STATISTICS IN TRANSITION-new series, Summer 2012
Vol. 13, No. 2, pp. 287—300

ESTIMATION FOR SHORT TERM STATISTICS

G. Dehnel[1]

ABSTRACT

Economic processes taking places in the EU markets require systematic changes that will facilitate the international exchange of socio-economic information. All this creates a need for modification of the business statistics information system. System modification is just one element of transformations undertaken by the Central Statistical Office aimed at reforming most of its surveys. These changes focus among other things on exploiting administrative sources for purposes of public statistics. Participation of Poland in the MEETS program offered a unique chance to take steps towards a reform of the present state of business statistics. The results obtained in the MEETS program study was the basis for this article. The objective of the paper was to present the possibility of using administrative registers to short term statistics in the light of DG1 survey.

Key words: Domain estimation, Short term statistics, Small business statistics.

1. Introduction

Changes in the information system of statistics are geared towards a greater reliance on administrative data. Information stored in administrative register is expected to improve the effectiveness of statistics. The changes are designed to reduce the response burden for companies owing to statistical reporting, reduce the cost of obtaining data by making direct use of administrative data in the case of non-response, and change the composition of the statistical office staff by increasing the number of employees responsible for data analysis at the expense of those involved in data collection.

All the benefits of using administrative data listed above are especially significant when it comes to short term statistics, which requires adequate estimation methods as well as systematic and timely provision of data by its administrators. This explains the need for continued efforts to find methodological solutions that will improve the effectiveness of the short term statistics research system.

[1] Poznan University of Economics, Department of Statistics. E-mail: g.dehnel@ue.poznan.pl.

A new system should enable quick access to basic measures about, for example, business activity.

Changes should take into account the STS Recommendations issued by Eurostat (cf. Methodology term business statistics, 2006).

They concern the following areas:

1. working-day adjustment[1],
2. seasonal adjustment,
3. data transmission[2],
4. common information policy on STS data revisions,
5. publishing of STS data,
6. treatment of data that should not be published by Eurostat.

The need to change the system of public statistics to enable a greater reliance on administrative registers is a good opportunity to modify short term business statistics methodology. Such modification should involve the use of modern estimation methods, which are based on “borrowing strength” from outside the sample by relying on auxiliary variables. Including administrative data will increase the information scope of databases. Such an approach is aimed at improving estimates and extending the range of levels at which results can be presented across different variables.

Work in the area of estimation can be conducted in at least two directions:

a) studies aimed at making use of information from administrative sources for calibration estimation in an effort to deal with non-response; and

b) studies focused on making use of information from administrative sources for small area estimation methodology.

A full definition of calibration approach was formulated by C.E. Särndal (2007). According to Särndal, the calibration approach to estimation for finite populations consists in:

(a) the computation of weights that incorporate specified auxiliary information and are restrained by calibration equation(s),

(b) the use of these weights to compute linearly weighted estimates of totals and other finite population parameters: weight times variable value, summed over a set of observed units,

(c) satisfying the objective of obtaining nearly design unbiased estimates given that non-response and other non-sampling errors are absent.

---

[1] The term 'working-day adjustment' concerns both calendar and working/trading day effect adjustments. The calendar effect is related to the fact that the economic activity varies around the special periods and dates in the year while the working/trading day effect originates from the varying number of days of the week in each month.

[2] A reliable and timely transmission of STS data from the Member States to Eurostat is of critical importance for the quality of the data and it’s processing without delays. Although many improvements can be stated – among which the implementation of GESMES/TS as a common data transmission protocol is the most important one – there are still issues to be solved.

The second area of research focuses on applications of small area estimation (SAE). Using administrative data to produce business statistics does not fully solve all the problems connected with economic statistics. Distributions of units by variables of interest are strongly right skewed as well as highly differentiated and heavily concentrated. This calls for the application of new, non-traditional methods of small area estimation. They are based on a more specific approach to produce estimates at different levels across variables. SAE comprises such methods as: robust estimation, modification of GREG estimation, Kernel Regression.

2. The aim of the study

Earlier studies on the subject indicate that the short term system of business statistics can rely on data from personal income tax and corporate tax registers and the social insurance register, as well as data from the VAT register. Information about VAT and income tax should be collected directly from businesses. This information should be obtained from accounts of monthly payments (for banks) and tax returns (for the Ministry of Finance). The response burden owing to statistical reporting should be limited to a minimum.

