Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

Generalisation of Singh and Srivastava's sampling schemes providing unbiased regression estimators

Janusz Wywial
Department of Econometrics, University of Economics,
Bogucicka 14, 40-226 Katowice, Poland; wywial@ae.katowice.pl

Summary

An estimation of a mean value in a fixed population is considered. Sampling schemes implementing the sampling design proportional to the sample generalised variance of auxiliary variables are proposed. The sampling schemes provide two unbiased regression estimators of a mean value dependent on auxiliary variables. The approximate variances of both sampling strategies are derived. Moreover, the unbiased estimators of the variances of the strategies are constructed. Accuracy of regression strategies under this sampling design and simple sample scheme are compared on the basis of the simulation studies.

Sarndal, Swensson, and Wretman (1992) proposed the general regression estimator the population average from a two-stage sample. The particular case of that estimator is considered. A population divided into disjoint and non-empty cluster is considered. Values auxiliary variables are attached to these clusters. On the first stage clusters sample is selected with probability proportional to the sample generalised variance of the auxiliary variables. On the second stage the simple sample is drawn without replacement from each previously selected clusters. Next, a regression type estimator from such two-stage sample is constructed. It is unbiased estimator of the population average of a variable under study. The approximate variance of the strategy is derived and unbiased estimator of the variance. This strategy can be applied in the cases when the number of first stage units is rather small.

KEY WORDS: generalised variance, two-stage sampling scheme, unequal probability sampling, unbiased estimation, regression estimator

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

1. Introduction

Let us assume that in a population the values of auxiliary variables are available. It is well known that the regression estimator has desirable properties but in the simple random sampling it is biased especially in the case when the size of the sample is small. Singh and Srivastava (1980) proposed two unbiased regression type strategies which depend on only one auxiliary variable. They proposed the two sampling designs constructed on the basis of sample variance of an auxiliary variable. This paper can be treated as a generalization of the Sen and Srivastava's results in the case when more than one auxiliary variable is included in the regression estimator. The first strategy consists of the usual multiple regression estimator and a sampling design proportionate to the order sample generalised variance of auxiliary variables. Second strategy consists of a modified regression estimator and another sampling design based on definition of a sample generalised variance. The last strategy consists of regression type estimator and sampling design implemented by two-stage sampling scheme. First stage units are selected with probability proportional to the sample generalised variance of the auxiliary variables. The second stage units are drawn without replacement from each previously selected clusters. All these strategies provide the unbiased estimates of a population mean.

2. Strategy I

Let U be a population consisting of N distinct and identifiable units. The vector $\mathbf{y}^T = [y_1 \ldots y_N]$ consists of all the values of a variable under study. Let $\mathbf{x} = [x_{ij}]$ be the matrix of the dimensions Nxk. The matrix $\mathbf{x}$ consists of all the values of a k-dimensional auxiliary variable. The element x_{ij} is an i-th value (i=1,...,N) of a j-th auxiliary variable (j=1,...,k≥1). Let $\mathbf{J}_N$ be the column vector of the dimensions Nx1. Each element of the vector $\mathbf{J}_N$ is equal to one. Let us define:

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$\bar{y} = \frac{1}{N}\boldsymbol{y}^T\boldsymbol{J}_N, \qquad \bar{\boldsymbol{x}} = \frac{1}{N}\boldsymbol{J}_N^T x$$

$$\boldsymbol{Y} = \boldsymbol{y} - \boldsymbol{J}_N \bar{y}, \qquad \boldsymbol{X} = \boldsymbol{x} - \boldsymbol{J}_N \bar{\boldsymbol{x}}$$

$$v_{yy} = \frac{1}{N}\boldsymbol{Y}^T\boldsymbol{Y}, \qquad \boldsymbol{V} = [v_{jt}] = \frac{1}{N}\boldsymbol{X}^T\boldsymbol{X}, \qquad \boldsymbol{v} = [v_{yj}] = \frac{1}{N}\boldsymbol{X}^T\boldsymbol{Y}$$

The population mean of the variable under study is denoted by $\bar{y}$. The row vector $\bar{\mathbf{x}}$ consists of the population means of the auxiliary variables. The population variance-covariance matrix of the auxiliary variables is denoted by **V**. The vector **v** consists of the population covariances of the auxiliary variables and the variable under study.

Let s be the sample of the size n drawn without replacement from a population U. Let $\boldsymbol{y}_s^T = [y_{i_1} \ldots y_{i_n}]$ be the vector of values of the variable under study observed in the sample. Similarly, the matrix

$$\boldsymbol{x}_s = \begin{bmatrix} x_{i_1 1} & \ldots & x_{i_1 k} \\ x_{i_2 1} & \ldots & x_{i_2 k} \\ \ldots & \ldots & \ldots \\ x_{i_n 1} & \ldots & x_{i_n k} \end{bmatrix}$$

consists of the values of the auxiliary variables observed in the sample s. Moreover, let us define:

$$\bar{y}_s = \frac{1}{n}\boldsymbol{y}_s^T\boldsymbol{J}_n, \qquad \bar{\boldsymbol{x}}_s = \boldsymbol{J}_n^T\boldsymbol{x}_s$$

$$\boldsymbol{Y}_S = \boldsymbol{y}_s - \boldsymbol{J}_n \bar{y}, \qquad \boldsymbol{X}_S = \boldsymbol{x}_s - \boldsymbol{J}_n\bar{\boldsymbol{x}}, \qquad \bar{Y}_S = \bar{y}_s - \bar{y}, \qquad \bar{\boldsymbol{X}}_S = \bar{\boldsymbol{x}}_s - \bar{\boldsymbol{x}}$$

$$\boldsymbol{V}_S = \frac{1}{n}\boldsymbol{x}_S^T\boldsymbol{x}_S - \bar{\boldsymbol{x}}_S^T\bar{\boldsymbol{x}}_S, \qquad \boldsymbol{v}_S = \frac{1}{n}\boldsymbol{x}_S^T\boldsymbol{y}_S - \bar{\boldsymbol{x}}_S^T\bar{y}_S$$

or

$$\boldsymbol{V}_S = \frac{1}{n}\boldsymbol{X}_S^T\boldsymbol{X}_S - \bar{\boldsymbol{X}}_S^T\bar{\boldsymbol{X}}_S, \qquad \boldsymbol{v}_S = \frac{1}{n}\boldsymbol{X}_S^T\boldsymbol{Y}_S - \bar{\boldsymbol{X}}_S^T\bar{\boldsymbol{Y}}_S$$

Hence, $\boldsymbol{V}_S = [v_{sij}]$ is the sample variance-covariance matrix between the auxiliary variables and $\boldsymbol{v}_S$ is the column vector of the sample covariances between the auxiliary variables and the variable under study.

