INCREASING THE ACCURACY OF IT-SILC ESTIMATES THROUGH THE USE OF AUXILIARY INFORMATION FROM LABOUR FORCE SURVEY

Claudio Ceccarelli[1]

Department for Environmental and Social Statistics, National Institute of Statistic, Italy

Alessio Guandalini

Department of Statistical Sciences, University "La Sapienza", Italy

Received: 17 July 2013 / Accepted: 11 February 2014

Abstract. *Since 2004 the European Statistics on Income and Living Condition (Eu-Silc) has been carried out every year in all the European Union member states. It has been harmonised to produce comparable estimates among the countries. The European Regulations suggests to calibrate on control totals of demographic variables from administrative sources. Apart from those, the Italian National Institute of Statistics (Istat) calibrates on estimated control totals from the Italian Labour Force Survey (LFS). This auxiliary information is strongly correlated with income but its sampling nature could have ambiguous effects on the efficiency of estimates. The aim of this paper is to show that using strongly correlated auxiliary population totals, even though survey-estimated, can improve the efficiency of estimates in It-Silc.*

Keywords: *Calibration estimator, Sampling error, Survey-estimated control totals, It-Silc.*

1. INTRODUCTION

Since 2004 the European Statistics on Income and Living Condition (Eu-Silc) has been replacing the European Community Household Panel Survey (ECHP).

The Eu-Silc's aim is equivalent to the ECHP's one, that is to be the reference source for income distribution, living conditions and social exclusion at Eu-

---

[1] Claudio Ceccarelli, clceccar@istat.it
The opinions expressed here are those of the authors and do not necessarily reflect the positions of the Italian National Institute of Statistics (Istat).

ropean Union level. The Commission regulations give technical background to harmonise the survey among the European Union member states and allow comparative statistics.

Commission regulation (EC) 1982/2003, on "sampling and tracing rules", enforces the use of a weighting system that takes into account design weights, nonresponse and auxiliary information on household and individuals in the target population, such as by sex, age (five-year age groups), household size and composition and NUTS II level.

To better face the nonresponse mechanism and the "under reporting" of income, the Italian National Istitute of Statistics (Istat) decided to add further auxiliary information to the weighting system. The additional variables are working status and educational level and their control totals are estimated from the Labour Force Survey (LFS). Information on working status and educational level are strongly correlated with income and can be used to recover the income of particular groups, mainly the one of self-employed individuals.

Auxiliary information is usually incorporated into the estimation procedure with the calibration estimator (see Särndal, 2007) or the regression estimator (see Fuller, 2002). Control totals are generally assumed to be accurately known, that is without sampling error. In the It-Silc case, this holds for demographic variables known by administrative sources, but it does not hold for working status and educational level information that are derived from LFS.

Deville (1999) proposes an optimal regression estimator that uses estimated control totals derived from independent surveys considered to be more precise, for instance, for their sample size. In practice, he suggests to constrain estimates not exactly on survey-estimated control totals. What Deville proposes can be seen as a ridge-regression estimator (Bardsley and Chambers, 1984; Chambers, 1996; Rao and Singh, 1997) or, more generally, as a penalised regression estimator (Goga and Shezad, 2010; Guggemos and Tillé, 2010). In fact, the calibration on auxiliary totals estimated from the external survey is relaxed by a penalisation term depending on the reliability of the survey estimates (in the matter of mitigated or penalized calibration see also Théberge, 1999, 2000).

However, consistency among survey-estimated control totals is an important matter for national agencies. It is required that estimates derived from different surveys on the same population are consistent with each other.

Statistics Netherlands (CBS) developed a technique, RW estimator, that realises consistency with given margins through an additional calibration step for a new adjustment of already calibrated weights (Knottnerus and van Duin, 2006). However, to achieve consistency with the margins, the calibration on the first-step variables could be, at least in part, relinquished (Traat and Särndal, 2011).

To meet needs of complete consistency it is proposed an estimator that uses control totals, both from registers and survey, in a unique calibration step (see also Traat and Särndal, 2011). The expression of the estimator for the variance is also derived.

