SMALL AREA STATISTICS AND QUALITY MANAGEMENT – THE POLISH PERSPECTIVE

Jan Kordos

Warsaw Management University
e-mail: jan1kor2@gmail.com
ORCID: 0000-0003-2000-8467

ŚLĄSKI PRZEGLĄD STATYSTYCZNY
Nr 16(22)

ISSN 1644-6739
e-ISSN 2449-9765

DOI: 10.15611/sps.2018.16.03
JEL Classification: C18

Abstract: The author begins with a discussion of the main factors affecting the efficiency of decision-making, and in particular the quality of the information used, the models applied and the knowledge of decision-makers in the field of the implementation. First, we considered the acronym GIGO ("*garbage in garbage out*") used in computer science, expressing the impact of the relationship between the quality of the information used in the model, and their outcome, and then the Box aphorism about the usefulness of models ("*All models are wrong but some are useful*"). Particular attention is paid to the statistics of small areas, discussing their source, size, and mainly their quality, giving the current definition of data quality. Next the author provided the development of estimation methods for small areas, apart from mathematical formulas and limited to the presentation of some flowcharts. The author then discusses the role of small area statistics in quality management, focusing on the evaluation of their quality, which is of particular importance when making decisions. The paper is prepared from the Polish perspective, referring mainly to Polish statisticians and some relevant publications in English.

Keywords: efficiency of decision-making, the quality of the information, model, the acronym GIGO, statistics of small areas, quality management, sources of statistical data.

1. Introduction

The effectiveness of decisions depends significantly on the quality of the information used, the models that combine a variety of information, the results obtained from the model based on the available information and the knowledge of decision-makers about the examined issues. The quality of the information used in decision-making plays a crucial role because on the basis of inadequate data we cannot draw the correct conclusions. At the beginning it is worth to quote an aphorism often used in computer science expressed with the acronym "GIGO", which means that in the model we put "rubbish" data, also the result is "rubbish". Bluntly expressed by Box (1976)[1], stating that

[1] See: https://en.wikiquote.org/wiki/George_E._P._Box.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

"*All models are wrong, but some may be useful*." The Box aphorism is often repeated in various publications, conferences, and in the Internet we can find on this subject a lot of different opinions. Box believed that each model is only an approximation of reality, and its usefulness in decision-making depends seriously on many factors, above all on the quality of the data used. For further considerations, in general form, the current definition of the quality of statistical data is cited.

The quality of official statistics is based on the definition of the quality of the European Statistical System and determined on the basis of six quality components [Eurostat 2007, pp. 9-10]: 1) utility, 2) accuracy, 3) timeliness and punctuality, 4) comparability, 5) consistency, 6) availability and transparency. In assessing the fulfilment of the recommendations for quality there are taken into account the costs and burdens that are associated with the creation of statistical data and confidentiality issues, transparency and data security. So it is possible to see that quality is a multidimensional phenomenon, requiring special attention in applications. With the components of data quality considerations, the component of accuracy, i.e. data used in decision-making, is taken into account.

2. Increased demand for statistical information

The steady growth in demand for statistical information describing the evolution of the processes of socio-economic systems of regional and local authorities is observed in all countries. This is due to a broader use of information in decision-making and the need for the monitoring, evaluation and analysis of the socio-economic development at regional and local levels.

So far official statistics has mainly relied on information collected from censuses and sample surveys, which are used to provide certain characteristics of the target population. Censuses are expensive survey tools, which are conducted every ten years; the resulting information is often out-of-date due to delays between the censuses and released results. On the other hand, sample surveys are constructed to reach a representative group of the target population and are focused on selected socio-economic phenomena.

Nowadays, administrative sources, such as registers, are becoming increasingly important as statistical data sources. The main characteristic that makes registers different from the traditional sources mentioned above is that they were not created for statistical purposes and

statisticians were not involved in the process (see: [Wallgren, Wallgren 2014]).

The general block diagram of the data sources in official statistics is shown in Figure 1.