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

Let ***S*** be the sample space of the unordered sample s of the size n selected without replacement from the population U. If the sample s is a simple one, its sampling design is defined by the probabilities $P_O(s) = \binom{N}{n}^{-1}$ for all s $\in$ ***S***.

Let us consider the sampling design $P_1(s) \propto det\boldsymbol{V}_s$. The sample s is selected with a probability proportionate to the sample generalised variance of the auxiliary variables. Wywial (1996, 1997) derived the following result (see the appendix 1, too):

$$P_1(s) = c_1 \frac{det\boldsymbol{V}_s}{det\boldsymbol{V}} \tag{1}$$

where:

$$c_1 = \binom{N-k-1}{n-k-1}^{-1} \left(\frac{n}{N}\right)^{k+1} \tag{2}$$

When k=1, the sampling plan $P_1(s)$ is reduced to the one of Singh and Srivastava (1980). Let s_{k+1} be the subset of the sample s. The size of the subset s_{k+1} is equal to k+1< n.

Let us define the following quantity:

$$q_1(s_{k+1}) = det^2\left[\boldsymbol{J}_{k+1} \quad \boldsymbol{x}_{s_{k+1}}\right] \tag{3}$$

where:

$$\boldsymbol{x}_{s_{k+1}} = \begin{bmatrix} x_{i_1 1} & \dots & x_{i_1 k} \\ . & \dots & . \\ x_{i_k 1} & \dots & x_{i_k k} \\ x_{i_{k+1} 1} & \dots & x_{i_{k+1} k} \end{bmatrix}$$

Let $\boldsymbol{x}_{r*} = \left[x_{i_r 1} \ \dots \ x_{i_r k}\right]$ be the r-th row of the matrix $\boldsymbol{x}_{s_{k+1}}$. After dropping the row $\boldsymbol{x}_{k+1*}$ in the matrix $\boldsymbol{x}_{s_{k+1}}$ we obtain the matrix $\boldsymbol{x}_{s_k}$. After subtracting the last row of the matrix $\left[\boldsymbol{J}_{k+1} \quad \boldsymbol{x}_{k+1}\right]$ from the previous rows of this matrix we have:

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$det\left[\boldsymbol{J}_{k+1} \quad \boldsymbol{x}_{k+1}\right] = det\begin{bmatrix} \boldsymbol{o} & \boldsymbol{x}_{s_k} - \boldsymbol{x}_{k+1*}\boldsymbol{J}_k \\ 1 & \boldsymbol{x}_{k+1*} \end{bmatrix}$$

This let us rewrite the expression (3) in the following way:

$$q_1(s_{k+1}) = det^2\left[\boldsymbol{x}_{s_k} - \boldsymbol{x}_{k+1*}\boldsymbol{J}_k\right] \qquad (4)$$

Let us note that $q_1(s_{k+1})$ is the k-dimensional measure (volume) of the parallelotop spanned by the vectors with their origins at the same point $\boldsymbol{x}_{k+1*}$ and the end points $\boldsymbol{x}_{1*},\ldots,\boldsymbol{x}_{k*}$, see e.g. Borsuk (1969). From the other point of view $q_1(s_{k+1})$ is proportionate to the k-dimensional volume of the simplex spanned by the points $\boldsymbol{x}_{1*},\ldots,\boldsymbol{x}_{k+1*}$.

In the appendix 1 it was proved that the following sampling scheme (implementing the sampling design $P_1(s)$) consists of the two following steps.

Step 1: Select k+1 units $s_{k+1} = \{i_1, i_2, ..., i_{k+1}\}$ with their probability of joint selection being proportional to $q_1(s_{k+1})$.

Step 2: Select (n-k-1) units from the remaining unit of the population by the simple random sampling without replacement.

The sampling design proportional to $q_1(s_{k+1})$ show the expressions (47) or (48) in the appendix 1.

Particularly, if k = 1, this sampling scheme is reduced to the sampling scheme I proposed by Singh and Srivastava (1980). In this case $q_1(s_2) = (x_{i_2} - x_{i_1})^2$.

The well known regression estimator of the population mean $\bar{y}$ is as follows

$$y_{RS} = y_S - (\boldsymbol{x}_S - \boldsymbol{x})\boldsymbol{B}_S \qquad (5)$$

where:

$$\boldsymbol{B}_S = \boldsymbol{V}_S^{-1}\boldsymbol{v}_S$$

Let us introduce the following matrix

$$\boldsymbol{A}_S = \begin{bmatrix} y_S - y & \boldsymbol{x}_S - \boldsymbol{x} \\ \boldsymbol{v}_S & \boldsymbol{V}_S \end{bmatrix} \qquad (6)$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

The well known property of the determinant of a block matrix let us rewrite the estimator y_{RS} in the following way:

$$y_{RS} = y + \frac{det\,\boldsymbol{A}_S}{det\,\boldsymbol{V}_S} \tag{7}$$

The determinant of the matrix $\boldsymbol{A}_S$ can be transformed into the following forms (see, the appendix 2):

$$det\,\boldsymbol{A}_S = n^{-k-1}\,det\begin{bmatrix} n(y_S - y) & n(\boldsymbol{x}_S - \boldsymbol{x}) \\ \boldsymbol{X}_S^T \boldsymbol{Y}_S & \boldsymbol{X}_S^T \boldsymbol{X}_S \end{bmatrix} \tag{8}$$

or

$$det\,\boldsymbol{A}_S = n^{-k-1}\,det\begin{bmatrix} \boldsymbol{J}_n^T \\ \boldsymbol{X}_S^T \end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_S\,\boldsymbol{X}_S\end{bmatrix} \tag{9}$$

In the appendix 3 it will be shown that

$$E(y_{RS}, P_1(s)) = y \tag{10}$$

Hence, $(y_{RS}, P_1(s))$ is the unbiased strategy of the population mean $\bar{y}$.

In the appendix 4 the derivation of the approximate value of the strategy $(\bar{y}_{Rs}, P_1(s))$ is going to be developed. When the sample size $n \to \infty$ and the population size $N \to \infty$ in such a way that N-$n \to \infty$, then

$$D^2(y_{RS}, P_1(s)) \approx \frac{1}{n}\left(v_{yy} - \boldsymbol{v}^T \boldsymbol{V}^{-1} \boldsymbol{v}\right)$$

(11)

Let $\boldsymbol{R}$ be the correlation matrix of auxiliary variables and $\boldsymbol{r}^T = \left[r_{y1} \ldots r_{yk}\right]$, where r_{yj} is the correlation coefficient between the j-th auxiliary variable and the variable under study. This lets us rewrite the expression (11) in the following way:

$$D^2(y_{RS}, P_1(s)) \approx \frac{1}{n} v_{yy}\left(1 - r_w^2\right)$$

(12)

where:

$$r_w = \sqrt{\boldsymbol{r}^T \boldsymbol{R}^{-1} \boldsymbol{r}}$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

is the multiple correlation coefficient between the auxiliary variables and the variable under study. Hence, in the asymptotic case, the precision of the strategy $(y_{RS}, P_1(s))$ increases when the value of the multiple correlation coefficient r_w increases too.

