The optimization problem of quantile and poverty measures estimation based on calibration.

S. Martínez[a], M. Rueda[b,], M. Illescas[a]

[a]*Department of Mathematics. University of Almería, Spain*
[b]*Department of Statistics and Operational Research. University of Granada, Spain*

Abstract

New calibrated estimators of quantiles and poverty measures are proposed. These estimators combine the incorporation of auxiliary information provided by auxiliary variables related to the variable of interest by calibration techniques with the selection of optimal calibration points under simple random sampling without replacement. The problem of selecting calibration points that minimize the asymptotic variance of the quantile estimator is addressed. Once the problem is solved, the definition of the new quantile estimator requires that the optimal estimator of the distribution function on which it is based verifies the properties of the distribution function. Through a theorem, the nondecreasing monotony property for the optimal estimator of the distribution function is established and the corresponding optimal estimator can be defined. This optimal quantile estimator is also used to define new estimators for poverty measures. Simulation studies with real data from the Spanish living conditions survey compares the performance of the new estimators against various methods proposed previously, where some resampling techniques are used for the variance estimation. Based on the results of the simulation study, the proposed estimators show a good performance and are a reasonable alternative to other estimators.

Keywords: Optimization, calibration technique, poverty measure estimates, survey sampling
2010 MSC: 62D05

1. Introduction

Quantile estimation is a issue of great interest because some measures and indicators depend on quantiles in many fields of research such as health science ([41]); anthropology ([6]) or economics ([14]). More specifically, in the field of economics,

studies on the analysis of poverty and social exclusion have an increasing importance for governments and society in general, since some poverty measures, like the proportion of people (or households) in poverty, are important measures of the country's overall economic welfare. Many indicators used in the poverty studies are based on quantiles, since they analyze variables with skewed distributions such as income, and in such cases the median is more suitable location measure than the mean. Thus, one of the commonly used measures in the poverty analysis is the poverty line that allows dividing the population into poor and nonpoor and that, for example, Eurostat fixes as 60% of the median of the equivalent net income. Additionally, poverty studies incorporate the analysis of wage inequality and income distribution, whose measurement is often based on percentile ratios, such as 50th/5th and 50th/25th ([16]); 50th/10th ([33],[24], [31], [8]); 95th/50th ([24], [8]) and 90th/10th; 95th/20th; and 80th/20th ([22]).

In official surveys of living conditions, in social surveys and in sample surveys in general, auxiliary information is often available through additional variables related to the study variable. When auxiliary information is available, there are several alternative methods for incorporating it into the estimation phase and obtaining more efficient estimators ([15]; [12]; [9]); [17]). These procedures have been applied to estimate the population mean ([35]; [38]), the distribution function ([9]; [17]; [40]; [30]) quantiles ([20]; [13]) and poverty measures [32]. Particularly, in the case of estimation of quantiles, the auxiliary information can be incorporated by means of indirect estimators. In this case, it is necessary to have the equivalent quantile of the auxiliary variable for a given quantile of the study variable ([23]; [34]). Another possibility considers the incorporation of the auxiliary information to obtain estimators of the distribution function and to obtain the estimation of the quantile through the inverse function ([9]; [17]). This procedure requires that the estimator of the distribution function fulfills the distribution function's properties. Thus, based on this option, [37] obtained quantile estimators based on calibration framework described in [36]. Similarly, also based on the same calibration framework, [26] developed post-stratified quantile estimators. The main advantage of the framework proposed in [36] is that the obtained estimators are genuine distribution functions[1] under some conditions. One drawback of these esti-

---

[1]For an estimator $\hat{F}(t)$ of $F(t)$ to be a genuine distribution function it should be monotonic increasing and such that $\hat{F}(-\infty) = 0$ and $\hat{F}(+\infty) = 1$

mators, is that their efficiency depends on the selection of some calibration points t_i. Recenlty, under simple random sampling, the problem of optimal selection points in order to obtain the best estimation is treated in [26]-[29]. Unfortunately, the quantile estimation through the estimation of the distribution function needs the estimation for all value t and the optimal selection of auxiliary points depends on the point t in which we want to estimate the distribution function. This implies that the distribution function estimators based on optimal choice, in general, are not monotonous non-decreasing and may take values beyond the range $[0, 1]$.

In this work, we will adapt and employ the optimal selection proposals in [29] in the estimation of quantiles. We show that the problem of optimizing the variance of a quantile estimator is equivalent to the optimization of the variance of the distribution function estimator at one point. We demonstrate that under certain conditions, the estimators obtained through the optimal selection proposed in [29] meet the distribution function properties and can be directly used in the quantile estimation. Due to the complexity of the quantile estimation and the optimal selection for calibration estimators, a practical mathematical expression for the variances of the quantile estimator could not be established. Thus, some resampling techniques will be employed to obtain variance estimation of the quantile estimators proposed. Finally, in this work we will define new percentile ratio estimators that can be applied in the estimation of poverty measures.

The remainder of the article is organized in four sections. After introducing the problem of quantile estimation in Section 2, in Section 3, new calibration quantiles estimators are proposed based on optimal selection points for the estimation of distribution function. In Section 4, we propose the use of resampling techniques for the variance estimation of the quantile estimators proposed in Section 4. The application of the optimal quantile estimators in poverty measures estimation is done in Section 5. Section 6 includes two simulation studies based on real survey data obtained from the Spanish living conditions survey in order to analyse the performance of quantile estimators and poverty measure estimators proposed in this work. Finally, Section 7 presents the concluding remarks.

2. Estimation of the distribution function and quantiles in survey sampling

Consider a finite population $U = \{1, \dots, N\}$ with N different units where a sampling design $p(\cdot)$ is defined with first and second-order inclusion probabilities $\pi_k > 0$

4

and $\pi_{kl} > 0 \ k, l \in U$. A random sample $s = \{1, 2, \ldots, n\}$ of fixed size n is selected according to the sampling design $p(\cdot)$ and $d_k = \pi_k^{-1}$ denotes the sampling design-basic weight for unit $k \in U$ which is known. We denote by y_k the study variable and by x_k a vector of auxiliary variables at unit k. The values x_k are assumed to be known for all population units but the value y_k is assumed to be known only if the sample s includes the kth unit. The distribution function $F_y(t)$ of the study variable y is given by

$$F_y(t) = \frac{1}{N} \sum_{k \in U} \Delta(t - y_k) \tag{1}$$

where

$$\Delta(t - y_k) = \begin{cases} 1 & \text{si } t \geq y_k \\ 0 & \text{si } t < y_k. \end{cases}$$

Based on $F_y(t)$, the finite population α-quantile of y is defined as minimum value of t for which at least $100 \cdot \alpha\%$ of the y's values are less than or equal to that value, that is

$$Q_y(\alpha) = \inf\{t : F_y(t) \geq \alpha\} = F_y^{-1}(\alpha).$$

A general procedure to obtain an indirect estimator for $Q_y(\alpha)$ is based on the incorporation of auxiliary information in the estimation of $F_y(t)$ to obtain an estimator $\widehat{F}_y(t)$ that fulfills the distribution function's properties, that is, $\widehat{F}_y(t)$ is a genuine distribution function. Under this assumption, the quantile $Q_y(\alpha)$ can be estimated by taking the inverse of $\widehat{F}_y(t)$ in the following way:

$$\widehat{Q}_y(\alpha) = \inf\{t : \widehat{F}_y(t) \geq \alpha\} = \widehat{F}_y^{-1}(\alpha).$$

The usual estimator of the distribution function $F_y(t)$ is the Horvitz-Thompson estimator given by:

$$\widehat{F}_{YHT}(t) = \frac{1}{N} \sum_{k \in s} d_k \Delta(t - y_k). \tag{2}$$

The estimator $\widehat{F}_{YHT}(t)$ is unbiased and under simple random sampling, it verifies the distribution function properties, but generally it is not a genuine distribution function and does not use the auxiliary information provided by the vector x.

Recently, to incorporate the auxiliary information in the estimation of $F_y(t)$, some authors (([20], [36], [37], [40] and [3])) have used the calibration method in the estimation of the distribution function and quantiles. Specifically,[36] modified the estimator $\widehat{F}_{YHT}(t)$ by the calibration method. To do so, they considered a pseudo-variable $g_k = \widehat{\beta}' \mathbf{x}_k$ for $k = 1, 2, \ldots N$, where

$$\widehat{\beta} = \left(\sum_{k \in s} d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \cdot \sum_{k \in s} d_k \mathbf{x}_k y_k \tag{3}$$

5

and they replaced the basic weights d_k by new calibrated weights ω_k by means of the minimization of the chi-square distance measure

$$\Phi_s = \sum_{k \in s} \frac{(\omega_k - d_k)^2}{d_k q_k} \tag{4}$$

subject to the calibration equations

$$\frac{1}{N} \sum_{k \in s} \omega_k \Delta(t_j - g_k) = F_g(t_j) \quad j = 1, 2, \ldots, P \tag{5}$$

where q_k are known positive constants unrelated to d_k, $F_g(t_j)$ denotes the finite distribution function of the pseudo-variable g_k evaluated at the points $t_j \quad j = 1, 2, \ldots, P$ and it is assumed, with no loss in generality, that $t_1 < t_2 < \ldots t_P$.

The resulting estimator ([36]) is given by

$$\widehat{F}_{yc}(t) = \widehat{F}_{YHT}(t) + \left(F_g(\mathbf{t_g}) - \widehat{F}_{GHT}(\mathbf{t_g}) \right)' \cdot \widehat{D}(\mathbf{t_g}) \tag{6}$$

where $\widehat{F}_{GHT}(\mathbf{t_g})$ is the Horvitz-Thompson estimator of $F_g(\mathbf{t_g})$ evaluated at $\mathbf{t_g} = (t_1, \ldots, t_P)'$ and

$$\widehat{D}(\mathbf{t_g}) = T^{-1} \cdot \sum_{k \in s} d_k q_k \Delta(\mathbf{t_g} - g_k)\Delta(t - y_k)$$

assuming that the matrix T, given by

$$\sum_{k \in s} d_k q_k \Delta(\mathbf{t_g} - g_k)\Delta(\mathbf{t_g} - g_k)'$$

is nonsingular.

