ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 175, 2004

*Tomasz Żądlo**

ON MEAN SQUARE ERROR OF SYNTHETIC REGRESSION ESTIMATOR

Abstract. The problem of estimation of the total value in a small domain is considered. Because of the small number (or the lack) of elements of the considered subpopulation in the sample, information on all drawn elements is used. The synthetic regression estimator is presented. The equations of the bias and the mean square error for any sample design are derived. The problem of the assumptions on the population and the domain's structure due to the bias and MSE reduction is considered. The importance of the bias influence on the accuracy of the estimation is presented. The possibility of the increase of the MSE and bias due to the increase of the sample size is shown. The approximate equations of the bias and mean square error for the simple random sampling without replacement are derived. The accuracy of the synthetic regression estimator (based on approximate equations and the simulation study) and the Horwitz-Thompson direct estimator is compared. The comparison is based on agricultural data from Dąbrowa Tarnowska region. The entire population consist of 8624 farms and it includes the domain of interest – Bolesław commune – with 588 farms.

Key words: small area statistics, indirect estimators, synthetic estimators.

1. INTRODUCTION

Synthetic estimation is considered both in Polish (e.g. Bracha 1994), Bracha 1996, Getka-Wilczyńska 2000 and in foreign literature (e.g. Särndal, Swensson, Wretman 1992). In this paper equations of the mean square error and the bias of the synthetic regression estimator are presented taking differences between the population and domain's structure into consideration.

The term "synthetic" means, that the estimator uses information both on surveyed small domain and period of time and also on other domains and/or periods of time. The estimator considered in this paper is domain indirect estimator. The idea of synthetic estimation is based on the assumption, that some relationship between the variable of interest and the auxiliary

* MA, Department of Statistics, University of Economics, Katowice.

variable observed in the entire population is the same in small area. Because most often the assumption is not met, it must be stressed, that synthetic estimators are biased.

2. MEAN SQUARE ERROR (MSE) FOR ANY SAMPLING DESIGN

Considerations are conducted for any sampling design. It is assumed, that the sample $S$is drawn from the entire population by sample design $P(S)$ with first order inclusion probabilities π_i, where $i = 1, ..., N$. For any sample S with size n drawn from the population ς with the size N, $S_d = s \cap \Omega_d$, where the d-th domain is denoted by ς_d. The size of S_d equals n_d (random variable) and the size of ς_d equals N_d. The set of elements of the population, which belong to d-th domain ς_d, could be written as $\Omega_d = S_d \cup \overline{S}_d$, where $\overline{S}_d$ denotes elements of the d-th domain, which were not drawn to the sample. Equations of the MSE and the bias for simple random sampling without replacement will also be derived in this paper.

Synthetic regression estimator of total value in small domain, which is most often more precise than synthetic ratio estimator, is as follows:

$$\hat{Y}_d^{SYN-regr} = N_d[\hat{\overline{Y}} + \hat{\beta}(\overline{x}_d - \hat{\overline{X}})] = \frac{N_d}{N}\hat{Y}^{regr} + N_d\hat{\beta}(\overline{x}_d - \overline{x}) \tag{1}$$

where

N_d – small domain size,

N – population size,

$\hat{\overline{Y}} = \frac{1}{N}\sum_{i \in S}\frac{y_i}{\pi_i}$ – Horwitz–Thompson (HT) estimator of the mean value of the variable of interest in the population,

$\hat{\overline{X}} = \frac{1}{N}\sum_{i \in S}\frac{x_i}{\pi_i}$ – HT estimator of the mean value of the auxiliary variable in the population,

$\overline{x}_d$ – the mean value of the auxiliary variable in d-th small domain,

$\overline{x}$ – the mean value of the auxiliary variable in the population,

$\hat{\beta} = \frac{\sum_{i \in S}(x_i - \hat{\overline{X}})(y_i - \hat{\overline{Y}})\frac{1}{\pi_i}}{\sum_{i \in S}(x_i - \hat{\overline{X}})^2\frac{1}{\pi_i}}$ – the estimator of regression coefficient in the population,

$\hat{Y}^{regr} = N(\hat{\bar{Y}} + \hat{\beta}(\bar{x} - \hat{\bar{X}})$ – regression estimator of the total value in the population.

Let the bias of the regression estimator of the total value in the population be denoted by $B^{regr} = \mathrm{E}(\hat{Y}^{regr} - Y)$. The bias B^{regr} for simple random sample without replacement is as follows (e.g. W y w i a ł 1992, s. 137):

$$B^{regr} = \mathrm{E}(\hat{Y}^{regr} - Y) = \frac{N}{n}\frac{N-n}{N-2}\sqrt{c_2(y)[B_2(x)-1]}\{k_{21}(x,y) - k_3(x)r(x,y)\} + \\ + O(N \cdot n^{-2}) \tag{2}$$

where

$$k_3(x) = \frac{c_3(x)}{\sqrt{c_2(x)[c_4(x) - c_2^2(x)]}}, \quad k_{21}(x,y) = \frac{c_{21}(x,y)}{\sqrt{c_2(y)[c_4(x) - c_2^2(x)]}},$$

$$B_2(x) = c_4(x) \cdot c_2^{-2}(x),$$

$$r(x,y) = \frac{c_{11}(x,y)}{\sqrt{c_2(x)c_2(y)}},$$

$$c_r(x) = \frac{1}{N}\sum_{i\in\Omega}(x_i - \bar{x})^r, \quad c_r(y) = \frac{1}{N}\sum_{i\in\Omega}(y_i - \bar{y})^r, \quad c_{rk}(x,y) = \frac{1}{N}\sum_{i\in\Omega}(x_i - \bar{x})^r(y_i - \bar{y})^k.$$

Let us derive the equation of synthetic regression estimator of the total value in small domain for any sample design:

$$\mathrm{E}(\hat{Y}_d^{SYN-regr}) - Y_d = \frac{N_d}{N}\mathrm{E}(\hat{Y}^{regr}) + \mathrm{E}(\hat{\beta})N_d(\bar{x}_d - \bar{x}) - Y_d - \frac{N_d}{N}Y + \frac{N_d}{N}Y =$$

$$= \frac{N_d}{N}B^{regr} + \mathrm{E}(\hat{\beta})N_d(\bar{x}_d - \bar{x}) + \left(\frac{N_d}{N}Y - Y_d\right) = \tag{3}$$

$$= \frac{N_d}{N}B^{regr} + \mathrm{E}(\hat{\beta})\left(\frac{N_d}{N}X - X_d\right) - \left(\frac{N_d}{N}Y - Y_d\right)$$

If the mean value of the variable of interest in the population equals the mean value of the variable of interest in d-th small domain and if the mean value of auxiliary variable in the population equals the mean value of auxiliary variable in d-th small domain, then $\mathrm{E}(\hat{Y}_d^{SYN-regr}) - Y_d = \frac{N_d}{N}B^{regr}$. In this case the bias of synthetic regression estimator of total value in small domain is the function of the bias of regression estimator of the total value in the population and it is the less, the less is the value of

ratio of domain size and population size. It is worth stressing, that for simple random sample it is of order $O(N \cdot n^{-1})$, so it decreases due to the increase of sample size.

