TEST
https://doi.org/10.1007/s11749-019-00685-z

DISCUSSION

Rejoinder on: Deville and Särndal's calibration: revisiting a 25-year-old successful optimization problem

Denis Devaud[1] · Yves Tillé[1]

We thank the discussants: Jean-François Beaumont and J.N.K Rao, Phillip S. Kott, Domingo Morales, Maria del Mar Rueda, Changbao Wu and Shixiao Zhang for their interest in our article. Their suggestions, criticisms and remarks provide very useful complements. They also report a list of interesting publications on calibration. Deville and Särndal's article had a considerable impact on the theory and practice of survey sampling. In October 2019, Scopus mentions that the article has been cited 614 times. It is therefore very difficult to present an exhaustive survey of all the contributions made by the original article on calibration.

1 Response to Jean-François Beaumont and J.N.K. Rao

We would like to thank Jean-François Beaumont and J.N.K. Rao for their particularly thoughtful comments. We have probably failed to mention a large number of interesting publications on calibration. It is true that we should have quoted the article of Huang and Fuller (1978) who already raised a large number of fundamental questions and proposed solutions especially for the problem of negative weights. We share the point of view of the authors. We just would like to add some general remarks. We deal with the problem of calibration in case of nonresponse in our response to Phillip Kott.

The limitation of the range of the weights is a vast subject. We would like to point out that there are no general recommendations for bounding the weights. It is obvious that we must avoid negative weights or even avoid w_k weights smaller than 1. Indeed, a statistical unit of the sample should represent itself at least. Many survey practitioners

This rejoinder refers to the comments available at: https://doi.org/10.1007/s11749-019-00682-2; https://doi.org/10.1007/s11749-019-00683-1; https://doi.org/10.1007/s11749-019-00684-0; https://doi.org/10.1007/s11749-019-00686-y; https://doi.org/10.1007/s11749-019-00687-x.

✉ Denis Devaud
denis.devaud@unine.ch

Yves Tillé
yves.tille@unine.ch

1 Institut de statistique, Université de Neuchâtel, Av. de Bellevaux 51, 2000 Neuchâtel, Switzerland

Published online: 22 November 2019

use the logistic calibration function but ultimately choose the bounds by trial and error without any clear methodology being developed. The q_k-weights can be used to take the heteroscedasticity of the data into account. An appropriate choice of the q_k-weights can often resolve most of the weight issues without necessarily imposing bounds. There are currently tools to minimize the ranges of the g-weights (Rebecq 2017, 2019; Graf 2016). Should we use them? Bounding weights is of particular interest for estimation in small areas. The few simulations using maximally bounded weights do not seem to show a significant increase in bias, but further work is still needed to assess the true impact of the boundary on bias and variance? These questions will probably have to be studied further and a discussion between practitioners and researcher may be interesting.

Nor is it obvious to give a general recommendation when a very large number of calibration variables are available. Admittedly, we have to try to focus on the variables most closely correlated with the variables of interest, but as pointed out by J.N.K. Rao and Jean-François Beaumont, the calibration is often used to make the results of a survey consistent with a census or with administrative data. Obviously, we should not calibrate too much, because as in any statistical problem, it is necessary to apply a parsimony principle so as to avoid to increase unnecessarily the variance. As in regression, one can use a ridge or lasso calibration. One can also make a soft calibration, with an underlying mixed model for some variables. Unfortunately, in this case, we lose the property of calibration. Therefore, we cannot have all at the same time: an accurate estimator that is consistent with all administrative sources for all known variables. At some point, we have to make an arbitration. It should be kept in mind that calibration is primarily an estimation method before being a tool to make the data consistent. Survey statistics is therefore also based on the experience of the survey managers.

2 Response to Philip S. Kott

In the original article of Deville and Särndal (1992), calibration is presented as a method for correcting the sampling error. The word "nonresponse" or "missing value" does not appear in the paper. Calibration was initially not proposed for estimating a probability of response.

The first publication where the estimation of the response probability is presented as a calibration problem is probably the article of Iannacchione et al. (1991). One year after the publication of Deville and Särndal's paper, Deville and Dupont (1993) also proposed to use calibration to estimate a probability of response. These authors conclude this paper by stating: "Developing a response model for reweighting requires a thorough analysis of the response mechanism. This must lead to the specification of a response model. Nevertheless, we must beware of too great complications. The way to estimate model parameters has little effect on the quality of the final estimates. Therefore, it is recommended to use a calibration criterion based on estimating equations having an obvious concrete meaning" (translated from French by the authors).