Singh and Srivastava (1980) derived the unbiased estimator of the variance $D^2(y_{RS}, P_1(s))$ in the case when k=1. Generalising their result for k≥1, we can construct the following unbiased estimator of this parameter.

$$\hat{D}_1^2 = y_{RS}^2 - \frac{N^{k-1}}{n^{k+1}} \frac{\prod_{h=1}^{k}(n-h)}{\prod_{h=1}^{k}(N-h)} \frac{\det \boldsymbol{V}}{\det \boldsymbol{V}_S} \left[\sum_{i\in S} y_i^2 + \frac{N-1}{n-1} \sum_{i\neq j\in S} y_i y_j \right] \quad (13)$$

This expression examines the derivation presented in the appendix 5.

3. Strategy II

Let us consider the sampling design $P_2(s) \propto \det \boldsymbol{V}_{\#S}$, where

$$\boldsymbol{V}_{\#S} = \frac{1}{n}\boldsymbol{X}_S^T \boldsymbol{X}_S = \frac{1}{n}(\boldsymbol{x}_S - \boldsymbol{J}_n \bar{\boldsymbol{x}})^T (\boldsymbol{x}_S - \boldsymbol{J}_n \bar{\boldsymbol{x}}) \quad (14)$$

is the sample variance-covariance matrix defined on the basis of the population means of the auxiliary variables. It can be proved that

$$P_2(s) = c_2 \frac{\det \boldsymbol{V}_{\#S}}{\det \boldsymbol{V}} \quad (15)$$

where:

$$c_2 = \frac{1}{\binom{N-k}{n-k}} \left(\frac{n}{N}\right)^k \quad (16)$$

When k=1, the defined sampling design is reduced to the one proposed by Singh and Srivastava (1980).

Let s_k be a subset of the sample s and k < n. Let us define the following:

$$q_2(s_k) = \det^2 \begin{bmatrix} 1 & \bar{\boldsymbol{x}} \\ \boldsymbol{J}_k & \boldsymbol{x}_{S_k} \end{bmatrix} \quad (17)$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

From the geometrical point of view $q_2(s_k)$ is the k-dimensional measure (volume) of the parallelotop spanned by the vectors with their origins at the same point $\bar{\boldsymbol{x}}$ and the end points which determine the rows of the matrix $\boldsymbol{x}_{S_k}$,see, e.g., Anderson (1958) or Borsuk (1969). The expression (17) can be transformed into the following one:

$$q_2(s_k) = det^2\left(\boldsymbol{X}_{S_k}\right) \tag{18}$$

It can be shown that the following sampling scheme implements the sampling design $P_2(s)$:

Step 1: Select k-units $s_k=\{i_1, ..., i_k\}$ with their probability of joint selection being proportional to $q_2(s_k)$.

Step 2: Select (n-k) units from the remaining units of the population by simple random sampling without replacement.

When k=1, the introduced sampling scheme is reduced to the sampling scheme II proposed by Singh and Srivastava (1980). In this case $q_2\left(s_1\right) = \left(x_j - \bar{x}\right)^2$.

Let us consider the following estimator:

$$y_{\#S} = \frac{n(N-k)}{N(n-k)}\left[y_S - \left(\bar{\boldsymbol{x}}_S - \bar{\boldsymbol{x}}\right)\boldsymbol{B}_{\#S}\right] \tag{19}$$

where:

$$\boldsymbol{B}_{\#S} = \boldsymbol{V}_{\#S}^{-1}\boldsymbol{v}_{\#S}$$

$$\boldsymbol{v}_{\#S} = \frac{1}{n}\boldsymbol{X}_S^T\boldsymbol{y}_S$$

The statistic $\bar{y}_{\#S}$ can be easily transformed into the following one:

$$\bar{y}_{\#S} = \frac{n(N-k)}{N(n-k)}\frac{det\,\boldsymbol{A}_{\#S}}{det\,\boldsymbol{V}_{\#S}} \tag{20}$$

where:

$$\boldsymbol{A}_{\#S} = \begin{bmatrix} \bar{\boldsymbol{y}}_S & \bar{\boldsymbol{x}}_S - \bar{\boldsymbol{x}} \\ \boldsymbol{v}_{\#S} & \boldsymbol{V}_{\#S} \end{bmatrix}$$

or:

$$\boldsymbol{A}_{\#S} = \frac{1}{n}\begin{bmatrix} \boldsymbol{J}_n^T \\ \boldsymbol{X}_S^T \end{bmatrix}\left[\boldsymbol{y}_S\boldsymbol{X}_S\right] \tag{21}$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

In the case when k = 1, the statistic $\bar{y}_{\#S}$ is reduced to the estimator proposed by Singh and Srivastava (1980).

It can be proved that $\left(\bar{y}_{\#S}, P_2(s)\right)$ is the unbiased strategy of the population mean $\bar{y}$. The approximate variance of the strategy $\left(y_{\#S}, P_2(s)\right)$ is expressed by the right side of the equation (12). The unbiased estimator of the variance of the sampling strategy $\left(\bar{y}_{\#S}, P_2(s)\right)$ is as follows:

$$\hat{D}_2^2 = y_{\#S}^2 - \frac{N^{k-3}}{n^{k-1}} \frac{\prod_{h=2}^{k}(n-h+1)}{\prod_{h=2}^{k}(N-h+1)} \frac{det\,\mathbf{V}_{\#}}{det\,\mathbf{V}_{\#S}} \left[\sum_{i\in S} y_i^2 + \frac{N-1}{n-1} \sum_{i\neq j\in S} y_i y_j\right] \qquad (22)$$

4. STRATEGY III

Let us assume that a population U is divided into disjoint and non-empty clusters U_g, $g=1,...,G$, and $U = \bigcup_{g=1}^{G} U_g$. The size of the sample U_g is denoted by $N_g > 1$ and $N = \bigcup_{g=1}^{G} N_g$. A multivariate auxiliary variable of dimension k is observed on all first stage units. The sample S, consisted of clusters, is selected on the first stage. The sampling design is proportional to the sample generalised variance of the auxiliary variables and it is determined by the expressions (1) and (2). On the second stage the simple samples $Q_{g_1},...,Q_{g_n}$ are drawn without replacement from the selected on the first stage clusters $U_{g_1},...,U_{g_n}$, respectively. The size of the sample Q_{g_j} is denoted by $1 < m_{g_j} \leq N_{g_j}$. The two stage sample will be denoted by $Q=\{S,Q_1,...,Q_n\}$ and its outcome *by* $q=\{s,q_1,...,q_n\}$. The sampling design is as follows:

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$P_3(q) = P_1(s) \prod_{g \in s} \binom{N_g}{m_g}^{-1} \tag{23}$$

where the sampling design $P_1(s)$ is determined by the equation (1).