Currently in It-Silc the auxiliary information on working status and educational level is used between the correction step for nonresponse and the calibration step to balance the sample on these characteristics. Their impact on sampling variance is overlooked.

This paper proposes to include the LFS-estimated control totals simultaneously with the exactly-known totals and to use a suitable estimator of the variance to evaluate properly the impact of LFS-estimated control totals on It-Silc estimates.

When survey-estimated control totals are considered, the assumption usually made is that any additional error associated with them is negligible and can be ignored, as if the control totals were known without error. On the contrary, they have an effect on the variance of estimates leading to the understimation when using traditional variance estimator (Berger et al., 2009; Dever and Valliant, 2010).

The outline of the article is as follows. Section 2 gives a concise summary of calibration estimator and regression estimator. In Section 3, the expression of an estimator by Ballin et al. (2000) and Rancourt (2001) that uses survey-estimated control totals along with its estimator for variance by Renssen and Nieuwenbroek (1997) are presented and adapted to It-Silc's weighting procedure. Section 4 describes the results of the application on It-Silc data of 2008, using LFS-estimated control totals from the fourth quarter of 2007. Finally, Section 5 contains concluding remarks.

2. CALIBRATION ON CONTROL TOTALS

The estimates in It-Silc, as those of many other surveys carried out by national statistical agencies, are derived computing weights calibrated on control totals.

Calibration on control totals is used to minimise errors. The decrease of errors is linked to the association of population control totals with undercoverage, pattern of non-ignorable nonresponse and, moreover, with the variable of interest (Kott, 2006).

Deville and Särndal (1992) formalised calibration in survey sampling (see also Särndal, 2007, for a review of the calibration approach). The basic idea of calibration is that if auxiliary variables strongly correlated with the study variables are available, it "means that the weights that perform well for the auxiliary variables also should perform well for the study variables" (Deville and Särndal, 1992, p. 376).

Consider to be interested in estimating the total of y variable, $t_Y = \sum_{k \in U} y_k$, in a population $U = \{1, ..., k, ..., N\}$ of size N.

From population U a probability sample s of fixed size n is drawn according to a sampling design $p(\cdot)$. The inclusion probabilities of first-order, $\pi_k = Pr(k \in s)$, and second-order, $\pi_{kl} = Pr(k, l \in s)$, are known and assumed stricly positive. Note that $\pi_{kk} = \pi_k$ and $\sum_s 1/\pi_k = N$.

Assume that to each element $k \in s$, besides the value of the interest variables, is associated an auxiliary vector value, $\mathbf{x}_k = \left(x_{k1} \cdots x_{kp} \cdots x_{kP}\right)^\top$. Complete response is also assumed. Moreover, the population total of $\mathbf{x}$, $\mathbf{t_x} = \sum_U \mathbf{x}_k$, is supposed to be accurately known. The calibration (CAL) estimator,

$$\hat{t}_{Y_{CAL}} = \sum_{k \in s} w_k \, y_k,$$

uses calibrated weights, w_k, computed from the design weights, $d_k = 1/\pi_k$, previously adjusted for nonresponse to recover the condition of complete response (Deville et al., 1993). The calibrated weights are derived by solving the system

$$\begin{cases} \min_{w_k} \left\{ \sum_{k \in s} G(w_k, \ d_k) / q_k \right\} \\ \sum_{k \in s} w_k \, \mathbf{x}_k = \mathbf{t_x} \end{cases}$$

This implies that w_k are as close as possible to the d_k, in average sense for a given metric $G(\cdot)$ (see Deville and Särndal, 1992; Singh and Mohl, 1996) and, when applied to the auxiliary variables, reproduce exactly the control totals. Furthermore, q_k, usually $q_k = 1$, is a positive weight which gives more or less importance to unit k.

Generally, in survey estimates the vector $\mathbf{x}_k$ is created so that it is the same for all individuals k in the same household. This guarantees that all the individuals of the same household have equal weights w_k (Lemaître and Dufour, 1987).

