Estimation of Poverty in Small Areas[1]

Agne Bikauskaite[2] | *Eurostat, Luxembourg*

Abstract

A qualitative techniques of poverty estimation is needed to better implement, monitor and determine national areas where support is most required. The problem of small area estimation (SAE) is the production of reliable estimates in areas with small samples. The precision of estimates in strata deteriorates (i.e. the precision decreases when the standard deviation increases), if the sample size is smaller. In these cases traditional direct estimators may be not precise and therefore pointless. Currently there are many indirect methods for SAE. The purpose of this paper is to analyze several different types of techniques which produce small area estimates of poverty.

Keywords	***JEL code***
Poverty, small area estimation, Horvitz-Thompson, Generalised Regression, Synthetic, Jack-Knife	*I32, C89*

INTRODUCTION

The focus of this analysis is persons and their income. Estimated parameters are the following: the average household income, the at-risk-of-poverty indicators and their variances. All parameters have been estimated using the Horvitz-Thompson, the Generalised Regression (GREG), and the Synthetic estimation methods. The Jack-Knife method has been used for the estimation of variances to indicate the precision of the estimates. The Absolute Relative Bias (ARB) was applied to compare the performance of the different estimators for 1 000 simulations.

1 DATA AND METHODOLOGY

1.1 Analysed population

Canadian household survey data[3] was used for the simulation. The analysed population $U = (1, ..., i, ..., N)$ consisted of 3 000 individuals from 1 024 households with income values obtained $(y_1,\dots,y_N)$. The gender[4] and age[5] of individuals have been used as auxiliary information. This population is actually a simple

1 This article represents the personal views of the author. It does not reflect in any way Eurostat's position or view in this topic. The article is based on a presentation at the European Conference on Quality in Official Statistics (Q2014) in Vienna, Austria, on 3–5 June 2014.

2 Eurostat, 5 rue Alphonse Weicker, 2721 Luxembourg. E-mail: agne.bikauskaite@ext.ec.europa.eu.

3 Canadian household survey data of the 1991 has been taken from Statvillage data base <http://www.lenato.eu/StatVillage/index.html>.

4 The population has been divided into two gender groups: males and females.

5 The population has been divided into seven age groups: less than 5, from 5 to 17, from 18 to 24, from 25 to 59, from 60 to 64, from 65 to 74, from 75 years old.

random sample but was treated as a population and has been divided into seven mutually exclusive strata of different size (see Table 1) for simulation purposes.

1.2 Stratified sampling

A simple random sample drawn from the population can be homogeneous. In order to have more precise estimates of the population the data set has been divided into $H = 7$ mutually exclusive strata $U_1, U_2,..., U_H$ randomly.

For the analysis a stratified simple random sample s composed of seven strata with n_h elements in each has been drawn and y_h values observed. The size of the sample s is $n = n_1 + ... + n_h$.

Table 1 Strata size

Number of strata	The population size N_h	The sample size n_h
1	496	50
2	333	33
3	177	18
4	119	12
5	92	9
6	794	79
7	989	99
Total	3 000	300

Source: Own computations

The sample design probability when element i belongs to strata h is $\pi_{ih} = \frac{n_h}{N_h}$; the sampling weight for selected person i from the h^{th} strata is $w_{ih} = \frac{1}{\pi_{ih}} = \frac{N_h}{n_h}$. The value of the observed variable y into h^{th} strata of the i^{th} element is y_{hi}, i = 1,2,..., N_h, h = 1,2,..., H. Then the sum of observed values y in h^{th} strata is $t_h = \sum_{i=1}^{N_h} y_{hi}$ and the mean $\mu_h = \frac{t_h}{N_h} = \frac{1}{N_h}\sum_{i=1}^{N_h} y_i$.