This assertion changed the calibration objective to use it as an estimation technique to estimate a model of response and no longer as a method of correcting a sampling error. This approach was next developed by Dupont (1994, 1996). Lundström and Särndal (1999), Särndal and Lundström (2005) and Folsom and Singh (2000) also advocated for the use of calibration for treatment of nonresponse." Next Jean-Claude Deville (1998, 2000, 2002, 2004) developed the generalized calibration by using instrumental variables that are different from the calibration variables. The interest in the generalized calibration was so important that it was quickly implemented and used in public statistics notably in the SAS macro CALMAR2 of the Insee (Le Guennec and Sautory 2002) and the gCALIBS SPSS package of Statistics Belgium (Vanderhoeft, 2001). Generalized calibration has been next studied by Kott (2006) and Chang and Kott (2008). Thus, we can say that calibration was first developed to decrease the sampling error. It has rapidly been applied to correct the error of unit nonresponse.

However, two approaches exist in order to deal with unit nonresponse. In the first one (called response model approach), the probability of response is first estimated by a nonresponse model (for instance a logistic regression). The original weights are then the inverses of the products of the inclusion probabilities and the estimated response probabilities. These weights are then calibrated on the totals. In the second one (called the one-step approach), calibration is directly used to correct at the same time the sampling error and the nonresponse error.

Haziza and Beaumont (2017) warn against the one-step approach. In this case, the calibration variables must be relevant to be able to model the nonresponse. Lesage et al. (2019) take strong reservations against the use of generalized calibration, which in some cases may prove counterproductive.

In any case, we completely share the point of view of Philip Kott for the estimation of variance. In the one-step approach, there are two sources of randomness coming from the sampling design and the nonresponse mechanism. The variance estimation is then specific (Kott and Liao 2012; Vallée and Tillé 2019).

3 Response to Maria del Mar Rueda

Thank you very much for these additional references. We comment the model calibration approach in the response to Changbao Wu. Concerning the calibration on other parameters of interest, there is also the interesting article of Lesage (2011) who explains that one can linearize the parameter to obtain an estimator calibrated on this one.

We strongly agree that the publications about sampling and calibration for functional data Cardot and Josserand (2011) and Boistard et al. (2017) are extremely interesting and presage the use of massive data. Finally, we are relatively reluctant to use calibration in data sets obtained without sampling designs. We are not convinced that calibration is an appropriate tool for correcting unstructured data.

4 Response to Domingo Morales

Thank you very much for these nice comments and for these additional references. The general pseudo-distance used in Deville and Särndal (1992), which is presented in our

Sect. 3.1.4, is directly related to the divergence proposed by Rényi (1961). The general pseudo-distance $G_k^\alpha(w_k, d_k)$ is also an extension of the power divergence studied in Cressie and Read (1984). The pseudo-distance $G_k^\alpha(w_k, d_k)$ admits values whose sum may be different from one. However, Cressie and Read (1984) used it to construct test of hypothesis and not to reweigh the observations. We strongly agree that this family of distances has a wide field of applications in statistics.

We completely agree that calibration is a tool that can be used in many other applications: treatment of nonresponse, coherence of estimators in small areas, estimation from multiple sources and fusion of samples (Guandalini and Tillé 2017).

5 Response to Changbao Wu and Shixiao Zhang

Thank you very much for these additional references. In our paper, we wanted to point out that the calibration function that derives from pseudo empirical likelihood method is already described in Deville and Särndal (1992). Unfortunately, this calibration function tends to produce even more extreme weights than the raking ratio method. We are convinced that this calibration function should be avoided or must be used with bounds.

The model calibration approach of Wu and Sitter (2001) [also used in Montanari and Ranalli (2005)] is interesting because it allows to take into account a nonlinear relation between the auxiliary variables and the variable of interest. However, in the practice of official statistics, it is often desired to construct weights applicable to all variables of interest. These weights are sometimes simply provided with a data file. Users can then apply these weights for any estimation problem. Calibration on auxiliary variables is often simply useful for constructing tables in which variables completely known by a register intervene (such as an age distribution). Relationships (correlations, contingency tables) between different variables must also be analyzed using a unique weighting system. The advantage of Deville and Särndal's approach was precisely to separate the weighting system from the variables of interest. The model calibration approach is therefore difficult to implement in the case where a large set of variables of interest must be coherently estimated.