Considering the fact that sampling survey is conducted on a monthly basis and depends on data timeliness, the available registers cannot be used as sources of up-to-date information. All enterprises have an account system, independent of statistical requirements. Thus, the ideal solution would be to collect statistical information from enterprises automatically, in the background, and transfer it to the statistical office without any extra effort on the part of the enterprise. However, this approach requires that a number of conditions are met, it entails certain technical difficulties and requires changes in legal regulations, which exceed the competence of public statistics (cf. MEETS[1] report).

One way to cope with this problem is to apply estimation methods relying on delayed register data. This approach should improve the completeness and quality of statistical data, which in turn will produce better estimates and help to correct information collected incrementally.

One of the objectives of the article is to highlight possibilities of using administrative register resources to improve estimation precision of variables describing short term business statistics. Another task is to extend the scope of estimates by taking into account estimation on a monthly basis at the low level of aggregation.

The necessity to produce estimates of business indicators for short term statistics, especially at a low level of aggregation, is a pressing need for the business sector, government and local authorities. Statistical institutes collect

[1] *Modernization of European Enterprise and Trade Statistics*, a project co-financed by the European Union (30121.2009.004-2009.807) was conducted by the Statistical Office in Poznan in cooperation with Poznan University of Economics.

information which can be used to generate business estimates. Even when register data is available, surveys often provide the only source of data in short term statistics. The question of how to estimate for short term statistics leads to the consideration of estimation methods. For present purposes, there are two of them:

- design-based methods – make use of survey weights and resulting inferences are based on the probability distribution introduced by the sampling design [cf. Rao, 2003 p.1]. The prime example of this approach is the HT estimator. In the design-based theory, although a model is used to assist estimation, no assumption is made that the population is actually defined by a model (Särndal et al., 1992, p.238-239). The approach is therefore also known as model-assisted estimation. The example of this kind of approach is the GREG estimator and synthetic estimator.
- model-based methods, where the actual finite population is regarded as a finite instance of an infinite superpopulation defined on a model incorporating distributional assumptions.

Under both approaches it is possible to strengthen estimates by using, in addition to survey data, other known data, which correlate with the variable of interest. These are known as auxiliary variables or covariates. Both design-based and model-based methods involve constructing a statistical model connecting the variable of interest to covariates. However, in both cases, models are used in different ways.

3. Assumptions of the research

The study made use of four types of estimators. Three of them represent the design-based approach: direct, GREG and regression synthetic. The fourth one – EBLUP - represents model-based methods. The EBLUP estimator is commonly used in for comparing estimator performance.

Direct estimator

The direct estimator **is** commonly used in small area estimation studies as a benchmark for comparing estimator performance.

$$\hat{\bar{Y}}_d^{DIRECT} = \frac{1}{\hat{N}_d} \sum_{i \in u_d} w_{id} y_{id} \tag{1}$$

where $\hat{N}_d = \sum_{i \in u_d} w_{id}$ $w_{id} = 1/\pi_{id}$

assuming that $\pi_{id,jd'} = 0$, for each $d \neq d'$ or $i \neq j$. Standard estimation error is calculated using the following formula:

$$M\hat{S}E(\hat{\bar{Y}}_d^{DIRECT}) = \left(\frac{1}{\hat{N}_d}\right)^2 \sum_{i \in u_d} w_{id}(w_{id} - 1)(y_{id} - \hat{\bar{Y}}_d^{DIRECT})^2 \tag{2}$$

It is characterised by high variability for most small areas; besides, its application does not guarantee estimates of the target variable for all domains – particularly with respect to cases of non-inclusion in the sample for a given domain. For this reason it is not very useful for estimation (cf. also Särndal et al. 1992, Ghosh, Rao, 1994, Lehtonen, Veijanen, 1998, Eurarea Documents: Standard Estimators, 2001).