The sum and the mean of y values observed through the whole population are accordingly $t = \sum_{h=1}^{H} t_h$ and $\mu = \frac{t}{N}$. (Krapavickaitė, Plikusas, 2005)

1.3 Estimated parameters

The average incomes, the at-risk-of-poverty threshold, the at-risk-of-poverty rate, the at-risk-of-poverty gap index, and the variances of these indicators have been calculated. 1 000 samples have been drawn to verify the best of three applied methods for small area estimation. The estimated indicators and variances have been compared with the real values. Parameters have been estimated for every strata separately and also for the total the population.

1.4 At-risk-of-poverty indicators

Persons or households with disposable income lower than at-risk-of-poverty threshold are considered as living in poverty or social exclusion because there is no possibility of participating fully in society life. In countries with high quality of life conditions not all residents below the at-risk-of-poverty threshold lack money. However, they have a significantly lower potential to meet their needs compared with the rest of community but they may live in good enough conditions.

The at-risk-of-poverty rate and the at-risk-of-poverty gap index are focused on those individuals below the at-risk-of-poverty threshold. The at-risk-of-poverty rate P_0 shows which part of society is below the poverty threshold. The at-risk-of-poverty gap shows the average lack of finance and how much the income has to increase so that the poverty threshold is reached.

1.4.1 The at-risk-of-poverty threshold

The at-risk-of-poverty threshold is defined as 60% of the median equivalised disposable income[6] $z = 60\%M$. This indicator depends on the income distribution in society and varies according to the changes of the general living conditions in the area.

1.4.2 The at-risk-of-poverty threshold estimation

To estimate the at-risk-of-poverty threshold, the median $\hat{M}$ of the income has to been estimated. Firstly units $y_1,\dots,y_n$ of s^{th} sample have been sorted in ascending order $y_{1:s} \le y_{2:s} \le \dots \le y_{n:s}$ and inclusion into s^{th} sample probabilities accordingly $\pi_{1:s};\pi_{2:s};\dots;\pi_{n:s}$. Accumulative totals of sampling weights have been counted $B_1 = \frac{1}{\pi_{1:s}}$, $B_2 = \frac{1}{\pi_{1:s}} + \frac{1}{\pi_{2:s}}$, …, $B_l = \sum_{j=1}^{l} \frac{1}{\pi_{j:s}}$ while one of the l satisfied the following condition $B_{l-1} < 0{,}5\hat{N}$ *and* $B_l > 0{,}5\hat{N}$.

Then the estimated number of the population is $\hat{N} = B_n = \sum_{j=1}^{n} \frac{1}{\pi_{j:s}} = \sum_{s} \frac{1}{\pi_i}$ and the median estimate is $\hat{M} = \begin{cases} y_{l:s}, & \text{if } B_{l-1} < 0{,}5\hat{N} < B_l \\ \frac{1}{2}(y_{l:s} + y_{l+1:s}), & \text{if } B_l = 0{,}5\hat{N} \end{cases}$.

Then the estimate of the poverty threshold is defined by formula $\hat{z} = 60\%\hat{M}$.

1.4.3 The at-risk-of-poverty rate

The at-risk-of-poverty rate is defined as the number of persons below the at-risk-of-poverty threshold divided by the population number $P_0 = \frac{1}{N}\sum_{i=1}^{N} I_{(y_i<z)} = \frac{N_q}{N}$, here $I_{(y_i<z)} = \begin{cases} 1, & y_i < z \\ 0 & y_i \ge z \end{cases}$. N_q defines the number of individuals whose income is below the at-risk-of-poverty threshold $N_q = \sum_{i=1}^{N} I_{(y_i<z)}$.

1.4.4 The at-risk-of-poverty rate estimation

The at-risk-of-poverty rate estimator is $\hat{P}_0 = \frac{1}{\hat{N}}\sum_{i=1}^{n} w_i I_{(y_i<z)} = \frac{\hat{N}_q}{\hat{N}}$, here $\hat{N}$ is the estimated number of the population elements; $\hat{N}_q$ is the estimated number of individuals in the population living in poverty or social exclusion.

