Tomasz Stachurski

University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, tstachurski93@gmail.com

Tomasz Żądło

University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, tomasz.zadlo@ue.katowice.pl

On Accuracy of Calibration Estimators Supported by Auxiliary Variables from Past Periods Based on Simulation Analyses

Abstract: In sample surveys there is often a need to estimate not only population characteristics, but subpopulation characteristics as well. We consider the problem of estimating the total value in domains (subpopulations). In this case, the Horvitz-Thompson estimator could be used. Nevertheless, it does not use any additional information about population units, which are usually known. To increase estimation accuracy we propose to use calibration estimators with auxiliary variables from the current and past periods. In the simulation studies based on real and generated data, we show the influence of using auxiliary information from past periods on the accuracy, and compare properties of two calibration estimators of domain totals in longitudinal surveys.

Keywords: calibration estimators, small area estimation, longitudinal surveys

JEL: C83

1. Introduction

Longitudinal data for periods $t = 1, 2, \ldots, M$ are considered. In the period t the population of size N_t is denoted by Ω_t. The population in the period t is divided into D disjoint subpopulations (domains) Ω_{dt} of size N_{dt}, where $d = 1, 2, \ldots, D$. The domain of interest in the period of interest is denoted by Ω_{d*t*}. Let the set of population elements for which observations are available in the period t be denoted by s_t and its size by n_t. The set of elements of d-th domain for which observations are available in the period t is denoted by s_{dt} and its size by n_{dt}. In the case of surveys conducted in one period (see section 2) the subscript t will be omitted. First and second order inclusion probabilities in the period t are denoted by π_{ti} and π_{tij} respectively. The value of the variable of interest and the vector of the p auxiliary variables for the i-th population element in the period t are denoted by y_{ti} and x_{ti} respectively.

Nowadays many surveys are conducted in more than one period which gives additional possibilities of increasing estimation accuracy. In the paper, we study properties of calibration estimators in the case of longitudinal surveys, where information on auxiliary variables is available not only for the current period but also for past periods. We consider the research hypothesis that the additional information provided by comparison with surveys conducted in another period will result in an increase in estimation accuracy. This must be verified in simulation studies because values of the same auxiliary variables in different periods are usually strongly correlated and hence similar information is included in the estimation process. The main purpose of this paper is to analyse the properties of certain calibration estimators of the domain total in longitudinal surveys in simulation studies, and to explore the influence of supporting estimations by auxiliary variables from past periods. Any results presented in the paper may be of direct interest to statistical offices, market research and opinion polling companies, and of indirect interest to decision-makers and all data users.

2. Calibration estimators – single wave

In this section we present calibration estimators used in surveys conducted over a particular period. The calibration estimator of the population total is given by (Deville, Särndal, 1992):

$$\hat{\theta}^{CAL} = \sum_{i \in s} w_{si} y_i \, ,$$
(1)

where weights w_{si} are solutions of:

$$\begin{cases} f_s(w_{si}, d_i, q_i) \rightarrow \min \\ \sum_{i \in s} w_{si} \mathbf{x}_i = \sum_{i \in \Omega} \mathbf{x}_i \end{cases}, \tag{2}$$

where $f_s(w_{si}, d_i, q_i)$ is some distance measure between weights of the calibration estimator w_{si} and the design weights (inverses of the first order inclusion probabilities) $d_i = \pi_i^{-1}$, where, for more generality, additional known weights q_i can be included. The minimization in (2) leads to the approximate design-unbiasedness of the calibration estimator. The equality in (2) is the condition of the model-unbiasedness of the estimator (1) under the linear model.

If in (2) we assume that:

$$f_s(w_{si}, d_i, q_i) = \sum_{i \in s} \frac{(w_{si} - d_i)^2}{d_i q_i}, \tag{3}$$

then the resulting calibration estimator called generalized regression estimator (GREG) is given by (Deville, Särndal, 1992; Särndal, Swensson, Wretman, 1992: 232; Rao, Molina, 2015: 13):

$$\hat{\theta}^{GREG} = \sum_{i \in s} d_i y_i + \left(\sum_{i \in \Omega} \mathbf{x}_i - \sum_{i \in s} d_i \mathbf{x}_i \right)^T \hat{\mathbf{B}}, \tag{4}$$

where

$$\hat{\mathbf{B}} = \left(\sum_{i \in s} d_i q_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i \in s} d_i q_i \mathbf{x}_i y_i. \tag{5}$$