References

Boistard H, Lopuhaä HP, Ruiz-Gazen A (2017) Functional central limit theorems for single-stage sampling designs. Ann Stat 45(4):1728–1758

Cardot H, Josserand E (2011) Horvitz-Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. Biometrika 98(1):107–118

Chang T, Kott PS (2008) Using calibration weighting to adjust for nonresponse under a plausible model. Biometrika 95:555–571

Cressie N, Read TRC (1984) Multinomial goodness-of-fit tests. J R Stat Soc Ser B (Methodol) 46(3):440–464

Deville JC (1998) La correction de la non-réponse par calage ou par échantillonnage équilibré. Technical report In see, Paris, recueil de la Section des méthodes d'enquête

Deville JC (2000) Generalized calibration and application to weighting for non-response. Compstat—proceedings in computational statistics: 14th symposium held in Utrecht. Springer, New York, pp 65–76

Deville JC (2002) La correction de la nonréponse par calage généralisé. Actes des Journées de Méthodologie Statistique. Insee-Méthodes, Paris, pp 3–20

Deville JC (2004) Calage, calage généralisé et hypercalage. Technical report Insee Paris

Deville JC, Dupont F (1993) Non-réponse: principes et méthodes. Actes des Journées de Méthodologie statistique. Insee, Paris, pp 53–70

Deville JC, Särndal CE (1992) Calibration estimators in survey sampling. J Am Stat Assoc 87:376–382

Dupont F (1994) Calibration used as a nonresponse adjustment, studies in classification, data analysis, and knowledge organization. In: Diday E (ed) New approaches in classification and data analysis. Springer, Berlin, pp 539–548

Dupont F (1996) Calage et redressement de la non-réponse totale : Validité de la pratique courante de redressement et comparaison des méthodes alternatives pour l'enquête sur la consommation alimentaire de 1989. In: Actes des journées de méthodologie statistique, 15 et 16 décembre 1993, Insee-Méthodes No 56-57-58, Complément, pp 9–42

Folsom RE, Singh AC (2000) The generalized exponential model for design weight calibration for extreme values, nonresponse and poststratification. In: Section on survey research methods, American Statistical Association, pp 598–603

Graf M (2016) Tight calibration of survey weights. institut de Statistique, Université de Neuchâtel

Guandalini A, Tillé Y (2017) Design-based estimators calibrated on estimated totals from multiple surveys. Int Stat Rev 85:250–269

Haziza D, Beaumont JF (2017) Construction of weights in surveys: a review. Stat Sci 32(2):206–226

Huang ET, Fuller WA (1978) Non-negative regression estimation for sample survey data. In: Proceedings of the social statistics section of the American Statistical Association, pp 300–305

Iannacchione VG, Milne JG, Folsom RE (1991) Response probability weight adjustments using logistic regression. In: Proceedings of the survey research methods section. American Statistical Association, pp 637–42

Kott PS (2006) Using calibration weighting to adjust for nonresponse and coverage errors. Surv Methodol 32:133–142

Kott PS, Liao D (2012) Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine. Surv Res Methods 6:105–111

Le Guennec J, Sautory O (2002) CALMAR2: une nouvelle version de la macro CALMAR de redressement d'échantillon par calage. Actes des Journées de Méthodologie. Insee, Paris, pp 33–38

Lesage É (2011) The use of estimating equations to perform a calibration on complex parameters. Surv Methodol 37(1):103–108

Lesage É, Haziza D, D'Haultfoeuille X (2019) A cautionary tale on instrumental calibration for the treatment of nonignorable unit nonresponse in surveys. J Am Stat Assoc 114:906–915

Lundström S, Särndal CE (1999) Calibration as a standard method for treatment of nonresponse. J Off Stat 15:305–327

Montanari GE, Ranalli MG (2005) Nonparametric model-calibration estimation in survey sampling. J Am Stat Assoc 100:1429–1442

Rebecq A (2017) icarus: Calibrates and Reweights Units in Samples, likelihood, estimation. R package version (3)

Rebecq A (2019) Méthodes de sondage pour les données massives. PhD thesis, Université Paris Nanterre

Rényi A (1961) On measures of entropy and information. In: Proceedings of the 4th Berkeley symposium on mathematical statistics and probability, University of California Press, Berkeley, vol 1, pp 547–561

Särndal CE, Lundström S (2005) Estimation in surveys with nonresponse. Wiley, New York

Vallée AA, Tillé Y (2019) Linearization for variance estimation by means of sampling indicators: application to nonresponse. Int Stat Rev 87(2):347–367

Vanderhoeft C (2001) Generalised calibration at Statistics Belgium SPSS module gCALIBS and current practices. Technical report, Statistics Belgium Working Paper No. 3

Wu C, Sitter RR (2001) A model-calibration approach to using complete auxiliary information from survey data. J Am Stat Assoc 96:185–193