Balancing and Calibration in Survey Sampling: A Simulation Study

Adriano Pareto and Antonio Pavone

Abstract The use of auxiliary information is a central issue in survey sampling from finite populations. Classical techniques making use of auxiliary information are balanced sampling and calibration weighting. Balanced sampling requires known variables values for each unit of the population, but it does not always produce a perfect calibration. On the other hand, calibration weighting can generate very unstable weights with high coefficient of variation. For this reason, balanced sampling and calibration are often used together. In this paper we propose a method to minimize the variance inflation due to unequal weighting when we want to draw a random sample which provides exact estimations for a set of auxiliary variables that are known for all the population units. An empirical comparison with some traditional strategies (simple random sampling and cube method with calibration weighting) is presented: the *Unequal Weighting Effect* (UWE) is, on average, always much lower for the proposed method as compared to the others.

Key words: sample design, calibration weighting

1 Introduction

When auxiliary data are known for the entire population, properties of estimators can be improved by incorporating the auxiliary information in the sample design.

A usual solution to obtain a planned sample is to stratify the population with strata given by cross-classification with respect to the available auxiliary variables. However, when the groups are defined from continuous auxiliary variables or a large number of

---

Adriano Pareto, Italian National Institute of Statistics; email: pareto@istat.it.

Antonio Pavone, Italian National Institute of Statistics; email: pavone@istat.it.

This work is the result of the common work of the authors. However, sections 1 and 2 are attributable to A. Pavone, sections 3 and 4 to A. Pareto.

grouping variables are crossed, an excessive number of strata can be needed and for small sample sizes may be impossible to cover all the cells. Moreover, stratification, for a continuous auxiliary variable satisfying a linear model, loses some information (Legg, 2007).

Another way to incorporate auxiliary information in a design is by probability proportional to size (PPS) sampling. If a linear relation between the study variable *y* and a known variable *x* (the “size” variable) is assumed for the population[1], the total of *y* may be estimated using first-order inclusion probabilities proportional to the *x* values. Unfortunately, due to the unidimensional nature of this method, it is not possible to use more than one auxiliary variable. However, using a set of auxiliary variables could be convenient when estimates for several variables of interest are required.

An alternative approach to incorporate information from auxiliary variables is to apply a balanced sampling design. A sample is said to be balanced, on a set of balancing variables, if the estimated totals of these variables equal to the population totals. Royall and Herson (1973) stressed the importance of balancing a sample for protecting inference against a misspecified model. Nevertheless, achieving exactly multivariate balanced sample is infeasible for most of data.

To increase the precision of the survey estimate, the initial weights are often adjusted in the estimation phase, using a sampling weights adjustment procedure that forces the estimates to equal the population totals (calibration weighting). If the design is approximately balanced, the sampling weights adjustment procedure consists of only rounding initial weights. However, if the sample is poorly balanced, adjusted sampling weights are unstable, with extreme values, and the variance of the estimator can be substantially inflated, depending on the relationship between the sampling weights and the survey variables.

We are looking for a strategy that, selecting randomly a sample, satisfies the total population constraint and achieves the lower-bound values for the weights inflation.

In this paper we propose to incorporate sample design in the estimation algorithm. Optimal weights and optimal sample are found by solving a mixed-integer nonlinear programming problem. The paper is organised as follows. Section 2 outlines the proposed algorithm. A set of simulations has been carried out in section 3, in order to evaluate the inflation effect on the final weights, comparing the suggested procedure with traditional strategies such as simple random sampling and cube method (Deville and Tillé, 2004; Rousseau and Tardieu, 2004) with calibration weighting. For the simulation we used data of municipalities in the Canton of Ticino of Tillé (2002, pp. 181–186). Conclusion is drawn in section 4.

2 Sampling and calibration weighting

Let us consider a finite population *U* of size *N* whose units can be identified by labels $k \in \{1, ..., N\}$. Given an inclusion probability $\pi_{(0)k}$, for each element *k* in the sample *S*, we can estimate the total of a variable of interest *y*, $T_y = \sum_{k \in U} y_k$, where y_k is the value of the *k*-th unit, with the following expansion estimator:

---

[1] The variable *x* should be such that the regression line of *y* on *x* pass through the origin.

$$\hat{T}_y = \sum\nolimits_{k \in U} y_k d_k I_k$$

in which: $d_k = \pi_{(0)k}^{-1}$ denotes the Horvitz-Thompson weight of the unit *k* calculated as reciprocal of the inclusion probability and I_k is a sample membership indicator such that $I_k = 1$ if $k \in S$ and 0 otherwise.