Estimator (4) can be also written as (1) where

$$w_{si} = g_{si} d_i, \tag{6}$$

$$g_{si} = 1 + (\sum_{i \in \Omega} \mathbf{x}_i - \sum_{i \in s} d_i \mathbf{x}_i)^T \left(\sum_{i \in s} d_i q_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \mathbf{x}_i q_i, \tag{7}$$

What is more, Deville and Särndal (1992) prove under some conditions that the calibration estimator (1) is design-consistent and asymptotically equivalent to the generalized regression estimator (4) in the sense that

$$N^{-1} \left(\hat{\theta}^{CAL} - \hat{\theta}^{GREG} \right) = O_p(n^{-1}). \tag{8}$$

Although (8) informs that the estimators are asymptotically equivalent, Singh and Mohl (1996) and Stukel, Hidiroglou and Särndal (1996) show that the values of the estimators are very similar even for small sample sizes.

To estimate the design-variance of (4) we can use one of the design-consistent estimators (Rao, Molina, 2003: 15) given by

$$\hat{D}_p^2(\hat{\theta}^{GREG}) = \sum_{j>i}^{n} \sum_{i}^{n} (\pi_i \pi_j - \pi_{ij}) \pi_{ij}^{-1} \left(d_i e_i - d_j e_j \right)^2 , \qquad (9)$$

where

$$e_i = y_i - \mathbf{x}_i^T \hat{\mathbf{B}} , \qquad (10)$$

or

$$\widehat{D}_p^2(\hat{\theta}^{GREG}) = \sum_{j>i}^{n} \sum_{i}^{n} (\pi_i \pi_j - \pi_{ij}) \pi_{ij}^{-1} \left(d_i g_{si} e_i - d_j g_{sj} e_j \right)^2 , \qquad (11)$$

where g_{si} is given by (7). Estimator (9) may lead to some underestimation of the true variance, but including g_{si} in the equation (11) reduces this underestimation.

The formula (1) can be adapted to estimate the subpopulation (domain) total instead of the population total. The generalized regression estimator of the $d*$-th domain total is given by (Rao, Molina, 2015: 17–18):

$$\hat{\theta}_{d*}^{GREG} = \sum_{i \in s} w_{si} a_{id*} y_i = \sum_{i \in s_{d*}} w_{si} y_i , \qquad (12)$$

where

$$a_{id*} = \begin{cases} 1 & for \quad i \in \Omega_{d*} \\ 0 & for \quad i \notin \Omega_{d*} . \end{cases} \qquad (13)$$

This means that the estimator (12) is given by (1) where y_i is replaced by $a_{id*} y_i$. Hence, the design-variance of the estimator (12) can be estimated based on (11) where y_i is replaced by $a_{id*} y_i$, but it means that the residuals (10) in (11) for $i \notin s_{d*}$

are given by $e_i = -\mathbf{x}_i^T \hat{\mathbf{B}}(a_{id*} y_i)$ which may lead to inefficiency. The estimator is approximately design-unbiased even if the expected domain sample size is small (Rao, Molina, 2015: 18), but it cannot be used if $s_{d*} = \varnothing$.

The domain total can also be estimated using modified generalized regression estimator (MGREG) proposed by Särndal (1981) and given by:

$$\hat{\theta}_{d*}^{MGREG} = \sum_{i \in s_{d*}} d_i \, y_i + \left(\sum_{i \in \Omega_{d*}} \mathbf{x}_i - \sum_{i \in s_{d*}} d_i \mathbf{x}_i \right)^T \hat{\mathbf{B}} =$$

$$= \left(\sum_{i \in \Omega_{d*}} \mathbf{x}_i \right)^T \hat{\mathbf{B}} + \sum_{i \in s_{d*}} \frac{e_i}{\pi_i} \,, \qquad (14)$$

where $\hat{\mathbf{B}}$ is given by (5) and e_i by (10). Rao and Molina (2015: 23) propose to estimate the design variance of (14) by (9), where e_i are replaced by $a_{id*}e_i$ assuming that the overall sample size is large (even if the domain sample size is small).