Under some conditions ([36]) the estimator $\widehat{F}_{yc}(t)$ is a genuine distribution function and based on this framework, [37] developed a new estimator for quantiles $Q_y(\alpha)$.

3. Optimal quantile estimators based on calibration estimation

In this section we will consider the search for quantile calibration estimators that are optimal in the sense of least error.

3.1. The optimization problem

A quantile estimator $\widehat{Q}_y(\alpha)$ can be expressed asymptotically as a linear function of the estimated distribution function evaluated at the quantile $Q_y(\alpha)$ by the Bahadur representation (see [9]):

$$\widehat{Q}_y = \frac{1}{f_y(Q_y(\alpha))}(\alpha - \widehat{F}_y(Q_y(\alpha))) + O(n^{-1/2}), \tag{7}$$

where $f_y(\cdot)$ denotes the derivative of the limiting value of $F_y(\cdot)$ as $N \longrightarrow \infty$. This linear approximation previously used by [23] and [4] helps to study the asymptotic properties of the estimator. Using this approximation we can express the asymptotic variance of $\widehat{Q}_y(\alpha)$ as

$$V_{asym}(\widehat{Q}_y(\alpha)) = \left(\frac{1}{f_y(Q_y(\alpha))} \right)^2 V(\widehat{F}_y(Q_y(\alpha)))$$

then the problem of minimizing the variance of the quantile's estimator $Q_y(\alpha)$ is the same as minimizing the variance of the estimator of the distribution function $\widehat{F}_y(Q_y(\alpha))$ on which it is based. Since the value $Q_y(\alpha)$ is unknown, it is not possible to obtain the optimal points for the estimation of $\widehat{F}_{yc}(Q_y(\alpha))$ following the approach developed in [29]. Consequently, for the optimal estimate of $Q_y(\alpha)$, we consider the optimal estimation of $F_y(t)$ for each point t.

Following [36], the asymptotic variance of $\widehat{F}_{yc}(t)$ is given by:

$$AV(\widehat{F}_{yc}(t)) = \frac{1}{N^2} \sum_{k \in U} \sum_{l \in U} \Delta_{kl}(d_k E_k)(d_l E_l) \tag{8}$$

where $E_k = \Delta(t - y_k) - \Delta(\mathbf{t_g} - g_k) \cdot D(\mathbf{t_g})$, with

$$D(\mathbf{t_g}) = \left(\sum_{k \in U} q_k \Delta(\mathbf{t_g} - g_k)\Delta(\mathbf{t_g} - g_k)' \right)^{-1} \cdot \left(\sum_{k \in U} q_k \Delta(\mathbf{t_g} - g_k)\Delta(t - y_k) \right).$$

Thus, the selection of the auxiliary vector $\mathbf{t_g}$ changes the precision of the calibration estimator $\widehat{Q}_y(\alpha)$.

Following [28], under simple random sampling without replacement, the minimization of asymptotic variance (8) is equivalent to the minimization of the function:

$$Q_t(\gamma_1, \ldots, \gamma_P) = 2NF_y(t) \cdot K_t(\gamma_P) - \sum_{j=1}^{P} \frac{\left(K_t(\gamma_j) - K_t(\gamma_{j-1}) \right)^2}{(F_g(\gamma_j) - F_g(\gamma_{j-1}))} - \left(K_t(\gamma_P) \right)^2$$

with $K_t(\gamma) = \sum_{k \in U} \Delta(\gamma - g_k)\Delta(t - y_k)$.

Under simple random sampling without replacement, the function $Q_t(\gamma_1, \ldots, \gamma_P)$ has its minimum at a vector $\mathbf{t_P} = (\gamma_{t1}, \ldots, \gamma_{tP})$ with $\gamma_{tj} \in A_t \cup B_t$, $j = 1, \ldots, P$ where

$$A_t = \{g_k : k \in U; y_k \leq t\} = \{a_1^t, a_2^t, \ldots, a_{M_t}^t\} \quad \text{with} \quad a_h^t < a_{h+1}^t \quad \text{for} \quad h = 1, \ldots, M_t - 1 \tag{9}$$

7

where M_t is the number of elements in the set A_t and

$$B_t = \{b_1^t, b_2^t, \ldots, b_M^t\}$$

with

$$b_1^t = \max_{l \in U_1}\{g_l\} \quad \text{where } U_1 = \{l \in U : g_l < a_1^t\}$$
$$b_h^t = \max_{l \in U_h}\{g_l\} \quad \text{where } U_h = \{l \in U : a_{h-1}^t \leq g_l < a_h^t\} \quad h = 2, 3, \ldots, M_t$$

and $b_h^t \leq b_{h+1}^t$ for $h = 1, \ldots, M_t - 1$.

Under simple random sampling without replacement [29] found that the auxiliary vector $\mathbf{t}_g$ has optimal dimension $P = 2m_t$ when b_1^t exists and for all $j = 2, \ldots, m_t$, $b_j^t \neq a_{j-1}^t$ and the optimal vector is given by

$$\mathbf{t_{OPT}}(t) = (b_1^t, a_1^t, \ldots, b_{m_t}^t, a_{M_t}^t). \tag{10}$$

In the case that for some values $j_1^t, j_2^t, \ldots j_{p_t}^t \in \{1, \ldots, m_t\}$; $a_{j_1-1}^t = b_{j_1}^t$ with $p_t \leq m_t$ and $j_h^t \neq j_q^t$ if $h \neq q$ the optimal dimension is given by $P = 2m_t - p_t$ and the optimal auxiliary vector $\mathbf{t_{OP}}$ is:

$$\mathbf{t_{OP}}(t) = (b_1^t, a_1^t, b_2^t, a_2^t, \ldots, b_{j_1-1}^t, a_{j_1-1}^t, a_{j_1}^t, b_{j_1+1}^t, \ldots, b_{j_h-1}^t, a_{j_h-1}^t, a_{j_h}^t, b_{j_h+1}^t, \ldots b_{M_t}^t, a_{M_t}^t).$$
$$\tag{11}$$

Generally, the optimal vector $\mathbf{t_{OPT}}(t)$ is unknown. Moreover, if its value is known, it can produce some problems when it is used with the data of a particular sample s (it can produce incompatible calibration restrictions in (5)). Thus, in a similar way to the previous cases, we consider a estimated vector $\widehat{\mathbf{t}}_{\mathbf{OP}}(t)$ based on the set A_{st} and B_{st} defined as:

$$A_{st} = \{g_k : k \in s; y_k \leq t\} = \{a_1^t, a_2^t, \ldots, a_{m_t}^t\}$$

with $a_h^t < a_{h+1}^t$ for $h = 1, \ldots, m_t - 1$ and B_{st} is defined, based on the sample s, in a similar way that B_t.

Then we define the calibration estimator for the distribution function estimator:

$$\widehat{F}_{YO}(t) = \widehat{F}_{YHT}(t) + \left(F_g(\widehat{\mathbf{t}}_{\mathbf{OP}}) - \widehat{F}_{GHT}(\widehat{\mathbf{t}}_{\mathbf{OP}})\right)' \cdot \widehat{D}(\widehat{\mathbf{t}}_{\mathbf{OP}}) \tag{12}$$

where

$$\widehat{D}(\widehat{\mathbf{t}}_{\mathbf{OP}}) = \frac{\displaystyle\sum_{k \in s} d_k q_k \Delta(\widehat{\mathbf{t}}_{\mathbf{OP}} - g_k)\Delta(t - y_k)}{\displaystyle\sum_{k \in s} d_k q_k \Delta(\widehat{\mathbf{t}}_{\mathbf{OP}} - g_k)}.$$

8

Since the optimal vector $\widehat{\mathbf{t}}_{\mathbf{OP}}$ depends on t, the estimator $\widehat{F}_{YO}(t)$ considers different calibration equations for each value of t. Consequently, the conditions developed in [36] for $\widehat{F}_{yc}(t)$ in general do not guarantee that $\widehat{F}_{YO}(t)$ is a genuine distribution function. In the next subsection we will see that $\widehat{F}_{YO}(t)$ meets the conditions of a true distribution function.

3.2. Defining the optimal quantile estimator.

In order to define the optimal quantile estimator, we must first demonstrate that the estimator $\widehat{F}_{YO}(t)$ is a genuine distribution function and a key property is nondecreasing monotony property. We consider the usual weights $q_k = 1$ (the uniform weighting is likely to dominate in applications [15]). The following theorem establish the nondecreasing monotony property for $\widehat{F}_{YO}(t)$.

Theorem. The calibration estimator $\widehat{F}_{YO}(t)$ is monotone nondecreasing.

Proof. If we consider values $t \leq z$ with $y_{[i]} \leq t \leq z < y_{[i+1]}$, and we denote by $B_{si} = B_{sy_{[i]}}$, it is clear that

$$A_{st} = A_{sz} = A_{si} \quad \text{and} \quad B_{st} = B_{sz} = B_{si}.$$

Consequently, $\widehat{\mathbf{t}}_{\mathbf{OP}}(t) = \widehat{\mathbf{t}}_{\mathbf{OP}}(z) = \widehat{\mathbf{t}}_{\mathbf{OP}}(y[i])$ and calibration weights ω_k in (5) are obtained with the same auxiliary vector for t and z and following ([36]), since $q_k = 1$ for all $k \in s$, we have $\widehat{F}_{YO}(t) \leq \widehat{F}_{YO}(z)$.

Now, we consider the case where $t \leq z$ with $y_{[i]} \leq t < y_{[i+1]}$ and $y_{[i+1]} \leq z < y_{[i+2]}; i = 1, \ldots, l-2$. For $y_{[i]} \leq t < y[i+1]$, we have:

$$A_{si} = \{a_1^i, \ldots, a_{m_i}^i\} \quad ; \quad B_{si} = \{b_1^i, \ldots, b_{m_i}^i\}.$$

We denote by $R_{si} = \{j : b_j^i = a_{j-1}^i\}$ and $\bar{R}_{si} = \{j : b_j^i \neq a_{j-1}^i\}$. It is clear that $\{1, \ldots, m_i\} = R_{si} \cup \bar{R}_{si}$.

