STATISTICS IN TRANSITION NEW SERIES, September 2016
Vol. 17, No. 3, pp. 411–428

ESTIMATION OF MEAN ON THE BASIS OF CONDITIONAL SIMPLE RANDOM SAMPLE

Janusz Wywiał[1]

ABSTRACT

Estimation of the population mean in a finite and fixed population on the basis of the conditional simple random sampling design dependent on order statistics (quantiles) of an auxiliary variable is considered. Properties of the well-known Horvitz-Thompson and ratio type estimators as well as the sample mean are taken into account under the conditional simple random sampling designs. The considered examples of empirical analysis lead to the conclusion that under some additional conditions the proposed estimation strategies based on the conditional simple random sample are usually more accurate than the mean from the simple random sample drawn without replacement.

Key words: conditional sampling design, order statistic, concomitant, sample quantile, auxiliary variable, Horvitz-Thompson statistic, inclusion probabilities, sampling scheme, ratio estimator.

1. Introduction

Sampling designs dependent on an auxiliary variable are constructed in order to improve accuracy of population parameters estimation. Application of auxiliary information to construction of conditional versions of sampling designs are considered, e.g. by Royall and Cumberland (1981), Tillé (1998, 2006) and Wywiał (2003).

The fixed population of size N denoted by U will be taken into account. The observation of a variable under study and an auxiliary variable are identifiable and denoted by y_i and $x_i, i = 1,\ldots,N$, respectively. We assume that $x_i \leq x_{i+1}, i = 1,\ldots,N-1$. Our general purpose is estimation of the population average: $\bar{y} = \frac{1}{N}\sum_{k\in U} y_k$ where y_i, $i = 1,...,N$, are values of the variable under study.

The well-known simple random sampling design is defined as follows: $P_0(s) = \binom{N}{n}^{-1}$ for all $s \in \mathbf{S}$ where $\mathbf{S}$ is the sample space of the samples s with fixed effective size $1 < n < N$.

Let $s = \{s_1, i, s_2\}$ where $s_1 = \{i_1,...,i_{r-1}\}$, $s_2 = \{i_{r+1},...,i_n\}$, $i_j < i$ for $j = 1,...,r$, $i_r = i$ and $i_j > i$ for $j = r+1,...,n$. Thus, x_i is one of the possible observations of order statistic $X_{(r)}$ of rank r $(r = 1,...,n)$ from sample s. Let $\mathbf{S}(r,i) =$

[1]Department of Statistics, Faculty of Management, University of Economics in Katowice. E-mail: janusz.wywial@ue.katowice.pl

$\{s : X_{(r)} = x_i\}$ be the set of all samples whose r -th order statistic of the auxiliary variable is equal to x_i where $r \le i \le N-n+r$. Hence, $\bigcup_{i=r}^{N-n+r} \mathbf{S}(r,i) = \mathbf{S}$.

The size of the set $\mathbf{S}(r,i)$ is denoted by $g(r,i) = Card(\mathbf{S}(r,i))$ and

$$g(r,i) = \binom{i-1}{r-1}\binom{N-i}{n-r}. \tag{1}$$

The conditional version of the order statistic distribution is as follows:

$$P\left(X_{(r)} = x_i | x_u \le X_{(r)} \le x_w\right) = \frac{P\left(X_{(r)} = x_i\right)}{P\left(x_u \le X_{(r)} \le x_w\right)} = \frac{g(r,i)}{z(r,u,w)} \tag{2}$$

where

$$P\left(X_{(r)} = x_i\right) = \frac{g(r,i)}{\binom{N}{n}}, \quad i = r,...,N-n+r. \tag{3}$$

$$P\left(x_u \le X_{(r)} \le x_w\right) = \frac{z(r,u,w)}{\binom{N}{n}}, \tag{4}$$

$$z(r,u,w) = \sum_{t=u}^{w} g(r,t). \tag{5}$$

Wywiał (2014) proposed the following conditional version of the simple random sampling design:

$$P_0\left(s|r,u,w\right) = P_0\left(s|x_u \le X_{(r)} \le x_w\right) = \frac{1}{z(r,u,w)}. \tag{6}$$

$P_0\left(s|r,u,w\right)$ provide such the simple random samples that r-th order $X_{(r)}$ takes a value from interval $[x_u;\ x_w]$ where $u \le r \le w$. Let us note that in the particular case when $u = r$ and $w = N-n+r$ sampling design $P_0(s|r,u,w)$ becomes ordinary simple random sample design $P_0(s)$.

Wywiał (2014) derived the first and second order inclusion probabilities for the sampling design. Moreover, Wywiał proposed the following sampling scheme implementing $P_0(s|r,u,w)$. Firstly, population elements are ordered according to the increasing values of the auxiliary variable. Next, the i-th element of the population where $i = u, u+1, ..., w$ and $r = [n\alpha]+1$, is drawn with probability:

$$P(X_{(r)} = x_i | x_u \le X_{(r)} \le x_w) = \frac{g(r,i)}{\sum_{j=u}^{w} g(r,j)}. \tag{7}$$

Finally, two simple samples $s_1(i)$ and $s_2(i)$ are drawn without replacement from subpopulations $U_1 = \{1,...,i-1\}$ and $U_2 = \{i+1, i+2, ..., N\}$, respectively. Sample $s_1(i)$ is of size $r-1$ and sample $s_2(i)$ is of size $n-r$. The sampling designs of these samples are independent and

$$P_0(s_1(i)) = \binom{i-1}{r-1}^{-1}, \quad P_0(s_2(i)) = \binom{N-i}{n-r}^{-1}. \tag{8}$$

2. Strategies dependent on conditional simple random sample

2.1. The Horvitz-Thompson estimator

The well known Horvitz-Thompson (1952) estimator is given by:

$$\bar{y}_{HT,s} = \frac{1}{N}\sum_{k\in s}\frac{y_k}{\pi_k} \tag{9}$$

Estimation strategy $(\bar{y}_{HT,s}, P(s))$ is unbiased for $\bar{y}$ if $\pi_k > 0$ for $k = 1,...,N$, where π_k is the inclusion probability of sampling design $P(s)$. The variance of the strategy is:

$$V_0\left(\bar{y}_{HT,s}, P(s)\right) = \frac{1}{N^2}\left(\sum_{k\in U}\sum_{l\in U}\Delta_{k,l}\frac{y_k y_l}{\pi_k\pi_l}\right), \qquad \Delta_{k,l} = \pi_{k,l} - \pi_k\pi_l. \tag{10}$$

Particularly, under simple random sampling design $P_0(s)$ the strategy $(t_{HT,s}, P(s))$ reduces to the simple random sample mean denoted by $(\bar{y}_s, P_0(s))$, where

$$\bar{y}_s = \frac{1}{n}\sum_{k\in s} y_k. \tag{11}$$

It is the unbiased estimator of the population mean and its variance is:

$$V_0(\bar{y}_s) = \frac{N-n}{Nn}v_*(y), \qquad v_*(y) = \frac{1}{N-1}\sum_{k\in U}(y_k - \bar{y})^2. \tag{12}$$

Moreover, let us note that in the case of the unconditional simple random scheme the Horvitz-Thompson strategy reduces to the simple random sample mean.

