A note on joint calibration estimators for totals and quantiles*

Abstract

In this paper, we combine calibration for population totals proposed by Deville and Särndal (1992) with calibration for population quantiles introduced by Harms and Duchesne (2006). We also extend the pseudo-empirical likelihood method proposed by Chen, Sitter, and Wu (2002). This approach extends the calibration equations for totals by adding relevant constraints on quantiles of continuous variables observed in the data. The proposed approach can be easily applied to handle non-response and data integration problems, and results in a single vector of weights. Furthermore, it is a multipurpose solution, i.e. it makes it possible to improve estimates of means, totals and quantiles for a variable of interest in one step. In a limited simulation study, we compare the proposed joint approach with standard calibration, calibration using empirical likelihood and the correctly specified inverse probability weighting estimator. Open source software implementing the proposed method is available.

Keywords: calibration approach, quantile estimation, inverse probability weighting estimation, pseudo-empirical likelihood method.

---

*Maciej Beręsewicz ((1) Poznań University of Economics and Business, Institute of Informatics and Quantitative Economics, Department of Statistics, Al. Niepodległości 10, 61-875 Poznań, Poland; (2) Statistical Office in Poznań, ul. Wojska Polskiego 27/29 60-624 Poznań, Poland. E-mail: maciej.beresewicz@ue.poznan.pl), Marcin Szymkowiak ((1) Poznań University of Economics and Business, Institute of Informatics and Quantitative Economics, Department of Statistics, Al. Niepodległości 10, 61-875 Poznań, Poland; (2) Statistical Office in Poznań, ul. Wojska Polskiego 27/29 60-624 Poznań, Poland

1 Introduction

Calibration weighting is a method commonly used in survey sampling to adjust original design weights for sampled elements to reproduce known population totals for all auxiliary variables (Deville & Särndal, 1992). Following the calibration paradigm, it can also be used to reproduce known population quantiles for all benchmark variables (Harms & Duchesne, 2006). This technique is also used in surveys to compensate for nonsampling errors, such as nonresponse or coverage errors (Särndal & Lundström, 2005). By appropriately adjusting the weights, it is not only possible to ensure consistency with known structures for key variables from other data sources, such as censuses or registers, but also to reduce the bias and improve the precision of final estimates.

In this article, we propose a joint calibration approach to estimate the total or quantile of order α for the variable of interest y. Final calibration weights w_k reproduce known population totals and quantiles for all auxiliary variables. At the same time, they help to reduce the bias and improve the precision of estimates. The proposed method is based on the classic approach to calibration and simultaneously takes into account calibration equations for totals and quantiles of all auxiliary variables.

The paper has the following structure. In section 2 we introduce the basic setup: notation, calibration for totals and quantiles. In section 3 we describe the procedure of joint calibration of totals and quantiles. In section 4 we describe results of a simulation study based on Chen, Li, and Wu (2020). Section 5 summarises the paper.

2 Theoretical basis

2.1 Calibration estimator for a total

In most applications the goal is to estimate a finite population total $\tau_y = \sum_{k \in U} y_k$ or the mean $\bar{\tau}_y = \tau_y/N$ of the variable of interest y, where U is the population of size N. The well-known estimator of a finite population total is the Horvitz-Thompson estimator, which is expressed as $\hat{\tau}_{y\pi} = \sum_{k=1}^{n} d_k y_k = \sum_{k \in s} d_k y_k$, where s denotes a probability sample of size n, $d_k = 1/\pi_k$ is a design weight and π_k is the first-order inclusion probability of the i-th element of the population U. This estimator is unbiased for τ_Y i.e. $E(\hat{\tau}_{y\pi}) = \tau_Y$.

