(E) = \mathbf{1}'(\boldsymbol{L})\,\mathbf{1} + 2\hat{Y}\, diag(\boldsymbol{F})'\boldsymbol{I}[j]$, (3.4) can be written:

$$
var(\hat{Y}_{\widehat{PV}}) \cong var(\hat{Y}_{GREG}) + var(E), \tag{3.5}
$$

where $var(E)$ is the exogenous variance, or equivalently the adding error we introduce when considering sampling-known constraints.

4. SELECTION OF AUXILIARY VARIABLES IN CAL ESTIMATOR

The choice of variables to insert into the constraint system of CAL estimator is a relevant problem. Besides sampling error reduction, in official statistics, we also need to have a consistent estimate system, that is to have the same estimate for the same aggregate provided by simultaneous surveys.

As already said, our aim is to build a new tool, based on sampling error, computationally efficient and able to give mathematical support to the decision of adding further auxiliary variables in the constraint system of CAL estimator.

Moreover, our tool should provide, with a good approximation, a measure of the expected gain of efficiency when considering further auxiliary variables.

It is known that from the HT estimator to $GREG$ (equivalently to CAL), using auxiliary variables, we have a reduction of sampling error proportional to the correlation among auxiliary variables and interest variable (Bethlehem & Keller, 1987, 146).

However, our main interest consists in studying how sampling error changes when M (≥ 1) auxiliary variables are added to a CAL estimator already including J (≥ 1) auxiliary variables. Hereafter we denote CAL^* the calibrated estimator with $J + M$ auxiliary variables.

Looking at (2.2), we can see that this is a function of residuals of regression model based on variables used as constraints. Intuitively, we may think that introducing more variables in the constraints system, and in the joint regression model, produces a sampling error reduction due to model fitting improvement.

Théberge (1999, 2000) presents a generalization of CAL estimator. He shows a method able to tell us under what conditions the constrained minimum problem admits solution. He also determines, in non-iterative way, a weight system equivalent to that of the CAL estimator.

Our proposal is to construct a quicker tool, useful especially in the planning stage of surveys, that permits to choose constraints for the estimator. This tool does not depend by the computation of final weights, that can take long time, especially for large size samples and complex system of constraints as in the Italian Labour Force Survey. Instead, it should depend only by the fitting of the model based on the considered auxiliary variables.

The proposed tool consists in a function that relates the fitting of two models, Model 1 (M_1) and Model 2 (M_2). The M_1 is the model on which CAL estimator is based, whilst M_2 is the model including only the additional constraints that we want include and the Y variable. The output of this function is the reduction of sampling error that we can expect considering the additional auxiliary variables.

This reduction is evaluated by $\alpha = \frac{CV(CAL^*)}{CV(CAL)}$, that is the ratio between the variation coefficient of CAL estimator and that of CAL^* estimator. We can see α as the constraints effect.

4.1. A SIMULATION STUDY

To determine this function we employed a simulation study. Population of 5500 units has been first generated. The variable of interest Y has been simulated from a binomial distribution and ten auxiliary variables X_{00}, $X_{01}, \dots, X_{10}$ have been generated from a binomial distribution with different correlation with the variable Y. In this way we have the known totals we will use as constraints in CAL estimator.

As a measure of correlation between the interest variable and auxiliary variables the coefficient $R = |r_{Y,X}|$ has been used. It can be considered as an index of a regression model fitting. We denote it in their subscript, for instance X_{03} is an auxiliary variable that has correlation with Y, measured with R, equal to 0.3.

Since the sampling error may also depend on the size of aggregate to estimate, we have simulated several Y (and relatives $X_{00}, X_{01}, \dots, X_{10}$) for several proportions of population units that present the given feature of Y. From now on, it will be called "incidence" and denoted by $p(Y)$.

With a simple random sampling without replacement, 600 units are selected from our population. On this sample we estimated, by CAL estimator, the total of the variable Y and its sampling error, through all possible pairs of X. We repeated this procedure for 1000 samples (see Appendix, Simulation results).

In Figure 1, (a) and (b) show two extreme scenarios: simulation with Y with incidence in the population of 40% and with incidence of 10%.

(a) $p(Y) = 0.40$ (b) $p(Y) = 0.10$

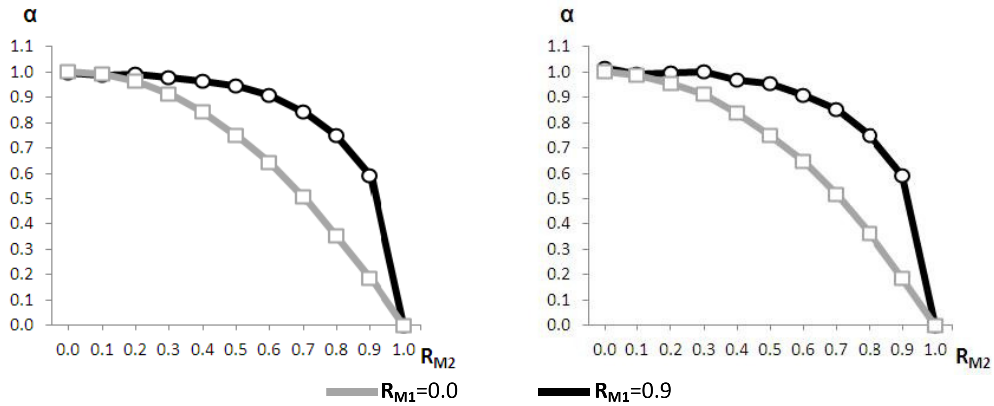

Figure 1 Reduction of variation coefficient (α) when we use CAL estimator based on a model with R_{M1} equal to 0.0 or 0.9 and we add variables that form a model with several R_{M2} (from 0.0 to 1.0) for Y incidence in the population of 40% and with incidence of 10%.