Example 2.1. *In the book by Särndal C. E., B. Swensson, J. Wretman (1992) the data about Sweden municipalities are presented. The size of the population of municipalities is* $N = 284$. *We take into account two variables. The first is revenues from the 1985 municipal taxation (in millions of kronor) and it is treated as the variable under study denoted by y. The second one is the 1975 population of municipalities (in thousands) and it is treated as the auxiliary variable denoted by x. Our purpose is the estimation of population mean* $\bar{y}$. *The mean of the auxiliary variable is* $\bar{x} = 28.810$. *The population mean of the variable under study is estimated by means of strategy* $(\bar{y}_{HT,s}, P_0(|r,u,w))$. *The relative efficiency is denoted by:*

$$deff(r,u,w|n) = V(\bar{y}_{HT,s}, P_0(s|r,u,w))/V_0(\bar{y}_s)$$

Particulary, we have $deff(3,260,270|3) = 0.045$*,* $deff(11,270,280|15) = 0.14$ *and* $deff(22,267,277|29) = 0.147$*. Thus, in all the considered cases the mean from the conditional simple random sample is several times more accurate than the simple random sample mean.*

2.2. Conditional simple random sample mean

Let H and T be statistics dependent on observations of the variable under study and the auxiliary variable observed in the sample s drawn according to sampling design $P_0(s|r,u,w)$. The basic moments of statistics H and T are as follows:

$$E_0(H|r,u,w) = \sum_{s\in\mathbf{S}(r,u,w)} hP_0(s|r,u,w),$$

$$E_0(HT|r,u,w) = \sum_{s\in\mathbf{S}(r,u,w)} htP_0(s|r,u,w).$$

$$V_0(H,T|r,u,w) = E_0(HT|r,u,w) - E_0(H|r,u,w)E_0(T|r,u,w).$$

Now, let H and T be statistics dependent on order statistic $X_{(r)}$ or its concomitant $Y_{[r]}$. The basic moments of the statistics H and T are denoted as follows:

$$E(H|r,u,w) = \sum_{i=u}^{w} h_i P(X_{(r)} = x_i|x_u \leq X_{(r)} \leq x_w),$$

$$E(HT|r,u,w) = \sum_{i=u}^{w} h_i t_i P(X_{(r)} = x_i|x_u \leq X_{(r)} \leq x_w),$$

$$V(H,T|r,u,w)) = E(HT|r,u,w) - E(H|r,u,w)E(H|r,u,w),$$

$$V(H|r,u,w)) = V(H,H|r,u,w)).$$

Let random variable I_r have the following probability function:

$$P\left(I_r = i|u \leq r \leq w\right) = P\left(X_{(r)} = x_i|x_u \leq X_{(r)} \leq x_w\right) = \frac{z(r,u,w)}{\binom{N}{n}} \tag{13}$$

where $i = u,...,w$ and $z(r,u,w)$ explains equation (5) and $r \leq u < w \leq N-n+r$. Let $\bar{x}(1,i-1)$, $i = u,...,w$, be the following population mean of the left-truncated (in the point x_i) distribution of the auxiliary variable:

$$\bar{x}(1,i-1) = \frac{1}{i-1}\sum_{k=1}^{i-1} x_k, \quad 1 < r \leq n. \tag{14}$$

Let us note that $\bar{x}(1,i-1)$ is a value of the random variable denoted by $\bar{X}(1,I_r-1)$ and

$$P(I_r=i|u\leq r\leq w)=P\left(X_{(r)}=x_i|x_u\leq X_{(r)}\leq x_w\right). \quad (15)$$

Similarly, we define the following random variables: $\bar{X}(I_r+1,N)$, $\bar{Y}[1,I_r-1]$, $\bar{Y}[I_r+1,N]$, $\overline{XY}[1,I_r-1]$ and $\overline{XY}[I_r+1,N]$, $V_{x,y}[1,I_r-1]$ and $V_{x,y}[I_r+1,N]$ which take values equal to the following moments, respectively:

$$\bar{x}(i+1,N)=\frac{1}{N-i}\sum_{k=i+1}^{N}x_k, \quad 1\leq r<n, i<N \quad (16)$$

$$\bar{y}[1,i-1]=\frac{1}{i-1}\sum_{k=1}^{i-1}y_k, \quad 1<r\leq n, \quad (17)$$

$$\bar{y}[i+1,N]=\frac{1}{N-i}\sum_{k=i+1}^{N}y_k \quad 1\leq r<n, i<N, \quad (18)$$

$$\overline{xy}[1,i-1]=\frac{1}{i-1}\sum_{k=1}^{i-1}x_k y_k, \quad 1<r\leq n, \quad (19)$$

$$\overline{xy}[i+1,N]=\frac{1}{N-i}\sum_{k=i+1}^{N}x_k y_k, \quad 1\leq r<n, i<N, \quad (20)$$

$$v_{x,y}[1,i-1]=\overline{xy}[1,i-1]-\bar{x}(1,i-1)\bar{y}[1,i-1], \quad 1<r\leq n, \quad (21)$$

$$v_{x,y}[i+1,N]=\overline{xy}[i+1,N]-\bar{x}(i+1,N)\bar{y}[i+1,N], \quad 1\leq r<n, i<N. \quad (22)$$

Particularly, $v_x[1,i-1]=v_{x,x}[1,i-1]$ and $v_y[i+1,N]=v_{y,y}[i+1,N]$. Parameters of sample means $\bar{x}_s$, $\bar{y}_s$ under the conditional simple random sample design are considered in the following theorem.