Let $\mathbf{x}_k^{\circ}$ be a J_1-dimensional vector of auxiliary variables (benchmark variables) for which $\tau_{\mathbf{x}} = \sum_{k \in U} \mathbf{x}_k^{\circ} = \left(\sum_{k \in U} x_{k1}, \ldots, \sum_{k \in U} x_{kJ_1} \right)^T$ is assumed to be known. In most cases in practice the d_k weights do not reproduce known population totals for benchmark variables $\mathbf{x}_k^{\circ}$. It means that the resulting estimate $\hat{\tau}_{\mathbf{x}\pi} = \sum_{k \in s} d_k \mathbf{x}_k^{\circ}$ is not equal to $\tau_{\mathbf{x}}$. The main idea of calibration is to look for new calibration weights w_k which are as close as possible to original design weights d_k and reproduce known population totals $\tau_{\mathbf{x}}$ exactly. In other words, in order to find new calibration weights w_k we have to minimise a distance function $D\left(\mathbf{d}, \mathbf{v}\right) = \sum_{k \in s} d_k G\left(\frac{v_k}{d_k}\right) \to \min$ to fulfil calibration equations $\sum_{k \in s} v_k \mathbf{x}_k^{\circ} = \sum_{k \in U} \mathbf{x}_k^{\circ}$, where $\mathbf{d} = \left(d_1, \ldots, d_n\right)^T$, $\mathbf{v} = \left(v_1, \ldots, v_n\right)^T$ and $G\left(\cdot\right)$ is a function which must satisfy some regularity conditions: $G\left(\cdot\right)$ is strictly convex and twice continuously differentiable, $G\left(\cdot\right) \geq 0$, $G\left(1\right) = 0$, $G'\left(1\right) = 0$ and $G''\left(1\right) = 1$. Examples of $G\left(\cdot\right)$ functions are given by Deville and Särndal (1992). For instance, if $G\left(x\right) = \frac{(x-1)^2}{2}$, then using the method of Lagrange multipliers the final calibration weights w_k can be expressed as $w_k = d_k + d_k \left(\tau_{\mathbf{x}} - \hat{\tau}_{\mathbf{x}\pi}\right)^T \left(\sum_{j \in s} d_j \mathbf{x}_j^{\circ} \mathbf{x}_j^{\circ T}\right)^{-1} \mathbf{x}_k^{\circ}$. It is worth adding that in order to avoid negative or large w_k weights in the process of minimising the $D\left(\cdot\right)$ function, one can consider some boundary constraints $L \leq \frac{w_k}{d_k} \leq U$, where $0 \leq L \leq 1 \leq U$, $k = 1, \ldots, n$. The final calibration estimator of a population total τ_y can be expressed as $\hat{\tau}_{y\mathbf{x}} = \sum_{k \in s} w_k y_k$, where w_k are calibration weights obtained under a specific chosen $G\left(\cdot\right)$ function.

2.2 Calibration estimator for a quantile

Harms and Duchesne (2006) considered the estimation of quantiles using the calibration approach in a way very similar to the what Deville and Särndal (1992) proposed for a finite population total τ_y. By analogy, in their approach it is not necessary to know values for all auxiliary variables for all units in the population. It is enough to know the corresponding quantiles for the benchmark variables. We will briefly discuss the problem of finding calibration weights in this setup.

We want to estimate a quantile $Q_{y,\alpha}$ of order $\alpha \in (0, 1)$ of the variable of interest y, which can be expressed as $Q_{y,\alpha} = \inf \left\{t \,|\, F_y\left(t\right) \geq \alpha\right\}$, where $F_y\left(t\right) = N^{-1} \sum_{k \in U} H\left(t - y_k\right)$ and the

Heavyside function is given by

$$H\left(t-y_k\right)=\begin{cases}1,&t\geq y_k,\\[2mm]0,&t<y_k.\end{cases}\tag{1}$$

We assume that $\mathbf{Q}_{\mathbf{x},\alpha}=\left(Q_{x_1,\alpha},\ldots,Q_{x_{J_2},\alpha}\right)^T$ is a vector of known population quantiles of order α for a vector of auxiliary variables $\mathbf{x}_k^*$, where $\alpha\in(0,1)$ and $\mathbf{x}_k^*$ is a J_2-dimensional vector of auxiliary variables. It is worth noting that, in general, the numbers J_1 and J_2 of the auxiliary variables are different. It may happen that for a specific auxiliary variable its population total and the corresponding quantile of order α will be known. However, in most cases quantiles will be known for continuous auxiliary variables, unlike totals, which will generally be known for categorical variables. In order to find new calibration weights w_k which reproduce known population quantiles in a vector $Q_{\mathbf{x},\alpha}$, an interpolated distribution function estimator of $F_y\left(t\right)$ is defined as $\hat{F}_{y,cal}(t)=\frac{\sum_{k\in s}w_k H_{y,s}(t,y_k)}{\sum_{k\in s}w_k}$, where the Heavyside function in formula (1) is replaced by the modified function $H_{y,s}\left(t,y_k\right)$ given by