1.4.5 The at-risk-of-poverty gap index

The at-risk-of-poverty gap G_n is defined as an amount of difference between the at-risk-of-poverty threshold and income value y_i of i^{th} person living in poverty or social exclusion $G_i = (z - y_i)\,I_{(y_i<z)}$. The at-risk-of-poverty gap index is a proportion of the at-risk-of-poverty gap and the at-risk-of-poverty threshold $P_1 = \frac{1}{N}\sum_{i=1}^{q}\frac{G_i}{z} = \frac{1}{N}\sum_{i=1}^{N}\frac{z-y_i}{z} I_{(y_i<z)}$, here q is number of individuals in poverty or social exclusion.

1.4.6 The at-risk-of-poverty gap index estimation

Then the direct estimate of the at-risk-of-poverty gap index is defined by formula:

$$\hat{P}_1 = \frac{1}{\hat{N}}\sum_{i=1}^{n}\frac{\hat{z}-y_i}{\hat{z}} w_i I_{(y_i<\hat{z})}.$$

---

[6] Equivalised disposable income of person is calculated by dividing the disposable household income by the equivalised household size. All members of the same household are assigned the same equivalised disposable income.

1.5 Direct and indirect estimators

1.5.1 Small Area Estimation

An area is regarded as large if the sample drawn from that area is large enough to get direct estimates of adequate precision. An area is regarded as small if the sample is not large enough to get simple direct estimates of adequate precision. The variance of the estimate decreases through enlarging the size of the sample (Rao, 2010).

In order to have better quality estimates in areas, unbiased auxiliary variables have to be used from the same areas. This kind of estimation is defined as direct. For indirect estimation the auxiliary information has to be taken from adjacent areas.

1.5.2 The Horvitz-Thompson estimator

The Horvitz-Thompson estimator of the sum is $\hat{t}_\pi = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{n} w_i y_i$.

For a stratified simple random sample the Horvitz-Thompson variance of the sum estimate is $D\hat{t}_\pi = \sum_{i,j \in s} (\pi_{ij} - \pi_i \pi_j) \frac{y_i y_j}{\pi_i \pi_j}$. The Horvitz-Thompson variance estimate of the sum estimate is $\hat{D}\hat{t}_\pi = \sum_{i,j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{y_i y_j}{\pi_i \pi_j}$, here $\pi_i = \frac{n_i}{N_i}, i \in U_h$; $\pi_{ij} = \frac{n_h}{N_h} \cdot \frac{n_h - 1}{N_h - 1}$, *when* $i, j \in U_h$ and $\pi_{ij} = \frac{n_h}{N_h} \cdot \frac{n_s}{N_s} = \pi_i \pi_j$, when $i \in U_h$, $j \in U_s$. π_{ij} is the inclusion into the sample probability of two elements *(i, j)*. If $i = j$ *then* $\pi_{ii} = \pi_i$ (Krapavickaitė, Plikusas, 2005).

1.5.3 The Generalised Regression Model (GREG)

y_i, is the values of the income and the value of the vector $\boldsymbol{x}$ is defined as the auxiliary information $\mathbf{x}_i = (x_{1i}, \ldots, x_{ji}, \ldots, x_{Ji})'$.

The sum of the dominant elements y is the GREG estimator of the sum t_y defined by the following formula $\hat{t}_{y,GREG} = \hat{t}_{y\pi} + \sum_{j=1}^{J} \hat{B}_j \left(t_{X_j} - \hat{t}_{X_j \pi} \right)$, where j is the number of several auxiliary pieces of information about the individual. The Horvitz-Thompson estimator of the sum t_{x_j} is $\hat{t}_{x_j \pi} = \sum_{i=1}^{n} \frac{\mathbf{x}_j}{\pi_i}$. $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_J$ are the estimated components of the vector $\mathbf{x}$ $\hat{\mathbf{B}} = \left(\sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}'_k q_k}{\pi_i} \right)^{-1} \sum_{i=1}^{n} \frac{\mathbf{x}_i y_i q_i}{\pi_i}$.