Suppose also that *p* auxiliary variables $\mathbf{x}_k = (x_{k1} \ldots x_{kj} \ldots x_{kp})$ are available for the entire population, a sampling design is said to be balanced with respect to auxiliary variables $\mathbf{x}_k$ if it satisfies the balancing equations given by $\hat{T}_{x_j} = T_{x_j}$, which can also be written as $\sum\nolimits_{k \in U} x_{kj} d_k I_k = \sum\nolimits_{k \in U} x_{kj}$, for j = 1, … , *p*.

Classical techniques that use auxiliary information in a sampling design are stratification (Neyman, 1934; Tschuprow, 1923) and unequal probability sampling or sampling proportional to size (Hansen & Hurwitz, 1943; Madow, 1949). In the other cases, nonzero slacks are usually detected in multidimensional balancing equation. This means that equalities in balancing are not fully satisfied. If the sample size is small, the deviation could be significant. In such case the Horvitz-Thompson weights are often adjusted using calibration weighting.

Deville and Särndal (1992) coined the term “calibration estimator” to describe an estimator of the form:

$$\hat{T}_{x_j}^c = \sum\nolimits_{k \in U} x_{kj} w_k I_k = \sum\nolimits_{k \in U} x_{kj} = T_{x_j}$$

for some row vector of *p* auxiliary variables $\mathbf{x}_k$, about which T_{x_j} is known, where $w_k = \pi_{(1)k}^{-1}$ and $\pi_{(1)k}$ is the modified inclusion probability of the unit *k*. They required that the difference between $\{w_k | k \in s\}$ and $\{d_k | k \in s\}$ minimize some specified loss function. However, this procedure inflates sampling variance and mean square error, due to calibrated weights variation and it depends on how "unbalanced" is the sample.

To protect against instability of the weights, we suggest to combine in the same mathematical framework sampling design and calibration.

The idea is straightforward: we want to find a suitable random sample that achieves the minimum distance between Horvitz–Thompson weights and calibrated weights. The proposed algorithm, denoted as Sam.&.Cal. (sampling and calibration weighting), is the following:

$$\begin{cases} \underset{w_k,\, s_k}{\text{Min}} \sum\nolimits_{k \in U} G(w_k; d_k) & \\ \text{s.t.} \quad \sum\nolimits_{k \in U} s_k = n & s_k \in \{0,1\} \\ \qquad \sum\nolimits_{k \in U} w_k s_k = N & w_k \in \Re^+ \\ \qquad \sum\nolimits_{k \in U} x_{kj} w_k s_k = \sum\nolimits_{k \in U} x_{kj} = T_{x_j} & j = 1, \ldots, p \end{cases} \tag{1}$$

where $G(w_k; d_k)$ is a distance function between w_k and d_k .

This is a mixed-integer nonlinear programming problem because it contains nonlinear expressions and an integer variable.

To initialize the algorithm, we set $w_k = d_k$ and the binary variable $s_k = I_k$ is rounding off randomly to 1 or 0 for all the elements k of the vector $\boldsymbol{\pi}_{(0)}$, or using some previous assigning probability classification. When the local search is processed, the algorithm scans through the neighbourhood of a starting solution and stop only if convergence criteria are met. According to the calibration constrains, a local minimum solution is found, in which $w_k \neq d_k$ when $s_k = 1$ and $w_k = d_k$ if $s_k = 0$.

The algorithm could be extended, fixing the sample size also to sub partition level (for example in case we need a sample that guarantee a given number of units for each administrative district).

In the next section, three types of random sampling strategies will be empirically evaluated in terms of weights variation.

3 A simulation study

Using the data of municipalities in the Canton of Ticino coming from the 2000 Swiss federal census of population, the three strategies SRS+Cal. (simple random sampling with calibration weighting), Cube+Cal. (cube method with calibration weighting) and Sam.&Cal.[2] (sampling and calibration weighting) are compared; see Tillé (2002) for more details.

Ticino has 245 municipalities. The municipalities are selected with equal inclusion probabilities and the sample size is 50. The used auxiliary variables are: ONE (constant variable that always takes the value 1)[3], POP (number of men and women), ARE (area of the municipality in hectares) and HOU (number of households).