$$H_{y,s}\left(t,y_k\right)=\begin{cases}1,&y_k\leqslant L_{y,s}(t),\\[2mm]\beta_{y,s}\left(t\right),&y_k=U_{y,s}\left(t\right),\\[2mm]0,&y_k>U_{y,s}\left(t\right),\end{cases}\tag{2}$$

where $L_{y,s}\left(t\right)=\max\left\{\{y_k,k\in s\mid y_k\leqslant t\}\cup\{-\infty\}\right\}$, $U_{y,s}\left(t\right)=\min\left\{\{y_k,k\in s\mid y_k>t\}\cup\{\infty\}\right\}$ and $\beta_{y,s}\left(t\right)=\frac{t-L_{y,s}(t)}{U_{y,s}(t)-L_{y,s}(t)}$ for $k=1,\ldots,n,\ t\in\mathbb{R}$. A calibration estimator of quantile $Q_{y,\alpha}$ of order α for variable y is defined as $\hat{Q}_{y,cal,\alpha}=\hat{F}_{y,cal}^{-1}(\alpha)$, where a vector $\mathbf{w}=\left(w_1,\ldots,w_n\right)^T$ is a solution of optimization problem $D\left(\mathbf{d},\mathbf{v}\right)=\sum_{k\in s}d_k G\left(\frac{v_k}{d_k}\right)\to\min$ subject to the calibration constraints $\sum_{k\in s}v_k=N$ and $\hat{\mathbf{Q}}_{\mathbf{x},cal,\alpha}=\left(\hat{Q}_{x_1,cal,\alpha},\ldots,\hat{Q}_{x_{J_2},cal,\alpha}\right)^T=\mathbf{Q}_{\mathbf{x},\alpha}$ or equivalently $\hat{F}_{x_j,cal}\left(Q_{x_j,\alpha}\right)=\alpha$, where $j=1,\ldots,J_2$.

As in the previous case, if $G\left(x\right)=\frac{(x-1)^2}{2}$ then using the method of Lagrange multipliers the final calibration weights w_k can be expressed as $w_k=d_k+d_k\left(\mathbf{T_a}-\sum_{k\in s}d_k\mathbf{a}_k\right)^T\left(\sum_{j\in s}d_j\mathbf{a}_j\mathbf{a}_j^T\right)^{-1}\mathbf{a}_k,$

where $\mathbf{T_a} = (N, \alpha, \ldots, \alpha)^T$ and the elements of $\mathbf{a}_k = (1, a_{k1}, \ldots, a_{kJ_2})^T$ are given by

$$
a_{kj} = \begin{cases}
N^{-1}, & x_{kj} \leq L_{x_j,s}\left(Q_{x_j,\alpha}\right), \\
N^{-1}\beta_{x_j,s}\left(Q_{x_j,\alpha}\right), & x_{kj} = U_{x_j,s}\left(Q_{x_j,\alpha}\right), \\
0, & x_{kj} > U_{x_j,s}\left(Q_{x_j,\alpha}\right),
\end{cases} \tag{3}
$$

with $j = 1, \ldots, J_2$.

Assuming that $y_1 \leq y_2 \ldots \leq y_n$ it can be shown that if there exists $p \in \{1, \ldots, n-1\}$ such that $\hat{F}_{y,cal}(y_p) \leq \alpha$, $\hat{F}_{y,cal}(y_{p+1}) > \alpha$ and $\hat{F}_{y,cal}$ is invertible at point $\hat{Q}_{y,cal,\alpha}$ then the calibration estimator $\hat{Q}_{y,cal,\alpha}$ of quantile $Q_{y,\alpha}$ of order $\alpha \in (0,1)$ can be expressed as $\hat{Q}_{y,cal,\alpha} = y_p + \frac{N\alpha - \sum_{i=1}^p w_i}{w_{p+1}}(y_{p+1} - y_p)$.