Lemma 2.1. *Under the sampling design defined by expression (6) the basic parameters of $\bar{x}_s$, $\bar{y}_s$ are as follows:*

$$E_0(\bar{x}_s|r,u,w))=\frac{r-1}{n}E\left(\bar{X}(1,I_r-1)|r,u,w\right)+\frac{1}{n}E\left(X_{(r)}|r,u,w\right)+ \\ +\frac{n-r}{n}E\left(\bar{X}(I_r+1,N)|r,u,w\right) \quad (23)$$

where

$$E\left(\bar{X}(1,I_r-1)|r,u,w\right)=\sum_{i=u}^{w}\bar{x}(1,i-1)P\left(X_{(r)}=x_i|r,u,w\right), \tag{24}$$

$$E\left(X_{(r)}|r,u,w\right)=\sum_{i=u}^{w}x_iP\left(X_{(r)}=x_i|r,u,w\right), \tag{25}$$

$$E\left(\bar{X}(I_r+1,N)|r,u,w\right)=\sum_{i=u}^{w}\bar{x}(i+1,N)P\left(X_{(r)}=x_i|r,u,w\right). \tag{26}$$

$$E_0(\bar{y}_s|r,u,w)=\frac{r-1}{n}E\left(\bar{Y}[1,I_r-1]|r,u,w\right)+ \\ +\frac{1}{n}E\left(Y_{[r]}|r,u,w\right)+\frac{n-r}{n}E\left(\bar{Y}[I_r+1,N]|r,u,w\right) \tag{27}$$

where

$$E\left(\bar{Y}[1,I_r-1]|r,u,w\right)=\sum_{i=u}^{w}\bar{y}[1,i-1]P\left(X_{(r)}=x_i|r,u,w\right), \tag{28}$$

$$E\left(Y_{[r]}|r,u,w\right)=\sum_{i=u}^{w}y_iP\left(X_{(r)}=x_i|r,u,w\right), \tag{29}$$

$$E\left(\bar{Y}[I_r+1,N]|r,u,w\right)=\sum_{i=u}^{w}\bar{y}[i+1,N]P\left(X_{(r)}=x_i|r,u,w\right). \tag{30}$$

$$V_0(\bar{x}_s,\bar{y}_s|r,u,w)= \\ =\frac{(r-1)^2}{n^2}V_0(\bar{x}_{s_1},\bar{y}_{s_1}|r,u,w)+\frac{r-1}{n^2}V_0(\bar{x}_{s_1},Y_{[r]}|r,u,w)+ \\ +\frac{(r-1)(n-r)}{n^2}V_0(\bar{x}_{s_1},\bar{y}_{s_2}|r,u,w)+\frac{r-1}{n^2}V_0(X_{(r)},\bar{y}_{s_1}|r,u,w)+ \\ +\frac{1}{n^2}V_0(X_{(r)},Y_{[r]}|r,u,w)+\frac{n-r}{n^2}V_0(X_{(r)},\bar{y}_{s_2}|r,u,w)+ \\ +\frac{(r-1)(n-r)}{n^2}V_0(\bar{x}_{s_2},\bar{y}_{s_1}|r,u,w)+\frac{n-r}{n^2}V_0(\bar{x}_{s_2},Y_{[r]}|r,u,w)+ \\ +\frac{(n-r)^2}{n^2}V_0(\bar{x}_{s_2},\bar{y}_{s_2}|r,u,w) \tag{31}$$

where

$$V_0(\bar{x}_{s_1},\bar{y}_{s_1}|r,u,w)=\frac{1}{r-1}E\left(\frac{I_r-r}{I_r-1}V_{xy}[1,I_r-1]|r,u,w\right)+ \\ +V\left(\bar{X}(1,I_r-1),\bar{Y}[1,I_r-1]|r,u,w\right), \tag{32}$$

$$V_0(\bar{x}_{s_1},Y_{[r]}|r,u,w)=V(Y_{[r]},\bar{X}(1,I_r-1)|r,u,w), \quad (33)$$

$$V_0(X_{(r)},\bar{y}_{s_1}|r,u,w)=V(X_{(r)},\bar{Y}[1,I_r-1]|r,u,w)= \\ =E(X_{(r)}\bar{Y}[1,I_r-1]|r,u,w)-E(X_{(r)}|r,u,w)E(\bar{Y}[1,I_r-1]|r,u,w), \quad (34)$$

$$V_0(\bar{x}_{s_1},\bar{y}_{s_2}|r,u,w)=V\left(\bar{X}(1,I_r-1),\bar{Y}[I_r+1,N]|r,u,w\right)= \\ =E\left(\bar{X}(1,I_r-1)\bar{Y}[I_r+1,N]|r,u,w\right)+ \\ -E\left(\bar{X}(1,I_r-1)|r,u,w\right)E\left(\bar{Y}[I_r+1,N]|r,u,w\right), \quad (35)$$

$$V_0(\bar{x}_{s_2},\bar{y}_{s_1}|r,u,w)=V\left(\bar{X}(I_r+1,N),\bar{Y}[1,I_r-1]|r,u,w\right)= \\ =E\left(\bar{X}(I_r+1,N)\bar{Y}[1,I_r-1]|r,u,w\right)+ \\ -E\left(\bar{X}(I_r+1,N)|r,u,w\right)E\left(\bar{Y}[1,I_r-1]|r,u,w\right), \quad (36)$$

$$V(X_{(r)},Y_{[r]}|r,u,w)=E(X_{(r)},Y_{[r]}|r,u,w)-E(X_{(r)}|r,u,w)E(_{[r]}|r,u,w), \quad (37)$$

$$E(X_{(r)},Y_{[r]}|r,u,w)=\sum_{i=u}^{w}x_iy_iP\left(X_{(r)}=x_i|r,u,w\right),$$

$$V_0(\bar{x}_{s_2},Y_{[r]}|r,u,w)=V(Y_{[r]},\bar{X}(I_r+1,N)|r,u,w), \quad (38)$$

$$V_0(X_{(r)},\bar{y}_{s_2}|r,u,w)=V(X_{(r)},\bar{Y}[I_r+1,N]|r,u,w), \quad (39)$$

$$V_0(\bar{x}_{s_2},\bar{y}_{s_2}|r,u,w)=\frac{1}{n-r}E\left(\frac{N-n+r-I_r}{N-I_r}V_{xy}[I_r+1,N]|r,u,w\right)+ \\ +V(\bar{X}(I_r+1,N),\bar{Y}(I_r+1,N)|r,u,w)). \quad (40)$$

The proof is presented in the Appendix.Let $v_{xy}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})$ and $v_x=v_{xx}$, $v_y=v_{yy}$.