When $G(\cdot)$ is the chi-squared distance, the resulting calibration estimator is equal to the generalised regression (GREG) estimator. Moreover, when the sample size is large, this holds for other metrics, too (see Deville and Särndal, 1992, p. 378-379). Therefore, the GREG estimator,

$$\hat{t}_{Y_{GREG}} = \hat{t}_{Y_\pi} + \left(\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}_\pi}\right)^\top \hat{\beta},$$

merely comes down doing a calibration on the regressors, where

$$\hat{\beta} = \left(\sum_{k\in s} \frac{q_k \mathbf{x}_k \mathbf{x}_k^\top}{\pi_k \sigma_k^2}\right)^{-1} \left(\sum_{k\in s} \frac{q_k \mathbf{x}_k y_k}{\pi_k \sigma_k^2}\right),$$

$\hat{t}_{Y_\pi} = \sum_{k\in s} y_k \; d_k$ is the Horvitz-Thompson estimator of t_Y and $\hat{\mathbf{t}}_{\mathbf{x}_\pi} = \sum_{k\in s} \mathbf{x}_k \; d_k$ is the vector of Horvitz-Thompson estimator of the auxiliary variables.

By definition, when using the GREG estimator, a linear regression superpopulation model between the interest variable y and the auxiliary variables can be implicitly assumed

$$\xi : y_k = \mathbf{x}_k^\top \beta + \varepsilon_k \qquad k \in U,$$

where the ε_k are uncorrelated random variables with null expectations and variances σ_k^2 (Bethlehem and Keller, 1987; Cassel et al., 1976, 1979; Fuller, 2002; Isaki and Fuller, 1982; Robinson and Särndal, 1983; Särndal, 1980; Särndal et al., 1989; Wright, 1983; Zieschang, 1990).

The expression of the estimator for the variance of $\hat{t}_{GREG}$ is also the asymptotic estimator for variance of $\hat{t}_{CAL}$,

$$\hat{V}\left(\hat{t}_{Y_{GREG}}\right) = \sum_{k\in s}\sum_{l\in s} \frac{\Delta_{kl}}{\pi_{kl}} \left(w_k \; e_k\right)\left(w_l \; e_l\right) \tag{1}$$

with $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$.

The expression (1) is a function of the residuals estimated on the sample, $e_k = y_k - \mathbf{x}_k^\top \hat{\beta}$. Then, considering auxiliary variables strongly correlated with the variable of interest can improve the efficiency of estimates.

3. CALIBRATION ON SURVEY-ESTIMATED CONTROL TOTALS

Control totals of auxiliary variables, useful to improve the efficiency of estimates, are not always available from registers or administrative sources, so survey-estimated totals are considered. However, they must refer to the same period and to the same population and the variables considered must have the same definitions. Moreover, it is not possible to use the traditional variance formula (i.e. expression (1)) as if the control totals were known without error.

For a sake of simplicity, only one external survey is considered, but the extension to the case in which more external surveys are considered easily follows.

Assume that another survey, carried out in the same reference period on the same population U, is already completed. Moreover, the estimates of control totals with high precision are derived by design-consistent and nearly design-unbiased estimator for this survey.

Therefore, let us define $\tilde{\mathbf{t}}_{\mathbf{x}} = (\tilde{t}_{x_1}\ \tilde{t}_{x_2} \cdots \tilde{t}_{x_m} \cdots \tilde{t}_{x_M})^\top$ the vector of M control totals estimated from the external survey.

Ballin et al. (2000) and Rancourt (2001) define the *extended regression estimator* (as called by Berger et al., 2009)

$$\hat{t}_{Y_E} = \hat{t}_{Y_\pi} + \left(\tilde{\mathbf{t}}_{\mathbf{x}} - \hat{\mathbf{t}}_{\mathbf{x}_\pi}\right)^\top \hat{\beta} \tag{2}$$

that uses survey-estimated control totals.