To make the figure more readable we have drawn only the extreme cases, where R_{M1} is equal to 0.0 and 0.9, respectively. Indeed the other cases ($0.1 \leq R_{M1} \leq 0.8$) draw curves nested within those represented.

In a saturated model with a high R_{M1}, the introduction of constraints, even if with high correlation with interest variable, has less effect on the gain of efficiency of our estimate. Hence, the relative curve is less steep.

The cases where, in our estimate, α can be larger than one are more common for low incidence of Y and when M_1 does already have a high fitting.

Grouping together the results for R_{M1} we see that for several incidence levels of Y there are unimportant variation of α. Therefore, we have decided to summarize the several curves through the median.

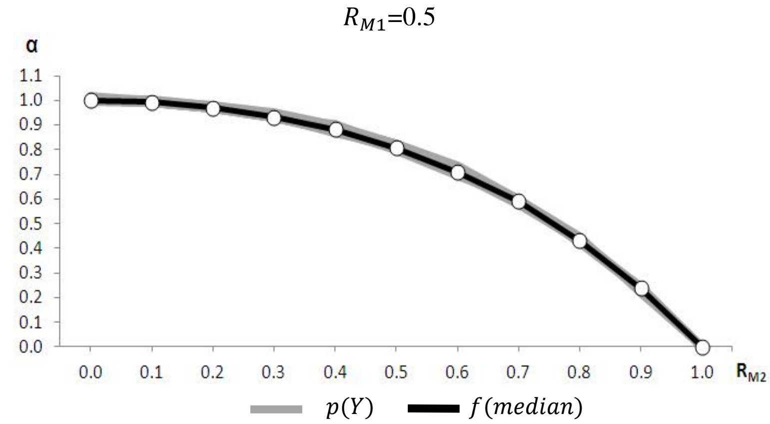

Figure 2 Reduction of variation coefficient (α) when we use CAL estimator based on a model with R_{M1}=0.5 and we add variables that form a model with several R_{M2} (from 0.0 to 1.0) for the different level of $p(Y)$ (incidence of Y on population) and median of α.

In Figure 2 the grey "wake" represents the α trend for the CAL estimator with several constraints that set up a model with Y with R_{M1} equal to 0.5 for each incidence level of Y and for each R_{M2} change. Whilst the black curve is the interpolation curve of the median of α values for each incidence level of Y.

Removing the incidence effect on the sampling error, that has an unimportant influence variation of α, leaves the α values for R_{M1} and R_{M2}, reported in Table 1.

Table 1
Reduction of variation coefficient percent (α) when we use CAL estimator based on a model with R_{M1} and we add variables that form a model with R_{M2}.

α		R_{M2}										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
R_{M1}	0.0	0.999	0.989	0.959	0.909	0.839	0.749	0.639	0.513	0.360	0.190	0.000
	0.1	1.000	0.990	0.961	0.910	0.841	0.751	0.643	0.514	0.364	0.191	0.000
	0.2	1.000	0.991	0.961	0.915	0.847	0.757	0.651	0.523	0.370	0.195	0.000
	0.3	1.001	0.991	0.964	0.917	0.855	0.768	0.662	0.537	0.383	0.206	0.000
	0.4	1.000	0.992	0.966	0.925	0.865	0.787	0.682	0.556	0.403	0.216	0.000
	0.5	1.001	0.993	0.970	0.931	0.883	0.807	0.709	0.591	0.432	0.238	0.000
	0.6	0.999	0.992	0.975	0.940	0.895	0.831	0.745	0.623	0.473	0.267	0.000
	0.7	0.999	0.992	0.980	0.952	0.913	0.860	0.783	0.682	0.540	0.316	0.000
	0.8	1.000	0.997	0.987	0.967	0.943	0.899	0.839	0.761	0.626	0.399	0.000
	0.9	0.997	0.994	0.989	0.981	0.963	0.942	0.907	0.850	0.757	0.565	0.000

An important result in Table 1 is that, even when we introduce variable with low correlation with Y, we obtain a reduction, albeit small. This completely agrees with the remarks by Deville and Särndal (1992, p. 367). Adding constraints, we set up a weight system that performs better for the Y variable.

Using the model $\frac{a+bx}{1+cx+dx^2}$, α can be written as a function of R_{M1} and R_{M2}:

$$\alpha = \frac{1.004 - 1.003\, R_{M2}}{1 - (0.126\, R_{M1} + 0.849)R_{M2} - (0.426\, R_{M1}^2 + 0.014 R_{M1} - 0.403)R_{M2}^2}, \qquad (4.1)$$

(see Appendix, Interpolation of α values in Table 1).

By (4.1) we obtain the reduction of variation coefficient if we move from CAL estimator with J constraints to CAL^* with $J + M$. As already said, this can be considered the constraints effect.

The expected coefficient of variation for CAL^* ($\widetilde{CV}^*$), so based on $J + M$ constraints, is simply:

$$\widetilde{CV}^* = CV\,\alpha. \qquad (4.2)$$

5. APPLICATION ON REAL DATA

In the application on real data, we aim at computing the real sampling error for the estimates of the Italian Labour Force Survey (It-LFS), and check if our model, constructed in Section 4, is useful to predict the reduction of sampling error due to the addition of further constraints in the CAL estimator.