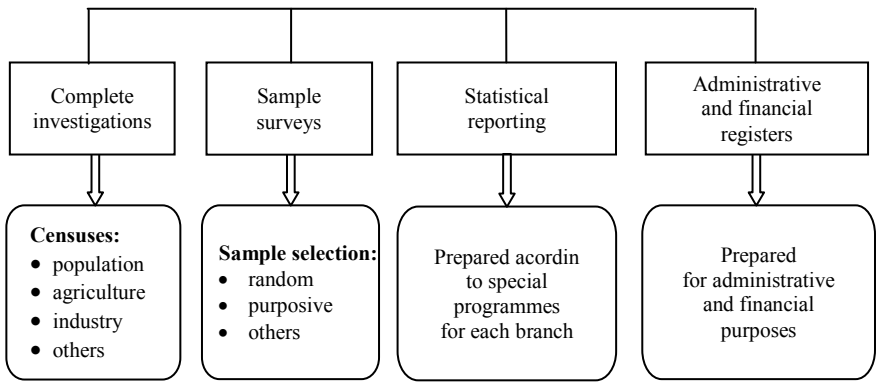

Fig. 1. Data Sources in Official Statistics

Sources: according to: http://stat.gov.pl/sprawozdawczosc.

The above data sources of the official statistics take into account the point of view of small area statistics, which focuses to some extent on data obtained from samples selected at random according to a certain probabilistic pattern, but also uses various sources of data obtained from different sources at different time and space.

Sampling survey methods (or representative methods as we still call them in Poland) are used quite extensively in the statistical offices of various countries, as well as in Poland.

When data are available from complete investigations, such as censuses, agricultural censuses, and relevant records or complete statistical reporting, the data can be developed in any cross-territorial and demographic pattern. It just depends on data processing capabilities, the timeliness of obtaining them and the available financial resources. However, complete investigations, due to financial and organizational reasons are rarely being carried out, and sample surveys are typically used based on samples selected from the general population, according to the sampling plan [Bracha 1996; Kish 1965; 1987; Yates 1980; Zasępa 1972].

In a sample survey, in which units are selected at random, the problem arises whether the obtained estimates are reliable, especially when they are based on small numbers of units. The precision of such

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

assessments is usually very low and their informative value negligible. This is typically in small areas, such as municipalities, counties, cities, and even some regions when the sample, based on which the results are estimated, is too small. This left out the problems associated with non-sampling errors, which often play an important role in the results from the sample and the complete investigations [Kordos 1988; Zarkovich 1966]. The problem arises of how to get more accurate estimates of a different break-down, based on available data from various sources to produce results useful to the user.

However, due to cost cuts and increasing non-response in sample surveys statisticians have started to search for new sources of information, such as registers, Internet data sources (IDSs, i.e. web portals) or big data [Beręsewicz 2016].

3. Development of estimation methods for small areas

In the past, however, when the necessary information was requested for small area, and the necessary data were not available, then statisticians used different kinds of assessments. Such assessments have played an important role in statistics in the world, as well as in Poland, in a variety of analyses and applications. Personally, in the late 1950s and later I worked on the construction of such assessments in the range of distributions of income population [Kordos 1959; 1963], in which I used a variety of data sources, including the results of censuses, sampling surveys of income and expenditure of the population and data from statistical reporting. In construction such assessments they used a similar approach, which now is used in the estimation for small areas. Usually they assumed a certain procedure to be followed, which combined components of data from different sources together, i.e. created a model that was capable of providing the desired assessments. However, not much attention was paid to the accuracy of those assessments, although some attempts were undertaken [Kordos 1960].

While the assessment for small areas referred to some time ago, however, a wider interest of statisticians of these problems began just over thirty years ago. They began to publish scientific and methodological papers, and started organizing international conferences and seminars which considered the various problems associated with reliability for small areas statistics, particularly the methods of estimation. It is impossible to present here, even in general terms, a list of significant items which would take dozens of pages. Therefore we

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

resort to giving available publications mainly of Polish statisticians and some publications in English, which show the development of small area estimation methods. However, it is worth mentioning the first international symposium, held at Ottawa in 1985 [Platek et al. 1987]. This publication had a significant impact on addressing these issues at the beginning of the transformation of Polish statistics [Kordos 1991]. There were also attempts to use estimation methods for small areas in the Labour Force Survey [Witkowski 1992; Szarkowski, Witkowski 1994].

In Poland, we realized quite early the difficulties that occur during the transformation of our economy in the evaluation of socio-economic systems at regional and local levels [Kordos 1991; 2004]. Complete statistical reports were severely limited, and new sampling surveys for organizational and financial resources might not be sufficient in obtaining information at regional and local levels. Therefore, in 1992, the Polish statisticians decided to organize in Warsaw, with the support of Eurostat and the international conference devoted to small areas of statistics and research design (small area statistics and survey designs). The results of that conference were published in two volumes [Kalton et al. 1993]. That conference had a significant impact on addressing the issues estimation for small areas in Poland and other countries in the transition period.