Generalised REGression estimator – GREG

The Greg estimator is treated as a specific case of direct estimator. The direct estimator for a given small area is adjusted for differences between the sample and population area means of covariates. Auxiliary variables are transformed and adapted to the value of the target valuable. For this purpose, various models are used, which describe the relationship between the target variable Y and the auxiliary variable X. The standard approach is to use the ordinary regression model:

$$\hat{Y}_d^{GREG} = \frac{1}{\hat{N}_d}\sum_{i\in s_d}\frac{y_i}{\pi_i} + \left(\overline{\mathbf{X}}_d^T - \frac{1}{\hat{N}_d}\sum_{i\in s_d}\frac{\mathbf{x}_i}{\pi_i}\right)^T \hat{\boldsymbol{\beta}} \tag{3}$$

where $\hat{N}_d = \sum_{i\in s_d}\frac{1}{\pi_i}$ and $\hat{\boldsymbol{\beta}}$ are estimated using the least square method.

When a domain contains no data, the GREG estimator reduces to a synthetic estimator, $\overline{\mathbf{X}}_d^T\hat{\boldsymbol{\beta}}$. The formula for the MSE estimator is:

$$M\hat{S}E(\hat{\overline{Y}}_d^{GREG}) = \sum_{i\in u_d}\sum_{j\in u_d}\frac{\pi_{ijd} - \pi_{id}\pi_{jd}}{\pi_{ijd}\pi_{id}\pi_{jd}} g_{id} r_{id} g_{jd} r_{jd} \tag{4}$$

The use of the auxiliary variable X can be justified by its strong correlation with the target variable Y. In this case, the variance of the GREG estimator is lower than the variance of the direct estimator. A small sample size in a domain is conducive to an increase in variance, but with increasing correlation between variables Y and X, variance is considerably reduced. One advantage of GREG estimator is its lack of bias. Assuming that multiple samples are drawn, the expected value of the GREG estimator for a domain is close to the real value of the variable for this domain in the population.

Synthetic estimator

In synthetic estimation for a population divided into homogenous categories it is assumed that the means computed for units belonging to each category are identical. Estimation for domains is the weighted mean of estimated means determined on the basis of sampled units. The weight depends on the share of a small area within a category. The synthetic estimator is unbiased provided the assumption is met. In reality, however, this happens extremely rarely. The regression synthetic estimator is constructed on the basis of a two-level model for unit data of the variable Y, accounting for the correlation with the values of covariates X at the level of individual units and territorial units:

$$y_{id} = x_{id}^{T}\beta + u_{d} + e_{id} \tag{5}$$

where: $u_d \sim iid\ N(0,\sigma_u^2)$, $e_{id} \sim iid\ N(0,\sigma_e^2)$ described by the formula:

$$\hat{\bar{Y}}_{d}^{SYNTH} = \overline{\mathrm{X}}_{.d}^{T}\hat{\beta} \tag{6}$$

The estimator does not account for sampling weights, and MSE can be estimated using the formula:

$$M\hat{S}E(\hat{\bar{Y}}_{d}^{SYNTH}) = \hat{\sigma}_{u}^{2} + \overline{\mathbf{X}}_{.d}\hat{\mathbf{V}}\overline{\mathbf{X}}_{.d}^{T} \tag{7}$$

where $\hat{\mathrm{V}}$ is the covariance matrix of auxiliary variables.

The EBLUP estimator

Empirical Best Linear Unbiased Predictors (EBLUP) can be explained in the following manner. They are predictors for small areas, and are the best in the sense of having the least model variance; they are linear in the sense of having a linear function of the sample values y; they are unbiased in the sense of lacking model-based bias. EBLUP is a composite estimator, combining direct linear estimators and regression synthetic estimators with weights depending on the value of MSE estimators. In the case of unit-level model, EBLUP can be defined as a weighted mean of the synthetic and GREG estimators. In the area-level model, EBLUP is a weighted mean of the direct and regression synthetic estimators. The EBLUP estimator is constructed by replacing the unknown value of variance with its estimate. The general formula of the EBLUP estimator takes the following form:

$$\hat{\bar{Y}}_{d}^{EBLUP} = w_{d}^{EBLUP}\hat{\bar{Y}}_{d}^{GREG} + (1 - w_{d}^{EBLUP})\hat{\bar{Y}}_{d}^{SYNTH} \tag{8}$$