The GREG estimation method is appropriate to estimate parameters in non-responses. Then the GREG estimator of the sum is $\hat{t}_{wy} = \sum_r w_i y_i$, where r is the set of the respondents. The calibrated weights are $w_i = \left(1 + \left(\mathbf{t}_x - \hat{\mathbf{t}}_x \right)' \left(\sum_r \frac{\mathbf{x}_i \mathbf{x}'_i q_i}{\pi_i \hat{\theta}_i} \right)^{-1} \mathbf{x}_i q_i \right) \times \frac{1}{\pi_i \hat{\theta}_i} = \frac{g_i}{\pi_i \hat{\theta}_i}$, where $\hat{\theta}_k$ is the estimator of element's i response to the survey probability.

The calibrated estimate of the sum $\hat{t}_{wy}$ is biased. When N is large but sampling rate $\frac{n}{N}$ small then the bias estimate is slight.

1.5.4 Simple Synthetic estimator

The stratified population U_h splits up into k mutually exclusive groups $G_1, \ldots, G_K$, $U_h = G_{h1} \cup \ldots \cup G_{hK}$ and $U = G_1 \cup \ldots \cup G_K$.

The mean of the elements from h^{th} strata and k^{th} group is $\mu_{yhk} = \frac{\sum_{i=0}^{n_{hk}} w_i y_i}{\sum_{i=0}^{n_{hk}} w_i}$, here $w_0 y_0 = 0$ when $n_{hk} = 0$, i.e. if the element from h^{th} strata and k^{th} group in the population does not exist then the sum is . The sum of population in strata h is $t_{yh} = \sum_{k=0}^{n_k} \mu_{yhk} N_{hk}$. The sum estimator of the sample is $\hat{t}_{yh}^{\sin t} = \sum_{k=0}^{K} \hat{\mu}_{yhk} N_{hk}$ and its variance is $D\hat{t}_{yh}^{\sin t} = D(\sum_{k=1}^{K} \hat{\mu}_{yhk} N_{hk})$.

The synthetic estimator is unbiased when $\mu_{yhk} = \mu_{yh}$, here $h = 1,\ldots,H$, $k = 1,\ldots K$. If this is the opposite, it is biased.

1.6 The variance estimation

To estimate the precision of estimated parameters the Jack-Knife variance estimation method has been used.

The Jack-Knife method's idea is to divide stratified sample S_h into K_h mutually exclusive subgroups. If $\hat{\theta}_h$ is the estimate of the parameter θ_h of the primary stratified sample s_h, then $\hat{\theta}_{(hk)}$ is parameter's θ estimator obtained by estimating the sample composed of h^{th} strata elements apart units from k^{th} ($k = 1,\ldots,K_h$) group. The modified sampling weights were used to estimate $\hat{\theta}_{(hk)}$:

$$w_{i(hk)} = \begin{cases} w_i, & \text{when } i^{th} \text{ element does not belong to } h^{th} \text{ stratum,} \\ 0, & \text{when } i^{th} \text{ element belongs to } h^{th} \text{ stratum and } k^{th} \text{ subgroup,} \\ \frac{n_i}{n_i - 1} w_i, & \text{when } i^{th} \text{ element belongs to } h^{th} \text{ stratum.} \end{cases} \quad (1)$$

Then the Jack-Knife variance estimator of θ estimate is equal to

$$\hat{D}_{JACK}\hat{\theta}_{(hk)} = \sum_{h=1}^{H}\left[\frac{K_h - 1}{K_h}\right]\sum_{k=1}^{K_h}\left(\hat{\theta}_{(hk)} - \hat{\bar{\theta}}_{(hk)}\right)^2, \text{ here } \hat{\bar{\theta}}_{(hk)} = \frac{1}{K_h}\sum_{k=1}^{K_h}\hat{\theta}_{(hk)}$$

1.7 The Absolute Relative Bias

The Absolute Relative Bias (ARB) assessed the accuracy of the estimates $ARB = \left|\frac{1}{K}\sum_{k=1}^{K}\frac{\hat{\theta}_h - \theta_h}{\theta_h}\right|$, where K is the number of drawn samples; $\hat{\theta}_h$ is the estimate of the parameter in the strata h; θ_h is real value of parameter in the strata h.