The estimator of variance of $\hat{t}_{Y_E}$ (Berger et al., 2009; Renssen and Nieuwenbroek, 1997) is equal to

$$\hat{V}\left(\hat{t}_{Y_E}\right) = \hat{V}\left(\hat{t}_{Y_{GREG}}\right) + \hat{V}_1 + \hat{V}_2 - \hat{V}_3$$

with $\hat{V}\left(\hat{t}_{Y_{GREG}}\right)$ given by (1) and

$\hat{V}_1 = \beta^\top\ \hat{V}\left(\tilde{\mathbf{t}}_{\mathbf{x}}\right)\ \beta,$

$\hat{V}_2 = 2\ \beta^\top\ \hat{C}\left(\hat{\mathbf{t}}_{Y_\pi}, \tilde{\mathbf{t}}_{\mathbf{x}}\right),$

$\hat{V}_3 = 2\ \beta^\top\ \hat{C}\left(\tilde{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}}\right) \beta$

where $\hat{C}(\cdot,\cdot)$ denotes a sampling covariance matrix in which the covariances depend on the sampling dependence between the samples of the surveys.

When the samples of the surveys are drawn independently, as those of It-Silc and LFS, $\hat{V}_2$ and $\hat{V}_3$ are equal to 0 and $\hat{V}\left(\hat{t}_{Y_E}\right)$ reduces to

$$\hat{V}\left(\hat{t}_{Y_E}\right) = \hat{V}\left(\hat{t}_{Y_{GREG}}\right) + \hat{V}_1. \tag{3}$$

However, in It-Silc both control totals from administrative sources and survey-estimated control totals from external surveys have to be considered.

Let it is $\check{\mathbf{t}}_{\mathbf{x}} = \left(\mathbf{t}_{\mathbf{x}}{}^{\top}, \tilde{\mathbf{t}}_{\mathbf{x}}{}^{\top}\right)^{\top}$. It is composed of two elements: the former of P control totals from administrative sources (with $p = 1, \cdots, P$) and the latter of M estimated control totals from the external survey (with $m = 1, \cdots, M$). Including it in (2)

$$\hat{t}_{Y_E} = \hat{t}_{Y_\pi} + \left(\check{\mathbf{t}}_{\mathbf{x}} - \hat{\mathbf{t}}_{\mathbf{x}_\pi}\right)^{\top} \hat{\beta} \tag{4}$$

is equivalent to calibrate the estimates in a unique step on both types of variables, therefore, the consictency on survey-estimated control totals is also guaranteed. Moreover, the estimator of variance in (3) still holds, but

$$\begin{aligned} \hat{V}_1 &= \beta^{\top} \hat{V}\left(\check{\mathbf{t}}_{\mathbf{x}}\right) \beta \\ &= \left[\begin{array}{cc} \beta_P & \beta_M \end{array}\right] \left[\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{V}\left(\tilde{\mathbf{t}}_{\mathbf{x}}\right) \end{array}\right] \left[\begin{array}{c} \beta_P \\ \beta_M \end{array}\right] \\ &= \beta_M^{\top} \hat{V}\left(\tilde{\mathbf{t}}_{\mathbf{x}}\right) \beta_M \end{aligned}$$

where β_P and β_M are the regression coefficients related respectively to the P control totals and the M survey-estimated control totals. Therefore, the expression of the estimator for variance of (4) can be written as

$$\hat{V}\left(\hat{t}_{Y_E}\right) = \sum_{k \in s} \sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \left(w_k\, e_k\right) \left(w_l\, e_l\right) + \sum_{m=1}^{M} \sum_{m'=1}^{M} \beta_m\, \beta_{m'}\, \hat{C}\left(\tilde{t}_{x_m}, \tilde{t}_{x_{m'}}\right). \tag{5}$$

The variables with estimated control totals, if strongly correlated with the variable of interest, are expected to contribute to the increase of efficiency of estimates, reducing the errors (first element in (5)). However, it is not obvious whether increased precision is realised by achieving consistency with survey-estimated control totals, because the addictional sampling variance plays an important role. In fact, the reliability of the survey, where the control totals are estimated is crucial for the magnitude of the second element in (5).

4. APPLICATION

Weighting procedure in It-Silc takes into account three steps: determination of the design weights, correction for nonresponse and calibration on control totals to obtain final weights.