The equation (3) could also be written as follows:

$$\mathrm{E}(\hat{Y}_d^{SYN-regr}) - Y_d = \left[\frac{N_d}{N} B^{regr} - (\mathrm{E}(\hat{\beta}) - \beta)\left(\frac{N_d}{N} X - X_d\right)\right] + \\ + \left[\left(\frac{N_d}{N} Y - Y_d\right) - \beta\left(\frac{\mathrm{N}_d}{\mathrm{N}} \mathrm{X} - \mathrm{X}_d\right)\right].$$

Let us notice, that the value of $\left[\left(\frac{N_d}{N} Y - Y_d\right) - \beta\left(\frac{N_d}{N} X - X_d\right)\right]$ does not depend on sample size. Let us consider the expression $\left[\frac{N_d}{N} B^{regr} - (\mathrm{E}(\hat{\beta}) - \beta) - \beta\left(\frac{N_d}{N} X - X_d\right)\right]$. The absolute value of the bias of regression estimator of total value in the population denoted by the expression B^{regr} reduces due to the increase of sample size. The same property has the bias of estimator of regression coefficient denoted by $(\mathrm{E}(\hat{\beta}) - \beta)$. Because the limit of the sum of sequences equals the sum of limits of these sequences, the value of $\left[\frac{N_d}{N} B^{regr} - (\mathrm{E}(\hat{\beta}) - \beta) - \beta\left(\frac{N_d}{N} X - X_d\right)\right]$ decreases due to the increase of sample size. The absolute value of this expression can decrease monotonically due to the increase of sample size. It should be mentioned, that if values of both considered elements, into which the bias of synthetic regression estimator was decomposed, have opposite signs, then the absolute value of the bias of synthetic regression estimator can grow due to the increase of sample size. It is also possible, that the absolute value of the bias of synthetic regression estimator does not change monotonically due to the increase of sample size.

Analysing the equation (3), it is worth stressing, that using synthetic regression estimators it should be assumed, that:

$$\frac{N_d}{N} = \frac{Y_d}{Y} \quad \text{and} \quad \frac{N_d}{N} = \frac{X_d}{X} \tag{4}$$

what can be written as:

$$\frac{N_d}{N} = \frac{X_d}{X} = \frac{Y_d}{Y}$$

Comparing aforementioned assumption (4) with assumption using for synthetic ratio estimator given by expression $\frac{X_d}{X} = \frac{Y_d}{Y}$, it should be mentioned, that assumption for synthetic regression estimators is more restrictive.

Let us derive equation of the mean square error of the estimator $\hat{Y}^{SYN-regr}$ taking the assumption (4) into consideration.

$$\mathrm{MSE}(\hat{Y}^{SYN-regr}) = \mathrm{E}\left(\frac{N_d}{N}\hat{Y}^{regr} + N_d\hat{\beta}(\bar{x}_d - \bar{x}) - Y_d\right)^2 =$$

$$= \mathrm{E}\left(\frac{N_d}{N}\hat{Y}^{regr} + \frac{N_d}{N}Y - \frac{N_d}{N}Y + N_d\hat{\beta}(\bar{x}_d - \bar{x}) - Y_d\right)^2 =$$

$$= \mathrm{E}\left(\frac{N_d}{N}(\hat{Y}^{regr} - Y) + \frac{N_d}{N}Y - Y_d - (\hat{\beta} - \mathrm{E}(\hat{\beta}) + \mathrm{E}(\hat{\beta}))\left(\frac{N_d}{N}X - X_d\right)\right)^2 =$$

$$= \left(\frac{N_d}{N}\right)^2 \mathrm{MSE}(\hat{Y}^{regr}) + \left(\frac{N_d}{N}Y - Y_d\right)^2 + (\mathrm{D}^2(\hat{\beta}) + \mathrm{E}^2(\hat{\beta}))\left(\frac{N_d}{N}X - X_d\right)^2 +$$

$$+ 2\frac{N_d}{N}\left(\frac{N_d}{N}Y - Y_d\right)B^{regr} - 2\frac{N_d}{N}\left(\frac{N_d}{N}X - X_d\right) \cdot$$

$$\cdot \mathrm{E}\{((\hat{Y}^{regr} - \mathrm{E}(\hat{Y}^{regr})) + (\mathrm{E}(\hat{Y}^{regr}) - Y))(\hat{\beta} - \mathrm{E}(\beta)) + \mathrm{E}(\hat{\beta}))\} +$$

$$- 2\mathrm{E}(\hat{\beta})\left(\frac{N_d}{N}Y - Y_d\right)\left(\frac{N_d}{N}X - X_d\right) =$$

$$= \left(\frac{N_d}{N}\right)^2 \mathrm{MSE}(\hat{Y}^{regr}) + \left(\frac{N_d}{N}Y - Y_d\right)^2 + (\mathrm{D}^2(\hat{\beta}) + \mathrm{E}^2(\hat{\beta}))\left(\frac{N_d}{N}X - X_d\right)^2 +$$

$$+ 2\frac{N_d}{N}\left(\frac{N_d}{N}Y - Y_d\right)B^{regr} - 2\frac{N_d}{N}\left(\frac{N_d}{N}X - X_d\right)(\mathrm{cov}(\hat{Y}^{regr}, \hat{\beta}) + \mathrm{E}(\hat{\beta})B^{regr}) +$$

$$- 2\mathrm{E}(\hat{\beta})\left(\frac{N_d}{N}Y - Y_d\right)\left(\frac{N_d}{N}X - X_d\right). \qquad (5)$$

If the assumption (4) is met, then $\mathrm{MSE}(\hat{Y}_d^{SYN-regr}) = \left(\frac{N_d}{N}\right)\mathrm{MSE}(\hat{Y}^{regr})$, and it should be pointed out, that $\left(\frac{N_d}{N}\right)^2 < \frac{N_d}{N} < 1$.