Theorem 2.1. *When $y_k\approx\bar{y}+a(x_k-\bar{x})$ for all $k=1,\ldots,N$ where $a=\frac{v_{xy}}{v_x}$ and $E_0(\bar{x}_s|r,u,w)=\bar{x}$ where $E_0(\bar{x}_s|r,u,w)$ is expressed by (23), then strategy $(\bar{y}_s,P_0(|r,u,w))$ is approximately unbiased for $\bar{y}$.*

The expressions (31)-(41) of Lemma 4.1 determine approximate variance:

$$V(\bar{y}_s|r,u,w) =$$
$$= \frac{1}{n^2}\left\{(r-1)E\left(\frac{I_r-r}{I_r-1}V_y[1,I_r-1]|r,u,w\right) + (r-1)^2V(\bar{Y}[1,I_r-1]|r,u,w)+\right.$$
$$+2(r-1)V(Y_{[r]},\bar{Y}(1,I_r-1)|r,u,w)+$$
$$+2(r-1)(n-r)V\left(\bar{Y}(1,I_r-1),\bar{Y}[I_r+1,N]|r,u,w\right)+V(Y_{[r]}|r,u,w)+$$
$$+2(n-r)V(Y_{[r]},\bar{Y}(I_r+1,N)|r,u,w)+(n-r)^2V(\bar{Y}(I_r+1,N)|r,u,w))+$$
$$\left.+(n-r)E\left(\frac{N-n+r-I}{N-I}V_y[I_r+1,N]|r,u,w\right)\right\}. \quad (41)$$

The proof is presented in the Appendix. Hence, if sample size n and parameters u and w are fixed then parameter r has to be determined in such a way that $|E_0(\bar{x}_s|r,u,w)-\bar{x}| = minimum$.

Example 2.2. *Let us consider the same data as those taken into account in Example 2.1. Our purpose is estimation population mean $\bar{y}$ by means of $(\bar{y}_s, P_0(s|r,u,w))$. The relative bias of the strategy is denoted by:*

$$\delta(r,u,w|n) = b(\bar{y}_s, P_0(s|r,u,w))/\sqrt{V(\bar{y}_s, P_0(s|r,u,w))}$$

where $b(r,u,w) = \bar{y}_s - \bar{y}$ is the bias of $(\bar{y}_s, P_0(s|r,u,w))$. The relative efficiency is denoted by

$$deff(r,u,w|n) = MSE(\bar{y}_s, P_0(s|r,u,w))/V_0(\bar{y}_s)$$

where

$$MSE(\bar{y}_s, P_0(s|r,u,w)) = V(\bar{y}_s, P_0(s|r,u,w)) + b^2(r,u,w)$$

and $V_0(\bar{y}_s)$ is the variance of the mean from the simple random sample drawn without replacement. After some computations we have:
$\delta(3,195,205|3) = -0.915$, $deff(3,195,205|3) = 0.372$,
$\delta(11,170,230|15) = -0.022$, $deff(11,170,230|15) = 3.458$,
$\delta(22,203,212|29) = -0.124$, $deff(22,203,212|29) = 5.74$.
Hence, only in the case of the small sample size $n = 3$ the mean from the conditional simple sample is more accurate than the simple sample mean.

David and Nagaraja (2003), p. 145 show that

$$E(Y_{[r]}) = \bar{y} + \frac{v_{xy}}{v_x}(E(X_{(r)}) - \bar{x}), \qquad r = 1,...,n \quad (42)$$

This let us consider concomitant $Y_{[r]}$ as the estimator of the population mean. Equation (42) straightforwardly leads to the following theorem.

Theorem 2.2. *Under the sampling design defined by expression (6), concomitant* $Y_{[r]}$ *is unbiased estimator of the population mean when*

$$E(X_{(r)}) = \bar{x} \quad and \quad V(Y_{[r]}) = \sum_{i=r}^{N-n+r} (y_i - \bar{y})^2 P(X_{(r)} = x_i). \tag{43}$$

Example 2.3. *Let us consider the same data as those taken into account in Example 2.1. Our purpose is estimation* $\bar{y}$. *The relationship between the variable under study and the auxiliary one is strict because their correlation coefficient is equal to* 0.967. *The population mean of the variable under study we estimate by means of* $(Y_{[r]}, P_0(s))$. *The range of* $Y_{[r]}$ *is the same as the range of* $X_{(r)}$ *where range* r *minimizes quantity* $|E(X_(r)) - \bar{x}|$.

The relative efficiency of the strategy is determined according the following expression: $deff(r|n) = V(Y_{[r]}, P_0(s))/V_0(\bar{y}_s)$ *After appropriate calculation, we have* $deff(2|3) = 0.235$, $deff(11|15) = 0.370$, $deff(22|29) = 0.430$. *Thus, in all the considered cases the mean from the conditional simple sample is more accurate than the simple sample mean.*

Example 2.4. *We still consider the problem formulated in Example 2.3. Now the population mean of municipal taxation is estimated on the basis of strategy* $(Y_{[r]}, P_0(s|r,u,w))$. *The relative bias of the strategy is denoted by* $\delta(r,u,w|n) = b(\bar{y}_s, P_0(s|r,u,w))/\sqrt{V(\bar{y}_s, P_0(s|r,u,w))}$ *where* $b(r,u,w)$ *is the bias of* $(\bar{y}_s, P_0(s|r,u,w))$. *The relative efficiency is defined as* $deff(r,u,w|n) = MSE(Y_{[r]}, P_0(s|r,u,w))/V_0(\bar{y}_s)$. *After some calculations we have:*
$\delta(3,213,222|3) = -0.796$, $deff(3,213,222|3) = 0.009$,
$\delta(11,213,222|15) = -0.777$, $deff(11,213,222|15) = 0.05$,
$\delta(22,200,210|29) = -1.604$, $deff(22,200,210|29) = 0.092$.
Thus, in all the considered cases the mean from the conditional simple sample is more accurate than the simple random sample mean.

2.3. Conditional ratio strategy

Let us consider the following ratio-type estimator:

$$\hat{y}_{r,u,w,s} = \bar{y}_s \frac{E_0(\bar{x}_s|r,u,w)}{\bar{x}_s} \tag{44}$$

where $E_0(\bar{x}_s|r,u,w)$ is explained by (23)-(26).