Let us introduce the following notation:

$$z_g = \sum_{i \in U_g} y_i \qquad z = \frac{1}{G}\sum_{g=1}^{G} z_g \qquad v_{zz} = \frac{1}{G-1}\sum_{g=1}^{G}\left(z_g - z\right)^2$$

$$y_{U_g} = \frac{1}{N_g} z_g \qquad v_{U_g} = \frac{1}{N_g - 1}\sum_{i \in U_g}\left(y_i - y_{U_g}\right)^2$$

$$z_{Q_g} = \sum_{i \in Q_g} y_i \qquad y_{Q_g} = \frac{1}{m_g} z_{Q_g}$$

$$\tilde{z}_{Q_g} = N_g y_{Q_g} \tag{24}$$

$$v_{Q_g} = \frac{1}{m_g - 1}\sum_{i \in Q_g}\left(y_i - y_{Q_g}\right)^2 \tag{25}$$

The statistics $\tilde{z}_{Q_g}$ and v_{Q_g} are unbiased estimators of the cluster total z_g and the cluster variance v_{U_g}, because:

$$E_{Q_g/S}\left(\tilde{z}_{Q_g}\right) = z_g \tag{26}$$

$$E_{Q_g/S}\left(v_{Q_g}\right) = v_{U_g} \tag{27}$$

The conditional variance of the estimator $\tilde{z}_{Q_g}$ is as follows:

$$D^2_{Q_g/S}\left(\tilde{z}_{Q_g}\right) = \frac{N_g\left(N_g - m_g\right)}{m_g} v_{U_g} \tag{28}$$

Let us consider the following estimator of the parameter z:

$$z_{RQ} = z_Q - \left(\bar{\mathbf{x}}_S - \bar{\mathbf{x}}\right)\mathbf{B}_Q \tag{29}$$

where:

$$z_Q = \frac{1}{n}\sum_{g \in S} \tilde{z}_{Q_g} \tag{30}$$

$$\mathbf{B}_Q = \mathbf{V}_S^{-1}\mathbf{w}_Q \tag{31}$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$\boldsymbol{w}_Q^T = \left[c_Q(z,x_1) \quad \dots \quad c_Q(z,x_k) \right]$$

$$c_Q(z,x_j) = \frac{1}{n-1} \sum_{g \in S} \left(\tilde{z}_{Q_g} - z_Q \right)\left(x_{gj} - x_{jS} \right)$$

or $\quad \boldsymbol{w}_Q = \frac{1}{n} \boldsymbol{X}_S^T \tilde{\boldsymbol{z}}_Q - \bar{\boldsymbol{X}}_S^T z_Q \quad$ or $\quad \boldsymbol{w}_Q = \frac{1}{n} \boldsymbol{x}_S^T \tilde{\boldsymbol{z}}_Q - \bar{\boldsymbol{x}}_S^T z_Q \quad$ (32)

where: $\quad \tilde{\boldsymbol{z}}_Q = \begin{bmatrix} \tilde{z}_{Q_1 / S} \\ \dots\dots\dots \\ \tilde{z}_{Q_1 / S} \end{bmatrix}$

The expressions (26), (29) and (30) lead to the following one:

$$E_{Q/S}\left(z_{RQ} \right) = z_{RS} = z_S - \left(\bar{\boldsymbol{x}}_S - \bar{\boldsymbol{x}} \right)\boldsymbol{B}_S \qquad (33)$$

This and the results of the first paragraph lead to the following expression:

$$E\left(z_{RQ} \right) = E_S E_{Q/S}\left(z_{RQ} \right) = E_S\left(z_{RS} \right) = z \qquad (34)$$

Hence, the regression estimator z_{RQ} from the two-stage sample is the unbiased estimator of the parameter z.

The derivation of the variance is based on the following decomposition:

$$D^2\left(z_{RQ} \right) = E_S\left(D_{Q/S}^2\left(z_{RQ} \right) \right) + D_S^2\left(E_{Q/S}\left(z_{RQ} \right) \right) \qquad (35)$$

This and the expressions (29)-(33) lead to the following:

$$\begin{aligned} D^2\left(z_{RQ} \right) = \frac{1}{n^2} E_S \Bigg\{ & \sum_{g \in S} D_{Q_g/S}^2\left(\tilde{z}_{Q_i} \right) + \\ & \bar{\boldsymbol{X}}_S \boldsymbol{V}_S^{-1} \boldsymbol{X}_S^T diag\left(\boldsymbol{D}_{Q/S}^2\left(\tilde{\boldsymbol{z}}_Q \right) \right) \boldsymbol{X}_S \boldsymbol{V}_S^{-1} \bar{\boldsymbol{X}}_S^T + \\ & - 2 \bar{\boldsymbol{X}}_S \boldsymbol{V}_S^{-1} \boldsymbol{X}_S^T \boldsymbol{D}_{Q/S}^2\left(\tilde{\boldsymbol{z}}_Q \right)\left(1 + \bar{\boldsymbol{X}}_S \boldsymbol{V}_S^{-1} \bar{\boldsymbol{X}}_S^T \right) + \\ & + \bar{\boldsymbol{X}}_S \boldsymbol{V}_S^{-1} \bar{\boldsymbol{X}}_S^T \left(2 + \bar{\boldsymbol{X}}_S \boldsymbol{V}_S^{-1} \bar{\boldsymbol{X}}_S^T \right) \sum_{g \in S} D_{Q_g/S}^2\left(\tilde{z}_{Q_g} \right) \Bigg\} + D_S^2\left(z_{RS} \right) \end{aligned} \qquad (36)$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

where: $\boldsymbol{D}^2_{Q/S}(\tilde{\boldsymbol{z}}_Q) = \begin{bmatrix} D^2_{Q_1/S}(\tilde{z}_{Q_1}) \\ \dots\dots \\ D^2_{Q_n/S}(\tilde{z}_{Q_n}) \end{bmatrix}$ and the elements of this vector are showed by the expression (28). The approximate value of the variance $D^2_S(z_{RS})$ is as follows (see the equation (12)):

$$D^2(z_{RS}, P_3(s)) \approx \frac{1}{n} v_{zz}\left(1 - r_w^2\right) \tag{37}$$

where r_w is the multiple correlation coefficient between the auxiliary variables and the variable *z*.