Equation (5) could also be written as follows:

$$\mathrm{MSE}(\hat{Y}_d^{SYN-regr}) = \left(\frac{N_d}{N}\right)^2 \mathrm{D}^2(\hat{Y}^{regr}) + \mathrm{D}^2(\hat{\beta})\left(\frac{N_d}{N}X - X_d\right)^2 +$$

$$-2\frac{N_d}{N}\left(\frac{N_d}{N}X - X_d\right)\mathrm{Cov}(\hat{Y}^{regr}, \hat{\beta}) +$$

$$+\left[\frac{N_d}{N}B^{regr} - \mathrm{E}(\hat{\beta})\left(\frac{N_d}{N}X - X_d\right) + \left(\frac{N_d}{N}Y - Y_d\right)\right]^2 \qquad (6)$$

First three elements of above-mentioned sum (6) form the variance of $\hat{Y}_d^{SYN-regr}$ estimator. It can be written by alternative expression:

$$\mathrm{D}^2(Y_d^{SYN-regr}) = \mathrm{E}\left(\frac{N_d}{N}(\hat{Y}^{regr} - \mathrm{E}(\hat{Y}^{regr})) - (\hat{\beta} - \mathrm{E}(\hat{\beta}))\left(\frac{N_d}{N}X - X_d\right)\right)^2.$$

The fourth element of the sum (6) is squared bias of $\hat{Y}_d^{SYN-regr}$ estimator, which is given by the equation (3).

The value of $\mathrm{D}^2(\hat{Y}_d^{SYN-regr})$ decreases due to the increase of sample size. The value of squared bias of synthetic regression estimator given by expression $\left[\frac{N_d}{N}B^{regr} - \mathrm{E}(\beta)\left(\frac{N_d}{N}X - X_d\right) + \left(\frac{N_d}{N}Y - Y_d\right)\right]^2$ can increase due to the increase of sample size. If the increase of value of squared bias due to the increase of sample size is higher than the decrease of value of $\mathrm{D}^2(\hat{Y}_d^{SYN-regr})$ due to the increase of sample size, then value of $\mathrm{MSE}(\hat{Y}^{SYN-regr})$ will increase due to the increase of sample size. It is also possible, that the value of the MSE of synthetic regression estimator does not change monotonically due to the increase of sample size.

3. MEAN SQUARE ERROR FOR SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT

The accuracy of synthetic ratio estimator will be compared with the accuracy of Horwitz–Thompson estimator of total value in the domain, which for simple random sampling without replacement is given by equation:

$$\hat{Y}_d = \frac{N}{n}\sum_{i=1}^{n_d} y_i.$$

Its variance is as follows (S ä r n d a l, S w e n s s o n, W r e t m a n 1992):

$$D^2(\hat{Y}_d) = N^2 \frac{N-n}{Nn} \frac{N_d}{N} \left(S^2_{*y,d} + \frac{N-N_d}{N} \bar{y}_d^2 \right),$$

where

$$S^2_{*y,d} = \frac{1}{N_d - 1} \sum_{i=1}^{N_d} (y_i - \bar{y}_d)^2.$$

Synthetic regression estimator given by the equation (1) for any sample design, for simple random sampling without replacement is as follows:

$$\hat{Y}_d^{SYN-regr} = N_d \left[\bar{y}_s + \frac{\hat{S}_{*xy}}{\hat{S}^2_{*x}} (\bar{x}_d - \bar{x}_s) \right],$$

where

$$\hat{S}^2_{*x} = \frac{1}{n-1} \sum_{i \in S} (x_i - \bar{x}_S)^2, \quad \hat{S}^2_{*xy} = \frac{1}{n-1} \sum_{i \in S} (x_i - \bar{x}_S)(y_i - \bar{y}_S),$$

$$\bar{x}_S = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{y}_S = \frac{1}{n} \sum_{i=1}^{n} y_i.$$

Using following expressions: $S^2_{*x} = \frac{1}{N-1} \sum_{i \in \Omega} (x_i - \bar{x})^2$, $S^2_{*y} = \frac{1}{N-1} \sum_{i \in \Omega} (y_i - \bar{y})^2$, $S^2_{*xy} = \frac{1}{N-1} \sum_{i \in \Omega} (x_i - \bar{x})(y_i - \bar{y})$, $\hat{S}^2_{*x} = \frac{1}{n-1} \sum_{i \in S} (x_i - \bar{x}_S)^2$, $\hat{S}^2_{*y} = \frac{1}{n-1} \sum_{i \in S} (y_i - \bar{y}_S)^2$, $S_{*xy} = \frac{1}{n-1} \sum_{i \in S} (x_i - \bar{x}_S)(y_i - \bar{y}_S)$ let us specify above-mentioned elements of equations (5) of $MSE(\hat{Y}_d^{SYN-regr})$ for simple random sampling without replacement.

First (see e.g. B r a c h a 1996),

$$MSE(\hat{Y}^{regr}) = N^2 \frac{N-n}{Nn} S^2_{*y} (1 - r^2(x, y)) + O(N^2 n^{-2}).$$

Second,

$$D^2(\hat{\beta}) = MSE(\hat{\beta}) - (E(\hat{\beta}) - \beta)^2,$$

where

$$\mathrm{MSE}(\hat{\beta}) = \mathrm{MSE}\left(\frac{\hat{S}_{\bullet xy}}{\hat{S}^2_{\bullet x}}\right) = \frac{1}{(S^2_{\bullet x})}\left[\mathrm{D}^2(\hat{S}^{\bullet}_{xy}) - 2\frac{S_{\bullet xy}}{S^2_{\bullet x}}\mathrm{cov}(\hat{S}_{\bullet xy}, \hat{S}^2_{\bullet x}) + \right.$$

$$\left. + \left(\frac{S_{\bullet xy}}{S^2_{\bullet x}}\right)^2 \mathrm{D}^2(\hat{S}^2_{\bullet x})\right] + O(n^{-2}),$$