2 RESULTS

2.1 Estimates of parameters

The real values of the average income and the at-risk-of-poverty indicators have been calculated. All parameters have been estimated using Horvitz-Thompson, Generalised Regression, and Synthetic methods (see Tables 2, 4, and 6).

Table 2 Estimates of the average income

Strata	Sample size	Real average income	Horvitz–Thompson estimate of average income	Generalised Regression estimate of average income	Synthetic estimate of average income	The minimum value of average income estimate	The maximum value of average income estimate
Population	300	19 377.85	19 365.35	19 396.90	19 361.57	17 413.25 (H-T) 17 443.45 (GREG) 17 493.11 (S)	21 829.21 (H-T) 21 867.56 (GREG) 21 857.35 (S)
1	50	20 594.44	20 528.29	20 529.53	20 523.78	15 433.70 (H-T) 15 201.54 (GREG) 14 996.18 (S)	27 192.90 (H-T) 27 567.32 (GREG) 27 576.73 (S)
2	33	17 484.20	17 479.58	17 481.74	17 472.13	12 126.88 (H-T) 11 768.48 (GREG) 11 871.44(S)	23 512.79 (H-T) 23 345.83 (GREG) 23 316.08 (S)
3	18	18 513.46	18 599.66	18 615.62	18 588.24	11 272.39 (H-T) 10 939.35 (GREG) 10 954.14 (S)	26 250.29 (H-T) 26 672.41 (GREG) 26 764.36 (S)
4	12	19 184.40	19 027.93	19 015.18	19 058.99	10 000.23 (H-T) 90 672.53 (GREG) 8 738.67 (S)	30 455.87 (H-T) 30 431.29 (GREG) 30 625.30 (S)
5	9	17 757.50	17 843.75	17 848.21	17 844.91	9 201.61 (H-T) 93 491.07 (GREG) 8 953.184 (S)	33 616.91 (H-T) 33 693.56 (GREG) 33 730.83 (S)
6	79	20 150.24	20 121.66	20 123.23	20 120.18	16 389.60 (H-T) 16 401.59 (GREG) 16 420.89 (S)	26 257.12 (H-T) 26 635.74 (GREG) 26 294.45 (S)
7	99	19 113.91	19 129.06	19 131.30	19 149.96	15 538.40 (H-T) 15 573.20 (GREG) 15 350.62 (S)	22 981.10 (H-T) 23 139.04 (GREG) 23 316.41(S)

Note: Here (H–T) value of Horvitz-Thompson estimate, (GREG) value of Generalised Regression estimate; (S) value of Synthetic estimate.
Source: Own computations

The best ARB, estimating the average income and the at-risk-of-poverty gap index for the whole population, was through using the Horvitz-Thompson method. The at-risk-of-poverty rate estimates obtained the least ARB applying the GREG method.

The purpose of the paper was to choose the most accurate method for the estimation in small areas. The results show that in the smallest, third and fourth strata which consist accordingly of 9 and 12 elements in the sample, the Synthetic estimates of the average income are closest to the real values (see Table 3).