In a more developed form, the models can be described as:

$$\hat{\bar{Y}}_{d} = \gamma_{d}(\bar{y}_{.d} - \bar{\mathrm{x}}_{.d}^{\mathrm{T}}\hat{\beta}) + \overline{\mathrm{X}}_{.d}^{T}\hat{\beta} \tag{9}$$

where: $$w_{d}^{EBLUP} = \gamma_{d} = \frac{\hat{\sigma}_{u}^{2}}{\hat{\sigma}_{u}^{2} + \hat{\sigma}_{e}^{2}/n_{d}} \tag{10}$$

$\bar{y}_{.d}$ and $\bar{\mathbf{x}}_{.\mathbf{d}}^{\mathbf{T}}$ are mean sample values y and covariates for area *d* respectively, and $\hat{\beta}, \hat{\sigma}_e^2, \hat{\sigma}_u^2$ are parameters estimated on the basis of the standard linear two-level model. The MSE estimator can then be estimated using the formula:

$$M\hat{S}E(\hat{\hat{Y}}_d) = \frac{\gamma_d \hat{\sigma}_e^2}{n_d} + (1-\gamma_d)^2 \overline{X}_{.d}^T \hat{V} \overline{X}_{.d} \tag{11}$$

where $\hat{V}$ is the covariance matrix of auxiliary variables.

A description of estimators used in the study can also be found in the book by R. Chambers and A. Saei (2003). The process of estimating statistics relied on the findings of the EURAREA project[1] [cf. Eurarea_Project_Reference_Volume, 2004] and Report of the MEETS program[2].

The study focused on monthly-based statistics. Information for the study came from the survey conducted by the Statistical Office in Poznan. The survey is conducted in the form of monthly reports submitted by all large and medium-sized enterprises and a 10% sample of small enterprises. Auxiliary data came from administrative registers made available by the Central Statistical Office, the Ministry of Finance and the Social Insurance Institution.

Registers provided by the Ministry of Finance included: National Register of Taxpayers, National Register of VAT Payers, a database of Personal Income Tax (PIT) payers and Corporate Income Tax (CIT) payers. In addition, there were two other databases from the Social Insurance Institution (ZUS): the register of natural persons and the register of legal persons.

All variables from databases were used as auxiliary variables:
- value added tax from the VAT database,
- number of employees from the ZUS register,
- revenue from the PIT or CIT register,
- cost from the PIT or CIT register [3],
- profit from the PIT or CIT register.

Average revenue was the target variable. Since administrative register data were only available as an annual release, they could only be used in the model as

[1] The European project entitled EURAREA IST-2000-26290 *Enhancing Small Area Estimation Techniques to meet European needs* was part of the Fifth framework programme of the European Community for research, technological development and demonstration activities. The project was coordinated by ONS – Office for National Statistics, UK) with the participation of six countries: The United Kingdom, Finland, Sweden, Italy, Spain and Poland. Poland was represented by a team of statisticians from the Chair of Statistics at the Academy of Economics in Poznan, (http://www.statistics.gov.uk/eurarea).

[2] The MEETS program research was conducted using the EBLUPGREG program (Veijanen A., Djerf K., Sőstra K., Lehtonen R., Nissinen K., 2004, EBLUPGREG.sas, a program for small area estimation borrowing Strength Over Time and Space using Unit level model, Statistics Finland, University of Jyväskylä).

[3] *Cost* was not included in the version of the CIT database made available for purposes of the project. Therefore, to complete information coming from this database, the value of cost had to be computed using the following algorithm: Cost = Revenue – Profit + Loss.

delayed variables. Short term statistics requires a certain timeliness of data, which in effect leads to the necessity of performing estimation based on variables from periods prior to the study period.

The level of aggregation adopted for the study was economic activity classification (NACE Rev.2). Estimation was conducted for two consecutive months: *t* and *t+1* (using administrative register data with time delay).

According to the second study objective, estimation was also conducted at a lower level of aggregation. The level of aggregation adopted for the study was a combination of economic activity classification (NACE Rev.2) and the territorial division by province.

The population consisted of business units which participated in the survey sampling and for which it was possible to match data from any of the administrative registers. Estimation was conducted as a simulation study to evaluate the performances of the estimators. 1,000 samples were drawn, which were then used to estimate basic characteristics of enterprises by economic activity sections (NACE Rev.2) and province.

The analysis refers to data from the years 2008 and 2009 for which databases were made available by the Polish Central Statistical Office (CSO).