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}\left(\frac{\hat{S}_{\bullet xy}}{\hat{S}^2_{\bullet x}}\right) = \frac{S_{\bullet xy}}{S^2_{\bullet x}} + \frac{1}{(S^2_{\bullet x})^2}\left[\frac{S_{\bullet xy}}{S^2_{\bullet x}}\mathrm{D}^2(\hat{S}^2_{\bullet x}) - \mathrm{cov}(\hat{S}_{\bullet xy}, \hat{S}^2_{\bullet x})\right] + O(n^{-2}) \tag{7}$$

where (see e.g. W y w i a ł 2000)

$$\mathrm{D}^2(\hat{S}^2_{\bullet xy}) = \frac{1}{n}(\mathrm{cov}_{\bullet 22}(x, y) - \mathrm{cov}^2_{\bullet 11}(x, y)) + O(N^{-1}) + O(n^{-2}),$$

$$\mathrm{D}^2(\hat{S}^2_{\bullet x}) = \frac{1}{n}(\mathrm{cov}_{\bullet 4}(x) - \mathrm{cov}^2_{\bullet 2}(x)) + O(N^{-1}) + O(n^{-2}),$$

and (the equation is derived in the Part A.3 of the Appendix)

$$\mathrm{cov}(\hat{S}^2_{\bullet xy}, \hat{S}^2_{\bullet x}) = \frac{1}{n}(\mathrm{cov}_{\bullet 31}(x, y) - \mathrm{cov}_{\bullet 2}(x)c_{\bullet 11}(x, y) + O(N^{-1}) + O(n^{-2}).$$

Third (the equation is derived in the Part A.1 of the Appendix),

$$\mathrm{cov}(\hat{Y}^{regr}, \hat{\beta}) = \frac{N}{n}\frac{N-n}{N-2}\left(\frac{\mathrm{cov}_{\bullet 12}(x, y)}{\mathrm{cov}_{\bullet 2}(x)} - 2\frac{\mathrm{cov}_{\bullet 11}(x, y)\mathrm{cov}_{\bullet 21}(x, y)}{\mathrm{cov}^2_{\bullet 2}(x)} + \frac{\mathrm{cov}^2_{\bullet 11}(x, y)\mathrm{cov}_{\bullet 3}(x)}{\mathrm{cov}^3_{\bullet 2}(x)}\right) +$$

$$- \frac{\mathrm{cov}_{\bullet 11}(x, y)}{\mathrm{cov}_{\bullet 2}(x)}\left(\frac{N}{n}\frac{N-n}{N-2}\sqrt{\mathrm{cov}_2(y)[B_2(x) - 1]}\{k_{21}(x, y) - k_3(x)r(x, y)\}\right) + O(N \cdot n^{-2}) + O(n^{-1}).$$

Fourth, B^{regr} for simple random sampling without replacement is given by the equation (2).

The bias of synthetic regression estimator given for any sample design by the equation (3) for simple random sampling without replacement using the equation (2) is as follows:

$$\mathrm{E}(\hat{Y}^{SYN-regr}) - Y_d = \frac{N_d}{N}\left(\frac{N}{n}\frac{N-n}{N-2}\sqrt{\mathrm{cov}_2(y)[B_2(x) - 1]}\{k_{21}(x, y) - k_3(x)r(x, y)\}\right) +$$

$$-\left(\frac{\mathrm{cov}_{\bullet 11}(x,y)}{\mathrm{cov}_{\bullet 2}(x)}-\frac{1}{n}\frac{1}{\mathrm{cov}^2_{\bullet 2}(x)}\left((\mathrm{cov}_{\bullet 31}(x,y)-\mathrm{cov}_{\bullet 2}(x)\mathrm{cov}_{\bullet 11}(x,y))+\right.\right.$$

$$\left.\left.+\frac{\mathrm{cov}_{\bullet 11}(x,y)}{\mathrm{cov}_{\bullet 2}(x)}(\mathrm{cov}_{\bullet 4}(x)-\mathrm{cov}^2_{\bullet 2}(x))\right)\right)\cdot\left(\frac{N_d}{N}X-X_d\right)+\left(\frac{N_d}{N}Y-Y_d\right)+O(N^{-1})+O(Nn^{-2}).$$

4. SIMULATION STUDY

The comparison is based on agricultural data on 8624 farms from Dąbrowa Tarnowska region. Approximate equations of the bias and the MSE derived in the Section 3will be used. The comparison will also be supported by simulation study in which 500 samples will be drawn at random. Data includes information on sowing area (in 100 square meters) – the variable of interest, and arable area (in 100 square meters), which is auxiliary variable. The value of the correlation coefficient between these variables equals 0.974. Bolesław commune is treated as the domain of interest. It includes 588 farms. Two sample sizes are considered – 86 and 259 elements, what equals 1 and 3% of population size. Let us analyse the assumptions given by equation (1). The value of $\frac{N_d}{N}=0.0682$, $\frac{X_d}{X}=0.0922$ and the value $\frac{Y_d}{Y}=0.0968$. The difference between first ratio and other ratios is significant.

Let us analyse results presented in the Tab. 1. At the beginning high relative efficiency of synthetic estimator (the ratio of the MSE of synthetic estimator and the variance of Horwitz–Thomson estimator) must be stressed, but it also should be pointed out, that it decreases due to the increase of sample size. The reason is that the MSE of synthetic ratio estimator includes the element, which does not depend on sample size. The value of the MSE of synthetic regression estimator equals ca. 1% of the variance of Horwitz-Thomson estimator for sample size of 86 elements and 3% for sample size of 259 elements. Because of the possibility of low efficiency of estimation, synthetic regression estimation should be used only for small domain estimation purposes. The value of the bias has strong influence on the accuracy of estimation. It does not exceed 5% of real mean value and it decreases due to the increase of sample size for analysed sample sizes. Simulation study shows that for sample size of 86 elements variance equals ca. 12% of MSE, for sample size of 259 elements – only ca. 3%.

Table 1

Results

	Horwitz–Thompson estimator		Synthetic ratio estimator simulation		Synthetic ratio estimator – Taylor	
	sample size		sample size		sample size	
	86	259	86	259	86	259
Square root of MSE	127 462.72	72 700.505	13 545.723	12 954.143	12 194.238	11 308.607
Relative square root of MSE (in %)	47.9434	27.3453	5.0950	4.8725	4.5867	4.2536
Standard error	127 462.72	72 700.505	4 709.3590	2 739.6764	3 009.1891	1 808.8175
Relative standard error (in %)	47.9434	27.3453	1.7714	1.0305	1.1319	0.6804
Bias	0	0	–12 700.73	–12 661.12	–11 817.12	–11 163.01
Relative bias (in %)	0	0	–4.7772	–4.7623	–4.4448	–4.1988
Ratio of variance and MSE	1	1	0.1209	0.0447	0.0609	0.0256
Ratio of MSE and variance of HT estimator	1	1	0.0113	0.0317	0.0091	0.0242

5. CONCLUSION

Summing up, in the paper the mean square error of synthetic regression estimator of total value in small domain is considered. The error was presented in convenient way as the function of inter alia the bias of ratio estimator of population mean value and the bias resulted from the fact, that the assumption (4) is not met. The accuracy of the estimator was compared in simulation studies with the accuracy of direct Horwitz–Thompson estimator for simple random sampling. They confirmed known fact, that synthetic estimator should be recommended for small samples.