Estimators (12) and (14) have a property called benchmarking – they sum up to the estimator of the population total (in this case GREG):

$$\sum_{d=1}^{D} \hat{\theta}_d^{GREG} = \sum_{d=1}^{D} \hat{\theta}_d^{MGREG} = \hat{\theta}^{GREG}. \qquad (15)$$

It is also possible to estimate the domain total by generalized regression estimators obtained by conditional minimization (2) (assuming (3)) where the set s is replaced by s_{d*} (see Rao, Molina, 2015: 19), but the resulting estimator is not approximately design-unbiased unless the domain size tends to infinity and it does not have the benchmarking property.

3. Calibration estimators – multiple waves

In the section we study the problem of estimation of the subpopulation total in longitudinal surveys. We assume, that we use not only information from the period of interest t^* but also from some k past periods: $t^* - k, t^* - k + 1, …, t^* - 1$. Hence, estimators proposed in this chapter are simple generalizations of the estimators proposed in Żądło (2011) and Żądło (2015: 219), where it is assumed that the information from all of the periods of the longitudinal survey is used.

Firstly, we propose the following generalized regression estimator of the total in the domain of interest, in the period of interest Ω_{d*t*} (compare Żądło, 2011):

$$\hat{\theta}_{d*t*}^{GREG-L} = \sum_{i \in s_{t*}} w_{st*i} a_{id*t*} y_{t*i} \,, \qquad (16)$$

where $a_{id^*t^*} = \begin{cases} 1 & for \quad i \in \Omega_{d^*t^*} \\ 0 & for \quad i \notin \Omega_{d^*t^*} \end{cases}$, and $w_{s^*t^*i}$ are obtained as the solution of:

$$\begin{cases} \sum_{i \in s_{t^*}} \frac{\left(w_{st^*i} - d_{t^*i}\right)^2}{d_{t^*i}q_{t^*i}} \rightarrow \min \\ \forall_{t \in \{t^*-k, \dots, t^*-1, t^*\}} \sum_{i \in s_{t^*}} w_{st^*i}\mathbf{x}_{ti} = \sum_{i \in \Omega_t} \mathbf{x}_{ti} \end{cases} . \tag{17}$$

Hence,

$$w_{st^*i} = g_{st^*i}d_{t^*i}, \tag{18}$$

$$g_{st^*i} = 1 + \left(col_{t^*-k \leq t \leq t^*}(\sum_{i \in \Omega_t} \mathbf{x}_{ti} - \sum_{i \in s_{t^*}} d_{t^*i}\mathbf{x}_{ti}) \right)^T \times$$

$$\times \left(\sum_{i \in s_{t^*}} d_{t^*i}q_{t^*i} \, col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{ti})\left(col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{ti})\right)^T \right)^{-1} \times$$

$$\times col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{ti})q_{t^*i}, \tag{19}$$

where $i \in s_{t^*}$, and $col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{ti}) = \begin{bmatrix} \mathbf{x}_{t^*-k\,i} & \dots & \mathbf{x}_{ti} & \dots & \mathbf{x}_{t^*i} \end{bmatrix}^T$. It must be stressed that the estimator (16) cannot be used if $s_{d^*t^*} = \varnothing$. We also assume that $\mathbf{x}_{ti}$ are known for $i \in s_{t^*}$ and $t = \{t^* - k, \dots, t^*\}$ (which is true e.g. when all of the values of the auxiliary variables are known or in the case of panel surveys) and that inclusion probabilities for the period of interest are known (in the case of complex sample designs it is possible to compute them based on simulation techniques – see e.g. Fattorini, 2006 or Gamrot, 2014). To estimate the design-variance of (16) we propose (similarly to the estimator (12)) to use (11) for the period t^* where g_{si} are replaced by (19) and e_i by:

$$e_{it^*}^{GREG-L} = a_{id^*t^*}y_{t^*i} - \left(col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{ti})\right)^T \hat{\mathbf{B}}_{d^*L}, \tag{20}$$

for $i \in s_{t^*}$, and

$$\hat{\mathbf{B}}_{d^*L} = \left(\sum_{i \in s_{t^*}} d_{t^*i}q_{t^*i} \, col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{ti})\left(col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{ti})\right)^T \right)^{-1} \times$$

$$\times \sum_{i \in s_{t^*}} d_{t^*i}q_{t^*i} \, col_{t^*-k \leq t \leq t^*}(\mathbf{x}_{tt})a_{id^*t^*}y_{t^*i}. \tag{21}$$