Estimation was conducted for two consecutive months: *t* and *t+1.*

4. Precision assessment methods

As a consequence of adopting a composite estimation method, the results are biased, owing to sampling error and non-sampling error. Both components are affected by non-sampling error. As for the sampling error, it can only be determined with respect to the estimation component.

For each of the estimators used in the study, expected values were computed based on results obtained in 1,000 samples to determine estimator variance, relative estimation error and relative bias. Measures of estimation precision were determined for each domain and for all domains combined. Thus, it was possible to make both a synthetic assessment of estimator properties and one that accounted for domain size and their unique characteristics. Mean values of estimator properties were estimated following 1,000 samples during the simulation study. In addition, distribution characteristics of the estimators were presented where possible.

The mean value of estimates after 1,000 samples can be calculated from:

$$\bar{\hat{Y}}_d = \frac{1}{1000}\sum_{p=1}^{1000}\hat{Y}_{dp} \tag{12}$$

Where:
d - domain
p=1, ... , 1000 – denotes the sample number;

The approximate value of the variance estimator was thus expressed as [cf. Bracha, 1994 p.33]:

$$\hat{V}(\hat{Y}_d) = \frac{1}{999}\sum_{p=1}^{1000}(\hat{Y}_{dp} - \hat{\bar{Y}}_d)^2 \tag{13}$$

The approximate value of the MSE was computed using the following formula [cf. Choudhry, Rao, 1993 p. 276]:

$$M\hat{S}E(\hat{Y}_d) = \frac{1}{999}\sum_{p=1}^{1000}(\hat{Y}_{dp} - Y_d)^2 \tag{14}$$

where Y_d denotes the known domain characteristic. Root MSE is a measure which combines variance and squared bias. Its estimate is defined on the basis of MSE:

$$\hat{S}(\hat{Y}_d)_{MSE} = \sqrt{M\hat{S}E(\hat{Y}_d)} \tag{15}$$

Relative Error of the Estimate (REE) was calculated on the basis of the value of MSE:

$$R\hat{E}E(\hat{Y}_d)_{MSE} = \frac{\sqrt{M\hat{S}E(\hat{Y}_d)}}{Y_d} \tag{16}$$

Absolute bias of the estimator was defined as the difference between the expected and real value.

$$\left|BIAS(\hat{Y}_d)\right| = \left|\hat{\bar{Y}}_d - Y_d\right| = \left|\frac{1}{1000}\sum_{p=1}^{1000}\hat{Y}_d - Y_d\right| \tag{17}$$

On the basis of the above characteristics computed for each domain it was possible to assess estimation precision for a domain, accounting for its specific nature, especially, the number of units.

5. Estimation results and assessment of their precision

<u>Estimating *revenue* by NACE sections</u>

The results of estimating *revenue* at the level of selected economic activity sections (NACE Rev.2) for period *t* are presented in Tables 1-6. Table 1 contains expected values obtained in the simulation study after 1,000 samples. The last column contains mean revenue within each section. It is used as the benchmark to assess the convergence of estimates. The actual assessment of estimation precision and bias is possible using information presented in tables 2–6.

Table 1. The expected value of estimators for revenue, period *t*

SECTION	Estimator				POPULATION MEAN
	DIRECT	GREG	SYNTHETIC	EBLUP	
Manufacturing	54585.85	54625.55	54768.17	54661.80	54576.28
Construction	34855.68	34836.24	34559.73	34703.67	34898.88
Trade	80320.49	80244.88	79884.69	80201.53	80280.19
Transport	63016.47	63255.07	63625.85	63386.54	63028.05

Source: Own tabulation based on the MEETS real dataset.

Table 2. Variance of estimators for revenue, period *t*

SECTION	Estimator			
	DIRECT	GREG	SYNTHETIC	EBLUP
Manufacturing	89293.41	37769.69	35070.17	21969.24
Construction	744019.25	70269.94	42931.32	47628.88
Trade	3042933.29	231401.36	1290654.38	271080.04
Transport	646764.21	1136134.85	56900.57	686058.41

Source: Own tabulation based on the MEETS real dataset.