The It-LFS is a continuous survey carrying out all the weeks of the year. It represents the official source of the labour market in Italy (for more information see Istat, 2006).

The sampling design for each quarterly sample is complex. It has two-stages (municipalities and families) with stratification of the primary units on the basis of location and demographic size. Moreover, the sampling design provides a rotation scheme of sampling $(2_Q, 2_Q, 2_Q)$. That is a rotation group remains in the sample for two consecutive quarters, then leaves the sample for two quarters, and then it re-enters the sample for another two consecutive quarters. It is then dropped from the sample completely. The Italian National Institute of Statistics (Istat), in agreement with Eurostat, organizes also a time sampling design to provide a monthly estimate and to guarantee an accurate representation of monthly samples.

However, monthly samples have one-third of the size of quarterly samples. Therefore, Istat, following the experience of the Canadian Labour Force Survey, adopted longitudinal constraints to provide less volatile estimates through monthly samples steadier and more similar in structure.

Because of the rotation scheme, the overlapping sample, between a quarter and four quarters (one year), should be about 50%, but it is really less than 30% for both, due to nonresponse (especially for attrition). So, the impact of imputation for high fraction of previous period missing value for the present month's respondent can be serious (Singh, Kennedy & Wu, 2001, p. 34).

The calibration strategy for $It\text{-}LFS$ consists in 302 constraints, 206 exactly-known by administrative sources (sex, age and location) and 96 sampling-known, such as estimate of work status (employed, unemployed and inactive) three months before and twelve months (one year) before. These are called longitudinal constraints. Therefore, for this particular sampling design we have OS constraints.

We apply (3.4) and our model to the sample of September 2009, with 20,928 families and 49,114 units. The sampling constraints are the estimates of work status three and twelve months before (June 2009 and September 2008).

We start from CAL estimator that uses only the 206 administrative constraints. The estimates and their relative error for several interest domains are in Table 2.

Table 2
Estimates of Employed, Unemployed and Inactive and their sampling error for Italy and for Males and Females with CAL estimator. $It\text{-}LFS$, September 2009.

	Estimate [1]	Variance [2]	CV% [3]
ITALY			
Employed	22,786,251	10,045,544,438	0.440
Unemployed	2,021,889	2,930,847,610	2.678
Inactive	34,982,481	9,287,085,077	0.275
MALES			
Employed	13,599,617	4,467,486,520	0.491
Unemployed	1,092,265	1,468,480,200	3.508
Inactive	14,371,974	3,914,412,178	0.435
FEMALES			
Employed	9,186,634	5,785,715,715	0.828
Unemployed	929,624	1,332,023,546	3.926
Inactive	20,610,507	5,518,925,815	0.360

Now, before compute CAL estimator including longitudinal constraints, it is of interest to evaluate how much gain in efficiency we may expect.

Using the model in the previous section, and mainly (4.1), we want predict the improvement in efficiency. We need the CV of estimates achieved with only the administrative constraints, $X_1 - X_{206}$ (Table 2[3]) and the fitting coefficients for each parameter and each domain we aim to estimating. Therefore, the first model (M_1) is $Y = f(X_1 - X_{206})$, namely the model based only on the 206 constraints with exactly-known totals from administrative sources. The second model (M_2) does have the form $Y = f(X_{207} - X_{302})$. It only considers the constraints from sampling sources that we would add.

Results in Table 3 show the reduction of the sampling error we can expect adding these 96 constraints to our estimator.

The percent variation coefficients ($\widetilde{CV}\%^{*}$) in Table 3 are very close to those actually computed by considering constraints as known from administrative sources. To compute values of $\widetilde{CV}\%^{*}$ we have avoid to use expression (2.2), whilst used for $CV\%^{*}$ in Table 4[3], which would require the computation of all sampling weights.

Table 3
Results of application of our model to $It\text{-}LFS$ of September 2009.

	ITALY	MALES	FEMALES
Employed			
R_{M1}	0.708	0.758	0.626
R_{M2}	0.637	0.612	0.654
α	0.766	0.814	0.708
$\widetilde{CV}\%^{*}$	0.337	0.400	0.586
Unemployed			
R_{M1}	0.224	0.232	0.217
R_{M2}	0.345	0.366	0.320
α	0.888	0.875	0.902
$\widetilde{CV}\%^{*}$	2.377	3.069	3.542
Inactive			
R_{M1}	0.738	0.791	0.657
R_{M2}	0.618	0.594	0.633
α	0.789	0.846	0.745
$\widetilde{CV}\%^{*}$	0.220	0.368	0.268

The $\widetilde{CV}\%^{*}$ drawn with our model differ by 5% at most from $CV\%^{*}$ value actually computed (Table 4[3]). The difference is greater than 5% only in the case of unemployed females (8.7%).

Table 4 shows the actual evaluation of sampling error using CAL estimator with OS constraints. We relate the variance and the percent variation coefficient, both computed by not considering exogenous variance (Table 4[2] and 4[3]), and including the exogenous variance (Table 4[4] and 4[5]). When we say that we not considering exogenous variance, it means that the additional 96 constraints are considered as known from administrative sources.

The introduction of longitudinal constraints leads to stable estimates guaranteeing the same work status of three months before and of one year before to all those who, meanwhile, in fact do not change their condition.