Table 3 The ARB of the average income estimates

Strata	Horvitz-Thompson estimate's ARB (%)	Generalised Regression estimate's ARB (%)	Synthetic estimate's ARB (%)
Population	**–0.06447544**	0.098310539	–0.08398375
1	–0.3211974	**–0.31518106**	–0.34310121
2	–0.02643092	**–0.014056**	–0.06902109
3	0.465571393	0.551799055	**0.403882282**
4	–0.81562095	–0.88208503	**–0.65375062**
5	**0.485715332**	0.510841272	0.492216146
6	–0.1417938	**–0.13401672**	–0.14913289
7	**0.079252793**	0.090945055	0.188597999

Source: Own computations

The Synthetic at-risk-of-poverty rate estimate's ARB in the smallest fifth strata is least (see Table 5).

Table 5 The ARB of the at-risk-of-poverty rate estimates

Strata	Horvitz-Thompson estimate's ARB (%)	Generalised regression estimate's ARB (%)	Synthetic estimate's ARB (%)
Population	0.36396329	**0.147665664**	0.152247869
1	**–3.51959494**	–3.7958481	–3.8266288
2	1.468493151	1.192029888	**1.003491015**
3	**4.644761905**	4.757144543	5.058255185
4	2.859782609	**2.601086957**	2.877924901
5	–2.80634921	–2.90370419	**–1.68042706**
6	**–0.63675717**	–0.78252971	–0.8622043
7	1.344097079	**1.068357786**	1.298860988

Source: Own computations

Table 4 Estimates of the at-risk-of-poverty rate

Strata	Sample size	Real value of the at-risk-of-poverty rate	Horvitz-Thompson estimate of the at-risk-of-poverty rate	Generalised Regression estimate of the at-risk-of-poverty rate	Synthetic estimate of the at-risk-of-poverty rate	The minimum value of the at-risk-of-poverty rate estimate	The maximum value of the at-risk-of-poverty rate estim
Population	300	0.194333333	0.193626	**0.1940464**	0.194037465	0.133518 (HT) 0.133518 (GREG) 0.136865 (S)	0.250143 (HT) 0.253482 (GREG) 0.251659 (S)
1	50	0.159274194	0.16488	0.16532	0.165369026	0.04 (HT) 0.04 (GREG) 0.040064 (S)	0.36 (HT) 0.36 (GREG) 0.40 (S)
2	33	0.219219219	0.216	0.2166061	0.217019374	0.030303 (HT) 0.030303 (GREG) 0.029412 (S)	0.454545 (HT) 0.454545 (GREG) 0.469924812 (S)
3	18	0.197740113	0.188556	0.1883333	0.187737913	0.03 (HT) 0.0 (GREG) 0.0 (S)	0.611111 (HT) 0.611111 (GREG) 0.60 (S)
4	12	0.193277311	0.18775	0.18825	0.187714935	0.0 (HT) 0.0 (GREG) 0.0 (S)	0.666667 (HT) 0.666667 (GREG) 0.6875 (S)
5	9	**0.22826087**	0.234667	0.2348889	0.232096627	0.0 (HT) 0.0 (GREG) 0.0 (S)	0.666667 (HT) 0.666667 (GREG) 0.75 (S)
6	79	0.164987406	0.166038	0.1662785	0.166409934	0.063291 (HT) 0.063291 (GREG) 0.060401 (S)	0.303797 (HT) 0.303797 (GREG) 0.303273427 (S)
7	99	0.223458038	0.220455	0.2210707	0.220555629	0.10101 (HT) 0.10101 (GREG) 0.100816 (S)	0.323232 (HT) 0.323232 (GREG) 0.330827068 (S)

Note: Here (H–T) value of Horvitz-Thompson estimate, (GREG) value of Generalised Regression estimate; (S) value of Synthetic estimate.
Source: Own computations