Table 3. MSE of estimators for revenue, period *t*

SECTION	Estimator			
	DIRECT	GREG	SYNTHETIC	EBLUP
Manufacturing	89384.99	40198.74	71926.17	29289.69
Construction	745887.78	74198.57	158070.54	85775.56
Trade	3044559.05	232648.72	1447231.29	277272.30
Transport	646898.32	1187722.65	414625.08	814702.77

Source: Own tabulation based on the MEETS real dataset.

Table 4. Root MSE of estimators for revenue, period *t*

SECTION	Root MSE			
	DIRECT	GREG	SYNTHETIC	EBLUP
Manufacturing	298.97	200.50	268.19	171.14
Construction	863.65	272.39	397.58	292.87
Trade	1744.87	482.34	1203.01	526.57
Transport	804.30	1089.83	643.91	902.61

Source: Own tabulation based on the MEETS real dataset.

Table 5. REE of estimators for revenue, period *t*

SECTION	REE (%)			
	DIRECT	**GREG**	**SYNTHETIC**	**EBLUP**
Manufacturing	0.55	0.37	0.49	0.31
Construction	2.47	0.78	1.14	0.84
Trade	2.17	0.60	1.50	0.66
Transport	1.28	1.73	1.02	1.43

Source: Own tabulation based on the MEETS real dataset.

Table 6. Absolute bias of estimators for revenue, period *t*

SECTION	Absolute bias of estimators			
	DIRECT	**GREG**	**SYNTHETIC**	**EBLUP**
Manufacturing	9.57	49.26	191.88	85.52
Construction	43.20	62.65	339.15	195.21
Trade	40.30	35.30	395.50	78.65
Transport	11.58	227.02	597.80	358.49

Source: Own tabulation based on the MEETS real dataset.

To assess the estimation one can use REE (cf. Table 5). This measure is based on estimates of MSE, which can be compared with its known domain characteristic, thus accounting for estimation precision and bias. The GREG estimator and *EBLUP estimator* yielded similar estimates for each of NACE sections. A significant improvement in estimation precision could be observed. For *manufacturing*, where the best results were obtained, REE is at 0.3% of the known domain characteristic. The bias of the GREG estimator is considerably lower than that of the EBLUP estimator, which often yields better general results owing to its lower variance. In the case of the *transport* section, however, none of the estimators used produced better results than those obtained by means of direct estimation.

Tables 7–8 present selected characteristics of the estimation of revenue for period *t+1*. The currently available data from administrative databases, updated annually or quarterly, can be used to inform a model based on delayed auxiliary data.

Table 7. The expected value of estimators for revenue, period $t + 1$

SECTION	Expected values of estimators				POPULA-TION MEAN
	DIRECT	**GREG**	**SYNTHETIC**	**EBLUP**	
Manufacturing	4140.10	4130.69	4146.73	4131.16	4138.36
Construction	4325.28	4310.89	4111.81	4300.93	4319.91
Trade	704.06	724.35	993.19	733.68	704.23
Transport	6630.63	6636.30	6597.27	6632.67	6632.82

Source: Own tabulation based on the MEETS real dataset.

Table 8. REE of estimators for revenue, period $t + 1$

SECTION	REE (%)			
	DIRECT	GREG	SYNTHETIC	EBLUP
Manufacturing	0.79	0.47	0.51	0.45
Construction	2.51	1.71	4.88	1.70
Trade	1.64	6.32	42.57	7.14
Transport	1.03	0.90	0.66	0.82

Source: Own tabulation based on the MEETS real dataset.

Estimating *revenue* by NACE section and province

In order to analyse estimates at the level of NACE sections and province one needs to choose a method of presentation. Owing to limited space, the results are confined to the expected value of revenue for two NACE sections (for all provinces).

Table. 9. REE of estimators for revenue, in the *manufacturing* section by province

Province	REE (%)			
	DIRECT	GREG	SYNTHETIC	EBLUP
Dolnośląskie	30.19	13.23	37.09	17.00
Kujawsko-Pomorskie	39.33	25.08	32.09	17.68
Lubelskie	54.80	27.54	32.00	17.34
Lubuskie	150.60	11.81	14.12	8.29
Łódzkie	49.21	12.85	24.74	11.05
Małopolskie	32.36	16.27	25.61	12.18
Mazowieckie	36.54	53.83	47.79	45.07
Opolskie	70.01	17.84	20.21	10.58
Podkarpackie	37.93	24.66	28.29	14.25
Podlaskie	41.01	35.82	36.14	22.95
Pomorskie	39.52	34.41	24.26	16.72
Śląskie	23.77	19.77	22.46	11.76
Świętokrzyskie	64.00	23.87	26.32	16.35
Warmińsko-Mazurskie	112.50	35.42	17.89	14.72
Wielkopolskie	36.02	11.37	17.77	9.27
Zachodniopomorskie	34.42	17.83	30.30	14.31

Source: Own tabulation based on the MEETS real dataset.