Three simulations are performed increasing the number of auxiliary variables: simulation 1 (ONE and POP), simulation 2 (ONE, POP and ARE), simulation 3 (ONE, POP, ARE and HOU). Each simulation consists in drawing, for every strategy, 100 random samples which provide exact estimations for the totals of the auxiliary variables. Therefore, the variances of the estimators of the total of all the study variables will be reduced, depending on the correlations of these variables with the auxiliary variables.

The analysis is focused on the obtained weights without any reference to the variables of interest since they do not enter directly into the algorithms for computing weights. Naturally, the resulting samples will provide very accurate estimators for “all the variables” that are correlated with the balancing variables (Tillé, 2002).

Table 1 shows the summary statistics for the distributions of the coefficient of variation of the sampling weights.

Note that 100 replications, for each simulation, are more than sufficient to test whether the proposed strategy is better than another in terms of variability of the weights as the Sam.&.Cal. method is designed, differently from the others, to provide samples with the least possible variability.

The distributions obtained are not normal and have different variances, but they are similar in shape (positive skew). In order to compare each strategy with each other, we

---

[2] We used the squared *Euclidean distance* as the function G in (1).

[3] This variable assures that the sum of the sampling weights is equal to the population size.

Table 1: Coefficient of variation of the sampling weights (%)

STATISTICS	SRS+Cal.	Cube+Cal.	Sam.&Cal.
	Simulation 1		
Mean	15.75	5.06	0.30
Standard deviation	17.34	6.10	0.32
Minimum	0.24	0.02	0.00
Median	10.83	2.63	0.24
Maximum	92.52	25.18	2.77
Range	92.28	25.17	2.76
	Simulation 2		
Mean	19.06	7.49	0.58
Standard deviation	11.57	8.38	0.59
Minimum	1.07	0.29	0.03
Median	16.69	4.86	0.49
Maximum	54.31	46.24	3.49
Range	53.23	45.95	3.46
	Simulation 3		
Mean	28.74	13.11	1.73
Standard deviation	16.96	8.43	2.06
Minimum	3.00	1.43	0.05
Median	23.31	10.89	0.62
Maximum	78.26	33.82	9.65
Range	75.26	32.39	9.60

used the Mann-Whitney statistic U[4] and all the tests are largely significant.

Results show that, the higher the number of auxiliary variables is, the larger becomes the mean of the coefficient of variation of the sampling weights. However, for each simulation, the greatest average coefficient of variation is obtained under SRS+Cal. strategy, while the Sam.&Cal. method produces the least average coefficient of variation. For example, in simulation 2 (3 auxiliary variables), the mean value is 19.1% for SRS+Cal., 7.5% for Cube+Cal. and only 0.6% for Sam.&Cal. method.

Therefore, the potential inflation due to weighting that can results in estimates of standard errors in Sam.&Cal. method is much less than the others.

A measure of variance inflation, often refer to as *Unequal Weighting Effect* (UWE), defined by Kish (1992) is the following:

$$\text{UWE} = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2} = 1 + ((\text{cv}(w))^2$$

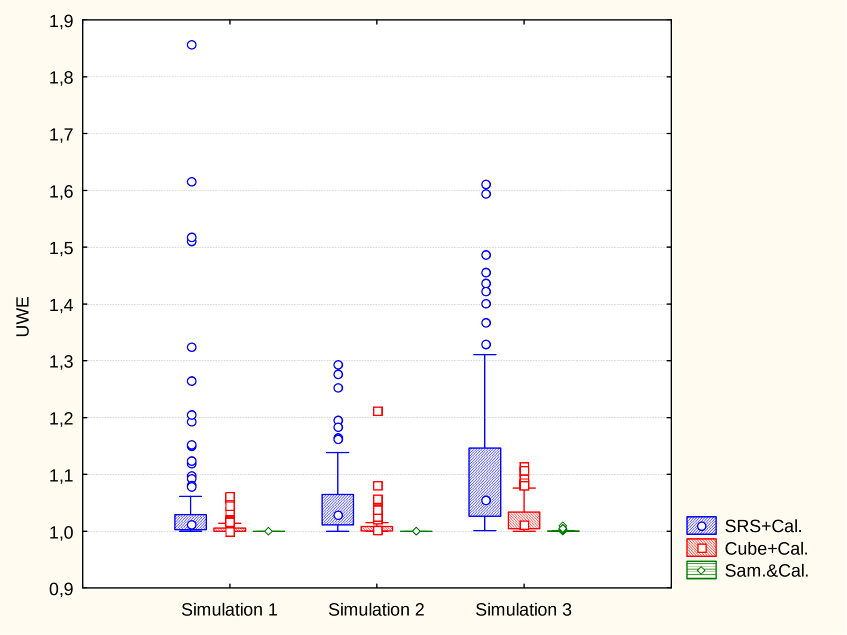
where w_i represents the weight of the i-th unit, n is the sample size and cv(w) is the coefficient of variation of the weights.