Let us consider the following statistic:

$$\breve{z}_{Q_g} = \frac{m_g(N_g - 1)}{N_g(m_g - 1)}\left(\tilde{z}^2_{Q_g} - \frac{N_g^2}{m_g}\frac{N_g - m_g}{N_g - 1} M_{Q_g}\right) \tag{38}$$

where:

$$M_{Q_g} = \frac{1}{m_g}\sum_{i \in Q_g} y_i^2$$

It is easy to show that $\breve{z}_{Q_g}$ is unbiased estimator of the squared total value in the g-th cluster, then

$$E_{Q_g/S}\left(\breve{z}_{Q_g}\right) = z_g^2 \quad \text{for } g=1,...,G \tag{39}$$

This and the expression (13) and the results of the appendix 5 let us prove that the following statistic is the unbiased estimator of the variance $D^2(z_{RQ})$:

$$\hat{D}^2_{3,Q} = z^2_{RQ} - \frac{N^{k-1}}{n^{k+1}}\frac{\prod_{h=1}^{k}(n-h)}{\prod_{h=1}^{k}(N-h)}\frac{\det \boldsymbol{V}}{\det \boldsymbol{V}_S}\left[\sum_{g \in S}\breve{z}_{Q_g} + \frac{N-1}{n-1}\sum_{g \neq t \in S}\tilde{z}_{Q_g}\tilde{z}_{Q_t}\right] \tag{40}$$

Finally, the unbiased estimator of population total is as follows:

$$t_{RQ} = G z_{RQ} \tag{41}$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

The unbiased estimator of population average is determined by the expression:

$$\bar{t}_{RQ} = \frac{G}{N} z_{RQ} \tag{42}$$

The variances of these estimators can be easy derived on the basis of the expression (36). The equation (40) leads to construction of the estimators of these variances.

Let us note that the statistics t_{RQ}, defined by the expression (41), can be treated, in some sense, as particular case of the regression estimator of the total value from two stage sample considered by Sarndal, Swensson, and Wretman (1992).

5. Example of simulation analysis of accuracy

Let us consider the example of average estimation by means of regression strategies: $(y_{RS}, P_0(s))$, $(y_{RS}, P_1(s))$, $(y_{\#S}, P_0(s))$ and $(y_{\#S}, P_2(s))$. The variable under study is: the revenues from 1985 municipal taxation (in million of kronor) - *y*, the auxiliary variables are: 1) the number of Conservative seats in municipal council –x_1, the number of Social-Democratic seats in municipal council - x_2 and the real estate values according to 1984 assessment (in milions of kronor) - x_3. These variables are observed in the population of Swedish municipalities. The population is divided into eight strata according to geographical region of Sweden. The data are published by Sarndal at all (1992). We are going to consider only data consisted of 15 municipalities in seventh stratum. The tables 1-6 let compare accuracy of the estimation of the mean of the revenues from 1985 municipal taxation. The relative efficiencies are determined by the expression:

$$e = \frac{D^2(y_{RS}, P_1)}{D^2(y_{RS}, P_0)} 100\% \qquad \text{or} \qquad e = \frac{D^2(y_{\#S}, P_2)}{D^2(y_{\#S}, P_0)} 100\%$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

Table 1. The accuracy of strategies $(y_{RS}, P_0(s))$ and $(y_{RS}, P_1(s))$. The auxiliary variable x_2.

The size of the sample	3	4	5	6	7
The bias under the plan P_0	-45	-28	-20	-16	-11
The % share of the squared bias in the mean square error under the plan P_0	12.1	13.0	10.6	8.6	6.7
The variance under the plan P_0	14867	5508	33230	2415	1785
The variance under the plan P_1	6814	4288	2976	2160	1606
The relative efficiency e	45.8	77.9	89.4	89.5	90.0

Analysis of the tables 1-3 leads us to the following conclusions. The bias of the strategy $(y_{RS}, P_0(s))$ is not greater than 13%. The absolute value of the bias decreases when the number of variables increases. Similarly, the absolute value of the bias decreases when size of sample increases for particular sets of auxiliary variables. The variances of the both strategies decreases when sample size became greater and greater. In the all cases precision of the strategy $(y_{RS}, P_1(s))$ is better than

Table 2. The accuracy of strategies $(y_{RS}, P_0(s))$, $(y_{RS}, P_1(s))$. The auxiliary variables: x_1, x_2.

The size of the sample	4	5	6	7
The bias under the plan P_0	-16	-12	-6	-5
The % share of the squared bias in the mean square error under the plan P_0	4.0	4.9	1.3	1.2
The variance under the plan P_0	6021	2847	2537	1561
The variance under the plan P_1	2780	1739	1434	1020
The relative efficiency e	46.2	61.1	56.5	65.3

strategy $(y_{RS}, P_0(s))$. The relative efficiency coefficient increases when the sample size increases. Hence, we can expect that accuracy of the strategy $(y_{RS}, P_1(s))$ is not much better than the accuracy of the strategy $(y_{RS}, P_0(s))$ in the large sample.

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

Table 3. The accuracy comparison of strategies $\left(y_{RS}, P_0(s)\right)$ and $\left(y_{RS}, P_1(s)\right)$. The auxiliary variables: x_1, x_2 and x_3.

The size of the sample	5	6	7
The bias under the plan P_0	0	-4	-3
The % share of the squared bias in the mean square error under the plan P_0	0.0	0.4	0.8
The variance under the plan P_0	33769	3214	1499
The variance under the plan P_1	2036	1242	855
The relative efficiency e	6.0	38.6	57.0

Table 4. The accuracy of strategies $\left(y_{\#S}, P_0(s)\right)$ and $\left(y_{\#S}, P_2(s)\right)$. The auxiliary variable: x_2.

The size of the sample	2	3	4	5	6	7
The bias under the plan P_0	-39.2	-29	-22	-18	-14	-12
The % share of the squared bias in the mean square error under the plan P_0	4.9	5.4	5.3	5.1	4.8	4.4
The variance under the plan P_0	29508	14355	8689	5778	4045	2912
The variance under the plan P_1	42285	17499	9727	6104	4084	2831
The relative efficiency e	143.3	121.9	112.0	105.7	101.0	97.3

Table 5. The accuracy of strategies $\left(y_{\#S}, P_0(s)\right)$, $\left(y_{\#S}, P_2(s)\right)$. The auxiliary variables x_1, x_2.

The size of the sample	3	4	5	6	7
The bias under the plan P_0	-39	-32	-26	-21	-16
The % share of the squared bias in the mean square error under the plan P_0	4.2	6.3	7.0	6.8	6.1
The variance under the plan P_0	33461	15014	8846	5813	4003
The variance under the plan P_1	33912	14026	7692	4717	3051
The relative efficiency e	101.4	93.4	87.0	81.1	76.2

The analysis of tables 1-6 let us infer that in small samples the regression estimator $y_{\#s}$ from the simple sample can be more precise than the strategy $\left(y_{\#s}, P_2(s)\right)$. The bias of the strategy $\left(y_{\#S}, P_0(s)\right)$ is larger than the bias of the strategy $\left(y_{RS}, P_1(s)\right)$ in the cases of two or three

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

auxiliary variables. In the all cases the variance of the strategy $(y_{RS}, P_1(s))$ is smaller than the variances of the strategies $(y_{\#S}, P_2(s))$, $(y_{\#S}, P_0(s))$ and $(y_{RS}, P_0(s))$. The same is with their mean square errors.