Table. 10. REE of estimators for revenue, in the *construction* section by province

Province	REE (%)			
	DIRECT	**GREG**	**SYNTHETIC**	**EBLUP**
Dolnośląskie	32.09	19.79	17.02	9.25
Kujawsko-Pomorskie	40.01	15.49	23.71	14.08
Lubelskie	42.32	18.34	20.47	13.85
Lubuskie	70.40	21.34	21.93	11.31
Łódzkie	42.68	18.56	28.84	14.56
Małopolskie	53.21	14.27	22.15	12.68
Mazowieckie	54.81	20.02	13.77	9.01
Opolskie	56.66	22.50	30.17	17.60
Podkarpackie	39.10	18.79	39.15	23.01
Podlaskie	58.30	73.16	22.77	19.41
Pomorskie	91.56	19.28	24.54	18.47
Śląskie	29.52	17.92	24.65	11.71
Świętokrzyskie	136.00	34.22	29.27	25.34
Warmińsko-Mazurskie	43.70	12.70	25.19	14.78
Wielkopolskie	106.50	27.77	24.94	24.76
Zachodniopomorskie	54.24	19.28	21.37	13.22

Source: Own tabulation based on the MEETS real dataset.

Measures of precision and bias presented in Tables 9 and 10 show an evident improvement in efficiency due to the use of estimation using auxiliary data from administrative databases.

6. Conclusion

The use of data from the tax system and the system of social insurance in short term statistics of small, medium and big enterprises in the proposed range undoubtedly shall have a positive influence on the completeness and accuracy of statistical data as far as full surveys are concerned, and in the case of representative surveys it shall improve the quality of estimates. In consequence, it shall contribute to the decrease of responsibilities of enterprises resulting from statistical reporting.

A basic problem in direct use of administrative data in short term statistics of enterprises is too long period of their elaboration and distant time limits of data transfer by administrators, what makes impossible to carry out responsibilities in the scope of time-limits of providing result data to users. Cooperation with administrators aiming at minimising the period of elaboration of administrative data and shortening time limits of their transfer to statistics may be the solution to this question.

The use of information from various data sources about business activity of enterprises can also help to improve the quality of short term statistics of small, medium-sized and large enterprises. This is because these databases are a rich source of potential auxiliary variables, which can be used to increase estimation precision by reducing the negative effect of missing data in statistical reporting.

REFERENCES

BRACHA, C., 1994. Metodologiczne aspekty badania małych obszarów, „Studia i Materiały. Z Prac Zakładu Badań Statystyczno-Ekonomicznych" No. 43, CSO, Warsaw.

CHAMBERS, R., SAEI, A., 2003. Linear Mixed Model with Spatial Correlated Area Effect in Small Area Estimation.

Eurarea_Project_Reference_Volume, 2004.

GHOSH, M., RAO J.N.K., 1994. *Small Area Estimation*, *An Appraisal*, „Statistical Science", Vol. 9, No.1.

LEHTONEN, R., VEIJANEN, A., 1998. *On multinomial logistic generalized regression estimators*, Department of statistics, University of Jyväskylä, No. 22, Jyväskylä.

Report of the MEETS program, 2011. CSO, Warsaw.

Methodology term business statistics, Associated documents, Luxembourg: Office for Official Publications of the European Communities, 2006.

RAO, J.N.K., 2003. Small Area Estimation, John Wiley & Sons, Inc., Hoboken, New Jersey.

SÄRNDAL, C.E., 2007. The Calibration Approach in Survey Theory and Practice, Survey Methodology. Vol. 33. No. 2. 99–119.

SÄRNDAL, C.E., SWENSSON B., WRETMAN J., 1992. *Model Assisted Survey Sampling*, Springer Verlag, New York Inc.