Table 4
Estimate of Employed, Unemployed and Inactive and their sampling error for Italy and for Males and Females with CAL and $\widehat{CAL}^*$ estimators. It-LFS, September 2009.

		without $var(E)$		with $var(E)$	
		$\hat{Y}_{CAL^*}$		$\hat{Y}_{\widehat{CAL}^*}$	
	Estimate [1]	Variance [2]	$CV\%^*$ [3]	Variance [4]	$CV\%^*$ [5]
		ITALY			
Employed	22,886,373	5,729,943,806	0.331	5,798,793,460	0.333
Unemployed	2,031,044	2,107,238,156	2.260	2,111,420,352	2.262
Inactive	34,873,584	5,694,253,968	0.216	5,913,857,533	0.221
		MALES			
Employed	13,647,567	2,753,880,461	0.385	2,793,283,490	0.387
Unemployed	1,093,438	1,144,596,083	3.094	1,146,978,094	3.097
Inactive	14,323,053	2,385,261,041	0.341	2,448,691,593	0.345
		FEMALES			
Employed	9,238,806	3,173,204,562	0.610	3,201,923,547	0.612
Unemployed	937,606	933,831,700	3.259	935,497,533	3.262
Inactive	20,550,531	3,187,456,800	0.275	3,338,568,872	0.281

Small variations in estimates occur. For all the domains the number of inactive units decreases, and there is a redistribution of these units towards labour force (employed and unemployed). However, we have an improvement in efficiency of estimates, as stressed by the decrease of percent relative error (Table 2[3] vs. 4[5]).

In every domain, the gain of efficiency is larger for unemployed. It reaches a maximum of 0.7% for females.

This efficiency gain is due to better specification of the regression model with the adding of longitudinal sampling constraints. This widely offsets the increasing of error due to the sampling source of constraints, which is never bigger than 0.006% (Table 4[3] vs. 4[5]).

6. CONCLUSION AND REMARKS

This work suggests two strongly innovative elements, mainly for the use of calibrated estimators for surveys with complex design and in complex contests such as those of the National Institute of Statistics.

Nowadays the improvement of estimates efficiency and consistency with other sampling surveys are key elements in the production of official statistics. But, another equally important aim is to provide estimates for unplanned domains for which we have not information from administrative sources.

However, besides this, there is the need to evaluate the amount of error that we introduce using auxiliary variables from sampling sources and how, on the whole, the procedure can be more or less statistically profitable.

This work aimed to meet these demands with a novel contribute. First, it provides an analytical quantification of sampling error of a total estimate through calibrated estimator with sampling constraints. Moreover, it introduces a strong simplification of the computation of the relative error with several combination of constraints, that can be used simply and quickly in design phase of an estimator. In addition it help us to decide which scenarios to adopt, in terms of variance.

The choice of auxiliary variables, obviously, is not conditioned only by the decrease of relative error, but it is a decision that has to take account of other considerations like, for instance, the opportunity to produce more accurate estimates consistent with unplanned domains.

Therefore having a tool, like the proposal model, strongly simplifies the operations leading to the choice of constraint system of the calibrate estimator.

Appendix

Proof 1.

The $GREG$ estimator with sample constraints, from Ballin, Falorsi and Russo (2000, p. 48) can be written as $\hat{Y}_{\widehat{GREG}} = \hat{Y}_{HT} + \boldsymbol{B}'(\hat{\hat{\boldsymbol{X}}} - \hat{\boldsymbol{X}}_{HT})$. Adding and deducting the vector of real values of auxiliary variables totals, $\boldsymbol{X}$, we have $\hat{Y}_{\widehat{GREG}} = \hat{Y}_{HT} + \boldsymbol{B}'\left[\left(\boldsymbol{X} - \hat{\boldsymbol{X}}_{HT}\right) + \left(\hat{\hat{\boldsymbol{X}}} - \boldsymbol{X}\right)\right]$. This expression can be decomposed in two elements $A_1 = \hat{Y}_{HT} + \boldsymbol{B}'(\boldsymbol{X} - \hat{\boldsymbol{X}}_{HT})$ that is the expression of the $GREG$ estimator (Fuller, 2002, p. 6) and $A_2 = \boldsymbol{B}'\left(\hat{\hat{\boldsymbol{X}}} - \boldsymbol{X}\right)$. If $\hat{\hat{\boldsymbol{X}}}$ are total estimates of NOS constraints, A_1 and A_2 are independent and $cov(A_1, A_2) = 0$. When $\hat{\hat{\boldsymbol{X}}}$ are total estimates of OS constraints, A_1 and A_2 are dependent. By denoting $Cov(A_1, A_2) = E[A_1 A_2] - E[A_1]E[A_2]$, because A_1 and A_2 are dependent, $E[A_1 A_2]$ is equal to $E[A_1] + E[A_2] - E[A_1]E[A_2]$, so $Cov(A_1, A_2) = E[A_1] + E[A_2] - 2E[A_1]E[A_2]$. A_1 is asymptotically unbiased so, for large size population, his expected value is Y, moreover considering that $\hat{\hat{\boldsymbol{X}}}$ is an unbiased estimate for $\boldsymbol{X}$ and $\boldsymbol{B}$ is constant, $Cov(A_1, A_2) = 2Y$. Therefore, an unbiased estimate is $cov(A_1, A_2) = 2\hat{Y}$.

Simulation results.