Lemma 2.2. *Under the sampling design defined by (6):*

$$E_0(\hat{y}_{r,u,w,s}|r,u,w) \approx E_0(\bar{y}_s|r,u,w), \tag{45}$$

and

$$V_0(\hat{y}_{r,u,w,s}|r,u,w) \approx V_0(\bar{y}_s|r,u,w) - 2h(r,u,w)V_0(\bar{x}_s,\bar{y}_s|r,u,w)+ \\ +h^2(r,u,w)V_0(\bar{x}_s|r,u,w) \quad (46)$$

where

$$h(r,u,w) = \frac{E_0(\bar{y}_s|r,u,w)}{E_0(\bar{x}_s|r,u,w)}, \quad (47)$$

Expected values $E_0(\bar{y}_s|r,u,w)$ *and* $E_0(\bar{x}_s|r,u,w)$ *are explained by (23)-(30). Expressions (31)-(41) of Lemma 2.1. let approximate variances* $V_0(\bar{x}_s|r,u,w)$, $V_0(\bar{y}_s|r,u,w)$ *and covariance* $V_0(\bar{x}_s,\bar{y}_s|r,u,w)$. The proof is in the Appendix.

Theorem 2.3. *If* $E_0(\bar{x}_s|r,u,w) = \bar{x}$, *then* $(\hat{y}_{r,u,w,s}, P_0(s|r,u,w))$ *is approximately unbiased for* m_y. *Hence:* $E_0(\hat{y}_{r,u,w,s}|r,u,w) \approx m_y$. The proof is similar to the proof of Theorem 2.1 presented in Appendix.

Example 2.5. *We continue the problem formulated in the previous examples. Now the population mean of municipal taxation is estimated on the basis of ratio strategy* $(\hat{y}_{r,u,w,s}, P_0(s|r,u,w))$. *Some calculations lead to*
$\delta(3,243,252|3) = -1.354$, $deff(3,243,252|3) = 0.010$,
$\delta(11,203,212|15) = 0.119$, $deff(11,203,212|15) = 0.111$,
$\delta(22,203,212|29) = -0.338$, $deff(22,203,212|29) = 0.115$.
In the considered cases the ratio estimator from the conditional simple sample is more accurate than the simple random sample mean. The simpler version of $\hat{y}_{r,u,w,s}$ is as follows:

$$\tilde{y}_{r,u,w,s} = Y_{[r]}\frac{E(X_{(r)}|r,u,w))}{X_{(r)}}, \quad (48)$$

where $E(X_{(r)}|r,u,w)$ is given by (25).

Corollary 2.1. *Under the sampling design defined by expression (6) strategy* $(\tilde{y}_{r,u,w,s}, P_0(s|r,u,w))$ *is approximately unbiased for* m_y *and*

$$V_0(\tilde{y}_{r,u,w,s}|r,u,w) \approx V(Y_{[r]}|r,u,w) - 2hV(X_{(r)},Y_{[r]}|r,u,w)+ \\ +h^2V(X_{(r)}|r,u,w) \quad (49)$$

where

$$h = h(r,u,w) = \frac{E(Y_{[r]}|r,u,w)}{E(X_{(r)}|r,u,w)}$$

and $V(X_{(r)},Y_{[r]}|r,u,w)$ *are explained by (25), (29) (37).* The proof is almost the same as the proof of Theorem 2.1. Strategy $(\tilde{y}_{r,u,w,s}, P_0(s|r,u,w))$ does not depend on the shortest or largest values of the auxiliary variable. Hence, the strategy is e.g. useful when there are right or left censored observations of the auxiliary variable.

Example 2.6. *Now the population mean of municipal taxation is estimated on the basis of the strategy* $(\tilde{y}_{r,u,w,s}, P_0(s|r,u,w))$*. After appropriate calculations, we have:*
$\delta(3,200,210|3) = -1.560$, $deff(3,200,210|3) = 0.009$,
$\delta(11,213,223|15) = -0.628$, $deff(11,213,223|15) = 0.047$,
$\delta(22,220,230|29) = 0.209$, $deff(22,220,230|29) = 0.086$.
In the considered cases ratio estimator $\tilde{y}_{r,u,w,s}$ *from the conditional simple sample is more accurate than the simple random sample mean.*

3. Conclusions

Let M_s be the sample median of the auxiliary variable. Thus, when we assume that the distribution of the auxiliary variable is symmetric then $\bar{x} = Me$, where Me is the population median of the auxiliary variable. When we assume that the distribution of the sample median is an approximation of the distribution of the sample mean $\bar{x}_s$ then $P_0(s|x_u \leq M_s \leq x_w)$ can be treated as an approximation of the conditional simple random sampling design denoted by $P_0(s|x_u \leq \bar{x}_s \leq x_w)$, considered by Royall and Cumberland (1981). This consideration can be generalized to the case when the distribution of the auxiliary variable is not necessary symmetric. It is possible to find such rank r that $|E(X_{(r)}) - \bar{x}| = minimum$. Thus, when we assume that the distribution of $\bar{x}_s$ is sufficiently approximated by the distribution of $X_{(r)}$ then $P_0(s|x_u \leq \bar{x}_s \leq x_w)$ can be approximated by $P_0(s|x_u \leq X_{(r)} \leq x_w)$. We can expect that the sampling design can be useful in the case when there are censored observations of the auxiliary variable as well as when outliers exist.

Acknowledgement

The research was supported by the grant number N N111 434137 from the Ministry of Science and Higher Education.

REFERENCES

DAVID, H. A., NAGARAJA, H. N., (2003). Order statistics. John Wiley & Sons.

HORVITZ, D. G., THOMPSON, D. J., (1952). A generalization of the sampling without replacement from finite universe. Journal of the American Statistical Association, Vol. 47, pp. 663–685.

ROYALL, R. M., CUMBERLAND, W. G., (1981). An empirical study of the ratio estimator and estimators of its variance. Journal of the American Statistical Association, Vol. 76.

SÄRNDAL, C. E., SWENSSON, B., WRETMAN, J., (1992). Model Assisted Survey Sampling. Springer Verlag, New York-Berlin-Heidelberg-London-Paris-Tokyo-Hong Kong-Barcelona-Budapest.

TILLé, Y., (1998). Estimation in surveys using conditional inclusion probabilities: Simple random sampling. International Statistical Review, 66, pp. 303–322.

TILLé, Y., (2006). Sampling Algorithms. Springer.