Table 6 Estimates of the at-risk-of-poverty rate

Strata	Sample size	Real value of the at-risk-of-poverty gap index	Horvitz-Thompson estimate of the at-risk-of-poverty gap index	Generalised Regression estimate of the at-risk-of-poverty gap index	Synthetic estimate of the at-risk-of-poverty gap index	The minimum value of the at-risk-of-poverty gap index estimate	The maximum value of the at-risk-of-poverty gap index estimate
Population	300	0.07252193	0.072638	0.0727797	0.072819745	0.045454 (HT) 0.045542 (GREG) 0.046188 (S)	0.106624 (HT) 0.107743 (GREG) 0.109839 (S)
1	50	0.052977463	0.053704	0.0538235	0.053812347	0.004695624 (HT) 0.00481718 (GREG) 0.005271356 (S)	0.136808961 (HT) 0.13717423 (GREG) 0.138750613 (S)
2	33	0.06671773	0.06736	0.0675437	0.067818474	0.0041682 (HT) 0.00417481 (GREG) 0.004052023 (S)	0.21001472 (HT) 0.21065701 (GREG) 0.217354998 (S)
3	18	0.073036323	0.0734	0.0735076	0.073541574	0.00 (HT) 0.00 (GREG) 0.00 (S)	0.304812082 (HT) 0.30481208 (GREG) 0.312076845 (S)
4	12	0.102809342	0.104061	0.1041447	0.10430057	0.00 (HT) 0.00 (GREG) 0.00 (S)	0.34929022 (HT) 0.34967569 (GREG) 0.502667938 (S)
5	9	0.088601259	0.089251	0.0894485	0.088580889	0.00 (HT) 0.00 (GREG) 0.00 (S)	0.336332409 (HT) 0.33700796 (GREG) 0.406565083 (S)
6	79	0.058100608	0.057692	0.0578231	0.057892628	0.008760654 (HT) 0.00877026 (GREG) 0.010559962 (S)	0.135791463 (HT) 0.13738105 (GREG) 0.13721441 (S)
7	99	0.090623887	0.090446	0.0906023	0.090371823	0.036089369 (HT) 0.03669861 (GREG) 0.036437574 (S)	0.175506287 (HT) 0.17515759 (GREG) 0.175058725 (S)

Note: Here (H–T) value of Horvitz-Thompson estimate, (GREG) value of Generalised Regression estimate; (S) value of Synthetic estimate.
Source: Own computations

In the same fifth strata the Synthetic at-risk-of-poverty gap index estimate has the smallest ARB (0.02%) compared with the Horvitz-Thompson and the GREG estimation methods.

Table 7 ARB of the at-risk-of-poverty gap index estimate

Strata	Horvitz-Thompson estimate's ARB (%)	Generalised regression estimate's ARB (%)	Synthetic estimate's ARB (%)
Population	**–0.1594528**	–0.35543944	–0.41065525
1	**–1.37126072**	–1.59705543	–1.57592157
2	**–0.9619282**	–1.23793553	–1.64985282
3	**–0.49766038**	–0.6453229	–0.69178013
4	**–1.21749012**	–1.2989069	–1.45047831
5	–0.73358855	–0.95628486	**0.02299011**
6	0.702989553	0.477610719	**0.357964671**
7	0.19625962	**0.02379632**	0.278143276

Source: Own computations

2.2 Estimated variances of parameters estimates

The largest over-estimations of the variance coefficients of averaged income estimates are in the smallest strata. Significantly better variance coefficients are obtained through the Horvitz-Thompson estimation (see Table 8). While the GREG and the Synthetic estimates are equally worse.

Table 8 Estimated variance coefficients of averaged income estimates

Strata	Sample size	Variance coefficient of the population	Horvitz-Thompson estimate's variance coefficient	GREG estimate's variance coefficient	Synthetic estimate's variance coefficient
Total	300	0.035	**0.039**	0.040	0.040
1	50	0.094	0.102	0.102	**0.101**
2	33	0.095	**0.104**	0.112	0.111
3	18	0.141	**0.135**	0.156	0.156
4	12	0.163	**0.181**	0.208	0.211
5	9	0.239	**0.252**	0.307	0.307
6	79	0.064	**0.068**	0.069	0.069
7	99	0.067	**0.072**	0.073	0.074

Source: Own computations

Concerning the variance coefficients of the at-risk-of-poverty rate and the at-risk-of-poverty gap index estimates, in most strata Horvitz-Thompson also produced the smallest overestimation (see Tables 9 and 10).