The design weights, d_k, are given by the inverse of inclusion probability, π_k, by a stratified design, with stratification by NUTS II level and demographic size

Table 1: Estimates and coefficient of variation (CV) of It-Silc 2008 estimates with and without LFS-estimated control totals of fourth quater of 2007, for income group and NUTS I level

	No LFS Constraints		LFS Constraints		
NUTS I level	Estimate of total (in bilion of e)	Estimate of total CV%	(in bilion of e)	No Random CV%	Random CV%
Household income					
North-East	222.792	1.229	222.573	1.010	1.106
North-West	156.728	1.044	156.737	0.965	1.087
Center	154.183	1.213	153.750	1.026	1.143
South	131.922	1.240	129.954	0.960	1.143
Islands	63.042	1.748	62.484	1.267	1.565
ITALY	728.667	0.575	725.497	0.478	0.538
Personal income					
North-East	221.674	1.127	221.455	0.910	1.027
North-West	155.520	1.042	155.538	0.898	1.038
Center	152.886	1.227	152.452	1.009	1.156
South	130.682	1.227	128.714	0.875	1.084
Islands	62.334	1.811	61.757	1.241	1.544
ITALY	723.096	0.557	719.916	0.444	0.515
Labour income					
North-East	147.170	1.546	146.627	1.246	1.481
North-West	104.429	1.448	104.151	1.218	1.355
Center	101.509	1.714	101.047	1.360	1.416
South	84.233	1.762	81.825	1.279	1.431
Islands	38.708	3.061	37.853	1.843	2.067
ITALY	476.049	0.789	471.503	0.615	0.695
Self-employed income					
North-East	43.440	5.024	43.405	3.724	4.252
North-West	30.554	4.858	30.612	3.640	4.308
Center	30.193	4.725	30.312	3.777	4.464
South	21.684	4.876	21.077	4.063	4.790
Islands	9.246	6.780	9.066	6.013	7.219
ITALY	135.119	2.399	134.472	1.851	2.158
Employee income					
North-East	103.730	1.345	103.221	0.922	1.334
North-West	73.874	1.413	73.539	0.878	1.411
Center	71.316	1.756	70.735	1.065	1.635
South	62.549	1.904	60.748	1.052	1.689
Islands	29.460	3.483	28.788	1.939	2.620
ITALY	340.930	0.780	337.031	0.479	0.723
Retirement income					
North-East	59.985	1.344	60.274	1.259	1.381
North-West	39.473	1.340	39.749	1.280	1.490
Center	42.347	1.549	42.391	1.442	1.621
South	37.426	1.507	37.835	1.337	1.531
Islands	19.079	2.124	19.305	1.875	2.131
ITALY	198.310	0.684	199.554	0.633	0.712

of the municipality. These are then transformed into a set of intermediate weights in order to adjust the sample to the teoretical extracted one, under the hyphotesis that the mechanism that generates the total nonresponse is non-ignorable (see Ceccarelli et al., 2008).

Finally, the weights already adjusted for nonresponse are calibrated on control totals. To prevent negative or extreme weights the truncated logarithmic metric with bounds 0.5 and 1.5 is used.

As suggested by Commission regulation (EC) 1982/2003 the control totals known from administrative sources are target population by sex, age (five-year age groups), household size and composition at NUTS II level.

In addition to that Istat decided to add an intermediate step, before considering the exactly-known control totals. The aim of this step is to achieve consistency with the estimates of working status and educational level by sex and NUTS I level from LFS estimates of the fourth quarter of the previous year, that is the reference period of income in It-Silc. Actually, the error of estimates in It-Silc is computed assuming that the additional sampling variance associated with LFS estimates is negligible.

Considering the exactly-known and the survey-estimated control totals in a unique calibration procedure allows the measurement, with expression (5), of the impact of LFS-constraint on the estimates and the evaluation of their convenience.

In particular, It-Silc data of 2008 (20.928 households and 52.433 individuals) is considered. In Table (1) the total estimates of several income categories by NUTS I level are derived using different weighting systems, with or without estimated control totals from fourth quarter of 2007 of LFS. On the contrary, control totals from administrative sources are always included.

When LFS-estimated control totals are considered in the weighting procedure the estimate for variance is computed with (1) and (5) [no random vs. random]. The expression (1) is actually implemented in It-Silc and it does not consider the additional sampling variance associated with LFS-estimated control totals. Instead, the expression (5), that here is proposed and firstly applied to It-Silc estimates, takes into account the additional sampling variance of LFS-estimated control totals and allows to evaluate properly the convenience of these further constraints.