$p(Y)$	R_{M1}	R_{M2} 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.90	0.0	1.018	1.002	0.968	0.964	0.835	0.742	0.633	0.507	0.358	0.194	0.000
0.90	0.1	1.004	0.985	0.965	0.943	0.844	0.742	0.624	0.506	0.351	0.179	0.000
0.90	0.2	0.987	0.995	0.947	0.942	0.825	0.741	0.635	0.516	0.372	0.192	0.000
0.90	0.3	0.984	0.960	0.946	0.908	0.825	0.746	0.643	0.506	0.353	0.195	0.000
0.90	0.4	1.002	0.979	0.970	0.975	0.866	0.762	0.685	0.563	0.391	0.209	0.000
0.90	0.5	1.008	1.005	0.971	0.968	0.880	0.770	0.715	0.594	0.436	0.249	0.000
0.90	0.6	0.984	0.997	0.978	0.967	0.919	0.817	0.731	0.597	0.463	0.265	0.000
0.90	0.7	1.017	1.008	0.971	0.963	0.893	0.861	0.722	0.693	0.524	0.335	0.000
0.90	0.8	1.057	1.061	1.051	1.006	1.004	0.908	0.866	0.766	0.666	0.410	0.000
0.90	0.9	1.067	1.003	1.080	1.036	0.992	1.035	0.896	0.889	0.782	0.589	0.000
0.80	0.0	1.003	1.003	0.955	0.909	0.847	0.761	0.639	0.507	0.350	0.184	0.000
0.80	0.1	1.001	0.981	0.957	0.921	0.842	0.750	0.650	0.516	0.371	0.191	0.000
0.80	0.2	1.000	0.997	0.960	0.907	0.837	0.745	0.655	0.518	0.366	0.194	0.000
0.80	0.3	0.991	0.989	0.976	0.915	0.854	0.754	0.650	0.518	0.380	0.201	0.000
0.80	0.4	0.991	0.980	0.947	0.911	0.842	0.753	0.669	0.539	0.391	0.205	0.000
0.80	0.5	0.991	0.992	0.968	0.939	0.877	0.814	0.700	0.579	0.436	0.229	0.000
0.80	0.6	0.990	0.984	0.960	0.908	0.860	0.820	0.733	0.601	0.452	0.248	0.000
0.80	0.7	1.000	1.003	0.978	0.953	0.932	0.869	0.785	0.681	0.518	0.319	0.000
0.80	0.8	0.986	0.956	0.962	0.954	0.901	0.909	0.806	0.725	0.639	0.385	0.000
0.80	0.9	1.066	1.015	1.026	1.065	0.997	0.983	0.937	0.880	0.778	0.606	0.000
0.75	0.0	1.006	0.992	0.971	0.920	0.835	0.752	0.639	0.508	0.373	0.195	0.000
0.75	0.1	1.003	0.987	0.962	0.910	0.838	0.756	0.644	0.511	0.361	0.191	0.000
0.75	0.2	0.999	0.984	0.966	0.905	0.841	0.759	0.649	0.519	0.367	0.197	0.000
0.75	0.3	1.004	0.989	0.965	0.907	0.837	0.771	0.641	0.544	0.387	0.205	0.000
0.75	0.4	1.004	0.999	0.969	0.927	0.874	0.773	0.705	0.562	0.408	0.220	0.000
0.75	0.5	0.988	0.982	0.956	0.901	0.858	0.767	0.685	0.600	0.423	0.236	0.000
0.75	0.6	0.993	0.991	0.976	0.929	0.913	0.818	0.754	0.629	0.466	0.270	0.000
0.75	0.7	1.014	0.999	0.986	0.968	0.932	0.903	0.793	0.676	0.529	0.310	0.000
0.75	0.8	1.007	0.997	0.989	0.968	0.950	0.912	0.833	0.762	0.652	0.411	0.000
0.75	0.9	0.997	1.000	1.000	0.984	0.974	0.974	0.896	0.824	0.757	0.579	0.000
0.70	0.0	0.996	0.991	0.952	0.904	0.831	0.754	0.638	0.514	0.367	0.195	0.000
0.70	0.1	1.005	0.997	0.962	0.907	0.851	0.754	0.652	0.518	0.365	0.200	0.000
0.70	0.2	0.986	0.981	0.951	0.904	0.839	0.758	0.646	0.522	0.364	0.197	0.000
0.70	0.3	1.007	0.991	0.972	0.923	0.856	0.769	0.676	0.540	0.391	0.205	0.000
0.70	0.4	1.005	1.000	0.994	0.929	0.881	0.815	0.690	0.559	0.407	0.222	0.000
0.70	0.5	0.995	0.984	0.973	0.929	0.893	0.805	0.700	0.582	0.433	0.232	0.000
0.70	0.6	0.994	0.991	0.965	0.942	0.893	0.826	0.735	0.620	0.469	0.266	0.000