Now, if we assume that $R_{si} = \emptyset$, then the optimal vector $\widehat{\mathbf{t}}_{\mathbf{OPT}}(t)$ is given by the sample-based version of (10) and following ([36]), the calibration estimator $\widehat{F}_{YO}(t)$ is given by:

$$\widehat{F}_{YO}(t) = \widehat{F}_{YHT}(t) + \sum_{j=1}^{2m_i} \left(F_g(t_j) - \widehat{F}_{GHT}(t_j) \right) \cdot A_i(t_j) \tag{13}$$

9

with

$$A_i(a_j^i) = \frac{\sum_{k \in s} d_k \Delta(a_j^i - g_k)\Delta(t - y_k) - \sum_{k \in s} d_k \Delta(b_j^i - g_k)\Delta(t - y_k)}{N(\widehat{F}_{GHT}(a_j^i) - \widehat{F}_{GHT}(b_j^i))}$$

$$-\frac{\sum_{k \in s} d_k \Delta(b_{j+1}^i - g_k)\Delta(t - y_k) - \sum_{k \in s} d_k \Delta(a_j^i - g_k)\Delta(t - y_k)}{N(\widehat{F}_{GHT}(b_{j+1}^i) - \widehat{F}_{GHT}(a_j^i))} =$$

$$\frac{(\widehat{k}_i(a_j^i) - \widehat{k}_i(b_j^i))}{N(\widehat{F}_{GHT}(a_j^i) - \widehat{F}_{GHT}(b_j^i))} = \frac{(\widehat{k}_i(a_j^i) - \widehat{k}_i(a_{j-1}^i))}{N(\widehat{F}_{GHT}(a_j^i) - \widehat{F}_{GHT}(b_j^i))}; \quad j = 1, \ldots, m_i \tag{14}$$

$$A_i(b_j^i) = \frac{\sum_{k \in s} d_k \Delta(b_j^i - g_k)\Delta(t - y_k) - \sum_{k \in s} d_k \Delta(a_{j-1}^i - g_k)\Delta(t - y_k)}{N(\widehat{F}_{GHT}(b_j^i) - \widehat{F}_{GHT}(a_{j-1}^i))}$$

$$-\frac{\sum_{k \in s} d_k \Delta(a_j^i - g_k)\Delta(t - y_k) - \sum_{k \in s} d_k \Delta(b_j^i - g_k)\Delta(t - y_k)}{N(\widehat{F}_{GHT}(a_j^i) - \widehat{F}_{GHT}(b_j^i))} = -A_i(a_j^i); \quad j = 1, \ldots, m_i \tag{15}$$

since it's easy to see that $\widehat{k}_i(a_{j-1}^i) = \widehat{k}_i(b_j^i)$ ($\widehat{k}$ is defined similarly to K but based on sample s) and where $\widehat{k}_i(a_0^i) = 0$ and $\widehat{k}_i(b_{m_i+1}^i) = \widehat{k}_i(a_{m_i}^i)$ as we consider $a_0^i < \min\{g_k : k \in U\}$ and $b_{m_i+1} > g_M$.

By replacing the values $A_i(a_j^i)$ and $A_i(b_j^i)$ in the equation (13), it could be easily seen how the estimator $\widehat{F}_{YO}(t)$ for $y_{[i]} \le t < y_{[i+1]}$ takes the following expression:

$$\widehat{F}_{YO}(t) = \sum_{j=1}^{m_i} \frac{(F_g(a_j^i) - F_g(b_j^i)) \cdot (\widehat{k}_i(a_j^i) - \widehat{k}_i(a_{j-1}^i))}{N \cdot (\widehat{F}_{GHT}(a_j^i) - \widehat{F}_{GHT}(b_j^i))}. \tag{16}$$

Now, if we suppose that $\bar{R}_{si} = \emptyset$, then the optimal vector $\mathbf{t_{OPT}}(t) = (a_1^i, a_2^i, \ldots, a_{m_i}^i)$ and $\widehat{F}_{YO}(t)$ take the following expression:

$$\widehat{F}_{YO}(t) = \widehat{F}_{YHT}(t) + \sum_{j=1}^{m_i} \left(F_g(a_j^i) - \widehat{F}_{GHT}(a_{j-1}^i)\right) \cdot A_i(a_j^i) \tag{17}$$

where for $j = 1, \ldots, m_i - 1$

$$A_i(a_j^i) = \frac{(\widehat{k}_i(a_j^i) - \widehat{k}_i(a_{j-1}^i))}{N(\widehat{F}_{GHT}(a_j^i) - \widehat{F}_{GHT}(a_{j-1}^i))} - \frac{(\widehat{k}_i(a_{j+1}^i) - \widehat{k}_i(a_j^i))}{N(\widehat{F}_{GHT}(a_{j+1}^i) - \widehat{F}_{GHT}(a_j^i))}. \tag{18}$$

From (17) and (18), the calibration estimator $\widehat{F}_{YO}(t)$ for $y_{[i]} \le t < y_{[i+1]}$ is:

$$\widehat{F}_{YO}(t) = \sum_{j=1}^{m_i} \frac{(F_g(a_j^i) - F_g(a_{j-1}^i)) \cdot (\widehat{k}_i(a_j^i) - \widehat{k}_i(a_{j-1}^i))}{N \cdot (\widehat{F}_{GHT}(a_j^i) - \widehat{F}_{GHT}(a_{j-1}^i))}. \tag{19}$$

Finally, we consider the case where $R_{si} \neq \emptyset$ and $\bar{R}_{si} \neq \emptyset$.

Be $R_{si} = \{j_1, \ldots j_{p_i}\}$, for all $j_h \in R_{si}$ we have:

$$A_i(a^i_{j_h}) = \frac{\left(\widehat{k}_i(a^i_{j_h}) - \widehat{k}_i(a^i_{j_h-1})\right)}{N\left(\widehat{F}_{GHT}(a^i_{j_h}) - \widehat{F}_{GHT}(a^i_{j_h-1})\right)}. \tag{20}$$

For $j \in \bar{R}_{si}$ and $j \neq j_h - 1$ for all $j_h \in R_{si}$, $A_i(a^i_j)$ and $A_i(b^i_j)$ are given by (14) and (15) while for $j = j_h - 1$ with $j \in \bar{R}_{si}$, $A_i(a^i_j)$ and $A_i(b^i_j)$ are given by (18) and (15).

Thus, in this case, the estimator $\widehat{F}_{YO}(t)$ for $y_{[i]} \leq t < y_{[i+1]}$ is given by:

$$\widehat{F}_{YO}(t) = \sum_{j \in R_{si}} \frac{\left(F_g(a^i_j) - F_g(a^i_{j-1})\right) \cdot \left(\widehat{k}_i(a^i_j) - \widehat{k}_i(a^i_{j-1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^i_j) - \widehat{F}_{GHT}(a^i_{j-1})\right)} +$$

$$\sum_{j \in \bar{R}_{si}} \frac{\left(F_g(a^i_j) - F_g(b^i_j)\right) \cdot \left(\widehat{k}_i(a^i_j) - \widehat{k}_i(a^i_{j-1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^i_j) - \widehat{F}_{GHT}(b^i_j)\right)}. \tag{21}$$

Now, for $y_{[i+1]} \leq z < y_{[i+2]}$, we consider the sets $A_{s(i+1)}$ and $B_{s(i+1)}$.

$$A_{s(i+1)} = \{a^{i+1}_1, \ldots, a^{i+1}_{m_{i+1}}\} \quad ; \quad B_{s(i+1)} = \{b^{i+1}_1, \ldots, b^{i+1}_{m_{i+1}}\}$$

and we define similarly the sets $R_{s(i+1)}$ and $\bar{R}_{s(i+1)}$.

Let be $A_{si} = A_{s(i+1)}$, then we have $B_{si} = B_{s(i+1)}$ and $\widehat{\mathbf{t}}_{\mathbf{OP}}(t) = \widehat{\mathbf{t}}_{\mathbf{OP}}(z) = \widehat{\mathbf{t}}_{\mathbf{OP}}(y[i])$. As in the previous case, we have $\widehat{F}_{YO}(t) \leq \widehat{F}_{YO}(z)$ because for both values t and z the weights ω_k in (5) are obtained with the same auxiliary vector. If we assume that $A_{si} \neq A_{s(i+1)}$, because $A_{si} \subset A_{s(i+1)}$ then exist a set

$$H_{si} = \{r_h : h = 1, \ldots, m_i\} \subset \{j : j = 1, \ldots m_{i+1}\}$$

such that

$$a^i_1 = a^{i+1}_{r_1}; \ldots; a^i_{m_i} = a^{i+1}_{r_{m_i}} \tag{22}$$

with $r_1 < r_2 < \cdots < r_{m_i}$ and $a^{i+1}_{r_{(h-1)}} \leq a^{i+1}_{r_h-1}$ for all $h \in \{1, \ldots, m_i\}$.
We denote by $\bar{H}_{si}$ the following set

$$\bar{H}_{si} = \{j : j = 1, \ldots, m_{i+1}\} - H_{si}.$$

On the other hand, since $a^{i+1}_{r_{(h-1)}} \leq a^{i+1}_{r_h-1}$ for all $h = 1, \ldots, m_t$

$$\{g_k : a^i_{h-1} \leq g_k < a^i_h\} = \{g_k : a^{i+1}_{r_{(h-1)}} \leq g_k < a^{i+1}_{r_h}\} =$$

11

$$\{g_k : a_{r_{(h-1)}}^{i+1} \le g_k \le a_{r_h-1}^{i+1}\} \cup \{g_k : a_{r_h-1}^{i+1} \le g_k < a_{r_h}^{i+1}\}$$

and therefore for all $h \in \{1, \dots, m_i\}$

$$b_h^i = \max\{g_k : a_{h-1}^i \le g_k < a_h^i\} = \max\{g_k : a_{r_h-1}^{i+1} \le g_k < a_{r_h}^{i+1}\} = b_{r_h}^{i+1}. \quad (23)$$