Secondly, we propose the following modified generalized regression estimator of the total in the domain of interest in the period of interest Ω_{d*t*} (compare Żądło, 2015: 219):

$$\hat{\theta}_{d*t*}^{MGREG-L} = \sum_{i \in s_{d*t*}} d_{t*i} y_{t*i} + \left(\sum_{i \in \Omega_{d*t*}} \mathbf{x}_{t*i} - \sum_{i \in s_{d*t*}} d_i \mathbf{x}_{t*i} \right)^T \hat{\mathbf{B}}_L , \quad (22)$$

where

$$\hat{\mathbf{B}}_L = \left(\sum_{t=t*-k}^{t*} \sum_{i \in s_t} d_{ti} q_{ti} \mathbf{x}_{ti} \mathbf{x}_{ti}^T \right)^{-1} \sum_{t=t*-k}^{t*} \sum_{i \in s_t} d_{ti} q_{ti} \mathbf{x}_{ti} y_{ti} . \quad (23)$$

Similarly to the estimation of the design-variance of the estimator (14), we propose to estimate the design variance of the estimator (22) using (9) for the period $t*$, where e_i are replaced by:

$$e_{it*}^{MGREG-L} = a_{id*t*} (y_{t*i} - \mathbf{x}_{t*i}^T \hat{\mathbf{B}}_L) . \quad (24)$$

For $s_{d*t*} = \varnothing$ estimator (22) simplifies to the synthetic estimator $\sum_{i \in \Omega_{d*t*}} \mathbf{x}_{t*i}^T \hat{\mathbf{B}}_L$,

but because of (24) it is not possible to estimate its design-variance in this case.

4. Simulation study – introduction

In sections 5 and 6 we present the results of design-based simulation studies conducted in R (R Development Core Team, 2016). The simulation study presented in section 5 is based on real data, while in section 6, generated data are used. In both simulation studies we analyze the accuracy of calibration estimators and the biases of estimators of their variances. We compute:
1) the relative bias (in %) as

$$100 \times \theta_d^{-1} B^{-1} \sum_{i=1}^{B} (\hat{\theta}_d^b - \theta_d) , \quad (25)$$

2) the ratio of mean square error of Horvitz-Thompson estimator (HT) given by $\sum_{i \in s_d} d_i y_i$ to calibration estimators given by (16) and (22) where MSE is given by

$$B^{-1} \sum_{i=1}^{B} \left(\hat{\theta}_d^b - \theta_d \right)^2 , \quad (26)$$

3) the relative biases of the estimator of the design-variance of calibration estimators as

$$100\% \cdot V^{-2} \frac{1}{B} \sum_{b=1}^{B} \left(\hat{V}_b^2 - V^2 \right),\tag{27}$$

where $\hat{V}_b^2$ is the estimator of the design-variance obtained in b-th iteration $b = 1, 2, …, B$, whereas V^2 is the simulation design-variance given by

$V^2 = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\theta}_d^b - \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_d^b \right)^2$, the number of samples drawn in the simulation

equals $B = 10.000$, θ_d is the total value in the d-th domain, $\hat{\theta}_d^b$ is the estimate of θ_d obtained in the b-th iteration ($b = 1, 2, …, B$).

5. Simulation study – real data

We use real data about Polish counties (NUTS 4). The database is taken from sources of The Central Statistical Office of Poland (https://bdl.stat.gov.pl/BDL). The study variable is value of investment outlays in enterprises and the auxiliary variable is the number of entities of the national economy. We consider data from $M = 7$ periods (years 2007–2013) describing 378 Polish counties. As a result, the data set contains $NM = 2646$ observations. Owing to missing data in some periods, Wałbrzych was omitted, and so was the capital city – Warsaw – as an outlier. We consider the problem of estimating the total value in $D = 6$ subpopulations (domains). Domains are regions in Poland under NTS nomenclature (NUTS 2). In the first period, the sample of size $n = 38$ which represents 10% of the population is drawn using simple random sampling without replacement and then the same sample is studied in other periods (the balanced panel sample).

In Table 1 we present the exact information for which estimators are calculated in the simulation study. We estimate domain totals in four periods from year 2010 up to 2013. For each year we use eight estimators. The purpose of using information on auxiliary variables from different numbers of periods (1 or 2 or 3 or 4 periods) is to check the influence of additional but autocorrelated variables on the accuracy of calibration estimators.