APPENDIX

A.1. DERIVATIONS

In this part of the appendix an equation of $\operatorname{cov}(\hat{Y}^{regr}, \hat{\beta})$ for simple random sampling without replacement will be derived. Following notations will be used:

$$\operatorname{cov}_r(x) = \frac{1}{N}\sum_{i\in\Omega}(x_i-\bar{x})^r,\ \operatorname{cov}_{\bullet r}(x) = \frac{1}{N-1}\sum_{i\in\Omega}(x_u-\bar{x})^r,\ \operatorname{cov}_r(y) = \frac{1}{N}\sum_{i\in\Omega}(y_i-\bar{y})^r,$$

$$\operatorname{cov}_{\bullet r}(y) = \frac{1}{N-1}\sum_{i\in\Omega}(y_i-\bar{y})^r,\ \operatorname{cov}_{rk}(x,y) = \frac{1}{N}\sum_{i\in\Omega}(x_u-\bar{x})^r(y_i-\bar{y})^k,$$

$$\operatorname{cov}_{\bullet rk}(x,y) = \frac{1}{N-1}\sum_{i\in\Omega}(x_i-\bar{x})^r(y_i-\bar{y})^k,$$

$$S_x^2 = \operatorname{cov}_2(x),\ S_y^2 = \operatorname{cov}_2(y),\ S_{xy} = \operatorname{cov}_{11}(x,y),\ S_{\bullet x}^2 = \operatorname{cov}_{\bullet 2}(x),\ S_{\bullet y}^2 = \operatorname{cov}_{\bullet 2}(y),$$

$$S_{\bullet xy} = \operatorname{cov}_{\bullet 11}(x,y),$$

$$\hat{S}_{\bullet x}^2 = \frac{1}{n-1}\sum_{i\in S}(x_i-\bar{x}_S)^2,\ \hat{S}_{\bullet y}^2 = \frac{1}{n-1}\sum_{i\in S}(y_i-\bar{y}_S)^2,\ \hat{S}_{\bullet xy} = \frac{1}{n-1}\sum_{i\in S}(x_i-\bar{x})(y_i-\bar{y}).$$

For simple random sampling without replacement $\hat{\bar{x}} = \bar{x}_S = \frac{1}{n}\sum_{i\in S}x_i$, $\hat{\bar{y}} = \bar{y}_S = \frac{1}{n}\sum_{i\in S}y_i$.

Let us notice that:

$$\operatorname{cov}(\hat{Y}^{regr}, \hat{\beta}) = \mathrm{E}((\hat{\beta} - \mathrm{E}(\hat{\beta})) = (\hat{Y}^{regr} - N\bar{y})) =$$

$$= \mathrm{E}[\hat{\beta}(N\bar{y}_S + \hat{\beta}(N\bar{x} - N\bar{x}_s) - N\bar{y})] - \mathrm{E}(\hat{\beta})\mathrm{B}^{regr}.$$

The bias of the regression estimator of the total value in the population denoted by $B^{regr} = \mathrm{E}(\hat{Y}^{regr} - N\bar{y})$ for simple random sampling without replacement is given by the equation (8).

For simple random sampling without replacement $\mathrm{E}(\hat{\beta})$ is given by the equation (7).

To derive an equation of $\operatorname{cov}(\hat{Y}^{regr}, \hat{\beta})$ let us notice, that:

$$\mathrm{E}[\hat{\beta}(N\bar{x}_S + \hat{\beta}(N\bar{x} - N\bar{x}_S) - N\bar{y})] = \mathrm{E}\left[\frac{\hat{S}_{\bullet xy}}{\hat{S}^2_{\bullet x}}\left(N\bar{y}_S + \frac{\hat{S}_{\bullet xy}}{\hat{S}^2_{\bullet x}}(N\bar{x} - N\bar{x}_S) - N\bar{y}\right)\right] =$$

$$= \mathrm{E}\left\{\left[\frac{S_{\bullet xy}}{S^2_{\bullet x}} + \frac{1}{S^2_{\bullet x}}(\hat{S}_{\bullet xy} - S_{\bullet xy}) - \frac{S_{\bullet xy}}{(S^2_{\bullet x})^2}(\hat{S}^2_{\bullet x} - S^2_{\bullet x})\right] \cdot \left[(N\bar{y}_S - N\bar{y}) + \right.\right.$$

$$\left.\left. + \left(\frac{S_{\bullet xy}}{S^2_{\bullet x}} + \frac{1}{S^2_{\bullet x}}(\hat{S}_{\bullet xy} - S_{\bullet xy}) - \frac{S_{\bullet xy}}{(S^2_{\bullet x})^2}(\hat{S}^2_{\bullet x} - S^2_{\bullet x})\right)(N\bar{x} - N\bar{x}_S)\right]\right\} + O(N \cdot n^{-2}) =$$

$$= \frac{N}{S^2_{\bullet x}}\operatorname{cov}(\hat{S}_{\bullet xy}, \bar{y}_S) - \frac{NS_{\bullet xy}}{(S^2_{\bullet x})^2}\operatorname{cov}(\hat{S}^2_{\bullet x}, \bar{y}_S) - \frac{NS_{\bullet xy}}{(S^2_{\bullet x})^2}\operatorname{cov}(\hat{S}^2_{\bullet xy}, \bar{x}_S) + \frac{N(S_{\bullet xy})^2}{(S^2_{\bullet x})^2}\operatorname{cov}(\hat{S}^2_{\bullet x}, \bar{x}_S) +$$

$$+ \frac{N}{(S^2_{\bullet x})^2}\mathrm{E}(\hat{S}^2_{\bullet x} - S_{\bullet xy})^2(\bar{x} - \bar{x}_S) - 2N\frac{S_{\bullet xy}}{(S^2_{\bullet x})^2}\mathrm{E}(\hat{S}_{\bullet xy} - S_{\bullet xy})(\hat{S}^2_{\bullet x} - S^2_{\bullet x})(\bar{x} - \bar{x}_S) +$$

$$+ N\frac{(S_{\bullet xy})^2}{(S^2_{\bullet x})}\mathrm{E}(\hat{S}^2_{\bullet x} - \hat{S}^2_{\bullet x})^2(\bar{x} - \bar{x}_S) + O(N \cdot n^{-2}).$$