$p(Y)$	R_{M1}	R_{M2}										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.70	0.7	1.001	0.998	0.971	0.959	0.894	0.858	0.792	0.669	0.523	0.325	0.000
0.70	0.8	1.006	1.009	0.977	0.936	0.941	0.907	0.831	0.731	0.624	0.395	0.000
0.70	0.9	0.988	1.007	1.042	1.022	1.000	0.939	0.937	0.879	0.806	0.575	0.000
0.60	0.0	0.998	0.993	0.960	0.910	0.844	0.753	0.633	0.508	0.364	0.194	0.000
0.60	0.1	0.997	0.992	0.958	0.909	0.840	0.751	0.633	0.516	0.365	0.194	0.000
0.60	0.2	1.000	0.988	0.956	0.909	0.849	0.756	0.651	0.524	0.374	0.197	0.000
0.60	0.3	0.990	0.988	0.962	0.909	0.855	0.763	0.649	0.516	0.386	0.214	0.000
0.60	0.4	1.003	0.983	0.975	0.924	0.862	0.789	0.680	0.541	0.406	0.222	0.000
0.60	0.5	0.999	0.999	0.962	0.934	0.883	0.803	0.696	0.559	0.431	0.236	0.000
0.60	0.6	1.003	0.995	0.977	0.938	0.890	0.822	0.733	0.621	0.479	0.273	0.000
0.60	0.7	0.982	0.991	0.978	0.928	0.904	0.827	0.777	0.680	0.537	0.323	0.000
0.60	0.8	1.005	1.007	0.993	0.982	0.930	0.920	0.847	0.783	0.629	0.406	0.000
0.60	0.9	0.978	0.997	1.001	0.962	0.933	0.943	0.915	0.849	0.744	0.549	0.000
0.50	0.0	0.989	0.977	0.943	0.894	0.826	0.750	0.664	0.461	0.355	0.192	0.000
0.50	0.1	0.998	0.992	0.948	0.898	0.808	0.748	0.663	0.462	0.357	0.190	0.000
0.50	0.2	0.998	0.993	0.955	0.899	0.818	0.759	0.673	0.478	0.364	0.201	0.000
0.50	0.3	0.998	0.994	0.952	0.910	0.824	0.775	0.679	0.481	0.382	0.203	0.000
0.50	0.4	0.999	0.993	0.960	0.912	0.854	0.789	0.702	0.513	0.405	0.226	0.000
0.50	0.5	0.995	0.993	0.964	0.919	0.852	0.802	0.728	0.529	0.427	0.243	0.000
0.50	0.6	1.000	0.999	0.968	0.930	0.869	0.824	0.766	0.564	0.460	0.266	0.000
0.50	0.7	1.001	0.993	0.969	0.951	0.906	0.867	0.807	0.714	0.558	0.334	0.000
0.50	0.8	1.023	0.998	1.017	0.975	0.945	0.903	0.866	0.719	0.630	0.419	0.000
0.50	0.9	1.028	1.009	1.002	1.021	0.990	0.947	0.932	0.849	0.796	0.550	0.000
0.40	0.0	0.999	0.989	0.957	0.911	0.836	0.740	0.633	0.508	0.377	0.200	0.000
0.40	0.1	0.998	0.985	0.957	0.911	0.832	0.741	0.637	0.519	0.376	0.200	0.000
0.40	0.2	1.002	0.990	0.959	0.913	0.846	0.751	0.644	0.526	0.384	0.207	0.000
0.40	0.3	1.001	0.992	0.960	0.912	0.844	0.763	0.653	0.532	0.398	0.211	0.000
0.40	0.4	1.004	0.992	0.966	0.917	0.862	0.777	0.673	0.554	0.413	0.232	0.000
0.40	0.5	0.958	0.982	0.961	0.929	0.879	0.806	0.704	0.587	0.440	0.255	0.000
0.40	0.6	0.997	1.004	0.972	0.942	0.888	0.835	0.754	0.612	0.498	0.290	0.000
0.40	0.7	0.995	1.000	0.978	0.953	0.857	0.761	0.760	0.661	0.530	0.330	0.000
0.40	0.8	0.985	0.989	0.981	0.958	0.915	0.880	0.827	0.715	0.607	0.401	0.000
0.40	0.9	0.990	0.960	0.983	0.970	0.952	0.907	0.882	0.846	0.760	0.604	0.000
0.30	0.0	1.001	0.990	0.953	0.906	0.832	0.743	0.640	0.496	0.351	0.178	0.000
0.30	0.1	1.027	1.017	0.982	0.931	0.857	0.767	0.663	0.508	0.357	0.181	0.000
0.30	0.2	0.999	0.994	0.961	0.909	0.840	0.754	0.652	0.515	0.370	0.180	0.000
0.30	0.3	1.002	1.000	0.965	0.917	0.846	0.763	0.668	0.525	0.376	0.190	0.000
0.30	0.4	0.994	0.988	0.956	0.916	0.853	0.779	0.676	0.539	0.394	0.202	0.000
0.30	0.5	1.005	1.000	0.975	0.928	0.880	0.808	0.711	0.577	0.426	0.219	0.000