Table 6. The accuracy comparison of strategies $(y_{\#S}, P_0(s))$ and $(y_{\#S}, P_2(s))$.

The auxiliary variables: x_1, x_2 and x_3.

The size of the sample	4	5	6	7
The bias under the plan P_0	-60	-45	-34	-25
The % share of the squared bias in the mean square error under the plan P_0	9.5	10.6	9.8	8.5
The variance under the plan P_0	34551	16940	10380	6764
The variance under the plan P_1	41342	16663	8844	5191
The relative efficiency e	119.7	98.4	85.2	76.7

Source: own elaboration

Let us note that in the case the designs $P_1(s)$ and $P_2(s)$ when size of the sample is large we have serious numerical problems with the sampling schemes implementing those designs. This and the above analysis lead to conclusion that the strategy $(y_{RS}, P_1(s))$ should be preferred to strategies $(y_{RS}, P_0(s))$, $(y_{\#S}, P_0(s))$ and $(y_{\#S}, P_2(s))$ in the case when the sample size is rather small.

Acknowledgement

The research was partially supported by the grants number 1 H02B 015 10 and 1 H02B 008 16 from the State Committee of Scientific Research.

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

6. Appendixes

Appendix 6.1. We are going to prove that the sampling scheme defined in the section 1 implements the sampling design $P_1(s)$ expressed by the formula (1). Let us multiply the column $\boldsymbol{J}_{k+1}$ by the vector $\bar{\boldsymbol{x}}$ and next let us subtract $\boldsymbol{J}_{k+1}\bar{\boldsymbol{x}}_s$ from the matrix $\boldsymbol{x}_{s_{k+1}}$. This operation let us transform the equation (3) to the following one:

$$q_1(s_{k+1}) = det^2\left[\boldsymbol{J}_{k+1} \quad \boldsymbol{H}_{s_{k+1}}\right] \tag{43}$$

where: $\boldsymbol{H}_{s_{k+1}} = \boldsymbol{x}_{s_{k+1}} - \boldsymbol{J}_{k+1}\bar{\boldsymbol{x}}_s$

Hence the equations (3) and (43) are equivalent.

Let each element of the column vector $\boldsymbol{o}_k$ be equal to zero. On the basis of the well known properties of the matrix determinant, see, e.g. Anderson (1958) we have:

$$\sum_{s_{k+1}\in s} q_1(s_{k+1}) = \sum_{s_{k+1}\subset s} det^2\left[\boldsymbol{J}_{k+1}\boldsymbol{H}_{s_{k+1}}\right] = det\begin{bmatrix}\boldsymbol{J}_n^T \\ \boldsymbol{H}_s^T\end{bmatrix}\left[\boldsymbol{J}_n\boldsymbol{H}_s\right] =$$

$$= det\begin{bmatrix} n & \boldsymbol{o}_k^T \\ \boldsymbol{o}_k & \boldsymbol{H}_s^T\boldsymbol{H}_s\end{bmatrix} = n\,det\left(\boldsymbol{H}_s^T\boldsymbol{H}_s\right)$$

where: $\boldsymbol{H}_s = \boldsymbol{x}_s - \boldsymbol{J}_n\bar{\boldsymbol{x}}_s$

Hence:

$$det\left(\boldsymbol{H}_s^T\boldsymbol{H}_s\right) = \frac{1}{n}\sum_{s_{k+1}\subset s} q_1(s_{k+1}) \tag{44}$$

The expression (44) can be rewritten in the following way:

$$\sum_{s_{k+1}\subset s} q_1(s_{k+1}) = n^{k+1}\, det\,\boldsymbol{V}_s \tag{45}$$

Similarly to the formula (44) it can be derived the following one:

$$\sum_{s_{k+1}\subset U} q_1(s_{k+1}) = N^{k+1}\, det\,\boldsymbol{V} \tag{46}$$

The expression (45) lead to the following one (see Wywial (1996,1997)):

$$\sum_{s\subset \boldsymbol{S}} det\,\boldsymbol{V}_s = \frac{1}{n^{k+1}}\sum_{s\subset \boldsymbol{S}}\ \sum_{s_{k+1}\in s} q_1(s_{k+1})$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

The above sum is divided into sub-sums of the same elements. For a fixed set s_{k+1} of the size $k+1$ the frequency of $q_1(s_{k+1})$ is equal to the number of the combinations s- s_{k+1} of the size n-k-1 chosen from the set U- s_{k+1} of the size N-k-1. Therefore:

$$\sum_{s\subset \boldsymbol{S}} det\boldsymbol{V}_s = \frac{1}{n^{k+1}}\binom{N-k-1}{n-k-1}\sum_{s_{k+1}\in U} q_1(s_{k+1}) = \binom{N-k-1}{n-k-1}\left(\frac{N}{n}\right)^{k+1} det\boldsymbol{V}$$

This lead to the expressions (1) and (2).

The sample s can be decomposed into such two subsets s_{k+1} and $s_{k+1}^{(c)}$ that $s = s_{k+1}\cup s_{k+1}^{(c)}$ and $s_{k+1}\cap s_{k+1}^{(c)} = \varnothing$. According to the sampling scheme and the expression (46) we have.

$$P(s_{k+1}) = \frac{q_1(s_{k+1})}{N^{k+1}\, det\boldsymbol{V}} \tag{47}$$

$$P(s_{k+1}^{(c)}) = \frac{1}{\binom{N-k-1}{n-k-1}}$$

Using the expression (45) the probability of the selection of the sample s can be derived in the following way:

$$P(s) = \sum_{s_{k+1}\in s} P(s_{k+1})P(s_{k+1}^{(c)}) = \frac{1}{\binom{N-k-1}{n-k-1}}\frac{1}{N^{k+1}\, det\boldsymbol{V}}\sum_{s_{k+1}\subset s} q_1(s_{k+1}) =$$

$$= \frac{1}{\binom{N-k-1}{n-k-1}}\left(\frac{n}{N}\right)^{k+1}\frac{det\boldsymbol{V}_s}{det\boldsymbol{V}} = P_1(s)$$

Hence, the considered sampling scheme implements the sampling design $P_1(s)$.