Using results presented in J. W y w i a ł (1992, p. 138–139) it is known, that:

$$\operatorname{cov}(\hat{S}_{\bullet xy}, \bar{y}_S) = \frac{1}{n}\frac{N-n}{N-2}\operatorname{cov}_{\bullet 12}(x, y),$$

$$\operatorname{cov}(\hat{S}_{\bullet xy}, \bar{x}_S) = \operatorname{cov}(\hat{S}^2_{\bullet x}, \bar{y}) = \frac{1}{n}\frac{N-n}{N-2}\operatorname{cov}_{\bullet 21}(x, y),$$

$$\operatorname{cov}(\hat{S}^2_{\bullet xy}, \bar{x}_S) = \frac{1}{n}\frac{N-n}{N-2}\operatorname{cov}_{\bullet 3}(x).$$

According to equations derived in the Part A.2 of the Appendix we have:

$$\mathrm{E}(\hat{S}_{\bullet xy} - S_{xy})^2(\bar{x}_S - \bar{x}) = O(n^{-2}) + O(n^{-1}N^{-1}).$$

Because $\mathrm{E}(\hat{S}_{\bullet xy} - S_{xy})(\hat{S}^2_{\bullet x} - S^2_x)(\bar{x}_S - \bar{x})$ and $\mathrm{E}(\hat{S}^2_{\bullet x} - S^2_x)(\bar{x}_S - \bar{x})$ are simplified forms of $\mathrm{E}(\hat{S}_{\bullet xy} - S_{xy})^2(\bar{x}_S - \bar{x})$, we have:

$$\mathrm{E}(\hat{S}_{\bullet xy} - S_{xy})(\hat{S}^2_{\bullet x} - S^2_x)(\bar{x}_S - \bar{x}) = O(n^{-2}) + O(n^{-1}N^{-1}),$$

$$\mathrm{E}(\hat{S}_{\bullet x} - S_x)(\bar{x}_S - \bar{x}) = O(n^{-2}) + O(n^{-1}N^{-1}).$$

Hence,

$$\mathrm{cov}(\hat{Y}^{\mathrm{regr}}, \hat{\beta}) = \frac{N}{n}\frac{N-n}{N-2}\left(\frac{\mathrm{cov}_{\bullet 12}(x,y)}{\mathrm{cov}_{\bullet 2}(x)} - 2\frac{\mathrm{cov}_{\bullet 11}(x,y)\mathrm{cov}_{\bullet 21}(x,y)}{\mathrm{cov}^2_{\bullet 2}(x)} + \frac{\mathrm{cov}^2_{\bullet 11}(x,y)\mathrm{cov}_{\bullet 3}(x)}{\mathrm{cov}^3_{\bullet 2}(x)}\right) +$$

$$-\left(\frac{N}{n}\frac{N-n}{N-2}\sqrt{\mathrm{cov}_2(x)[B_2(x)-1]}\{k_{21}(x,y) - k_3(x)r(x,y)\}\right)\cdot$$

$$\cdot\left(\frac{\mathrm{cov}_{\bullet 11}(x,y)}{\mathrm{cov}_{\bullet 2}(x)} - \frac{1}{n}\frac{1}{\mathrm{cov}^2_{\bullet 2}(x)}\left((\mathrm{cov}_{\bullet 31}(x,y) - \mathrm{cov}_{\bullet 2}(x)\mathrm{cov}_{\bullet 11}(x,y)) + \right.\right.$$

$$\left.\left. + \frac{\mathrm{cov}_{\bullet 11}(x,y)}{\mathrm{cov}_{\bullet 2}(x)}(\mathrm{cov}_{\bullet 4}(x) - \mathrm{cov}^2_{\bullet 2}(x))\right)\right) +$$

$$+ O(N\cdot n^{-2}) + O(n^{-1}) = \frac{N}{n}\frac{N-n}{N-2}\left(\frac{\mathrm{cov}_{\bullet 12}(x,y)}{\mathrm{cov}_{\bullet 2}(x)} - 2\frac{\mathrm{cov}_{\bullet 11}(x,y)\mathrm{cov}_{\bullet 21}(x,y)}{\mathrm{cov}^2_{\bullet 2}(x)} + \right.$$

$$\left. + \frac{c^2_{\bullet 11}(x,y)c_{\bullet 3}(x)}{c^3_{\bullet 2}(x)}\right) +$$

$$-\frac{\mathrm{cov}_{\bullet 11}(x,y)}{\mathrm{cov}_{\bullet 2}(x)}\left(\frac{N}{n}\frac{N-n}{N-2}\sqrt{\mathrm{cov}_2(y)[B_2(x)-1]}\{k_{21}(x,y) - k_3(x)r(x,y)\}\right) +$$

$$+\left(\frac{N}{n}\frac{N-n}{N-2}\sqrt{\mathrm{cov}_2(y)[B_2(x)-1]}\{k_{21}(x,y) - k_3(x)r(x,y)\}\right)\cdot$$

$$\cdot\left(\frac{1}{n}\frac{1}{\mathrm{cov}^2_{\bullet 2}(x)}\left((\mathrm{cov}_{\bullet 31}(x,y) - \mathrm{cov}_{\bullet 2}(x)\mathrm{cov}_{\bullet 11}(x,y)) + \right.\right.$$

$$\left.\left. + \frac{\mathrm{cov}_{\bullet 11}(x,y)}{\mathrm{cov}_{\bullet 2}(x)}\mathrm{cov}_{\bullet 4}(x) - \mathrm{cov}^2_{\bullet 2}(x))\right)\right) + O(N\cdot n^{-2}) + O(n^{-1}).$$

Because

$$\left(\frac{N}{n}\frac{N-n}{N-2}\sqrt{\operatorname{cov}_2(y)[B_2(x)-1]}\{k_{21}(x,y)-k_3(x)r(x,y)\}\right)\cdot$$

$$\cdot\left(\frac{1}{n}\frac{1}{\operatorname{cov}^2_{\bullet 2}(x)}\left((\operatorname{cov}_{\bullet 31}(x,y)-\operatorname{cov}_{\bullet 2}(x)\operatorname{cov}_{\bullet 11}(x,y))+\right.\right.$$

$$\left.\left.+\frac{\operatorname{cov}_{\bullet 11}(x,y)}{\operatorname{cov}_{\bullet 2}(x)}(\operatorname{cov}_{\bullet 4}(x)-\operatorname{cov}^2_{\bullet 2}(x))\right)\right)=O(N\cdot n^{-2})$$

finally we have:

$$\operatorname{cov}(\hat{\bar{Y}}^{regr},\hat{\beta})=\frac{N}{n}\frac{N-n}{N-2}\left(\frac{\operatorname{cov}_{\bullet 12}(x,y)}{\operatorname{cov}_{\bullet 2}(x)}-2\frac{\operatorname{cov}_{\bullet 11}(x,y)\operatorname{cov}_{\bullet 21}(x,y)}{\operatorname{cov}^2_{\bullet 2}(x)}+\frac{\operatorname{cov}^2_{\bullet 11}(x,y)\operatorname{cov}_{\bullet 3}(x)}{\operatorname{cov}^3_{\bullet 2}(x)}\right)+$$

$$-\frac{\operatorname{cov}_{\bullet 11}(x,y)}{\operatorname{cov}_{\bullet 2}(x)}\left(\frac{N}{n}\frac{N-n}{N-2}\sqrt{\operatorname{cov}_2(y)[B_2(x)-1]}\{k_{21}(x,y)-k_3(x)r(x,y)\}\right)+$$

$$+O(N\cdot n^{-2})+O(n^{-1}).$$

A.2. ADDITIONAL DERIVATIONS – PART ONE

In this part of the appendix an equation of $\mathrm{E}(\hat{S}_{\bullet xy}-S_{\bullet xy})^2(\bar{x}_S-\bar{x})$ for simple random sampling without replacement will be derived. Following derivations are based on the assumption, that $\bar{x}=\bar{y}=0$, what does not have influence on the generality of results:

$$\mathrm{E}(\hat{S}_{\bullet xy}-S_{\bullet xy})^2(\bar{x}_S-\bar{x})=\mathrm{E}(\hat{S}^2_{\bullet xy}\bar{x}_S-2S_{\bullet xy}\hat{S}_{\bullet xy}\bar{x}_S+S^2_{xy}\bar{x})=$$

$$=\mathrm{E}(\hat{S}^2_{\bullet xy}\bar{x}_S)-2S_{\bullet xy}\mathrm{E}(\hat{S}_{\bullet xy}\bar{x}_S).$$

Then, using results presented in J. W y w i a ł (1992), p. 138–139, we receive

$$\mathrm{E}(\hat{S}_{\bullet xy}-S_{\bullet xy})^2(\bar{x}_S-\bar{x})=\mathrm{E}(\hat{S}^2_{\bullet xy}\bar{x}_S)-\frac{2N-n}{nN-2}\operatorname{cov}_{\bullet 11}(x,y)\operatorname{cov}_{\bullet 21}(x,y). \quad (8)$$

Let us notice, that

$$E(\hat{S}^2_{*xy}\bar{x}_S) = \frac{1}{(n-1)^2}E\left(\sum_{i=1}^{N} x_i y_i a_i - n\bar{x}_S\bar{y}_S\right)^2 \bar{x}_S = \frac{1}{(n-1)^2}E\left[\frac{1}{n}\left(\sum_{i=1}^{N} x_i y_i a_i\right)^2\left(\sum_{i=1}^{N} x_i a_i\right) + \right.$$

$$\left. -\frac{2}{n^2}\left(\sum_{i=1}^{N} x_i y_i a_i\right)\left(\sum_{i=1}^{N} x_i a_i\right)^2\left(\sum_{i=1}^{N} y_i a_i\right) + \frac{1}{n^3}\left(\sum_{i=1}^{N} x_i a_i\right)^2\left(\sum_{i=1}^{N} y_i a_i\right)^2\right].$$

All of three elements of above-mentioned sum will be derived separately using following equations: $E(a_i) = \frac{n}{N}$, $E(a_i a_j) = \frac{n(n-1)}{N(N-1)}$, $E(a_i a_j a_k) = \frac{n(n-1)(n-2)}{N(N-1)(N-2)}$, $E(a_i a_j a_k a_l) = \frac{n(n-1)(n-2)(n-3)}{N(N-1)(N-2)(N-3)}$. Some simple transformations will be omitted.

First,

$$\frac{1}{n(n-1)^2}E\left(\sum_{i=1}^{N} x_i y_i a_i\right)^2\left(\sum_{i=1}^{N} x_i a_i\right) = \left(\frac{2}{n} - \frac{2}{N}\right)\mathrm{cov}_{*21}(x,y)\mathrm{cov}_{*11}(x,y) + O(n^{-2}).$$

Second,

$$\frac{-2}{n^2(n-1)^2}E\left(\sum_{i=1}^{N} x_i y_i a_i\right)^2\left(\sum_{i=1}^{N} x_i a_i\right)^2 = \left(\sum_{i=1}^{N} y_i a_i\right) = O(n^{-2}).$$

Third,

$$\frac{1}{n^3(n-1)^2}E\left(\sum_{i=1}^{N} x_i a_i\right)^3\left(\sum_{i=1}^{N} y_i a_i\right)^2 = O(n^{-3}).$$

After summing up the three elements derived above we receive:

$$E(\hat{S}^2_{*xy}\bar{x}_S) = \left(\frac{2}{n} - \frac{2}{N}\right)\mathrm{cov}_{*21}(x,y)\mathrm{cov}_{*11}(x,y) + O(n^{-2}).$$

Using the equation (8) finally we received:

$$E(\hat{S}_{*xy} - S_{*xy})^2(\bar{x}_S - \bar{x}) = E(\hat{S}^2_{*xy}\bar{x}_S) - \frac{2N-n}{nN-2}\mathrm{cov}_{11}(x)\mathrm{cov}_{*21}(x,y) =$$

$$= \left(\frac{2}{n} - \frac{2}{N}\right)\mathrm{cov}_{*21}(x,y)\mathrm{cov}_{*11}(x,y) - \frac{2N-n}{nN-2}\mathrm{cov}_{11}(x)\mathrm{cov}_{*21}(x,y) + O(n^{-2}) = O(n^{-2}).$$

A.3. ADDITIONAL DERIVATIONS – PART TWO

In Part A.2 of the appendix an equation of $\operatorname{cov}(\hat{S}_{\bullet xy}, \hat{S}^2_{\bullet xy})$ for simple random sampling without replacement will be derived. Following derivations are based on the assumption, that $\overline{x} = \overline{y} = 0$, what does not have influence on the generality of results.

$$\operatorname{cov}(\hat{S}_{\bullet xy}, \hat{S}^2_{\bullet x}) = \mathrm{E}(\hat{S}_{\bullet xy} - S_{\bullet xy})(\hat{S}^2_{\bullet x} - S^2_{\bullet x}) = \mathrm{E}(\hat{S}_{\bullet xy}, \hat{S}_{\bullet xy}) - S_{\bullet xy} S^2_{\bullet x} \tag{9}$$