By (22) and (23), then we have

$$\widehat{k}_i(a_{r_h}^{i+1}) = \widehat{k}_i(a_h^i) \quad ; \quad \widehat{k}_i(a_{r_{(h-1)}}^{i+1}) = \widehat{k}_i(a_{h-1}^i) = \widehat{k}_i(b_h^i) = \widehat{k}_i(b_{r_h}^{i+1}). \quad (24)$$

Now, we define Γ_h as:

$$\Gamma_h = \left(\widehat{k}_{i+1}(a_{r_h}^{i+1}) - \widehat{k}_{i+1}(a_{r_h-1}^{i+1})\right) - \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r_{(h-1)}}^{i+1})\right) \quad (25)$$

where

$$\widehat{k}_{i+1}(a_{r_h}^{i+1}) = \sum_{k \in s} d_k \Delta(a_{r_h}^{i+1} - g_k)\Delta(y_{[i+1]} - y_k) =$$

$$\widehat{k}_i(a_{r_h}^{i+1}) + \sum_{k \in s} d_k \Delta(a_{r_h}^{i+1} - g_k)I_{[i+1]}(y_k)$$

with

$$I_{[i]}(y_k) = \left\{ \begin{array}{ll} 0 & \text{if } y_k \ne y_{[i]} \\ 1 & \text{if } y_k = y_{[i]}. \end{array} \right.$$

We denote by $\widehat{q}_{i+1}(z) = \sum_{k \in s} d_k \Delta(z - g_k)I_{[i+1]}(y_k)$. Thus $\widehat{k}_{i+1}(a_{r_h}^{i+1}) = \widehat{k}_i(a_{r_h}^{i+1}) + \widehat{q}_{i+1}(a_{r_h}^{i+1})$.

Similarly

$$\widehat{k}_{i+1}(a_{r_h-1}^{i+1}) = \widehat{k}_{i+1}(b_{r_h}^{i+1}) = \widehat{k}_i(b_{r_h}^{i+1}) + \widehat{q}_{i+1}(b_{r_h}^{i+1}) = \widehat{k}_i(a_{r_{(h-1)}}^{i+1}) + \widehat{q}_{i+1}(b_{r_h}^{i+1}).$$

Since $b_{r_h}^{i+1} < a_{r_h}^{i+1}$ for all $h \in \{1, \dots, m_i\}$

$$\Gamma_h = \left(\widehat{k}_{i+1}(a_{r_h}^{i+1}) - \widehat{k}_{i+1}(a_{r_h-1}^{i+1})\right) - \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r_{(h-1)}}^{i+1})\right) = \widehat{q}_{i+1}(a_{r_h}^{i+1}) - \widehat{q}_{i+1}(b_{r_h}^{i+1}) \ge 0. \quad (26)$$

Based on the sets R_{si}, $\bar{R}_{si}$, $R_{s(i+1)}$ and $\bar{R}_{s(i+1)}$ we consider several cases:

Case 1) $R_{si} = \emptyset$.

In this case, if we assume and $R_{s(i+1)} \ne \emptyset$ and $\bar{R}_{s(i+1)} \ne \emptyset$, the set $\{j : j = 1, \dots, m_{i+1}\}$ is given by:

$$\{j : j = 1, \dots, m_{i+1}\} = C_1 \cup C_2 \cup C_3 \cup C_4 =$$

$$= (R_{s(i+1)} \cap H_{si}) \cup (R_{s(i+1)} \cap \bar{H}_{si}) \cup (\bar{R}_{s(i+1)} \cap H_{si}) \cup (\bar{R}_{s(i+1)} \cap \bar{H}_{si}).$$

For $h = 1, \ldots, m_i$ with $r_h \in C_1 = R_{s(i+1)} \cap H_{si}$

$$a^{i+1}_{r_h-1} = b^{i+1}_{r_h} = b^i_h \neq a^i_{h-1} = a^{i+1}_{r_{(h-1)}} \tag{27}$$

while for $h = 1, \ldots, m_i$ with $r_h \in C_3 = \bar{R}_{s(i+1)} \cap H_{si}$

$$a^{i+1}_{r_h-1} \neq b^{i+1}_{r_h} = b^i_h \neq a^i_{h-1} = a^{i+1}_{r_{(h-1)}}. \tag{28}$$

By (16); (27) and (28), $\widehat{F}_{YO}(t)$ is given by:

$$\widehat{F}_{YO}(t) = \sum_{h=1}^{m_i} \frac{\left(F_g(a^i_h) - F_g(b^i_h)\right)}{\left(\widehat{F}_{GHT}(a^i_h) - \widehat{F}_{GHT}(b^i_h)\right)} \cdot \frac{\left(\widehat{k}_i(a^i_h) - \widehat{k}_i(a^i_{h-1})\right)}{N} =$$

$$= \sum_{\substack{h=1 \\ r_h \in C_1}}^{m_i} \frac{\left(F_g(a^{i+1}_{r_h}) - F_g(a^{i+1}_{r_h-1})\right) \cdot \left(\widehat{k}_i(a^{i+1}_{r_h}) - \widehat{k}_i(a^{i+1}_{r_{(h-1)}})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^{i+1}_{r_h}) - \widehat{F}_{GHT}(a^{i+1}_{r_h-1})\right)} +$$

$$\sum_{\substack{h=1 \\ r_h \in C_3}}^{m_i} \frac{\left(F_g(a^{i+1}_{r_h}) - F_g(b^{i+1}_{r_h})\right) \cdot \left(\widehat{k}_i(a^{i+1}_{r_h}) - \widehat{k}_i(a^{i+1}_{r_{(h-1)}})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^{i+1}_{r_h}) - \widehat{F}_{GHT}(b^{i+1}_{r_h})\right)}. \tag{29}$$

From (21), $\widehat{F}_{YO}(z)$ takes the following expression:

$$\widehat{F}_{YO}(z) = \sum_{r_h \in C_1} \frac{\left(F_g(a^{i+1}_{r_h}) - F_g(a^{i+1}_{r_h-1})\right) \cdot \left(\widehat{k}_{i+1}(a^{i+1}_{r_h}) - \widehat{k}_{i+1}(a^{i+1}_{r_h-1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^{i+1}_{r_h}) - \widehat{F}_{GHT}(a^{i+1}_{r_h-1})\right)} +$$

$$+ \sum_{r_h \in C_3} \frac{\left(F_g(a^{i+1}_{r_h}) - F_g(b^{i+1}_{r_h})\right) \cdot \left(\widehat{k}_{i+1}(a^{i+1}_{r_h}) - \widehat{k}_{i+1}(a^{i+1}_{r_h-1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^{i+1}_{r_h}) - \widehat{F}_{GHT}(b^{i+1}_{r_h})\right)} + W_s(z) \tag{30}$$

where $W_s(z) = V_s(z) + T_s(z)$ and

$$V_s(z) = \sum_{j \in C_2} \frac{\left(F_g(a^{i+1}_j) - F_g(a^{i+1}_{j-1})\right) \cdot \left(\widehat{k}_{i+1}(a^{i+1}_j) - \widehat{k}_{i+1}(a^{i+1}_{j-1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^{i+1}_j) - \widehat{F}_{GHT}(a^{i+1}_{j-1})\right)} \geq 0$$

$$T_s(z) = \sum_{j \in C_4} \frac{\left(F_g(a^{i+1}_j) - F_g(b^{i+1}_j)\right) \cdot \left(\widehat{k}_{i+1}(a^{i+1}_j) - \widehat{k}_{i+1}(a^{i+1}_{j-1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a^{i+1}_j) - \widehat{F}_{GHT}(b^{i+1}_j)\right)} \geq 0.$$

Consequently, from (30) and (29), we have

$$\widehat{F}_{YO}(z) - \widehat{F}_{YO}(t) = W_s(z) + \sum_{r_h \in C_1} \frac{\left(F_g(a^{i+1}_{r_h}) - F_g(a^{i+1}_{r_h-1})\right) \cdot \Gamma_h}{N \cdot \left(\widehat{F}_{GHT}(a^{i+1}_{r_h}) - \widehat{F}_{GHT}(a^{i+1}_{r_h-1})\right)} +$$

13

$$\sum_{r_h \in C_3} \frac{\left(F_g(a_{r_h}^{i+1}) - F_g(b_{r_h}^{i+1})\right) \cdot \Gamma_h}{N \cdot \left(\widehat{F}_{GHT}(a_{r_h}^{i+1}) - \widehat{F}_{GHT}(b_{r_h}^{i+1})\right)} \geq 0.$$

On the other hand, if we suppose that $R_{s(i+1)} = \emptyset$, then $C_1 = C_2 = \emptyset$. For all $h \in \{1, \ldots, m_i\}$, we have (28) and $\widehat{F}_{YO}(t)$ is given by (29) with the sum in C_1 null. From (16), $\widehat{F}_{YO}(z)$ is given by (30) with $V_s(z) = 0$ and the summation based on the set C_1 null. Similarly, we can see that $\widehat{F}_{YO}(z) - \widehat{F}_{YO}(t) \geq 0$.

Identically, it can be shown that $\widehat{F}_{YO}(z) - \widehat{F}_{YO}(t) \geq 0$ in the case that $\bar{R}_{s(i+1)} = \emptyset$. From (19) (27) and because the sets C_3 and C_4 are empty, $\widehat{F}_{YO}(t)$ and $\widehat{F}_{YO}(z)$ are given by (29) and(30) respectively, with the summations based on null C_3 and $T_s(z) = 0$.