3 Proposed method

We propose a simple method that jointly calibrates weights for totals and quantiles. The resulting calibrated weights w_k will allow us to retrieve known population totals and quantiles of auxiliary variables simultaneously. In the case of a single scalar auxiliary variable x, the final calibration estimator based on weights w_k delivers an exact population total and quantile for variable y when the relationship between y and x is exactly linear i.e. when $y_k = \beta x_k$ for all $k \in U$.

Let us assume that we are interested in estimating a population total τ_y and/or quantile $Q_{y,\alpha}$ of order α, where $\alpha \in (0,1)$ for variable of interest y. Let $\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_k^\circ \\ 1 \\ \mathbf{x}_k^* \end{pmatrix}$ be a $J+1$-dimensional vector of auxiliary variables, where $J = J_1 + J_2$. We assume that for J_1 variables a vector of population totals $\tau_\mathbf{x}$ is known and for J_2 variables a vector $\mathbf{Q}_{\mathbf{x},\alpha}$ of population quantiles is known. In practice it may happen that for the same auxiliary variable we know its population total and quantile. We do not require that the complete auxiliary information described by the vector $\mathbf{x}_k$ is known for all $k \in U$; however, for some auxiliary variables unit-population data would be necessary, because accurate quantiles are not likely to be known from other sources (Särndal, 2007).

Our main aim is to find new calibration weights w_k which are as close as possible to original design weights d_k and for some auxiliary variables reproduce known population totals and for the remaining benchmark variables – reproduce known population quantiles exactly. In

our joint approach we are looking for a vector $\mathbf{w} = (w_1, \ldots, w_n)^T$ which is a solution of the optimization problem $D(\mathbf{d}, \mathbf{v}) = \sum_{k \in s} d_k\, G\left(\frac{v_k}{d_k}\right) \to \min$ subject to the calibration constraints $\sum_{k \in s} v_k = N$, $\sum_{k \in s} v_k \mathbf{x}_k^\circ = \tau_\mathbf{x}$ and $\hat{\mathbf{Q}}_{\mathbf{x},cal,\alpha} = \mathbf{Q}_{\mathbf{x},\alpha}$. In general, $J+1$ calibration equations have to be fulfilled. Alternatively, calibration equations can be expressed as $\sum_{k \in s} v_k \mathbf{x}_k^\circ = \tau_\mathbf{x}$ and $\sum_{k \in s} v_k \mathbf{a}_k = \mathbf{T_a}$ with possible boundary constraints on calibration weights.

Assuming a quadratic metric $D(\cdot)$, which is based on $G(x) = \frac{(x-1)^2}{2}$ function, an explicit solution of the above optimization problem can be derived. This solution is similar to the calibration weights for totals and quantiles. Let $\mathbf{h_x} = \binom{\tau_\mathbf{x}}{\mathbf{T_a}}$ and $\hat{\mathbf{h}}_\mathbf{x} = \binom{\sum_{k \in s} d_k \mathbf{x}_k^\circ}{\sum_{k \in s} d_k \mathbf{a}_k}$. Then the vector of calibration weights $\mathbf{w} = (w_1, \ldots, w_n)^T$ which solves the above optimization problem satisfies the relation:

$$w_k = d_k + d_k \left(\mathbf{h_x} - \hat{\mathbf{h}}_\mathbf{x}\right)^T \left(\sum_{j \in s} d_j \mathbf{x}_j \mathbf{x}_j^T\right)^{-1} \mathbf{x}_k. \tag{4}$$

Remark 1: In our proposed method we assume that we reproduce known population totals and known population quantiles for a set of benchmark variables simultaneously. This approach can be easily extended by assuming that for estimated totals or quantiles are reproduced for some auxiliary variables. Moreover, we assumed that the process of calibration is based on a particular quantile (of order α). For instance, it could be a median i.e. $\alpha = 0.5$. Nothing stands in the way of searching for calibration weights which reproduce population totals and a set of population quantiles (for example quartiles) for a chosen set of auxiliary variables. Moreover, the proposed method can be easily extended for the generalized calibration, in particular, for not missing at random non-response (Kott & Chang, 2010), or empirical likelihood by adding additional constraints on quantiles, i.e. $\sum_{k \in s} p_k a_{kj} = \frac{\alpha}{N}$, where $j = 1, \ldots, J_2$ and p_k are elements of the vector $\mathbf{p} = (p_1, \ldots, p_n)^T$ which is is a discrete probability measure over the sample s (Wu & Thompson, 2020).