Our study covers properties of the following estimators:
1) GREG1 given by (12),
2) GREG2 given by (16) where $k = 1$,
3) GREG3 given by(16) where $k = 2$,
4) GREG4 given by (16) where $k = 3$,

5) MGREG1 given by (14),
6) MGREG2 given by (22) where $k = 1$,
7) MGREG3 given by (22) where $k = 2$,
8) MGREG4 given by (22) where $k = 3$.

Table 1. Information used by calibration estimators considered in the simulation study

Estimator	Estimates for			
	2010	2011	2012	2013
GREG1	x from 2010 y from 2010	x from 2011 y from 2011	x from 2012 y from 2012	x from 2013 y from 2013
GREG2	x from 2009–2010 y from 2010	x from 2010–2011 y from 2011	x from 2011–2012 y from 2012	x from 2012–2013 y from 2013
GREG3	x from 2008–2010 y from 2010	x from 2009–2011 y from 2011	x from 2010–2012 y from 2012	x from 2011–2013 y from 2013
GREG4	x from 2007–2010 y from 2010	x from 2008–2011 y from 2011	x from 2009–2012 y from 2012	x from 2010–2013 y from 2013
MGREG1	x from 2010 y from 2010	x from 2011 y from 2011	x from 2012 y from 2012	x from 2013 y from 2013
MGREG2	x from 2009–2010 y from 2009–2010	x from 2009–2010 y from 2009–2010	x from 2011–2012 y from 2011–2012	x from 2011–2012 y from 2011–2012
MGREG3	x from 2008–2010 y from 2008–2010	x from 2009–2011 y from 2009–2011	x from 2009–2011 y from 2009–2011	x from 2011–2013 y from 2011–2013
MGREG4	x from 2007–2010 y from 2007–2010	x from 2008–2011 y from 2008–2011	x from 2009–2012 y from 2009–2012	x from 2010–2013 y from 2010–2013

Source: own preparation

In Figure 1, the first column displays the relative biases (see (25)) of the generalized regression estimator of the total in each of $D = 6$ domains, and the second column displays the relative biases of the modified generalized regression estimator in each of $D = 6$ domains. For the GREG estimator there are substantially bigger biases than for the MGREG estimator. For the GREG estimator, relative biases are from –16% to 23%.

For the MGREG estimator, the relative biases are on average closer to zero than in the case of the GREG estimator, where they span from –10% to 3%. There are some outliers. The number of outliers increases with the number of past periods of the auxiliary variable used in the supporting estimation. In each case, the outlier is the bias for the north region in Poland.

In Figure 2 are presented values of the ratio of the mean square error (MSE) of the Horvitz-Thompson estimator to the MSE of the calibration estimators given by (16) and (22). The aim of this comparison is to show a decrease in the MSE of the calibration estimators in comparison to the HT estimator. In the case of GREG estimators, in a little less than a half of the cases these ratios take values greater

than 1 (grey vertical line). It means that in nearly half of the cases the GREG estimator gets smaller values for its MSE than the Horvitz-Thompson estimator.
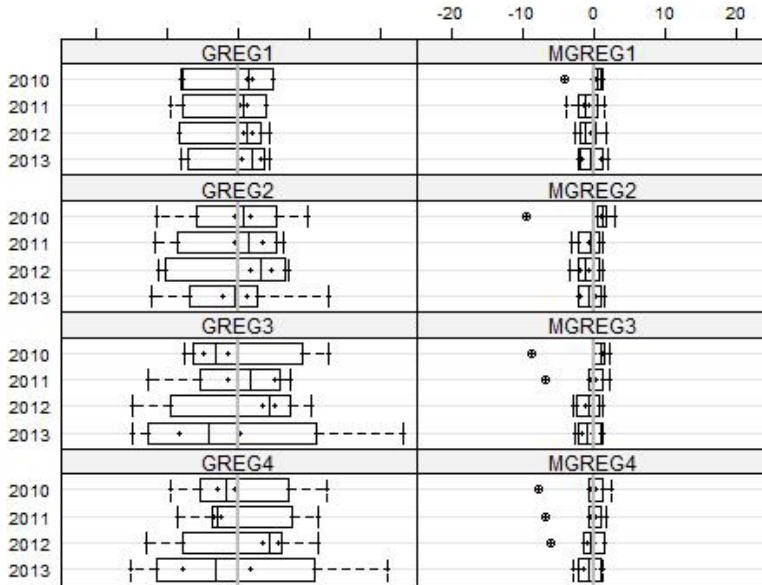

Figure 1. Relative biases of calibration estimators in % in $D = 6$ domains for years 2010–2013 for the real data

Source: own preparation

In the case of the MGREG estimator, better results are obtained. In each period, regardless of the number of periods in which the auxiliary variable was taken into consideration, ratios are larger than 1. This means that in each case the MSE of the MGREG is lower than the MSE of the HT estimator. In best cases, an astounding fifteenfold decrease in MSE was actually obtained. Using information from more periods does not significantly improve the estimation accuracy.