WYWIAŁ, J. L., (2003). On conditional sampling strategies. Statistical Papers, Vol. 44, 3, pp. 397–419.

WYWIAŁ, J. L., (2014). On conditional simple random sample. Statistics in Transition new series, Vol. 15, 4, pp. 525–534.

APPENDICES

Let $\mathbf{S}_1(r,i)=\mathbf{S}(U_1(i),s_1(i))$ and $\mathbf{S}_2(r,i)=\mathbf{S}(U_2(i),s_2(i))$ be the sample spaces of the samples $s_1(i)$ and $s_2(i)$ selected from the sets $U_1(i)=\{1,...,i-1\}$ and $U_2(i)=\{i+1,...,N\}$, respectively. The samples $s(i)$, $s_1(i)$ and $s_2(i)$, are of size n, $r-1$ and $n-r$, respectively, where $s(i)=s_1(i)\cup\{i\}\cup s_2(i)$ and the index i is fixed, $i=r,...,N-n+r$. Sample $s=s_1\cup\{i\}\cup s_2$ where index i is not fixed although $i=r,...,N-n+r$.
Thus, $\mathbf{S}(r,i)=\mathbf{S}(\{1,...,i-1\},s_1(i))\times\{i\}\times\mathbf{S}(\{i+1,...,N\},s_2(i))$
or $\mathbf{S}(r,i)=\mathbf{S}_1(r,i))\times\{i\}\times\mathbf{S}_2(r,i))$
and $\mathbf{S}(r;u,w)=\mathbf{S}(r,u)\times\mathbf{S}(r,u+1)\times...\times\mathbf{S}(r,i)\times...\times\mathbf{S}(r,w)$
where $\mathbf{S}(r,i)$ was defined in Introduction.

Proof of Lemma 2.1

Let us make the following derivation

$$E_0(\bar{x}_s|r,u,w)=E_0\left(\frac{r-1}{n}\bar{x}_{s_1}+\frac{1}{n}X_{(r)}+\frac{n-r}{n}\bar{x}_{s_2}|r,u,w\right)=$$
$$=\frac{r-1}{n}E_0\left(\bar{x}_{s_1}|r,u,w\right)+\frac{1}{n}E_0\left(X_{(r)}|r,u,w\right)+\frac{n-r}{n}E_0\left(\bar{x}_{s_2}|r,u,w\right),$$

On the basis of Definition 2.1 we have:

$$E_0\left(\bar{x}_{s_1}|r,u,w\right)=\sum_{s\in\mathbf{S}(r;u,w)}\bar{x}_{s_1}P_0(s|r,u,w)=\frac{1}{z(r,u,w)}\sum_{i=u}^{w}\sum_{s\in\mathbf{S}(r,i)}\bar{x}_{s_1(i)}=$$
$$=\frac{1}{(r-1)z(r,u,w)}\sum_{i=u}^{w}\sum_{s(i)\in\mathbf{S}_1(r,i)\times\{i\}\times\mathbf{S}_2(r,i)}\sum_{k\in s_1(i)}x_k=$$
$$=\frac{1}{(r-1)z(r,u,w)}\sum_{i=u}^{w}\sum_{s_1(i)\in\mathbf{S}_1(r,i)}\sum_{k\in s_1(i)}x_k=$$

$$
\begin{aligned}
&= \frac{1}{(r-1)z(r,u,w)} \sum_{i=u}^{w} \sum_{k=1}^{i-1} \binom{i-2}{r-2} \binom{N-i}{n-r} x_k = \\
&\qquad = \frac{1}{(r-1)z(r,u,w)} \sum_{i=u}^{w} \binom{i-1}{r-1} \binom{N-i}{n-r} \frac{r-1}{i-1} \sum_{k=1}^{i-1} x_k = \\
&\qquad = \frac{1}{z(r,u,w)} \sum_{i=u}^{w} \binom{i-1}{r-1} \binom{N-i}{n-r} \bar{x}(1,i-1) = \\
&\qquad = \sum_{i=u}^{w} \bar{x}(1,i-1) P\left(X_{(r)} = x_i | r,u,w\right) = E\left(\bar{X}(1,I_r-1)|r,u,w\right). \quad (50)
\end{aligned}
$$

The next derivation is:

$$
\begin{aligned}
E_0\left(X_{(r)}|r,u,w\right) &= \frac{1}{z(r,u,w)} \sum_{i=u}^{w} \sum_{s\in\mathbf{S}(r,i)} x_i = \\
&= \frac{1}{z(r,u,w)} \sum_{i=u}^{w} \sum_{s(i)\in\mathbf{S}_1(r,i)\times\{i\}\times\mathbf{S}_2(r,i)} x_i = \\
&= \frac{1}{z(r,u,w)} \sum_{i=u}^{w} \binom{i-1}{r-1} \binom{N-i}{n-r} x_i = \\
&= \sum_{i=u}^{w} x_i P\left(X_{(r)} = x_i|r,u,w\right) = E\left(X_{(r)}|r,u,w\right) \quad (51)
\end{aligned}
$$

Similar derivation of the parameter $E_0\left(\bar{x}_{s_2}|r,u,w\right)$ and expressions (50), (51) lead to (23).

$$
\begin{aligned}
&V_0(\bar{x}_s,\bar{y}_s|r,u,w) = \\
&\quad = E_0\Bigg(\Bigg(\frac{r-1}{n}\left(\bar{x}_{s_1} - E_0(\bar{x}_{s_1}|r,u,w)\right) + \frac{1}{n}(X_{(r)} - E_0(X_{(r)}|r,u,w)) + \\
&\quad + \frac{n-r}{n}(\bar{x}_{s_2} - E_0(\bar{x}_{s_2}|r,u,w))\Bigg)\Bigg(\frac{r-1}{n}\left(\bar{y}_{s_1} - E_0(\bar{y}_{s_1}|r,u,w)\right) + \\
&\quad + \frac{1}{n}(Y_{[r]} - E_0(Y_{[r]}|r,u,w)) + \frac{n-r}{n}(\bar{y}_{s_2} - E_0(\bar{y}_{s_2}|r,u,w))\Bigg)|r,u,w\Bigg) = \\
&\quad = \frac{(r-1)^2}{n^2} V_0(\bar{x}_{s_1},\bar{y}_{s_1}|r,u,w) + \frac{r-1}{n^2} V_0(\bar{x}_{s_1},Y_{[r]}|r,u,w) +
\end{aligned}
$$