When including the auxiliary information on working status and educational level in weighting system, without considering the additional variance, the efficiency of estimates increases considerably with respect to the case in which they are not considered. This means that the added control totals give an important contribution to increase the accuracy and the consistency of income estimates.

This even holds, with only few exceptions, when considering the additional sampling variance associated with LFS-estimated control totals (LFS constraints - random). In fact, the gain in efficiency of estimates due to the strong correlation between the LFS-estimated controls totals and income overcomes the additional sampling variance associated with them.

Very remarkable improvements occur for self-employed (with the exception of Islands) which is, usually, a difficult category in income surveys. On the contrary, a small increase of the CV of estimates occurs for retirement income.

Two are the possible causes that can explain these results. Educational level and working status are not as strongly correlated with retirement income as with the self-employed income. Furthermore, when survey-estimated control totals are added, the increase in efficiency is larger for the larger variances, that is related to small domains, and smaller for the smaller variances, related to larger domains (Merkouris, 2010; Traat and Särndal, 2011). Therefore, the decrease of the first element in (5) is not enough to balance the additional variance due to the LFS-estimated control totals.

In this case it is necessary to evaluate whether to accept an increase of the CV for retirement income to achieve more accurated and consistent estimates for self-employment income.

5. CONCLUSION AND REMARKS

In It-Silc weighting system, beside control totals of demographic variables from administrative sources, estimated control totals from LFS are considered.

Inserting the LFS-estimated control totals in the weighting procedure together with the exactly-known total and using a suitable estimator of the sampling variance allowed to evaluate the impact of these constraints on the efficiency of estimates. Their use considerably increases the efficiency of estimates. It allows to include information on working status and educational level – that are strongly correlated with income – which otherwise could not be taken into account.

The additional sampling variance associated with these controls is not negligible. The application on It-Silc 2008 data shows how using traditional variance estimators can understimate the resulting sampling variance. The results obtained agree with those of Berger et al. (2009) and Dever and Valliant (2010).

However, even evaluating properly the additional sampling variance, the estimates derived with LFS-estimated control totals are, almost always, more efficient than those derived with only demographic control totals. The reason of the increasing accuracy lies in the strong correlation between the auxiliary variables and the variable of interest, but also in the high reliability of the LFS-estimated control totals used.

Therefore, the advice in using survey-estimated control totals is to evaluate both the correlation between the auxiliary variables, which affects the first part in expression (5), and the reliability of the survey-estimated control totals, which affects the second part of the same expression.

However the use of survey-estimated control totals can have ambiguous effects, that is it increases the efficiency in some domains and decreases it in others. Therefore, it is necessary to evaluate which scenario is preferable. For instance, in the case of It-Silc it is preferable to include LFS-estimated control totals, because they yield a great increase in efficiency for self-employed income and just a small decrease for retirement income and for other few domains.

References

Ballin, M., Falorsi, P.D., and Russo, A. (2000). Condizioni di coerenza e metodi di stima per le indagini campionarie sulle imprese. In *Rivista di Statistica Ufficiale,* 2: 31–52.

Bardsley, P. and Chambers, R.L. (1984). Multipurpose estimations from unbalanced samples. In *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 33 (3): 290–299.

Berger, Y.G., Muñoz, J.F., and Rancourt, E. (2009). Variance estimation of survey estimates calibrated on estimated control totals - An application to the extended regression estimator and the regression composite estimator. In *Computational Statistics & Data Analysis,* 53: 2596–2604.

Bethlehem, J.G. and Keller, J.W. (1987). Linear weighting of sample survey data. In *Journal of Official Statistics,* 3 (2): 141–153.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite population. In *Biometrika,* 63 (3): 615–620.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1979). Prediction theory on finite population when model-based and design-based principle are combined. In *Scandinavian Journal of Statistics,* 6 (3): 97–106.

Ceccarelli, C., Coppola, L., Cutillo, A., and Di Laurea, D. (2008). Combining survey and administrative data in the Italian Eu-Silc experience: Positive and critical aspects. In *Conference of European Statisticians - Working Session on Statistical Data Editing,* 1–10. UNECE. URL `http://www.unece.org/fileadmin/DAM/stats/documents/2008/04/sde/wp.14.e.pdf`.