Let us notice, that:

$$\mathrm{E}(\hat{S}_{\bullet xy}, \hat{S}_{\bullet xy}) = \frac{1}{(n-1)} \mathrm{E}\left(\sum_{i=1}^{N} x_i y_i a_i - n\overline{x}_S \overline{y}_S \right)\left(\sum_{i=1}^{N} x_i^2 a_i - n\overline{x}_S^2 \right) =$$

$$= \frac{1}{(n-1)^2} \mathrm{E}\left[\left(\sum_{i=1}^{N} x_i y_i a_i \right)\left(\sum_{i=1}^{N} x_i^2 a \right) - \frac{1}{n}\left(\sum_{i=1}^{N} x_i y_i a_i \right)\left(\sum_{i=1}^{N} x_i a_i \right)^2 + \right.$$

$$\left. - \frac{1}{n}\left(\sum_{i=1}^{N} x_i^2 a_i \right)\left(\sum_{i=1}^{N} x_i a_i \right)\left(\sum_{i=1}^{N} y_i a_i \right) + \frac{1}{n^2}\left(\sum_{i=1}^{N} x_i a_i \right)^3 \left(\sum_{i=1}^{N} y_i a_i \right) \right].$$

All of four elements of above-mentioned sum will be derived separately using following equations: $\mathrm{E}(a_i) = \frac{n}{N}$, $\mathrm{E}(a_i a_j) = \frac{n(n-1)}{N(N-1)}$, $\mathrm{E}(a_i a_j a_k) = \frac{n(n-1)(n-2)}{N(N-1)(N-2)}$, $\mathrm{E}(a_i a_j a_k a_l) = \frac{n(n-1)(n-2)(n-3)}{N(N-1)(N-2)(N-3)}$. Some simple transformations will be omitted:

First,

$$\frac{1}{(n-1)^2} \mathrm{E}\left(\sum_{i=1}^{N} x_i y_i a_i \right)\left(\sum_{i=1}^{N} x_i^2 a \right) = \frac{1}{n} \operatorname{cov}_{\bullet 31}(x,y) + \left(1 + \frac{1}{n}\right) \operatorname{cov}_{\bullet 11}(x,y) \operatorname{cov}_{\bullet 2}(x) +$$

$$+ O(n^{-2}) + O(N^{-1}).$$

Second,

$$\frac{-1}{n(n-1)^2} \mathrm{E}\left(\sum_{i=1}^{N} x_i y_i a_i \right)\left(\sum_{i=1}^{N} x_i a \right)^2 = -\frac{1}{n} \operatorname{cov}_{\bullet 11}(x,y) \operatorname{cov}_{\bullet 2}(x) + O(n^{-2}) + O(N^{-1}).$$

Third,

$$-\frac{1}{n(n-1)^2} \mathrm{E}\left(\sum_{i=1}^{N} x_i^2 a_i \right)\left(\sum_{i=1}^{N} x_i a_i \right)\left(\sum_{i=1}^{N} y_i a_i \right) = \frac{1}{n} \operatorname{cov}_{\bullet 11}(x,y) \operatorname{cov}_{\bullet 2}(x) + O(n^{-2}) + O(N^{-1}).$$

Fourth,

$$\frac{1}{n^2(n-1)^2}\mathrm{E}\left(\sum_{i=1}^{N} x_i a_i\right)^3\left(\sum_{i=1}^{N} y_i a_i\right) = +O(n^{-2}) + O(N^{-1}).$$

After summing up the four elements derived above we receive:

$$\mathrm{E}(\hat{S}_{*xy}\hat{S}_{*x}) = \frac{1}{n}\mathrm{cov}_{*31}(x,y) + \left(1-\frac{1}{n}\right)\mathrm{cov}_{*11}(x,y)\mathrm{cov}_{*2}(x) + O(n^{-2}) + O(N^{-1}).$$

Using the equation (9) finally we received:

$$\mathrm{cov}(\hat{S}_{*xy},\hat{S}_{*x}) = \frac{1}{n}\left(\mathrm{cov}_{*31}(x,y) - cov_{*2}(x)cov_{*11}(x,y)\right) + O(N^{-1}) + O(n^{-2}).$$

REFERENCES

Bracha C. (1994), *Metodologiczne aspekty badania małych obszarów*, "Studia i Materiały. Z Prac Zakładu Badań Statystyczno-Ekonomicznych", nr 43, GUS, Warszawa.

Bracha C. (1996), *Teoretyczne podstawy metody reprezentacyjnej*, PWN, Warszawa.

Getka-Wilczyńska E. (2000), *Estymacja zjawisk rzadkich w populacji skończonej, PhD thesis*, Szkoła Główna Handlowa, Warszawa.

Särndal C. E., Swensson B., Wretman J. (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Wywiał J. (1992), *Statystyczna metoda reprezentacyjna w badaniach ekonomicznych (optymalizacja badań próbkowych)*, Akademia Ekonomiczna w Katowicach, Katowice.

Wywiał J. (2000), *Ocena parametrów cech populacji z wykorzystaniem danych o zmiennych dodatkowych*, sprawozdanie z realizacji Grantu KBN 1H02B 008 16.

Tomasz Żądło

O BŁĘDZIE ŚREDNIOKWADRATOWYM SYNTETYCZNEGO ESTYMATORA REGRESYJNEGO

W opracowaniu rozważa się problem estymacji wartości globalnej w małym obszarze. Ze względu na małą liczbę (lub brak) elementów populacji z rozważanej domeny, w próbie wykorzystywane są informacje o wszystkich elementach populacji. Zaprezentowany zostaje syntetyczny estymator regresyjny. Wyprowadzono też ogólne wzory na błąd średniokwadratowy i obciążenie tego estymatora dla dowolnego planu losowania. Omówiony zostaje problem założeń dotyczących struktury populacji i domeny z punktu widzenia redukcji błędu średniokwadratowego i obciążenia rozważanego estymatora. Zaprezentowane jest znaczenie wpływu

obciążenia na precyzję estymacji. Pokazana zostaje możliwość wzrostu wartości błędu średniokwadratowego i obciążenia wraz ze wzrostem liczebności próby. Wyprowadzone zostają przybliżone wzory na błąd średniokwadratowy i obciążenie estymatora dla próby prostej losowanej bezzwrotnie. Porównano również precyzję syntetycznego estymatora regresyjnego (wartości uzyskane na podstawie wzorów przybliżonych oraz symulacji) z precyzją bezpośredniego estymatora Horwitza–Thompsona. Porównanie bazuje na danych ze spisu rolnego dla powiatu Dąbrowa Tarnowska. Populacja składała się z 8624 gospodarstw rolnych i obejmowała rozważany mały obszar gminy Bolesław, na której terenie znajduje się 588 gospodarstw rolnych.