Figure 3 presents the relative biases of the estimators of design-variances of GREG and MGREG estimators. With regard to the biases of GREG estimators, it is visible that they are usually above zero. This means that the design-variance of GREG estimators is overestimated on average. When it comes to biases of the design-variance of MREG estimators, it turns out that they are usually below zero. This indicates that design-variance of MGREG estimators is underestimated on average.
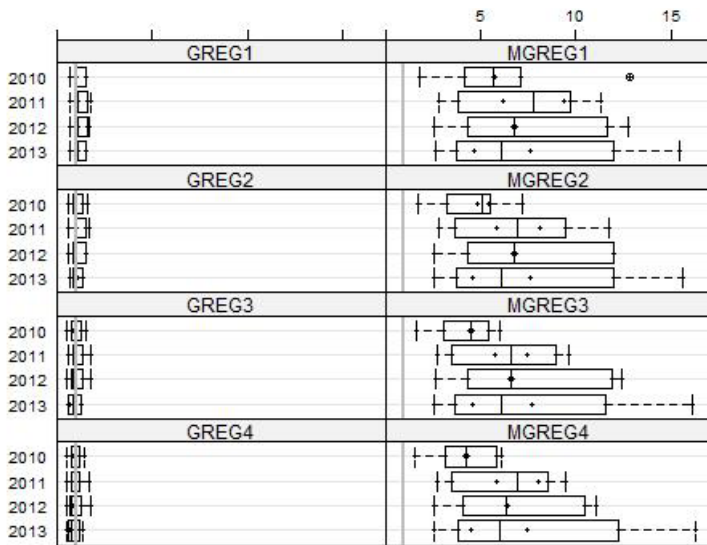
Figure 2. The ratio of the MSE of HT estimator to the MSE of studied calibration estimators in $D = 6$ domains for years 2010–2013 for the real data
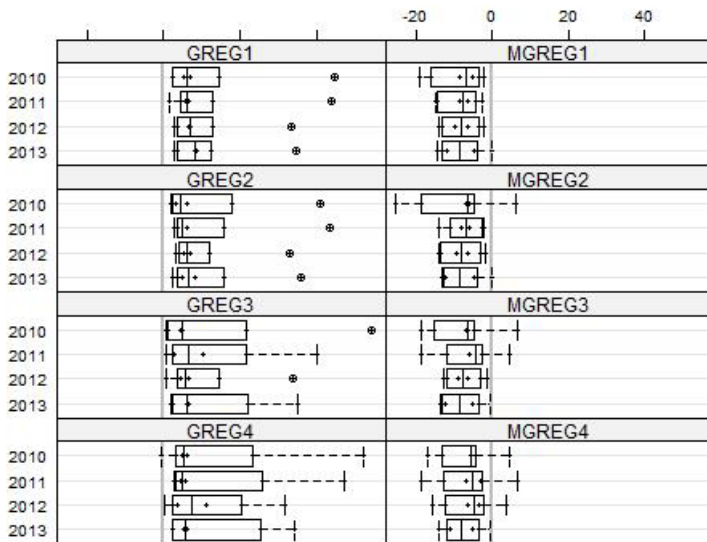
Source: own preparation

Figure 3. Relative biases in % of the design-variance of calibration estimators in $D = 6$ domains for years 2010–2013 for the real data