$p(Y)$	R_{M1}	R_{M2} 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.30	0.6	1.001	0.996	0.972	0.943	0.889	0.826	0.733	0.616	0.458	0.254	0.000
0.30	0.7	1.001	0.990	0.966	0.958	0.912	0.850	0.786	0.666	0.533	0.307	0.000
0.30	0.8	1.009	1.012	0.990	0.981	0.933	0.916	0.847	0.750	0.662	0.376	0.000
0.30	0.9	0.983	0.969	0.981	0.974	0.946	0.923	0.892	0.861	0.744	0.583	0.000
0.25	0.0	1.001	1.000	0.962	0.905	0.842	0.753	0.634	0.495	0.360	0.196	0.000
0.25	0.1	0.988	0.986	0.957	0.901	0.835	0.752	0.637	0.491	0.367	0.190	0.000
0.25	0.2	0.997	0.994	0.958	0.910	0.849	0.780	0.639	0.511	0.377	0.194	0.000
0.25	0.3	0.994	0.983	0.954	0.917	0.842	0.767	0.643	0.524	0.389	0.202	0.000
0.25	0.4	1.021	0.995	0.965	0.917	0.858	0.778	0.661	0.541	0.405	0.219	0.000
0.25	0.5	1.004	0.994	0.966	0.944	0.875	0.806	0.700	0.572	0.431	0.232	0.000
0.25	0.6	0.997	0.993	0.968	0.934	0.880	0.831	0.742	0.596	0.484	0.283	0.000
0.25	0.7	1.014	0.985	0.969	0.954	0.916	0.863	0.763	0.698	0.533	0.331	0.000
0.25	0.8	1.010	1.018	0.972	0.962	0.929	0.900	0.826	0.727	0.603	0.406	0.000
0.25	0.9	0.986	1.041	0.996	0.959	0.972	0.929	0.915	0.836	0.761	0.561	0.000
0.20	0.0	1.002	0.974	0.962	0.898	0.836	0.754	0.644	0.504	0.350	0.178	0.000
0.20	0.1	0.996	0.985	0.952	0.899	0.836	0.744	0.637	0.499	0.344	0.180	0.000
0.20	0.2	0.990	0.975	0.971	0.909	0.841	0.762	0.641	0.519	0.357	0.187	0.000
0.20	0.3	1.004	0.996	0.968	0.931	0.857	0.775	0.661	0.532	0.372	0.193	0.000
0.20	0.4	0.995	0.999	0.970	0.926	0.851	0.782	0.672	0.540	0.401	0.212	0.000
0.20	0.5	0.999	0.993	0.972	0.947	0.868	0.807	0.709	0.578	0.417	0.231	0.000
0.20	0.6	1.003	0.996	0.969	0.946	0.885	0.845	0.748	0.618	0.468	0.264	0.000
0.20	0.7	1.002	0.999	0.989	0.950	0.905	0.862	0.784	0.677	0.518	0.307	0.000
0.20	0.8	1.000	0.996	0.975	0.957	0.923	0.899	0.833	0.735	0.620	0.406	0.000
0.20	0.9	0.976	0.936	0.942	0.935	0.920	0.910	0.856	0.842	0.725	0.575	0.000
0.10	0.0	1.023	1.011	0.956	0.928	0.860	0.756	0.658	0.514	0.386	0.184	0.000
0.10	0.1	1.049	1.040	0.990	0.950	0.892	0.782	0.679	0.537	0.394	0.192	0.000
0.10	0.2	1.019	0.990	0.979	0.921	0.861	0.770	0.658	0.525	0.386	0.193	0.000
0.10	0.3	1.030	1.008	0.969	0.954	0.865	0.783	0.683	0.544	0.408	0.192	0.000
0.10	0.4	0.996	0.989	0.951	0.917	0.856	0.777	0.677	0.557	0.419	0.211	0.000
0.10	0.5	0.997	1.000	0.975	0.939	0.881	0.805	0.718	0.596	0.458	0.225	0.000
0.10	0.6	0.995	0.987	0.974	0.958	0.894	0.821	0.746	0.623	0.486	0.257	0.000
0.10	0.7	0.989	0.988	0.972	0.938	0.920	0.855	0.787	0.664	0.550	0.310	0.000
0.10	0.8	0.920	0.931	0.910	0.895	0.873	0.833	0.711	0.683	0.586	0.357	0.000
0.10	0.9	0.946	1.005	0.990	0.998	1.047	0.906	0.893	0.862	0.810	0.601	0.000
0.05	0.0	0.984	0.990	0.972	0.887	0.828	0.741	0.631	0.504	0.351	0.189	0.000
0.05	0.1	0.986	0.967	0.943	0.903	0.844	0.740	0.638	0.514	0.350	0.176	0.000
0.05	0.2	1.009	0.962	0.973	0.928	0.842	0.774	0.662	0.526	0.367	0.191	0.000
0.05	0.3	0.994	0.993	0.941	0.919	0.860	0.753	0.645	0.523	0.378	0.196	0.000
0.05	0.4	0.987	0.989	0.953	0.917	0.849	0.789	0.692	0.556	0.391	0.196	0.000

$p(Y)$	R_{M1}	R_{M2}										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	0.5	0.998	1.014	0.961	0.954	0.884	0.809	0.723	0.579	0.447	0.212	0.000
0.05	0.6	0.987	1.011	0.988	0.947	0.921	0.849	0.737	0.479	0.469	0.248	0.000
0.05	0.7	0.997	0.978	0.992	0.935	0.929	0.863	0.789	0.702	0.535	0.304	0.000
0.05	0.8	0.931	0.976	0.989	0.897	0.919	0.863	0.844	0.739	0.611	0.383	0.000
0.05	0.9	0.863	0.914	0.889	0.961	0.869	0.916	0.839	0.773	0.733	0.576	0.000
0.03	0.0	0.985	0.965	0.985	0.925	0.827	0.750	0.658	0.513	0.351	0.189	0.000
0.03	0.1	1.007	0.974	1.003	0.891	0.842	0.851	0.651	0.532	0.352	0.198	0.000
0.03	0.2	1.035	1.013	0.962	0.954	0.859	0.795	0.487	0.536	0.364	0.191	0.000
0.03	0.3	1.025	0.982	0.964	0.907	0.854	0.762	0.696	0.536	0.379	0.175	0.000
0.03	0.4	0.967	0.959	0.936	0.916	0.860	0.762	0.664	0.553	0.371	0.213	0.000
0.03	0.5	1.043	0.995	0.985	0.934	0.899	0.822	0.726	0.623	0.426	0.229	0.000
0.03	0.6	0.998	0.990	0.941	0.968	0.942	0.836	0.784	0.632	0.475	0.250	0.000
0.03	0.7	1.022	1.040	1.031	1.004	0.937	0.923	0.841	0.757	0.558	0.319	0.000
0.03	0.8	0.991	0.978	0.983	0.980	0.960	0.910	0.861	0.760	0.659	0.455	0.000
0.03	0.9	1.064	1.014	0.985	1.041	1.038	0.962	0.967	0.923	0.780	0.662	0.000
0.02	0.0	1.095	0.977	1.141	0.976	0.866	0.877	0.646	0.549	0.403	0.163	0.000
0.02	0.1	0.994	0.935	0.963	0.886	0.850	0.741	0.639	0.490	0.330	0.173	0.000
0.02	0.2	1.008	0.967	1.000	0.904	0.843	0.751	0.629	0.534	0.364	0.167	0.000
0.02	0.3	0.961	0.994	0.974	0.910	0.847	0.766	0.645	0.539	0.405	0.189	0.000
0.02	0.4	0.989	0.968	1.008	0.949	0.872	0.788	0.673	0.560	0.399	0.203	0.000
0.02	0.5	1.015	0.975	0.969	0.939	0.868	0.802	0.723	0.585	0.433	0.232	0.000
0.02	0.6	0.941	0.972	0.906	0.913	0.873	0.824	0.717	0.646	0.510	0.251	0.000
0.02	0.7	1.017	0.983	0.989	0.980	0.948	0.868	0.824	0.697	0.578	0.329	0.000
0.02	0.8	1.962	1.944	1.924	1.912	1.950	1.841	1.744	1.565	1.341	0.889	0.000
0.02	0.9	1.105	1.134	1.079	1.036	0.994	0.985	1.063	1.050	0.930	0.692	0.000