Chambers, R.L. (1996). Robust case-weighting for multipurpose establishment surveys. In *Journal of Official Statistics,* 12 (1): 3–32.

Dever, J.A. and Valliant, R. (2010). A comparison of variance estimators for poststratification to estimated control totals. In *Survey Methodology,* 36 (1): 45–56.

Deville, J.C. (1999). Simoultaneus calibration of several surveys. In *Proceedings of Statistics Canada Symposium 99 - Combining data from different sources,* 207–212. Statistics Canada, Ottawa, Canada.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. In *Journal of the American Statistical Association,* 87 (418): 376–382.

Deville, J.C., Särndal, C.E., and Sautory, O. (1993). Generalized raking procedure in survey sampling. In *Journal of the American Statistical Association,* 88 (423): 1013–1020.

Fuller,W.A. (2002). Regression estimation for survey samples. In *Survey Methodology,* 28 (1): 5–23.

Goga, C. and Shezad, M. (2010). Overview of ridge regression estimators in survey sampling. *Tech. rep.,* IMB, Université de Bourgogne, Dijon, France.

Guggemos, F. and Tillé, Y. (2010). Penalized calibration in survey sampling: design-based estimation assisted by mixed models. In *Journal of Statistical Planning and Inference,* 140 (11): 3199–3212.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *In Journal of the American Statistical Association,* 77 (377): 89–96.

Knottnerus, P. and van Duin, C. (2006). Variances in repeated weighting with an application to the Dutch Labour Force Survey. In *Journal of Official Statistics,* 22 (3): 565–584.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. In *Survey Methodology,* 32 (2): 133–142.

Lemaître, G. and Dufour, J. (1987). An integrated method for weighting persons and families. In *Survey Methodology,* 13 (2): 199–207.

Merkouris, T. (2010). Combining information from multiple surveys by using regression for efficient small domain estimation. In *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 72 (Part 1): 27–48.

Rancourt, E. (2001). La régression étendue: un ensemble de pratiques d'estimation qui poussent constamment la théorie. Recuil Enquêtes. In J.J. Droesbeke and I. Lebart, eds., *Modéles et Application,* 334–343. Dunod. (in French).

Rao, J.N.K. and Singh, A.C. (1997). A ridge-shrinkage method for rangerestricted weight calibration in survey sampling. In *Proceedings of the Section on Survey Research Method,* 207–212. American Statistical Association, Washington, D.C.

Renssen, R.H. and Nieuwenbroek, N.J. (1997). Aligning estimates for common variables in two or more sample surveys. In *Journal of the American Statistical Association,* 92 (437): 368–374.

Robinson, P.M. and Särndal, C.E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. In *Sankhya, Series B,* 45 (2): 240–248.

Särndal, C.E. (1980). On p-inverse weighting versus best linear unbiased weighting in probability sampling. In *Biometrika,* 67 (3): 639–650.

Särndal, C.E. (2007). The calibration approach in survey theory and practice. In *Survey Methodology,* 33 (2): 99–119.

Särndal, C., Swensson, B., and Wretman, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. In *Biometrika,* 76

(3): 527–537.

Singh, A.C. and Mohl, C.A. (1996). Understanding calibration estimator in survey sampling. In *Survey Methodology,* 22 (2): 107–115.

Théberge, A. (1999). Calibration and restricted weights. In *Survey Methodology,* 26 (1): 99–107.

Théberge, A. (2000). Estension of calibration in survey sampling. In *Journal of the American Statistical Association,* 94 (446): 635–644.

Traat, I. and Särndal, C.E. (2011). Domain estimators calibrated on information from other surveys. *Acta e commentationes universitatis tatuensis de mathemathica 2,* University of Tartu, Tartu, Estonia.

Wright, R.L. (1983). Finite population sampling with multivariate auxiliary information. In *Journal of the American Statistical Association,* 78 (384): 879–884.

Zieschang, K.D. (1990). Sample weighting methods and estimation of totals in the Consumer Expenditure Survey. In *Journal of the American Statistical Association,* 85 (412): 886–1001.