Case 2) $R_{si} \neq \emptyset$ and $\bar{R}_{si} \neq \emptyset$

For $h = 1, \ldots, m_i$ with $h \in R_{si}$,

$$a_h^i = a_{r_h}^{i+1} \quad ; \quad b_{r_h}^{i+1} = b_h^i = a_{h-1}^i = a_{r(h-1)}^{i+1}$$

since $a_{r(h-1)}^{i+1} \leq a_{r_h - 1}^{i+1} \leq b_{r_h}^{i+1} = a_{r(h-1)}^{i+1}$, we have

$$a_{r_h-1}^{i+1} = b_{r_h}^{i+1} = b_h^i = a_{h-1}^i = a_{r(h-1)}^{i+1} \tag{31}$$

and consequently, for $h = 1, \ldots, m_i$ with $h \in R_{si}$; $r_h \in R_{s(i+1)} \cap H_{si} = C_1$ and the set $R_{s(i+1)} \neq \emptyset$. Because in this case $R_{s(i+1)} \neq \emptyset$ if we assume that $\bar{R}_{s(i+1)} \neq \emptyset$, then as in Case 1) with $R_{s(i+1)} \neq \emptyset$ and $\bar{R}_{s(i+1)} \neq \emptyset$, the value $\widehat{F}_{YO}(z)$ is given by (30). Additionally, for $h = 1, \ldots, m_i$ with $h \in \bar{R}_{si}$, it is clear that $r_h \in C_1 \cup C_3$. For $h \in \bar{R}_{si}$ with $r_h \in C_1$ condition (27) is verified while for $h \in \bar{R}_{si}$ with $r_h \in C_3$; condition (28) is satisfied.

From (21); (31); (27) and (28); the value $\widehat{F}_{YO}(t)$ takes the following expression:

$$\widehat{F}_{YO}(t) = \sum_{h \in R_{si}} \frac{\left(F_g(a_{r_h}^{i+1}) - F_g(a_{r_h-1}^{i+1})\right) \cdot \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r(h-1)}^{i+1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a_{r_h}^{i+1}) - \widehat{F}_{GHT}(a_{r_h-1}^{i+1})\right)} +$$

$$+ \sum_{\substack{h \in \bar{R}_{si} \\ r_h \in C_1}} \frac{\left(F_g(a_{r_h}^{i+1}) - F_g(a_{r_h-1}^{i+1})\right) \cdot \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r(h-1)}^{i+1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a_{r_h}^{i+1}) - \widehat{F}_{GHT}(a_{r_h-1}^{i+1})\right)}$$

14

$$+ \sum_{\substack{h \in \bar{R}_{si} \\ r_h \in C_3}} \frac{\left(F_g(a_{r_h}^{i+1}) - F_g(b_{r_h}^{i+1})\right) \cdot \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r_{(h-1)}}^{i+1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a_{r_h}^{i+1}) - \widehat{F}_{GHT}(b_{r_h}^{i+1})\right)} =$$

$$+ \sum_{\substack{h=1 \\ r_h \in C_1}}^{m_i} \frac{\left(F_g(a_{r_h}^{i+1}) - F_g(a_{r_h-1}^{i+1})\right) \cdot \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r_{(h-1)}}^{i+1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a_{r_h}^{i+1}) - \widehat{F}_{GHT}(a_{r_h-1}^{i+1})\right)}$$

$$+ \sum_{\substack{h=1 \\ r_h \in C_3}}^{m_i} \frac{\left(F_g(a_{r_h}^{i+1}) - F_g(b_{r_h}^{i+1})\right) \cdot \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r_{(h-1)}}^{i+1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a_{r_h}^{i+1}) - \widehat{F}_{GHT}(b_{r_h}^{i+1})\right)}. \tag{32}$$

As in the Case 1) with $R_{s(i+1)} \neq \emptyset$ and $\bar{R}_{s(i+1)} \neq \emptyset$, the value $\widehat{F}_{YO}(z) - \widehat{F}_{YO}(t) \geq 0$.

If we assume that $\bar{R}_{s(i+1)} = \emptyset$, the sets C_3 and C_4 are empty. For $h \in \bar{R}_{si}$, then $r_h \in C_1$ and condition (27) is satisfied. From (19) and (27) it is easy to see that $\widehat{F}_{YO}(z)$ is given by (30) with the sum based on null C_3 and $T_s(z) = 0$.

Similarly, because $C_3 = \emptyset$; and the conditions(31) and (27) are satisfied, by (21) the value $\widehat{F}_{YO}(t)$ is given by (32) with the sum based on null C_3 . Thus, it is clear that $\widehat{F}_{YO}(z) - \widehat{F}_{YO}(t) \geq 0$.

Case 3) $\bar{R}_{si} = \emptyset$

In this case, in a similar way that in the previous case, for all $h = 1, \ldots, m_1$; the condition (31) is satisfied and consequently $r_h \in C_1$ and $C_3 = \emptyset$.

By (31) and (19), we have:

$$\widehat{F}_{YO}(t) = \sum_{\substack{h=1 \\ r_h \in C_1}}^{m_i} \frac{\left(F_g(a_{r_h}^{i+1}) - F_g(a_{r_h-1}^{i+1})\right) \cdot \left(\widehat{k}_i(a_{r_h}^{i+1}) - \widehat{k}_i(a_{r_{(h-1)}}^{i+1})\right)}{N \cdot \left(\widehat{F}_{GHT}(a_{r_h}^{i+1}) - \widehat{F}_{GHT}(a_{r_h-1}^{i+1})\right)}.$$

If we assume that $R_{s(i+1)} \neq \emptyset$, the value $\widehat{F}_{YO}(z)$ is given by (30) with the sum based on null C_3 while if we assume that $R_{s(i+1)} = \emptyset$ the value $\widehat{F}_{YO}(z)$ is given by (30) with the sum based on null C_3 and $T_s(z) = 0$. In any case, $\widehat{F}_{YO}(z) - \widehat{F}_{YO}(t) \geq 0$.

Definitely, in all cases, $\widehat{F}_{YO}(z) - \widehat{F}_{YO}(t) \geq 0$ if we consider $t \leq z$ with $y_{[i]} \leq t < y_{[i+1]}$ and $y_{[i+1]} \leq z < y_{[i+2]}$; $i = 1, \ldots, l-2$.

For $t \leq z$ with $y_{[i]} \leq t < y_{[i+1]}$ and $y_{[q]} \leq z < y_{[q+1]}$; $i = 1, \ldots, l-2$ and $q = 3, \ldots, l$ with $q > i+1$, by the previous cases, it is clear that:

$$\widehat{F}_{YO}(t) \leq \widehat{F}_{YO}(y_{[i+1]}) \leq \widehat{F}_{YO}(y_{[i+2]}) \leq \cdots \leq \widehat{F}_{YO}(y_{[q]}) \leq \widehat{F}_{YO}(z)$$

and the nondecreasing monotony of $\widehat{F}_{YO}(t)$ is proved. $\qquad\square$

Note

The estimator $\widehat{F}_{YO}(t)$ does not satisfy, in general, the condition $\lim_{t \to +\infty} \widehat{F}_{YO}(t) = 1$, but this condition is not strictly necessary as long as the following condition is satisfied

$$\max\{\widehat{F}_{YO}(y_i) : i \in s\} \geq \alpha. \tag{33}$$

Thus, we can define the following quantile estimator:

$$\widehat{Q}_{YO}(\alpha) = \inf\{t : \widehat{F}_{YO}(t) \geq \alpha\} = \widehat{F}_{YO}^{-1}(\alpha). \tag{34}$$

4. Variance estimation with resampling method

In this section we employ resampling techniques for the variance estimation of the quantile estimators proposed in Section 3 and the development of confidence intervals for quantiles associated with the calibration estimators proposed, because it is possible that a mathematic expression for their variance could be not establish due to the complexity of the estimators proposed (they are not linear functions of the data). More specifically, from a practical viewpoint we have considered the use of bootstrap techniques by their applicability in many cases and under different conditions.

Initially, the bootstrap method was developed by [18] under assumptions of an infinity population with unknown distribution and the data is independently and identically distributed. Due to the popularity of this technique, the classical framework has been adjusted for survey sampling and incorporate the sampling design in several studies ([19], [7], [10], [5], [10], [1] and [2]). Thus, [19], [7] and [5] developed bootstrap methods where artificial populations are created from the sample by repeating its units and bootstrap samples are selected with the original sampling design from the artificial population. On the other hand, [1] and [2] consider direct bootstrap techniques where the bootstrap samples are obtained by units directly selected from the original sample

16

under a completely different sampling scheme from the one which generated the original sample. In this study, we consider the frameworks proposed in [7], [1] and [2].

Given a generic quantile estimator $\widehat{Q}_y(\alpha)$, following [7], if $N = n \cdot c + a$ with $0 < a < n$, the artificial population is obtained by repeating c times the initial sample s and selecting by simple random sampling without replacement an additional sample of size a from the original sample s. The artificial population U_B is formed with this sample and the c replicates of s. Thus, let U_B^j with $j = 1, \ldots, M$ be M independent artificial populations obtained from s, for each pseudo population U_B^j we select K bootstrap samples $s_1^j, \ldots, s_K^j$ with sample size n. Next, following ([11]), we compute the bootstrap estimates $\widehat{Q}_y^*(\alpha)_h^j$ with the sample s_h^j for the population U_B^j and we consider:

$$\widehat{V}_j = \frac{1}{K-1} \sum_{h=1}^{K} (\widehat{Q}_y^*(\alpha)_h^j - \widehat{Q}_y^*(\alpha)^j)^2 \tag{35}$$

where

$$\widehat{Q}_y^*(\alpha)^j = \frac{1}{K} \sum_{h=1}^{K} \widehat{Q}_y^*(\alpha)_h^j.$$

Now, the variance estimation for the estimator $\widehat{Q}_y(\alpha)$ is given by:

$$\widehat{V}(\widehat{Q}_y(\alpha)) = \frac{1}{M} \sum_{j=1}^{M} .\widehat{V}_j. \tag{36}$$

Recently, [1] and [2] have proposed direct boostrap methods where it is not necessary to obtain an artificial population, since the boostrap samples are obtained from the original sample by means of a sampling design different from the original sampling design considered. Thus, when the original sample s is obtained with simple random sampling without replacement, [1] has proposed a mixture sampling design where two samples are selected from s, one obtained by simple random sampling without replacement and the other sample obtained with one-one sampling design (a sampling design defined by the authors for resampling). Similarly, when s is obtained by simple random sampling without replacement, [2] has proposed a mixture sampling scheme where the first sample is obtained with Bernoulli design while the second one is obtained with another sampling designed for resampling called double half sampling design by the authors. For more details on the two direct boostrap methods included in this study, see [1] and [2].