Remark 2: In our approach we use an interpolated distribution function $H_{y,s}$, which is a simple modification of the Heavyside function defined in (1). From a practical point of view a smooth approximation to the step function, based on the logistic function can be used i.e. $H(x) \approx \frac{1}{2} + \frac{1}{2}\tanh kx = \frac{1}{1+e^{-2kx}}$, where a larger value of k corresponds to a sharper transition at $x = 0$.

Remark 3: The proposed method can be easily applied to data from household surveys, where integrated calibration is applied, i.e. when the weights of particular household members should

be equal. This can be in particularly useful in the case of EU-SILC, where information from administrative data can be used to provide population distributions for auxiliary variables.

4   Simulation study based on Chen, Li, and Wu (2020)

In this section, we present results of a design-based simulation study using Chen, Li, and Wu (2020) setting. Our population U is equal to $N = 20,000$, and the observed set (non-probability sample), which suffers from selection bias, is equal to $n = 10,000$. We generate auxiliary variables $x_{k1} = z_{k1}$, $x_{k2} = z_{k2} + 0.3x_{k1}$, $x_{k3} = z_{k3} + 0.2(x_{k1} + x_{k2})$ and $x_{k4} = z_{k4} + 0.1(x_{k1} + x_{k2} + x_{k3})$ where $z_{k1} \sim \text{Bernoulli}(0.5)$, $z_{k2} \sim \text{Uniform}(0,2)$, $z_{k3} \sim \text{Exp}(1)$ and $z_{k4} \sim \chi^2(4)$ for all $k \in U$. The target variable y is generated using the following linear model

$$y_k = 2 + x_{k1} + x_{k2} + x_{k3} + x_{k4} + \sigma\varepsilon_k, \quad k = 1, 2, \ldots, N,$$

where ε_i's are independent and identically distributed random variables (i.i.d.) from normal distribution $\mathcal{N}(0,1)$. Values of σ are chosen such that the correlation coefficient ρ between y and the linear predictor $\mathbf{x}^T\beta$ is controlled at $0.3, 0.5$ and 0.8 (such variables y are denoted as y_1, y_2, and y_3 respectively). The true propensity scores π_k for the non-probability sample follow the logistic regression model:

$$\log\left(\frac{\pi_k}{1 - \pi_k}\right) = \theta_0 + 0.1x_{k1} + 0.2x_{k2} + 0.1x_{k3} + 0.2x_{k4},$$

where θ_0 is chosen such that $\sum_{k \in U} \pi_k = n$. The non-probability sample is selected by the Poisson sampling method with inclusion probabilities specified by π_k and the target sample size n.

We conducted a simulation study where population totals were known for all auxiliary variables $x_1 - x_4$, while population quartiles Q_1, Q_2 and deciles $D_1, \ldots, D_9$ were known for $x_2 - x_4$ only. For all calibration estimators, we used raking (exponential function) to ensure positive weights:

- $\hat{\mu}_{\text{Naive}}$ – sample mean or quantile based on a sample n,

- $\hat{\mu}_{IPW}$ – we used all auxiliary variables assuming logistic regression with calibration con-

straints (as in Chen, Li, and Wu (2020, sec 3.1)) ,

- $\hat{\mu}_{CAL}$ – calibration using population size N and totals for all benchmark variables i.e. $\tau_{\mathbf{x}} = (N, \tau_{x_1}, \tau_{x_2}, \tau_{x_3}, \tau_{x_4})^T$,

- $\hat{\mu}_{QCAL1}$ – calibration using population quartiles of order 0.25 and 0.75 and population deciles i.e. quantiles of order 0.1 to 0.9 for auxiliary variables x_2, x_3, x_4, and population size N,

- $\hat{\mu}_{QCAL2}$ – calibration using population size and totals for variables $x_1 - x_4$, population quartiles of order 0.25 and 0.75 and deciles for auxiliary variables $x_2 - x_4$,

- $\hat{\mu}_{EL}$ – calibration using empirical likelihood (EL) using constraints on means ($\sum_{k \in s} p_k \mathbf{x}_k^{\circ} = \frac{\tau_{\mathbf{x}}}{N}$) of auxiliary variables $x_1 - x_4$,

- $\hat{\mu}_{QEL1}$ – calibration using EL using constraints on quartiles of order 0.25 and 0.75 and deciles ($\sum_{k \in s} p_k a_{kj} = \frac{\alpha}{N}$) of auxiliary variables $x_2 - x_4$,

- $\hat{\mu}_{QEL2}$ – calibration using EL combining constraints described by $\hat{\mu}_{EL}$ and $\hat{\mu}_{QEL1}$.