The box-plot of the UWE distributions for the considered strategies in the different simulations is displayed in Figure 1.

---

[4] In a Mann-Whitney test to evaluate whether a variable x is stochastically smaller than a variable y, a sample size of 52 units is adequate for α=0.05, β=0.10 and δ=P(y>x)≥2/3 (Noether, 1987).

Figure 1: Box-plot of Unequal Weighting Effect (UWE)



The effects of unequal weighting increase, on average, with the number of auxiliary variables and they range from 1 to 1.86 for SRS+Cal. strategy, from 1 to 1.21 for Cube+Cal. strategy and from 1 to 1.01 for Sam.&Cal. method.

In order to verify the randomness of the samples drawn by the Sam.&Cal. method, for each simulation, a chi-square goodness-of-fit test was performed on the frequencies of the single municipalities in the overall of the 100 samples of 50 units (total frequency = 5,000, expected frequency for each municipality = 5,000/245=20.4).

Results are shown in Table 2.

Table 2: Chi-square goodness-of-fit test to uniform distribution

STATISTICS	SRS+Cal.	Cube+Cal.	Sam.&Cal.
	Simulation 1		
Chi-square	162.836	205.172	218.990
df	244	244	244
Sig.	1.000	0.966	0.873
	Simulation 2		
Chi-square	185.180	197.724	227.222
df	244	244	244
Sig.	0.998	0.987	0.772
	Simulation 3		
Chi-square	197.724	184.298	221.538
df	244	244	244
Sig.	0.987	0.998	0.846

In general, the chi-square value for the Sam.&Cal. method is greater than the other values, but it is always largely not significant. For example, in simulation 3 (4 auxiliary variables), the chi-square statistic is equal to 221.5 and it is well below the 0.95 cut-off point of 281.5, so we can accept the hypothesis of an underlying uniform distribution.

4 Conclusions

Survey weights must be calculated taking into account the potential inflation that can results in estimates of standard errors. When auxiliary information is available from several variables correlated with the study variables, balanced sampling, calibration weighting or a mixed strategy can be used.

In this work, we found that the variance inflation, due to unequal weighting, can be minimized using a simultaneous method of sampling and calibration weighting. The method can be applied when we want to draw a random sample which provides exact estimations for a set of auxiliary variables that are known for all the population units.

Considering the computational complexity of the proposed algorithm, it is particularly useful in cases where both sample size and population size are not too large.

Acknowledgement

The authors thank the anonymous referee for providing constructive comments to improve the contents of this paper. They are also grateful to Alessandra Palmieri for her useful suggestions.

References

1. Deville, J., Sarndal, C.: Calibration Estimation in Survey Sampling. *Journal of the American Statistical Association*, **87**, 376-382 (1992)
2. Deville, J., Tillé, Y., Efficient balanced sampling: The cube method. *Biometrika*, **91**, 4, 893-912 (2004)
3. Hansen, M., Hurwitz, H. W.: On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, **14**, 333–62 (1943)
4. Kish, L., Weighting for Unequal P_i. *Journal of Official Statistics*, **8**, 183-200 (1992)
5. Legg, J.: Comparison of sample set curtailment procedures. *Proceedings of the Survey Research Methods Section, ASA*, 2775-2781 (2007)
6. Madow, W.: On the theory of systematic sampling, II. *Annals of Mathematical Statistics*, **20**, 333–4 (1949)
7. Neyman, J.: On the two different aspects of representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, **97**, 558–606 (1934)
8. Noether, G.E.: Sample Size Determination for Some Common Nonparametric Tests. *Journal of the American Statistical Association*, **82**, 645-647 (1987)
9. Rousseau, S., Tardieu, F.: *La macro SAS CUBE d'échantillonnage équilibré. Documentation de l'utilisateur*. INSEE (2004)

10. Royall, R. M., Herson, J.: Robust estimation in finite populations, I. *Journal of the American Statistical Association*, **68**, 880–9 (1973)
11. Tillé, Y., *Sampling Algorithms*. Springer, New York (2002)
12. Tschuprow, A.: On the mathematical expectation of the moments of frequency distributions in the case of correlated observation. *Metron* **3**, 461–93, 646–80 (1923)