Let $\boldsymbol{x}_{s_{k+1}} = \frac{1}{k+1}\boldsymbol{J}_{k+1}^T\boldsymbol{X}_{s_{k+1}}$ be the vector of mean values in the set s_{k+1}. This let us rewrite the expression (3) in the following way:

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$q_1(s_{k+1}) = det^2\begin{bmatrix} \boldsymbol{J}_{k+1} & \boldsymbol{U}_{s_{k+1}} \end{bmatrix}$$

where: $\boldsymbol{U}_{s_{k+1}} = \begin{bmatrix} \boldsymbol{x}_{s_{k+1}} - \boldsymbol{J}_{k+1}\bar{\boldsymbol{x}}_{s_{k+1}} \end{bmatrix}$

Hence:

$$q_1(s_{k+1}) = det\begin{bmatrix} \boldsymbol{J}_{k+1}^T \\ \boldsymbol{U}_{s_{k+1}}^T \end{bmatrix}\begin{bmatrix} \boldsymbol{J}_{k+1} & \boldsymbol{U}_{s_{k+1}} \end{bmatrix} = det\begin{bmatrix} k+1 & \boldsymbol{o}^T \\ \boldsymbol{o} & \boldsymbol{U}_{s_{k+1}}^T \boldsymbol{U}_{s_{k+1}} \end{bmatrix}$$

And finally:

$$q_1(s_{k+1}) = (k+1)\, det\left(\boldsymbol{U}_{s_{k+1}}^T \boldsymbol{U}_{s_{k+1}} \right)$$

This result let us rewrite the expression (47) in the following way:

$$P(s_{k+1}) = \frac{(k+1)\, det\left(\boldsymbol{U}_{s_{k+1}}^T \boldsymbol{U}_{s_{k+1}} \right)}{N^{k+1}\, det\boldsymbol{V}} \qquad (48)$$

Appendix 6.2. The derivation of the expression (8):

The determinant of the matrix $\boldsymbol{A}_s$ defined by the expression (6) can be transformed in the following way:

$$det\, \boldsymbol{A}_s = det\left(\frac{1}{n}\begin{bmatrix} n\bar{Y}_s & n\bar{\boldsymbol{X}}_s \\ \boldsymbol{X}_s^T \boldsymbol{Y}_s - n\bar{\boldsymbol{X}}_s^T \bar{Y}_S & \boldsymbol{X}_s^T \boldsymbol{X}_s - n\bar{\boldsymbol{X}}_s^T \bar{\boldsymbol{X}}_s \end{bmatrix} \right)$$

Let the product of the vector $\begin{bmatrix} n\bar{Y}_S & n\bar{\boldsymbol{X}}_S \end{bmatrix}$ and the i-th element of the vector $\bar{\boldsymbol{X}}_S^T$ be added to the i-th row of the matrix $\begin{bmatrix} \boldsymbol{X}_S^T \boldsymbol{Y}_S - n\bar{\boldsymbol{X}}_S^T \bar{Y} & \boldsymbol{X}_S^T \boldsymbol{X}_S - n\bar{\boldsymbol{X}}_S^T \bar{\boldsymbol{X}}_S \end{bmatrix}$ for each *i*=1,...,*k*. Hence:

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$det\, A_s = n^{-k-1}\, det\begin{bmatrix} n\bar{Y}_s & n\bar{\boldsymbol{X}}_s \\ \boldsymbol{X}_s^T\boldsymbol{Y}_s & \boldsymbol{X}_s^T\boldsymbol{X}_s \end{bmatrix}$$

This equation leads to the expression (8).

Appendix 6.3. The expected value of the strategy I:

On the basis of the expressions (1), (7), (9) and the well known properties of matrix algebra we derive the expected value of the strategy $\left(\bar{y}_{Rs}, P_1(s)\right)$:

$$E\left(y_{Rs}, P_1(s)\right) = \sum_{s\in\boldsymbol{S}} y_{Rs}P_1(s) = y + \frac{c_1}{det\boldsymbol{V}}\sum_{s\in\boldsymbol{S}} det\,\boldsymbol{A}_s = y + \frac{c_1}{n^{k+1}\,det\boldsymbol{V}}\sum_{s\in\boldsymbol{S}} det\begin{bmatrix}\boldsymbol{J}_n^T \\ \boldsymbol{X}_s^T\end{bmatrix}\left[\boldsymbol{Y}_s \boldsymbol{X}_s\right]$$

Similarly, like in the case of the derivation of the expression (1) (see the appendix 1), we have:

$$E\left(y_{Rs}, P_1(s)\right) = y + \frac{c_1}{n^{k+1}\,det\boldsymbol{V}}\sum_{s\in\boldsymbol{S}}\;\sum_{s_k\in s} det\begin{bmatrix}\boldsymbol{J}_k^T \\ \boldsymbol{X}_{s_k}^T\end{bmatrix}\left[\boldsymbol{Y}_{s_k}\, \boldsymbol{X}_{s_k}\right]$$

Under the fixed set s_k of the size k the frequency of the element $det\begin{bmatrix}\boldsymbol{J}_k^T \\ \boldsymbol{X}_{s_k}^T\end{bmatrix}\left[\boldsymbol{Y}_{s_k} \quad \boldsymbol{X}_{s_k}\right]$ is equal to the number of the combinations s-s_k of the size (n-k) chosen from the set U-s_k of the size N-k. Then:

$$E\left(y_{Rs}, P_1(s)\right) = y + \frac{c_1}{n^{k+1}\,det\boldsymbol{V}}\binom{N-k}{n-k}\sum_{s_k\in U} det\begin{bmatrix}\boldsymbol{J}_k^T \\ \boldsymbol{X}_{s_k}^T\end{bmatrix}\left[\boldsymbol{Y}_{s_k}\, \boldsymbol{X}_{s_k}\right] =$$

$$= y + \frac{c_1}{n^{k+1}\,det\boldsymbol{V}}\binom{N-k}{n-k} det\begin{bmatrix}\boldsymbol{J}_N^T \\ \boldsymbol{X}_U\end{bmatrix}\left[\boldsymbol{Y}_U \boldsymbol{X}_U\right] =$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$= y + \frac{c_1}{n^{k+1} \det \boldsymbol{V}} \binom{N-k}{n-k} \det \begin{bmatrix} \boldsymbol{J}_N^T \boldsymbol{Y}_U & \boldsymbol{J}_N^T \boldsymbol{X}_U \\ \boldsymbol{X}_U^T \boldsymbol{Y}_U & \boldsymbol{X}_U^T \boldsymbol{X}_U \end{bmatrix} =$$

$$= y + \frac{c_1}{n^{k+1} \det \boldsymbol{V}} \binom{N-k}{n-k} \det \begin{bmatrix} o & \boldsymbol{o}_N^T \\ \boldsymbol{X}_U^T \boldsymbol{Y}_U & \boldsymbol{X}_U^T \boldsymbol{X}_U \end{bmatrix} = y$$

Appendix 6.4. Derivation of the variance of the strategy I.