Source: own preparation

6. Simulation study – generated data

Artificial data are generated for design-based simulation study purposes. Generated data which mimic or modify in some sense real data are widely used for simulation purposes in economic applications (see e.g. Białek, 2014; Krzciuk, 2014). We generate artificial data in order to get data without any outliers and also to obtain a stronger relationship between studied variables than can be observed in the real data. Values of the auxiliary variable for all of the periods are generated once using normal distribution with the mean and the standard deviation of the real auxiliary variable. Values of the variable of interest are generated once as follows:

$$y_{(art)ti} = a_{(r)}x_{(art)ti} + b_{(r)} + e_{ti}, \tag{28}$$

where e_{ti} are generated once using $N(0, 0.1S_u)$, $a_{(r)}$ and $b_{(r)}$ are estimates of regression parameters computed using the least squares methods based on real values of the variable of interest and real values of the auxiliary variable, S_u is the residual standard error of the regression based on the real data.
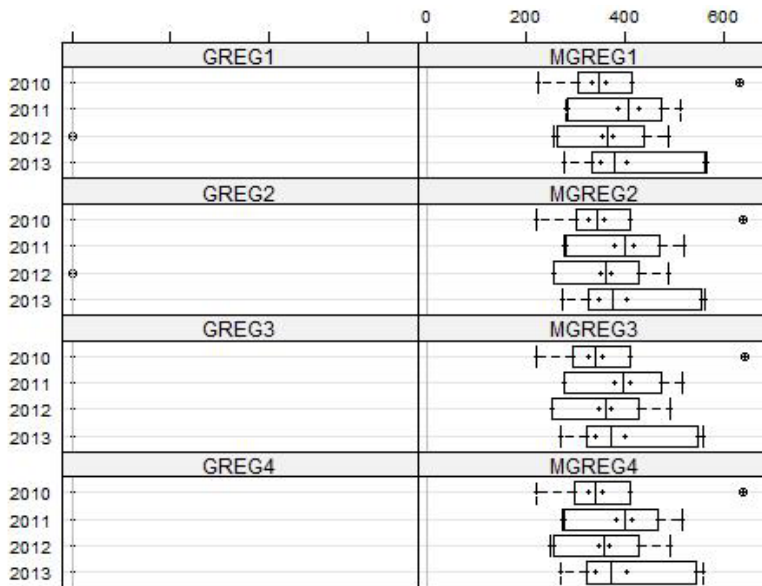

Figure 4. The ratio of the MSE of HT estimator to the MSE of studied calibration estimators in $D = 6$ domains for years 2010–2013 for the generated data

Source: own preparation

When it comes to the results obtained for the generated data, the relative biases of calibration estimators are smaller in comparison to the results obtained

in Figure 1. The relative biases for GREG1, GREG2, GREG3 and GREG4 estimators are from –4% to 6%. For MGREG1, MGREG2, MGREG3 and MGREG4 they presented much lower biases. Their absolute values are less than 1%.

Figure 4 shows ratios of the MSE of Horvitz-Thompson estimator to the MSE of calibration estimators for the generated data. Comparing results for the real and generated data it can be noticed that ratios for GREG estimators are close to 1, so it can be concluded that in the case of data with smaller diversity, the MSE for both GREG and HT estimators are quite similar. In regard to MGREG estimators, it can be observed that ratios are much higher in comparison to the real data. In this case, the use of MGREG instead of HT estimators is sometimes able to give a more than 600-fold decrease in the MSE. Similarly, as in Figure 2, the use of information from past periods results in is no significant improvement of estimation accuracy, even though in this case there is a stronger relationship between variables than in the real data.
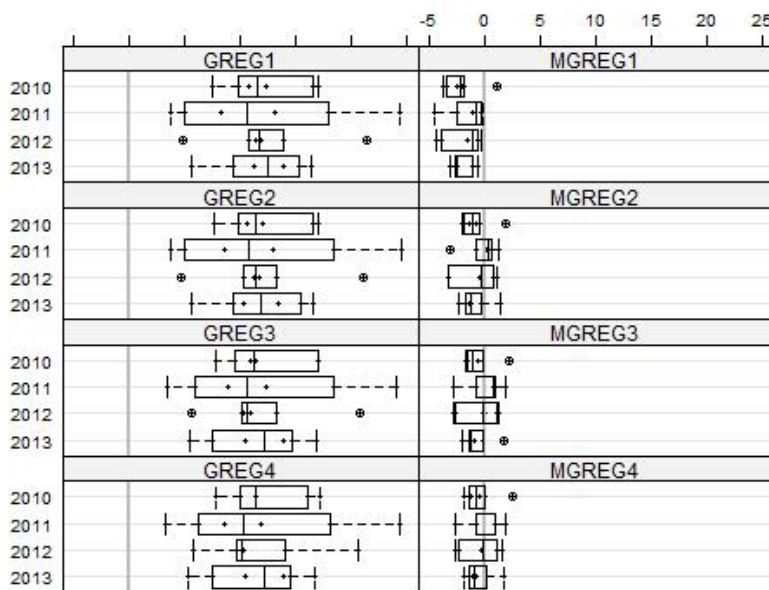