In both frameworks, given a generic quantile estimator $\widehat{Q}_y(\alpha)$, for the original sample

17

s, we select M bootstrap samples $s_1^*, \ldots, s_M^*$ according to the sampling schemes of [1] and [2] respectively. The variance bootstrap estimation for the estimator $\widehat{Q}_y(\alpha)$ is given by:

$$\widehat{V}(\widehat{Q}_y(\alpha)) = \frac{1}{M} \sum_{j=1}^{M} (\widehat{Q}_y(\alpha)_j^* - \bar{Q}_y(\alpha)^*)^2 \tag{37}$$

where $\widehat{Q}_y(\alpha)_j^*$ is the bootstrap estimator computed on the bootstrap sample s_j^* and

$$\bar{Q}_y(\alpha)^* = \frac{1}{M} \sum_{j=1}^{M} \widehat{Q}_y(\alpha)_j^*.$$

Finally, for a quantile estimator $\widehat{Q}_y(\alpha)$ with a variance estimation $\widehat{V}(\widehat{Q}_y(\alpha))$ obtained with a bootstrap method, we consider the $1 - \alpha$ level confidence interval based on the approximation by a standard normal distribution:

$$\left[\widehat{Q}_y(\alpha) - z_{1-\alpha/2} \cdot \widehat{V}(\widehat{Q}_y(\alpha)), \widehat{Q}_y(\alpha) + z_{1-\alpha/2} \cdot \widehat{V}(\widehat{Q}_y(\alpha)) \right] \tag{38}$$

where z_α denotes the α quantile of the standard normal distribution. For the three proposed bootstrap methods included in this study, we can obtain with this procedure the respective confident interval.

5. Application of the optimal quantile estimators in poverty measures estimation

For governments it is of high interest the estimation of poverty and wage inequality. Inequality and life condition indicators and many social indicators related to the measurement of poverty are based upon quantiles. Among the poverty measures commonly used by institutions in their reports on poverty, we can find the poverty line and the Head Count Index. For instance, Eurostat establishes poverty line as 60 percent of the median of the equivalized net income. Thus, the poverty line is defined as a threshold that divides the population into poor and nonpoor that depends on the median value. The Head Count Index (HCI) can be calculated as the proportion of persons (or households) with an equivalised disposable income below the poverty line. On the other hand, some measures for wage inequality employed in several studies are based on percentiles ratios like 50th/5th and 50th/25th ([16]); 50th/10th ([33],[24], [31], [8]); 95th/50th ([24], [8]) and 90th/10th; 95th/20th; and 80th/20th ([22]). In this study, we focus on the estimation of the poverty measures based on percentile ratios.

18

Thus, for a finite population $U = \{1, \ldots, N\}$ with distribution function $F_y(t)$ given by (1), the percentile ratio $R(\alpha_1, \alpha_2)$ is defined as follows:

$$R(\alpha_1, \alpha_2) = \frac{Q_y(\alpha_1)}{Q_y(\alpha_2)} = \frac{F_y^{-1}(\alpha_1)}{F_y^{-1}(\alpha_2)}$$

and evidently, it can be estimated with the quantile estimator $\tilde{Q}_{YO}(\alpha)$ as follow:

$$\tilde{R}_{YO}(\alpha_1, \alpha_2) = \frac{\tilde{Q}_{YO}(\alpha_1)}{\tilde{Q}_{Y1}(\alpha_2)}.$$

370　Obviously, the variance estimation of a percentile ratio estimator present similar drawbacks to the estimation of variance for quantile estimator and consequently, we can compute the estimation of variance for $\tilde{R}_{YO}(\alpha_1, \alpha_2)$ and confidence intervals for $\widehat{R}_{YO}(\alpha_1, \alpha_2)$ with the resampling techniques described in the previous section.

6. Simulation study

375　　This section provides numerical comparisons for some poverty measure estimators proposed in Sections 3 and 5. In two simulation studies the proposed estimators are compared with the corresponding poverty measures estimators derived from previous estimators of the distribution function: the Horvitz-Thompson estimator $\widehat{F}_{YHT}(t)$, the difference estimator $\widehat{F}_{YD}(t)$ (see [34]), the ratio estimator $\widehat{F}_{YR}(t)$ (see [34]),
380　the Chambers–Dunstan estimator $\widehat{F}_{YCD}(t)$ (see [9]) and the Rao,Kovar and Mantel estimator (see [34]) $\widehat{F}_{YRKM}(t)$. Additionally, we have included the quantile estimator and the estimator of poverty measures derived from the calibrated estimator $\widehat{F}_{yc}(t)$ of the distribution function proposed in [36], with auxiliary vector $\mathbf{t_g} = (Q_g(0.25), Q_g(0.5), Q_g(0.75))$ and we denoted by $\widehat{F}_{YQUAR}(t)$ the corresponding cal-
385　ibration estimator. Some of these estimators $\widehat{F}_y(t)$ included in the simulation study are not monotonically nondecreasing functions; for these estimators we have considered a general procedure described in [34] to obtain a monotonous nondecreasing version of the estimator $\tilde{F}_y(t)$.

For both simulation studies, the estimation of the variance provided by the bootstrap
390　methods included in Section 4 is also analyzed. All simulations included in this section have been developed with new code programmed in R.

　　In the first study we consider real data from the region of Cantabria of the 2008 Spanish living conditions survey carried out by the Instituto Nacional de Estadística (INE) of Spain. The survey data collected are considered as a population with size

19

$N = 377$ and samples are selected from it. In this study we obtain estimation of the poverty threshold L, where L is calculated following the criteria recommended by Eurostat, that is, the threshold L is set at 60% of the median of the equivalised net income (the study variable). We considered the attribute "Home with own computer" as the auxiliary variable. We selected $W = 1000$ samples with several sample sizes, n, under SRSWOR and for each estimator included in the simulation study, we computed estimates of the poverty threshold L. The performance of each estimator is measured by the relative bias (RB) and the relative efficiency (RE), given respectively by

$$\text{RB}(\widehat{L}) = \frac{1}{W} \sum_{w=1}^{W} \frac{(\widehat{L}_w - L)}{L} \tag{39}$$

$$\text{RE}(\widehat{L}) = \frac{\displaystyle\sum_{w=1}^{W} \left[\widehat{L}_w - L\right]^2}{\displaystyle\sum_{w=1}^{W} \left[\left(\widehat{L}_{HT}\right)_w - L\right]^2}, \tag{40}$$

where w indexes the wth simulation run; $\widehat{L}$ is a poverty threshold estimator and $\widehat{L}_{HT}$ is the poverty threshold estimator based in the Horvitz-Thompson $\widehat{F}_{YHT}(t)$ estimator.

From every simulation sample, 1000 bootstrap samples were selected using the three bootstrap methods considered in Section 4, for the variance estimation and confidence intervals. We computed the following measures: the coverage probability (CP), the lower (L) and the upper (U) tail error rates of the 95% confidence intervals, in percentage and the average length (AL) of the confidence intervals for each estimator and each bootstrap method, except for the Chambers-Dunstan estimator whose results are only obtained with the Booth method, since this estimator needs the whole population for its calculation and the techniques described do not obtain the whole artificial population. Results from this simulation study are presented in Table 1 and Table 2.

The results derived from this simulation study gave values for RB within a reasonable range. The proposed estimator significantly improve the results of the calibrated estimator $\widehat{L}_{YQUAR}$. With respect to efficiency, the best estimator for all sample sizes is $\widehat{L}_{YCO}$ whereas the usual calibrated estimator have an efficiency similar to $\tilde{L}_{HT}$.

With respect to the variance estimation, all estimators provide high coverages, with values very close to 99% in the three resampling methods considered. For the resam-

20

Table 1: RB and RE for several sample sizes of the estimators compared. SRSWOR from the 2008 SPANISH LIVING CONDITIONS SURVEY.

Estimator	RB	RE	RB	RE	RB	RE	RB	RE
	$n = 50$		$n = 60$		$n = 70$		$n = 80$	
$\widehat{L}_{HT}$	-0.0159	1.0000	-0.0103	1.0000	-0.0144	1.0000	-0.0118	1.0000
$\widehat{L}_{CD}$	0.0452	1.2561	0.0408	1.1163	0.0332	1.0551	0.0324	1.0732
$\tilde{L}_{d}$	-0.0038	0.9246	-0.0010	0.8975	-0.0056	0.8796	-0.0044	0.8873
$\tilde{L}_{r}$	-0.0069	1.7426	-0.0072	1.5946	-0.0057	1.4706	-0.00603	1.6292
$\tilde{L}_{RKM}$	-0.0043	0.9456	-0.0014	0.8979	-0.0058	0.8942	-0.0041	0.8971
$\widehat{L}_{YQUAR}$	-0.0159	1.0000	-0.0103	1.0000	-0.0144	1.0000	-0.0118	1.0000
$\widehat{L}_{YCO}$	-0.0039	0.9165	-0.0005	0.8974	-0.0056	0.8730	-0.0043	0.8849

pling methods proposed in [7] and [2], the proposed estimators present the best average length (AL) results for some sample sizes, whereas with the method proposed in [1], the proposed estimators present the best results for all the sample sizes, with the exception of size $n = 60$.