Table 9 Estimated variance coefficients of the at-risk-of-poverty rate estimates

Strata	Sample size	Real variation coefficient	Horvitz-Thompson variation coefficient's estimate	GREG variation coefficient's estimate	Synthetic variation coefficient's estimate
Total	300	0.104	**0.110**	0.115	0.117
1	50	0.422	**0.415**	0.410	0.440
2	33	0.408	0.477	0.477	**0.475**
3	18	0.544	0.483	0.484	**0.532**
4	12	0.698	0.602	**0.624**	0.818
5	9	0.172	**0.156**	0.206	0.186
6	79	0.232	**0.228**	0.266	0.284
7	99	0.171	0.217	0.217	**0.203**

Source: Own computations

Table 10 Estimated variance coefficients of the at-risk-of-poverty gap index estimates

Strata	Sample size	Real variation coefficient	Horvitz-Thompson variance coefficient's estimate	GREG variance coefficient's estimate	Synthetic variance coefficient's estimate
Total	300	0.141	**0.151**	0.162	0.166
1	50	0.420	**0.458**	0.461	0.472
2	33	0.421	**0.451**	0.462	0.467
3	18	0.645	**0.638**	0.637	0.613
4	12	0.666	**0.697**	0.747	0.733
5	9	0.792	**0.873**	0.982	1.051
6	79	0.332	0.362	**0.361**	0.362
7	99	0.226	**0.246**	0.247	0.256

Source: Own computations

CONCLUSIONS

Consequently we can see that to get good precise estimates would be better to apply different estimation methods for large and for small areas. Horvitz-Thompson method produces reliable estimates in large areas, but in most of the cases it does not suit for the poverty estimation in areas where sample size is small.

It is therefore suggested that if poverty estimation in small areas is to be made and if auxiliary information from the adjacent areas can be taken into account, the Synthetic method should be used. If, however, that auxiliary information is not available, then given the simulation results in general, the most appropriate estimation method for the analysed data would be Horvitz-Thompson.

The SAS programs text prepared for this simulation could be easily adjusted for other data to check how each of analysed techniques copes with your specific data taken from specific areas.

When comparing estimated variances of parameters estimates with real variances, large ARBs have been obtained. The best results of poverty indicator's estimation of population in small and in large areas are achieved by the Horvitz-Thompson method. This technique must be quite reliable enlarging the sample size, but in opposite, when sample size is reduced and goes to 0 the calculated estimates applying any direct method would be pointless.

Estimating the Jack-Knife variances calculation takes more time but the precision of the estimates increases when the group size is extremely small.

ACKNOWLEDGEMENT

The author would like to thank Morag Ottens who patiently edited the writings and offered invaluable advice. The presentation and publication of this paper would not be possible without her encouragement and support.

References

ANDERSSON, C., NORDBERG, L. *A User's Guide to CLAN 97 – a SAS – Program for Computation of Point and Standard Error Estimates in Sample Survey.* Statistics Sweden, 1998.

KRAPAVICKAITĖ, D., PLIKUSAS, A. *Imčių teorijos pagrindai.* Vilnius, Leidykla „Technika", 2005.

MOLINA, I., RAO, J. N. K. *Estimation of Poverty Measures in Small Areas* [online]. <http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S3P3_ESTIMATION_OF_POVERTY_MEASURES_MOLINA_RAO.pdf>.

RAO, J. N. K. *Small Area Estimation*. Wiley, 2010.

SÄRNDAL, S.-E., SWENSSON, B., WRETMAN, J. *Model Assisted Survey Sampling.* New York: Springer, 1992.

WOLTER, K. M. *Introduction to Variance Estimation.* New-York: Springer, 1985.