Figure 5. Relative biases in % of the estimators of design-variance of calibration estimators in $D = 6$ for years 2010–2013 for the generated data

Source: own preparation

Figure 5 shows the relative biases of the estimators of the design-variance of the considered calibration estimators which were calculated for the generated data. For GREG estimators the design variance was usually overestimated, since the obtained biases were between 3% and 25%. As regards MGREG estimators, the design vari-

ance was underestimated on average, as the relative biases were in the range of –5% and 3%. The properties of design-variance estimators for both kinds of considered calibration estimators are better for the generated data than for the real data. Nevertheless, the true cause of this improvement is not the greater relationship between the variables, but a different distribution of the auxiliary variable. This is concluded from another simulation study for which the results are not discussed in this paper.

7. Conclusion

We present some calibration estimators of a domain total in longitudinal surveys. We analyze their properties in a simulation study. In terms of the relative biases of estimation and relative biases of the design-variance, both for real and generated data the modified generalized regression estimator has better properties than the generalized regression estimator. It is also shown that the MSE of the MGREG estimator is notably smaller than the MSE of the Horvitz-Thompson estimator. It is also worth noting that using auxiliary variables from more than one period does not noticeably improve the estimation accuracy.

References

Białek J. (2014), *Simulation study of an original price index formula*, "Communications in Statistics – Simulation and Computation", vol. 43, issue 2, pp. 285–297, http://www.tandfonline.com/doi/abs/10.1080/03610918.2012.700367.

Deville J.C., Särndal C.E. (1992), *Calibration estimators in survey sampling*, "Journal of the American Statistical Association", no. 87, pp. 376–382.

Fattorini L. (2006), *Applying the Horvitz-Thompson criterion in complex designs: A computer-intensive perspective for estimating inclusion probabilities*, "Biometrika", vol. 93(2), pp. 269–278.

Gamrot W. (2014), *Estimators for the Horvitz-Thompson statistic based on some posterior distributions*, "Mathematical Population Studies", vol. 21(1), pp. 12–29.

Krzciuk M.K. (2014), On *the design accuracy of Royall's predictor of domain total for longitudinal data*, Conference Proceedings. 32nd International Conference on Mathematical Methods in Economics (MME 2014), Olomouc.

R Development Core Team (2016), *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.

Rao J.N.K., Molina I . (2015), *Small area estimation*, 2nd ed., John Wiley & Sons, Hoboken, New Jersey.

Särndal C.E. (1981), *Frameworks for Inference in Survey Sampling with Applications to Small Area Estimation and Adjustment for Nonresponse*, "Bulletin of the International Statistical Institute", no. 49, pp. 494–513.

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

Singh A.C., Mohl C.A. (1996), *Understanding calibration estimators in survey sampling*, "Survey Methodology", no. 22, pp. 107–115.

Stukel D.M., Hidiroglou M.A., Särndal C.E. (1996), *Variance estimation for calibration estimators: A comparison of jackknifing versus Taylor linearization*, "Survey Methodology", no. 22, pp. 177–125.

Żądło T. (2011), *On calibration estimators of subpopulation total for longitudinal data*, "Acta Universitatis Lodziensis. Folia Oeconomica", vol. 252, pp. 191–204.

Żądło T. (2015), *Statystyka małych obszarów w badaniach ekonomicznych. Podejście modelowe i mieszane*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice.

O estymacji kalibrowanej wspomaganej informacjami o zmiennych dodatkowych z okresów przeszłych

Streszczenie: W badaniach reprezentacyjnych nierzadko zachodzi potrzeba szacowania nie tylko parametrów populacji, ale także parametrów podpopulacji (domen). W artykule rozważany jest problem estymacji wartości globalnej w domenach. W takim przypadku może być stosowany estymator Horvitza-Thompsona. Niemniej jednak nie uwzględnia on informacji dodatkowych o elementach populacji, które zazwyczaj są dostępne. Dlatego podjęto próbę zbadania własności estymatorów kalibrowanych, w których będą wykorzystywane informacje o zmiennych dodatkowych z bieżącego oraz przeszłych okresów.

Słowa kluczowe: estymatory kalibrowane, statystyka małych obszarów, badania wielookresowe

JEL: C83