For the second simulation study, we consider real data from the region of Andalusia of 2016 Spanish living conditions survey carried out by the Instituto Nacional de Estadística (INE) of Spain. The survey data collected are considered as a population with size $N = 1442$ and samples are selected from it. The study variable y is the equivalised net income and the auxiliary variables included are the attribute "Can the home afford to go on vacation away from home, at least one week a year?", the attribute "Home with own computer" and the attribute "Home with own washing machine" as the auxiliary variables. Again, we selected $W = 1000$ samples with several sample sizes, $n = 75$, $n = 95$, $n = 115$ and $n = 135$, under SRSWOR and for each estimator included in the simulation study, we computed estimates of $R(\alpha_1, \alpha_2)$ for 95th/50th. The performance of each estimator is measured by the values RB and RE, given by

$$\text{RB}(\widehat{R}(\alpha_1, \alpha_2)) = \frac{1}{W} \sum_{w=1}^{W} \frac{(\widehat{R}(\alpha_1, \alpha_2))_w - R(\alpha_1, \alpha_2)}{R(\alpha_1, \alpha_2)} \tag{41}$$

$$\text{RE}(\widehat{R}(\alpha_1, \alpha_2)) = \frac{\sum_{w=1}^{W} \left[(\widehat{R}(\alpha_1, \alpha_2))_w - R(\alpha_1, \alpha_2) \right]^2}{\sum_{w=1}^{W} \left[(\widehat{R}_{HT}(\alpha_1, \alpha_2))_w - R(\alpha_1, \alpha_2) \right]^2}, \tag{42}$$

where $\widehat{R}(\alpha_1, \alpha_2)$ is a percentile ratio estimator and $\widehat{R}_{HT}(\alpha_1, \alpha_2)$ is the percentile ratio

Table 2: AL, CP %, L % and U % for several sample sizes and several resampling methods of the estimators compared. SRSWOR from the 2008 SPANISH LIVING CONDITIONS SURVEY.

Estimator	$n = 50$				$n = 60$				$n = 70$				$n = 80$			
	AL	CP%	L%	U%	AL	CP%	L%	U%	AL	CP%	L%	U%	AL	CP%	L%	U%
Booth et al., (1994)																
$\hat{L}_{HT}$	7355	99.9	0.0	0.1	4513	100.0	0.0	0	3554	100.0	0	0.0	3199	100.0	0.0	0
$\hat{L}_{CD}$	7323	99.9	0.1	0.0	3670	99.8	0.2	0	4451	100.0	0	0.0	4082	99.1	0.1	0
$\tilde{L}_d$	7383	100.0	0.0	0.0	4164	100.0	0.0	0	3501	100.0	0	0.0	3103	100.0	0.0	0
$\tilde{L}_r$	7397	99.8	0.2	0.0	6001	100.0	0.0	0	3834	99.9	0	0.1	3423	99.8	0.2	0
$\tilde{L}_{RKM}$	7096	100.0	0.0	0.0	4171	100.0	0.0	0	3418	100.0	0	0.0	3041	100	0.0	0
$\hat{L}_{YQUAR}$	7355	99.9	0.0	0.1	4513	100.0	0.0	0	3454	100.0	0	0.0	3115	100	0.0	0
$\hat{L}_{YCO}$	7301	100.0	0.0	0.0	4166	100.0	0.0	0	3507	100.0	0	0.0	3016	100	0.0	0
Antal, E, Tillé, Y. (2014)																
$\hat{L}_{HT}$	5198	98.9	0.1	1.0	3970	98.4	0.2	1.4	3507	98.0	0.4	1.6	3072	98.1	0.4	1.5
$\tilde{L}_d$	5204	98.9	0.4	0.7	3477	98.7	0.4	0.9	2914	98.3	0.6	1.1	2504	98.4	0.5	1.1
$\tilde{L}_r$	5518	99.0	0.2	0.8	4108	99.4	0.0	0.6	360	98.8	0.1	1.1	316898.7	0.2	1.1	
$\tilde{L}_{RKM}$	4659	99.1	0.2	0.7	4224	98.9	0.3	0.8	3001	98.0	0.7	1.3	2671	98.1	0.6	1.3
$\hat{L}_{YQUAR}$	5198	98.9	0.1	1.0	3970	98.4	0.2	1.4	3507	98.0	0.4	1.6	3044	98.3	0.3	1.4
$\hat{L}_{YCO}$	4754	98.7	0.5	0.8	4235	98.5	0.5	1.0	2938	98.1	0.6	1.3	2448	98.4	0.4	1.2
Antal, E, Tillé, Y. (2011)																
$\hat{L}_{HT}$	4417	98.9	0.1	1.0	4627	98.0	0.3	1.7	2488	98.4	0.2	1.4	2279	98.5	0.2	1.3
$\tilde{L}_d$	3219	99.1	0.1	0.8	3855	98.6	0.5	0.9	2511	98.9	0.5	0.6	2281	99.2	0.4	0.4
$\tilde{L}_r$	4329	99.3	0.2	0.5	4957	99.5	0.1	0.4	4177	98.7	0.3	1.0	3837	98.9	0.3	0.8
$\tilde{L}_{RKM}$	3846	98.9	0.0	1.1	4088	98.7	0.6	0.7	2724	98.8	0.5	0.7	2435	99.1	0.5	0.4
$\hat{L}_{YQUAR}$	4417	98.9	0.1	1.0	4627	98.0	0.3	1.7	2488	98.4	0.2	1.4	2173	98.6	0.2	1.2
$\hat{L}_{YCO}$	3059	98.9	0.2	0.9	4188	98.2	0.9	0.9	2309	98.7	0.5	0.8	2018	99.3	0.3	0.4

22

estimator based in the Horvitz-Thompson $\widehat{F}_{YHT}(t)$ estimator .

For the variance estimation and confidence intervals, we computed the coverage probability (CP), the lower (L) and the upper (U) tail error rates of the 95% confidence intervals, in percentage and the average length (AL) of the confidence intervals for each percentile ratio estimator and each bootstrap method.

Concerning the variance estimation and confidence intervals, we used 1,000 bootstrap replications from each initial sample with all bootstrap methods included in the study to compute CP, L, U and AL of the 95% confidence intervals for each percentile ratio considered. Result from this simulation study are presented in Table 3 and Table 4.

Table 3: RB and RE for several sample sizes of the estimators compared. SRSWOR from the 2016 SPANISH LIVING CONDITIONS SURVEY.

	RB	RE%	RB	RE	RB	RE	RB	RE
Estimator	$n = 75$		$n = 95$		$n = 115$		$n = 135$	
$\widehat{R}_{HT}$	0.0392	1	0.0358	1	0.0285	1	0.0234	1
$\widehat{R}_{CD}$	-0.1224	1.0838	-0.1152	1.0771	-0.1219	1.3093	-0.125	1.4158
$\tilde{R}_{d}$	0.0407	1.0394	0.0353	1.0467	0.0269	1.0673	0.0223	1.0689
$\tilde{R}_{r}$	0.0461	1.3042	0.0406	4.0376	0.018	4.4338	0.0161	3.2556
$\tilde{R}_{RKM}$	0.0333	1.0384	0.033	1.0475	0.0263	1.0325	0.0208	1.0266
$\widehat{R}_{YQUAR}$	0.0251	0.948	0.0208	0.9276	0.0183	0.9635	0.0143	0.9623
$\widehat{R}_{YCO}$	0.0174	0.921	0.0157	0.8871	0.0164	0.9458	0.0126	0.9390

These tables show:

- The percentile ratio estimator based on the Chambers–Dunstan estimator has a serious problem of bias. This is expected because the estimator $\widehat{F}_{YCD}(t)$ is biased when the relation between y and x is not linear. We found no evidence of any significant bias for the other estimators considered.

- In terms of efficiency the best overall performance is achieved by our proposed calibration estimator. This estimator performs remarkably better than the other estimators.

- The three methods of estimating the variances provide intervals with coverage below the nominal coverage. Although there is not much difference between the methods it seems that the first method ([7]) provides narrower intervals.

To sum up, these simulation show how the use of the auxiliary information by the

Table 4: AL, CP %,L % and U % for several sample sizes and several resampling methods of the estimators compared. SRSWOR from the 2016 SPANISH LIVING CONDITIONS SURVEY.

Estimator	AL	CP%	L%	U%	AL	CP%	L%	U%	AL	CP%	L%	U%	AL	CP%	L%	U%
	n = 75				n = 95				n = 115				n = 135			
Booth et al., (1994)																
$\widehat{R}_{HT}$	3.7901	87.7	3.8	8.5	0.8089	87.4	5.1	7.5	1.7979	86.9	5.6	7.5	0.7069	87.3	5	7.7
$\widehat{R}_{CD}$	3.5177	66	1	33	1.4043	63.3	1.2	35.5	0.2989	60.9	0.7	38.4	0.5982	56.8	0.5	42.7
$\widehat{R}_d$	3.8003	85.9	4.4	9.7	1.1235	87.9	5	7.1	1.7837	86.7	5.3	8	0.6935	86.5	5.4	8.1
$\widetilde{R}_r$	3.7722	82.7	7.7	9.6	1.8362	84.4	6.8	8.8	1.0592	83.4	6	10.6	0.991	83.6	5.3	11.1
$\widetilde{R}_{RKM}$	3.8459	87.5	3.5	9	1.2744	87.9	4.5	7.6	1.5897	86	5.2	8.8	0.6935	86.2	5.3	8.5
$\widehat{R}_{YQUAR}$	3.8042	86.4	3.8	9.8	1.2659	88	4.2	7.8	1.7837	85.9	4.8	9.3	0.6935	87.3	4.4	8.3
$\widehat{R}_{YCO}$	3.6703	85.6	4	10.4	1.2659	87.8	3.8	8.4	1.7837	85.4	5.3	9.3	0.6935	88.2	3.6	8.2
Antal, E., Tillé, Y. (2014)																
$\widehat{R}_{HT}$	3.8038	84.7	6	9.3	1.1358	82.6	8.6	8.8	1.7578	85.2	6.7	8.1	1.3928	82.8	8.8	8.4
$\widehat{R}_d$	3.9	85.3	5.2	9.5	1.1534	82.5	8.7	8.8	1.4795	84.5	6.5	9	1.2933	81.3	8.8	9.9
$\widetilde{R}_r$	4.4885	78.9	7.6	13.5	1.4523	80.8	6.7	12.5	2.1691	80.9	6.8	12.3	2.2027	77.9	8.9	13.2
$\widetilde{R}_{RKM}$	3.8207	84.8	5.2	10	1.2017	82.6	8.7	8.7	1.4795	84.6	7.2	8.2	1.4643	82.2	7.8	10
$\widehat{R}_{YQUAR}$	3.4395	83.4	5.4	11.2	1.1509	84.3	7.1	8.6	1.527	83.7	7.1	9.2	1.2797	82.9	7.3	9.8
$\widehat{R}_{YCO}$	3.4395	84.3	4.5	11.2	1.1336	84.8	6.5	8.7	1.8641	82.7	7.3	10	1.4264	83.1	7	9.9
Antal, E., Tillé, Y. (2011)																
$\widehat{R}_{HT}$	3.8276	83.4	6.3	10.3	0.5086	83.3	7.8	8.9	1.3731	85.4	6.6	8	1.1126	82.9	8.9	8.2
$\widetilde{R}_d$	4.7222	83	5.5	11.5	1.0117	83.7	8.2	8.1	1.3943	84.4	7	8.6	1.0308	81.3	9.1	9.6
$\widetilde{R}_r$	3.6417	78.9	8.7	12.4	3.5245	79.4	8	12.6	1.0688	77.8	9.3	12.9	1.2873	79.5	8.8	11.7
$\widetilde{R}_{RKM}$	4.6509	82.6	5.8	11.6	1.0117	83.4	7.6	9	1.3963	85.8	5.5	8.7	1.0308	81.9	8.8	9.3
$\widehat{R}_{YQUAR}$	4.7222	82.2	5.7	12.1	1.5538	84.7	6.4	8.9	1.3963	85.2	5.4	9.4	0.9372	81.6	8.4	10
$\widehat{R}_{YCO}$	3.8444	82.5	5.6	11.9	2.0123	83.8	6.5	9.7	1.3281	85.2	4.9	9.9	1.0308	82.6	7.2	10.2