In the results, we report bias, standard error (SE) and relative mean square error (RMSE) based on $R = 1,000$ Monte Carlo simulations for each y_1, y_2 and y_3 variables: Bias $= \bar{\hat{\mu}} - \mu$, SE $= \sqrt{\frac{\sum_{r=1}^{R} (\hat{\mu}^{(r)} - \bar{\hat{\mu}})^2}{R-1}}$ and RMSE $= \sqrt{\text{Bias}^2 + \text{SE}^2}$, where $\bar{\hat{\mu}} = \sum_{r=1}^{R} \hat{\mu}^{(r)}/R$ and $\hat{\mu^{(r)}}$ is an estimate of mean/quantile in r-th replication. We present results of our simulation study in Table 4.1 for the calibration estimator of the mean and quartiles of variable of interest y (for three variants).

The simulation results can be summarized as follows.

1. All considered calibration estimators, as expected, yield better results in terms of bias compared to $\hat{\mu}_{\text{Naive}}$. As regards the estimated mean, one can see that $\hat{\mu}_{QCAL2}$, $\hat{\mu}_{EL}$ and $\hat{\mu}_{QEL2}$ are nearly unbiased for all cases.

2. Of all calibration estimators based on the classical calibration paradigm ($\hat{\mu}_{QCAL}$, $\hat{\mu}_{QCAL1}$ and $\hat{\mu}_{QCAL2}$) the proposed joint calibration estimator $\hat{\mu}_{QCAL2}$ has the best statistical properties. The simultaneous consideration of population totals and quantiles for auxiliary variables contributes to the reduction of the estimator bias.

8

Table 4.1: Simulation results: bias, SE and RMSE for y_1, y_2 and y_3 (all numbers multiplied by 100)