$$+\frac{(r-1)(n-r)}{n^2}V_0(\bar{x}_{s_1},\bar{y}_{s_2}|r,u,w)+\frac{r-1}{n^2}V_0(X_{(r)},\bar{y}_{s_1}|r,u,w)+$$
$$+\frac{1}{n^2}V_0(X_{(r)},Y_{[r]}|P_0(r,u,w))+\frac{n-r}{n^2}V_0(X_{(r)},\bar{y}_{s_2}|r,u,w)+$$
$$+\frac{(r-1)(n-r)}{n^2}V_0(\bar{x}_{s_2},\bar{y}_{s_1}|r,u,w)+\frac{n-r}{n^2}V_0(\bar{x}_{s_2},Y_{[r]}|r,u,w)+$$
$$+\frac{(n-r)^2}{n^2}V_0(\bar{x}_{s_2},\bar{y}_{s_2}|r,u,w) \quad (52)$$

$$V_0(\bar{x}_{s_1},\bar{y}_{s_1}|r,u,w)=$$
$$=\frac{1}{z(r,u,w)}\sum_{s\in\mathbf{S}(r;u,w)}(\bar{x}_{s_1}-E_0(\bar{x}_{s_1}|r,u,w))(\bar{y}_{s_1}-E_0(\bar{y}_{s_1}|r,u,w))=$$
$$=\frac{1}{z(r,u,w)}\sum_{i=u}^{w}\sum_{s\in\mathbf{S}(r,i)}(\bar{x}_{s_1(i)}-E_0(\bar{x}_{s_1}|r,u,w))(\bar{y}_{s_1(i)}-E_0(\bar{y}_{s_1}|r,u,w)).$$

In order to simplify the notation let

$$E(H|r,u,w)=E(H),\qquad V(H,T|r,u,w)=V(H,T),$$

$$p_i=P(X_{(i)}=x_i|r,u,w)=\frac{1}{z(r,u,w)}\binom{i-1}{r-1}\binom{N-i}{n-r}.$$

Let

$$e_k=x_k-E_0(\bar{x}_{s_1}|r,u,w)=x_k-E(\bar{X}(1,I_r-1)|r,u,w)=x_k-E(\bar{X}(1,I_r-1)), \quad (53)$$

$$d_k=y_k-E_0(\bar{y}_{s_1})=y_k-E(\bar{Y}[1,I_r-1]), \quad (54)$$

$$\bar{e}_{s_1(i)}=\frac{1}{r-1}\sum_{k\in s_1(i)}e_k=\bar{x}_{s_1(i)}-E(\bar{X}(1,I_r-1)), \quad (55)$$

$$\bar{d}_{s_1(i)}=\frac{1}{r-1}\sum_{k\in s_1(i)}d_k=\bar{y}_{s_1(i)}-E(\bar{Y}[1,I_r-1]). \quad (56)$$

Thus,

$$V_0(\bar{x}_{s_1},\bar{y}_{s_1}|r,u,w) = \frac{1}{z(r,u,w)}\sum_{i=u}^{w}\sum_{s_1(i)\in\mathbf{S}_1(r,i)}\bar{e}_{s_1(i)}\bar{d}_{s_1(i)} =$$

$$= \frac{1}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\sum_{s_1(i)\in\mathbf{S}_1(r,i)}\sum_{k\in s_1(i)} e_k \sum_{h\in s_1(i)} d_h =$$

$$= \frac{1}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\sum_{s_1(i)\in\mathbf{S}_1(r,i)}\left(\sum_{k\in s_1(i)} e_k d_k + \sum_{k\in s_1(i)}\sum_{h\in s_1(i),h\neq k} e_k d_h\right) =$$

$$= \frac{1}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\sum_{s_1(i)\in\mathbf{S}_1(r,i)}\sum_{k\in s_1(i)} e_k d_k +$$

$$+ \frac{1}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\sum_{s_1(i)\in\mathbf{S}_1(r,i)}\sum_{k\in s_1(i)}\sum_{h\in s_1(i),h\neq k} e_k d_h =$$

$$= \frac{1}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\binom{i-2}{r-2}\binom{N-i}{n-r}\sum_{k\in U_1(i)} e_k d_k +$$

$$+ \frac{1}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\binom{i-3}{r-3}\binom{N-i}{n-r}\sum_{k\in U_1(i)}\sum_{h\in U_1(i),h\neq k} e_k d_h =$$

$$= \frac{r-1}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\binom{i-1}{r-1}\binom{N-i}{n-r}\frac{1}{i-1}\sum_{k\in U_1(i)} e_k d_k +$$

$$+ \frac{(r-1)(r-2)}{(r-1)^2 z(r,u,w)}\sum_{i=u}^{w}\binom{i-1}{r-1}\binom{N-i}{n-r}\frac{1}{(i-1)(i-2)}\sum_{k\in U_1(i)}\sum_{h\in U_1(i),h\neq k} e_k d_k. \quad (57)$$

Let

$$a_k(i) = x_k - \bar{x}(1,i-1), \qquad b_k(i) = y_k - \bar{y}[1,i-1]. \quad (58)$$

Thus, $\sum_{k\in U_1} a_k(i) = \sum_{k\in U_1} b_k(i) = 0$ and

$$e_k = a_k(i) + \bar{x}(1,i-1) - E(\bar{X}(1,I_r-1)), \qquad d_k = b_k(i) + \bar{y}[1,i-1] - E(\bar{Y}[1,I_r-1]). \quad (59)$$

$$V_0(\bar{x}_{s_1},\bar{y}_{s_1}|r,u,w) =$$
$$= \frac{1}{r-1}\sum_{i=u}^{w} p_i \frac{1}{i-1}\sum_{k\in U_1(i)}(a_k(i)+\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1)))$$
$$(b_k(i)+\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]|r,u,w))+$$
$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{p_i}{(i-1)(i-2)}\sum_{k\in U_1(i)}\sum_{h\in U_1(i),h\neq k}$$
$$(a_k(i)+\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1)))(b_h(i)+\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1])=$$

$$=\frac{1}{r-1}\sum_{i=u}^{w}\frac{p_i}{i-1}\sum_{k\in U_1(i)}a_k(i)b_k(i)+\frac{1}{r-1}\sum_{i=u}^{w}\frac{p_i}{i-1}$$
$$\sum_{k\in U_1(i)}(\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1)))(\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]))+$$