proposed estimators can reduce the error of the usual direct and indirect estimators.
Overall, the proposed estimators $\widehat{L}_{YCO}$ and $\widehat{R}_{YCO}$, appear to be good estimators.

It is also remarkable that the bootstrap method tends to overestimate the variance for the poverty threshold L, whenever that bootstrap variance is smaller than the variance in the case of the 95th / 50th percentile ratio. This is not surprising since the bootstrap technique for nonlinear parameters does not provide unbiased estimators of the variances and rescaling may be necessary to achieve exact unbiasedness (Wolter, 2007). This same problem appears in the results obtained in the simulations performed by [1, 2]. In these simulations the bootstrap variance estimators are also strongly biased when applied to quantiles and poverty measures.

7. Conclusions

In this paper we investigate the optimum estimation of the quantiles in the sense of minimum variance. We start from the calibration estimator proposed by [37] and transform the problem of minimizing the variance of this quantile estimator into a problem of minimizing the variance of the estimator of the associated distribution function. Besides, we obtain an optimal estimator of the distribution function $\widehat{F}_{YO}(t)$ that is a genuine function of distribution and therefore does not need the procedure to satisfy the non-decreasing monotony as the $\tilde{Q}_d$ and $\tilde{Q}_{RKM}$ estimators. The simulation studies indicate that calibrated estimators proposed in the present work are also a suitable option for the estimation of measures for wage inequality based on percentiles ratios and poverty lines, but the simulation also shows that the bootstrap estimators for the poverty measures does not provide unbiased estimators of the variances and rescaling may be necessary to achieve exact unbiasedness.

References

References

[1] Antal, E., Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. Journal of the American Statistical Association 106 (494), 534–543.

[2] Antal, E., Tillé, Y. (2014). A new resampling method for sampling designs without replacement: the doubled half bootstrap. Computational Statistics 29 (5), 1345–1363.

[3] Arcos, A., Martínez, S., Rueda, M., Martínez, H. (2017) Distribution function estimates from dual frame context, Journal of Computational and Applied Mathematics 318, 242–252.

[4] Arcos, A., Rueda, M., Muñoz, J.F. (2007) An improved class of estimators of a finite population quantile in sample surveys Applied Mathematics Letters 20(3), 312–315.

[5] Bickel, P. J., Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. The Annals of Statistics 12 (2), 470–482.

[6] Bogin, B., Sullivan, T. (1986). Socioeconomic status, sex, age, and ethnicity as determinants of body fat distribution for Guatemalan children. American Journal of Physical Anthropology, 69(4), 527-535.

[7] Booth, J. G., R. W. Butler, R. W., Hall, P. (1994). Bootstrap methods for finite populations. Journal of the American Statistical Association 89 (428), 1282–1289.

[8] Burtless, G. (1999). Effects of growing wage disparities and changing family composition on the US income distribution. European Economic Review, 43(4-6), 853-865.

[9] Chambers, R. L., Dunstan, R. (1986). Estimating distribution functions from survey data. Biometrika, 73 (3), 597-604.

[10] Chao, M. T., Lo, S.H. (1994). Maximum likelihood summary and the bootstrap method in structured finite populations. Statistica Sinica 4 (2), 389–406.

[11] Chauvet, G. (2007) Méthodes de bootstrap en population finie. PhD thesis, Université Rennes 2.

[12] Chen, J., Sitter, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. Statistica Sinica, 385-406.

26

[13] Chen, J., Wu, C., 2002. Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. Statist. Sinica 12, 1223-1129.

[14] Decker, R., Haltiwanger, J., Jarmin, R., Miranda, J. (2014). The role of entrepreneurship in US job creation and economic dynamism. Journal of Economic Perspectives, 28(3), 3-24.

[15] Deville,J.C., Särndal, C.E.(1992). Calibration estimators in survey sampling, J. Amer. Statist. Assoc. 87(418), 376-382.

[16] Dickens, R., Manning, A. (2004). Has the national minimum wage reduced UK wage inequality?. Journal of the Royal Statistical Society: Series A, 167(4), 613-626.

[17] Dorfman, A. H., Hall, P. (1993). Estimators of the finite population distribution function using nonparametric regression. The Annals of Statistics, 21(3), 1452-1475.

[18] Efron, B. (1979). Bootstrap methods: another look at the jackknife. The Annals of Statistics 7 (1), 1–26.

[19] Gross, S. (1980). Median estimation in sample surveys. In Proceedings of the Section on Survey Research Methods, American Statistical Association, 181–184.

[20] Harms, T., Duchesne, P. (2006). On calibration estimation for quantiles, Survey Methodology. 32, 37–52.

[21] INE (2015). Continuous Register Statistics. $https$ $:$ $//www.ine.es/dyngs/INEbase/en/operacion.htm?c$ $=$ $Estadistica\_C\&cid = 1254736177012\&menu = resultados$

[22] Jones Jr, A. F., Weinberg, D. H. (2000). The changing shape of the nation's income distribution, 1947-1998. Current Population Reports P60-204. Washington, DC:U.S. Census Bureau. http://www.census.gov/ftp/pub/hhes/www/p60204.html

[23] Kuk, A. Y., Mak, T. K. (1989). Median estimation in the presence of auxiliary information. Journal of the Royal Statistical Society: Series B (Methodological), 51(2), 261-269.

[24] Machin, S., Manning, A., Rahman, L. (2003). Where the minimum wage bites hard: introduction of minimum wages to a low wage sector. Journal of the European Economic Association, 1(1), 154-180.

[25] Martínez, S., Rueda, M., Arcos, A., Martínez, H. (2010). Optimum calibration points estimating distribution functions. Journal of computational and applied mathematics, 233(9), 2265-2277.

[26] Martínez, S., Rueda, M., Arcos, A., Martínez, H., Sánchez-Borrego, I. (2011). Post-stratified calibration method for estimating quantiles. Computational Statistics and Data Analysis, 55(1), 838-851.

[27] Martínez, S., Rueda, M., Arcos, A., Martínez, H., Muñoz, J. F. (2012). On determining the calibration equations to construct model-calibration estimators of the distribution function. Revista matemática complutense, 25(1), 87-95.

[28] Martínez, S., Rueda, M., Martínez, H., Arcos, A. (2015). Determining P optimum calibration points to construct calibration estimators of the distribution function. Journal of Computational and Applied Mathematics, 275, 281-293.

[29] Martínez, S., Rueda, M., Martínez, H., Arcos, A. (2017). Optimal dimension and optimal auxiliary vector to construct calibration estimators of the distribution function. Journal of Computational and Applied Mathematics, 318, 444-459.

[30] Mayor-Gallego, J. A., Moreno-Rebollo, J. L., Jiménez-Gamero, M. D. (2019). Estimation of the finite population distribution function using a global penalized calibration method. AStA Advances in Statistical Analysis, 103 (1), 1–35.

[31] Metcalf, D. (2008). Why has the British national minimum wage had little or no impact on employment?. Journal of Industrial Relations, 50(3), 489-512.

[32] Morales, D., Rueda, M., Esteban, D. (2018). Model-assisted estimation of small area poverty measures: An application within the Valencia Region in Spain. Social Indicators Research, 138, 873–900.

[33] Nickell, S. (2004). Poverty and worklessness in Britain. The Economic Journal, 114(494), C1-C25.

[34] Rao, J. N. K. and Kovar, J. G. and Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information, Biometrika, 77(2), 365–375.

28

[35] Rueda, M., Martinez, S., Martinez, H., Arcos, A. (2006). Mean estimation with calibration techniques in presence of missing data. Computational statistics and data analysis, 50(11), 3263-3277.

[36] Rueda, M., Martínez, S., Martínez, H., Arcos, A. (2007). Estimation of the distribution function with calibration methods. Journal of statistical planning and inference, 137(2), 435-448.

[37] Rueda, M., Martínez-Puertas, S., Martínez-Puertas, H., Arcos, A. (2007). Calibration methods for estimating quantiles. Metrika, 66(3), 355-371.

[38] Rueda, M., Martinez, S., Arcos, A., Munoz, J. F. (2009). Mean estimation under successive sampling with calibration estimators. Communications in Statistics—Theory and Method, 38(6), 808-827.

[39] Särndal C.E. (2007) The calibration approach in survey theory and practice. Survey Metholodogy, 33(2), 99–119.

[40] Singh, H. P., Singh, S., Kozak, M. (2008). A family of estimators of finite-population distribution function using auxiliary information. Acta applicandae mathematicae, 104(2), 115-130.

[41] Tellez-Plaza, M., Navas-Acien, A., Crainiceanu, C. M., Guallar, E. (2007). Cadmium exposure and hypertension in the 1999–2004 National Health and Nutrition Examination Survey (NHANES). Environmental health perspectives, 116(1), 51-56.