Interpolation of α values in Table 1.

To interpolate values in Table 1 we decide to use the following model function of R_{M2}:

$$\alpha = \frac{\mathrm{a} + b\, R_{M2}}{1 + c\, R_{M2} + d\, R_{M2}^2}.$$

The values of parameters in the model for several level of R_{M1} are:

Table 5
Parameters for several level of $\mathbf{R_{M1}}$.

R_{M1}	a	b	c	d	r^2
0.0	1.00541	-1.00377	-0.86171	0.39850	0.99987
0.1	1.00633	-1.00484	-0.86227	0.39304	0.99987
0.2	1.00605	-1.00472	-0.87274	0.38901	0.99989
0.3	1.00610	-1.00484	-0.87386	0.35816	0.99990
0.4	1.00408	-1.00336	-0.89716	0.34510	0.99994
0.5	1.00403	-1.00367	-0.90917	0.30187	0.99994
0.6	1.00188	-1.00160	-0.92486	0.25946	0.99996
0.7	1.00157	-1.00165	-0.92966	0.11138	0.99991
0.8	1.00208	-1.00217	-0.95799	0.13771	0.99992
0.9	0.99889	-0.99888	-0.96895	0.06070	0.99998

The values in the Table above can be written in function of R_{M1} like:

Table 6
Parameters of model as function of $\mathbf{R_{M1}}$.

Parameters	R_{M1}^2	R_{M1}	intercept	r^2
a		-0.00749	1.00701	0.84161
b		0.00530	-1.00534	0.71701
c		-0.12636	-0.84898	0.96264
d	-0.42592	-0.01395	0.40316	0.95402

so the model, replacing the value in the Table above, is given by (4.1).

REFERENCES

Ballin, M., Falorsi, P.D., Russo, A. (2000). Condizioni di Coerenza e Metodi di Stima per le Indagini Campionarie Sulle imprese. *Rivista di Statistica Ufficiale*, 2, 31-52.

Bethlehem, J.G., Keller, W.J. (1987). Linear Weighting of Sample Survey Data. *Journal of Official Statistics*, 3, 141-153.

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

Deville, J.C., Särndal, C.E., Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association* , 88, 1013-1020.

Deville, J.C. (1999). Simultaneous Calibration of Several Surveys. *Proceedings of Statistics Canada Symposium*. Combining data from different sources, 207-212.

Estevao, V., Hidiroglou, M.A., Särndal, C.E. (1995). Methodological Principles for a Generalized Estimate System at Statistics Canada. *Journal of Official Statistics*, 11, 181-204.

Fuller, W.A. (2002). Regression Estimator for Survey Samples. *Survey Methodology*, 28, 5-23.

Fuller, W.A., Rao, J.N.K. (2001). A Regression Composite Estimator with Application to the Labour Force Survey. *Survey Methodology*, 27, 45-51.

Istat (2006). La Rilevazione sulle Forze di Lavoro: Contenuti, Metodologie, Organizzazione. *Collana Metodi e norme*, 32.

Lemaître, G., Dufour, J. (1987). An Integrated Method for Weighting Persons and Families. *Survey Methodology*, 13, 199-207.

Särndal, C. E. (2007). The Calibration Approach in Survey Theory and Practice. *Survey Methodology*, 33, 99-199.

Singh, A.C., Mohl, C.A. (1996). Understanding Calibration Estimators in Survey Sampling. *Survey Methodology*, 22, 107-115.

Singh, A.C, Kennedy, B., Wu, S. (2001). Regression Composite Estimate for the Canadian Labour Force Survey with Rotating Panel Design. *Survey Methodology*, 27, 33-44.

Théberge, A. (1999). Extensions of Calibration Estimator in Survey Sampling. *Journal of the American Statistical Association*, 94, 635-644.

Théberge, A. (2000). Calibration and Restricted Weights. *Survey Methodology*, 26, 99-107.

Wolter, K.M. (1979). Composite Estimate in Finite Populations. *Journal of the American Statistical Association*, 74, 604-613.

Wright, R.L. (1983). Finite Population Sampling with Multivariate Auxiliary Information. *Journal of the American Statistical Association*, 78, 879-884.