$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{p_i}{(i-1)(i-2)}\sum_{k\in U_1(i)}\sum_{h\in U_1(i),h\neq k}a_k(i)b_h(i)+$$
$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{(\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]))p_i}{(i-1)(i-2)}\sum_{k\in U_1(i)}\sum_{h\in U_1(i),h\neq k}a_k(i)+$$
$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{(\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1))p_i}{(i-1)(i-2)}\sum_{k\in U_1(i)}\sum_{h\in U_1(i),h\neq k}b_h(i)+$$
$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{(\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1)))(\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]))p_i}{(i-1)(i-2)}$$
$$\sum_{k\in U_1(i)}\sum_{h\in U_1(i),h\neq k}1=$$

$$=\frac{1}{r-1}\sum_{i=u}^{w}\frac{i-2}{i-1}v_{xy}(1,i-1)p_i+\frac{1}{r-1}\sum_{i=u}^{w}p_i(\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1))$$
$$(\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]))\frac{1}{i-1}\sum_{k\in U_1(i)}1+$$

$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{p_i}{(i-1)(i-2)}\left(\sum_{k\in U_1(i)}a_k(i)\sum_{h\in U_1(i)}b_h(i)-\sum_{k\in U_1(i)}a_k(i)b_k(i)\right)+$$

$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]|r,u,w)}{(i-1)(i-2)}p_i\sum_{k\in U_1(i)}a_k(i)\sum_{h\in U_1(i),h\neq k}1+$$

$$+\frac{r-2}{r-1}\sum_{i=u}^{w}\frac{\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1)|r,u,w)}{(i-1)(i-2)}p_i\sum_{h\in U_1(i)}b_h(i)\sum_{k\in U_1(i),k\neq h}1+$$

$$+\frac{r-2}{r-1}\sum_{i=u}^{w}(\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1)))(\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]))p_i=$$

$$=\frac{1}{r-1}\sum_{i=u}^{w}\frac{i-2}{i-1}v_{xy}(1,i-1)p_i+$$

$$+\frac{1}{r-1}\sum_{i=u}^{w}(\bar{x}(1,i-1)-E(\bar{X}(1,I_r-1)))(\bar{y}[1,i-1]-E(\bar{Y}[1,I_r-1]))p_i+$$

$$+\frac{r-2}{r-1}\left(V\left(\bar{X}(1,I_r-1),\bar{Y}[1,I_r-1]\right)-\sum_{i=u}^{w}\frac{\sum_{k\in U_1(i)}a_k(i)b_k(i)}{(i-1)(i-2)}p_i\right)=$$

$$=\frac{1}{r-1}\sum_{i=u}^{w}\frac{i-r}{i-1}v_{xy}(1,i-1)p_i+V(\bar{X}(1,I_r-1),\bar{Y}(1,I_r-1)|r,u,w)=$$

$$=\frac{1}{r-1}E\left(\frac{I-r}{I-1}V_{xy}(1,I_r-1)\right)+V(\bar{X}(1,I_r-1),\bar{Y}(1,I_r-1)|r,u,w).$$

Derivations of other expression of Lemma 4.1 are similar to the above ones.

Proof of Theorem 2.1

$$E_0(\bar{y}_s|r,u,w)=$$

$$=\sum_{i=u}^{w}\left(\frac{r-1}{n}\bar{y}[1,i-1]+\frac{y_i}{n}+\frac{n-r}{n}\bar{y}[i+1,N]\right)P\left(X_{(r)}=x_i|r,u,w\right)=$$

$$=\sum_{i=u}^{w}\left(\frac{r-1}{n(i-1)}\sum_{k=1}^{i-1}y_k+\frac{y_i}{n}+\frac{n-r}{n(N-i)}\sum_{k=i+1}^{N}y_k\right)P\left(X_{(r)}=x_i|r,u,w\right)\approx$$

$$\approx\sum_{i=u}^{w}\left(\frac{r-1}{n(i-1)}\sum_{k=1}^{i-1}(\bar{y}+a(x_k-\bar{x}))+\frac{\bar{y}+a(x_i-\bar{x})}{n}+\right.$$

$$+\frac{n-r}{n(N-i)}\sum_{k=i+1}^{N}(\bar{y}+a(x_k-\bar{x}))\Bigg)P\left(X_{(r)}=x_i|r,u,w\right)=$$

$$=\bar{y}-a\bar{x}+a\sum_{i=u}^{w}\left(\frac{r-1}{n(i-1)}\sum_{k=1}^{i-1}x_k+\frac{x_i}{n}+\frac{n-r}{n(N-i)}\sum_{k=i+1}^{N}x_k\right)P\left(X_{(r)}=x_i|r,u,w\right)=$$

$$=\bar{y}-a\bar{x}+a\sum_{i=u}^{w}\left(\frac{r-1}{n}\bar{x}(1,i-1)+\frac{x_i}{n}+\frac{n-r}{n}\bar{x}(i+1,N)\right)P\left(X_{(r)}=x_i|r,u,w\right)=$$

$$=\bar{y}+a\left(E_0(\bar{x}_s|r,u,w)-\bar{x}\right).$$

Thus, the proof is completed.

Proof of Lemma 2.2

Estimator $\hat{y}_{r,u,w,s}=\bar{y}_s\frac{E_0(\bar{x}_s|r,u,w)}{\bar{x}_s}$ can be treated as the function of statistics $\bar{x}_s$ and $\bar{y}_s$ denoted by $f(\bar{x}_s,\bar{y}_s)$. The first derivative of $f(\bar{x}_s,\bar{y}_s)$ in points $\bar{x}_s=E_0(\bar{x}_s|r,u,w)$ and $\bar{y}_s=E_0(\bar{y}_s|r,u,w)$ are as follows: $f_x=\frac{\partial f}{\partial\bar{x}_s}=-h$ where $h=\frac{E_0(\bar{y}_s|r,u,w)}{E_0(\bar{x}_s|r,u,w)}$ and $f_y=\frac{\partial f}{\partial\bar{y}_s}=1$, respectively. This let us write the following Taylor's linearisation of $\hat{y}_{r,u,w,s}$:

$$\hat{y}_{r,u,w,s}-E_0(\bar{y}_s|r,u,w)\approx(\bar{y}_s-E_0(\bar{y}_s|r,u,w))-h(\bar{x}_s-E_0(\bar{x}_s|r,u,w))$$

This leads to the derivation of expressions (45) - (47).