Estimator	$y_1 (\rho = 0.3)$			$y_2 (\rho = 0.5)$			$y_3 (\rho = 0.8)$		
	Bias	SE	RMSE	Bias	SE	RMSE	Bias	SE	RMSE
				Mean					
$\hat{\mu}_{\text{Naive}}$	94.58	7.13	94.85	93.92	4.21	94.02	93.48	2.54	93.52
$\hat{\mu}_{IPW}$	-0.17	7.86	7.86	-0.09	4.26	4.26	-0.04	1.84	1.84
$\hat{\mu}_{CAL}$	1.10	7.74	7.82	0.60	4.20	4.24	0.26	1.81	1.83
$\hat{\mu}_{QCAL1}$	3.75	7.84	8.69	4.04	4.28	5.88	4.24	1.90	4.64
$\hat{\mu}_{QCAL2}$	0.12	7.85	7.85	0.07	4.25	4.25	0.03	1.84	1.84
$\hat{\mu}_{EL}$	0.06	7.82	7.82	0.03	4.24	4.24	0.01	1.83	1.83
$\hat{\mu}_{QEL1}$	3.37	7.88	8.57	3.84	4.29	5.76	4.16	1.91	4.58
$\hat{\mu}_{QEL2}$	-0.24	7.88	7.89	-0.13	4.28	4.28	-0.06	1.85	1.85
				1th Quartile (25%)					
$\hat{\mu}_{\text{Naive}}$	94.52	11.24	95.18	89.55	6.82	89.81	79.04	3.37	79.11
$\hat{\mu}_{IPW}$	0.35	12.16	12.16	-0.14	6.89	6.89	-0.17	3.62	3.62
$\hat{\mu}_{CAL}$	5.06	12.20	13.21	6.27	6.56	9.08	9.35	3.89	10.13
$\hat{\mu}_{QCAL1}$	4.91	12.37	13.31	4.21	6.73	7.94	4.84	3.77	6.14
$\hat{\mu}_{QCAL2}$	2.20	12.21	12.40	2.31	6.80	7.18	3.31	3.64	4.92
$\hat{\mu}_{EL}$	1.34	12.17	12.24	1.35	6.84	6.97	1.86	3.61	4.06
$\hat{\mu}_{QEL1}$	3.45	12.38	12.85	2.51	6.84	7.29	2.29	3.68	4.34
$\hat{\mu}_{QEL2}$	0.68	12.24	12.26	0.30	6.90	6.90	0.44	3.59	3.62
				2th Quartile (50%)					
$\hat{\mu}_{\text{Naive}}$	90.01	8.95	90.45	95.27	5.25	95.41	92.23	2.82	92.27
$\hat{\mu}_{IPW}$	-0.46	9.36	9.37	-0.34	5.07	5.08	-0.14	3.17	3.17
$\hat{\mu}_{CAL}$	2.54	9.19	9.53	2.75	4.94	5.65	5.67	2.82	6.34
$\hat{\mu}_{QCAL1}$	3.55	9.43	10.07	3.49	5.06	6.15	4.18	2.91	5.09
$\hat{\mu}_{QCAL2}$	0.92	9.38	9.42	0.95	5.00	5.09	1.99	3.01	3.61
$\hat{\mu}_{EL}$	0.39	9.27	9.28	0.80	4.99	5.06	2.31	2.99	3.78
$\hat{\mu}_{QEL1}$	2.52	9.40	9.73	2.93	5.06	5.85	3.43	2.96	4.53
$\hat{\mu}_{QEL2}$	-0.17	9.40	9.40	0.19	5.05	5.05	0.99	3.08	3.23
				3th Quartile (75%)					
$\hat{\mu}_{\text{Naive}}$	94.56	9.34	95.02	104.29	5.60	104.44	104.51	3.96	104.58
$\hat{\mu}_{IPW}$	-1.24	11.18	11.24	-0.44	5.78	5.80	0.09	3.13	3.14
$\hat{\mu}_{CAL}$	-1.86	11.27	11.42	-2.58	5.74	6.29	-2.20	3.02	3.74
$\hat{\mu}_{QCAL1}$	2.62	10.83	11.14	3.18	5.74	6.56	2.86	3.10	4.22
$\hat{\mu}_{QCAL2}$	-1.47	11.24	11.34	-1.22	5.78	5.91	-1.05	3.00	3.18
$\hat{\mu}_{EL}$	-1.19	11.19	11.25	-0.46	5.75	5.77	0.97	3.16	3.31
$\hat{\mu}_{QEL1}$	2.84	10.78	11.15	4.05	5.70	6.99	4.64	2.89	5.47
$\hat{\mu}_{QEL2}$	-1.28	11.21	11.28	-0.30	5.76	5.77	0.79	3.07	3.17

3. All considered estimators yield similar results in terms of the efficiency measured by SE. The standard errors decrease as the correlation coefficient ρ increases.

4. The proposed joint approach is nearly as efficient as the correctly specified IPW estimator.

5 Summary

In this article we have presented a simple and joint calibration estimator for totals and quantiles based on the calibration paradigm. More precisely, using a set of auxiliary variables we are able to create a single vector of calibration weights which reproduce known population totals for some benchmark variables and known population quantiles for other variables exactly. This vector can be used to directly find a total and a quantile of the variable of interest y.

In a small simulation study we compared the calibration estimator of the mean and quartiles of variable y based on calibration weights obtained under different settings with other estimators using pseudo-empirical likelihood and inverse probability weighting methods. The proposed estimator based on a single set of calibration weights which reproduce known population totals and reproduce exactly quantiles for auxiliary variables performed well in our simulation. By combining information about population totals and population quantiles for auxiliary variables in a process of calibration, it is possible to decrease the bias and RMSE in comparison to the classical estimator of a total and a quantile, where calibration equations are fulfilled for population totals and population quantiles separately.

References

Chen, J., Sitter, R., & Wu, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, *89*(1), 230–237.

Chen, Y., Li, P., & Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, *115*(532), 2011–2021.

Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, *87*(418), 376–382.

Harms, T., & Duchesne, P. (2006). On calibration estimation for quantiles. *Survey Methodology*, *32*(1), 37–52.

Kott, P. S., & Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association*, *105*(491), 1265–1275.

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

Särndal, C.-E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.

Wu, C., & Thompson, M. E. (2020). *Sampling theory and practice*. Springer.