$$D^2(y_{Rs}, P_1(s)) = \sum_{s \in \boldsymbol{S}} (\bar{y}_{Rs} - y)^2 P_1(s) = \sum_{s \in \boldsymbol{S}} (\bar{Y}_s - \bar{\boldsymbol{X}}_s \boldsymbol{B}_s)^2 P_1(s) =$$

$$= \frac{c_1}{\det \boldsymbol{V}} \sum_{s \in \boldsymbol{S}} (\bar{Y}_s - \bar{\boldsymbol{X}}_s \boldsymbol{B}_s)^2 \boldsymbol{V}_s =$$

$$= \frac{c_{*1}}{\det \boldsymbol{V}} E\left[(\bar{Y}_s - \bar{\boldsymbol{X}}_s \boldsymbol{B}_s)^2 \boldsymbol{V}_s, P_o(s)\right]$$

(49)

where:
$$c_{*1} = \frac{\prod_{h=0}^{k} (N-h)}{\prod_{h=0}^{k} (n-h)} \frac{n^{k+1}}{N^{k+1}}$$

or
$$c_{*1} = \frac{\prod_{h=1}^{k} \left(1 - \frac{h}{N}\right)}{\prod_{h=1}^{k} \left(1 - \frac{h}{n}\right)}$$

Let us note that:

$$\lim_{\substack{N \to \infty \\ n \to \infty}} c_{*1} = 1 \tag{50}$$

Under the assumption: $N \to \infty$, $n \to \infty$ and $N\text{-}n \to \infty$ the simple sample *s* drawn without replacement can be treated as a simple one drawn with replacement. Hence on the basis of the expressions (8), (49) and (50) we have:

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$D^2(y_{Rs},P_1(s)) \approx \frac{1}{\det \mathbf{V}} E\left\{\left[\overline{Y}_S^2 - 2\overline{Y}_S \mathbf{X}_S \mathbf{B}_S + \mathbf{B}_S^T \mathbf{X}_S^T \mathbf{X}_S \mathbf{B}_S\right]\det \mathbf{V}_S\right\} =$$

$$= \frac{1}{\det \mathbf{V}} E\left[\left(\overline{Y}_S^2 - 2\overline{Y}_S \mathbf{X}_S \mathbf{V}_S^{-1}\mathbf{v}_S + \mathbf{v}_S^T \mathbf{V}_S^{-1} \mathbf{X}_S^T \mathbf{X}_S \mathbf{V}_S^{-1}\mathbf{v}_S\right)\det \mathbf{V}_S\right]$$

$$= \frac{1}{\det \mathbf{V}}\left[E\left(\overline{Y}_S^2 \det \mathbf{V}_S\right) - 2E\left(\overline{Y}_S \mathbf{X}_S \mathbf{V}_S^{-1}\mathbf{v}_S \det \mathbf{V}_S\right) + \right.$$
$$\left. + E\left(\mathbf{v}_S^T \mathbf{V}_S^{-1} \mathbf{X}_S^T \mathbf{X}_S \mathbf{V}_S^{-1}\mathbf{v}_S \det \mathbf{V}_S\right)\right]$$

This result and the well known properties of the expected value of the rational functions of simple sample moments lead to the expression (11).

Appendix 6.5. Unbiasednees of the estimator of the variance

The expressions (1), (2) and (13) lead to the following derivation:

$$E\left(\hat{D}_1^2, P_1(S)\right) =$$

$$E\left(y_{RS}^2, P_1(S)\right) - \frac{N^{k-1}}{n^{k+1}} \frac{\prod_{h=1}^{k}(n-h)}{\prod_{h=1}^{k}(N-h)} \sum_{s\in \mathbf{S}} \frac{\det \mathbf{V}}{\det \mathbf{V}_S}\left[\sum_{i\in s} y_i^2 + \frac{N-1}{n-1}\sum_{i\neq j\in s} y_i y_j\right] P_1(s) =$$

$$= E\left(y_{RS}^2, P_1(S)\right) - \frac{1}{N^2} \frac{\prod_{h=1}^{k}(n-h)}{\prod_{h=1}^{k}(N-h)} \frac{(N-n)!(n-k-1)!}{(N-k-1)!} \sum_{s\in \mathbf{S}}\left[\sum_{i\in s} y_i^2 + \frac{N-1}{n-1}\sum_{i\neq j\in s} y_i y_j\right] =$$

$$= E\left(y_{Rs}^2, P_1(s)\right) - \frac{1}{nN}\frac{(N-n)!n!}{N!}\sum_{s\in \mathbf{S}}\left[\sum_{i\in s} y_i^2 + \frac{N-1}{n-1}\sum_{i\neq j\in s} y_i y_j\right] =$$

$$= E\left(y_{Rs}^2, P_1(s)\right) - \frac{1}{nN}\frac{1}{\binom{N}{n}}\sum_{s\in \mathbf{S}}\left[\sum_{i\in s} y_i^2 + \frac{N-1}{n-1}\sum_{i\neq j\in s} y_i y_j\right] =$$

Statistics in Transition
(Journal of the Polish Statistical Association)
1999, vol. 4 No. 2, p. 259-281

$$=E\left(y_{Rs}^2, P_1(s)\right) - \frac{1}{N^2}\sum_{i\in U} y_i^2 - \frac{1}{nN}\frac{1}{\binom{N}{n}}\frac{N-1}{n-1}\binom{N-2}{n-2}\sum_{i\neq j\in U} y_i y_j =$$

$$=E\left(y_{Rs}^2, P_1(s)\right) - \frac{1}{N^2}\left[\sum_{i\in U} y_i^2 + \sum_{i\neq j\in U} y_i y_j\right] =$$

$$=E\left(y_{Rs}^2, P_1(s)\right) - y^2 = D^2\left(y_{Rs}, P_1(s)\right)$$

REFERENCES

1. Anderson T.W. (1958): An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, New York.
2. Borsuk K. (1969): Multidimensional Analytic Geometry. PWN, Warsaw.
3. Sarndal C. E., B. Swensson, J. Wretman (1992): „Model Assisted Survey Sampling”. Springer Verlag, New York-Berlin-Heidelberg-London-Paris-Tokyo-Hong Kong- Barcelona-Budapest.
4. Singh P., Srivastava A.K. (1980): Sampling schemes providing unbiased regression estimators. *Biometrika*, vol. 67, 1, pp. 205-9.
5. Wywial J. (1996): Unbiased estimators of generalised variance from simple random sample drawn without replacement from finite population. *Journal of the Indian Statistical Association,* vol. 34, 1996r. pp. 125-129
6. . Wywial J. (1997): Sampling Design proportional to the Sample Generalized Variance of Auxiliary Variables. In: *Proceedings of 16th International Conference on Multivariate Statistical Analysis- MSA'97*. Edited by Cz. Domanski and D. Parys. November 27-29 1997. Department of Statistical Methods, Institute of Econometrics and Statistics, University of Łódź, Polish Statistical Association. November 27-29. 1997r., s. 129-143.