After the conference, the problems of small areas were taken up by research centres in Poland (see some authors' papers)[2]. Some Polish statisticians participated in different international conferences and projects. An important event devoted to small area was the international conference organized in September 2014 by the Department of Statistics from the Poznan University of Economics and Business, in cooperation with the Central Statistical Office of Poland. Key papers presented at that conference have recently been published in the *Statistics in Transition − new series* and *Survey Methodology*: part one[3] and part two[4].

[2] In Warsaw [Bracha 1994; 1996; 2003; Bracha et al. 2003; 2004; Kordos 1992; 1997; 1999, 2016a; 2016b), in Poznan [Dehnel 2010; Dehnel et al. 2004; Gołata 1996; 2004; 2015; Paradysz 1998; 2012; Klimanek 2012; Szymkowiak 2007; Beręsewicz 2015; 2016; Wawrowski 2014], in Lodz [Domanski, Pruska 1996; 1997; 2001; Kubacki 1997; 2003; 2004) and in Katowice [Żądło 2004; 2012; 2015].

[3] See: http://stat.gov.pl/en/sit-en/joint-issue-part-i-sae-poznan-2014/.

[4] See: http://stat.gov.pl/en/sit-en/joint-issue-part-ii-sae-poznan-2014/.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

4. Statistics of small areas and the quality assessment

It is worth to clarify what we mean by the statistical term *small area*, as it can be sometimes misunderstanding. It might seem that this is the size of the area examined, geographical or administrative. The term *small area* adopted here is more for historical reasons, but primarily concerned with such area or domain (also called *the area of study*), i.e. of the population of interest, for which, using estimation directly based on a sample, we cannot get sufficiently precise estimates. Sometimes in a given area, in general one cannot encounter any part of the sample, and information is needed for this area (or domain). Then other research methods are sought that will enable more accurate assessments.

In the classical sample surveys, in which individuals are randomly selected according to a certain sample design, the parameters are estimated on the basis of the results obtained directly from these surveys, using direct estimators. Direct estimators use the elements of the sample, but they can also use the additional data (e.g. quotient estimates or regression), depending on the availability of information from other sources.

Small area estimates are usually obtained by fitting statistical models to survey data and then applying these models to auxiliary information available for the small area population of interest. There are different possible sources of additional information for small area estimation (see Figure 2). Often a number of potential or candidate models are considered involving various combinations of the auxiliary variables. The most reliable of these candidate models is then chosen as the final model, on the basis of:

- plausibility of the model in light of previous studies or accepted wisdom;
- how well the model fits the observed data; and,
- accuracy of the small area estimates predicted from the model.

It should be stressed that, to properly judge the sample surveys, we should know not only the evaluation parameters, but also their average standard errors or confidence intervals; although in practice they are not always published together with the results. Sometimes even they are not calculated, which may mislead users in their decision-making, because they treated these assessments as accurate estimates.

In practice, statistical offices usually apply *direct estimators*, which are unbiased or approximately unbiased, according to the theory

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

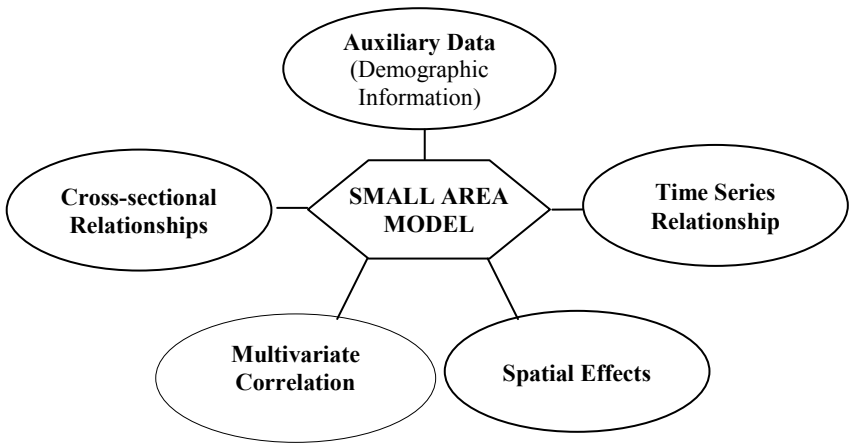

Fig. 2. Possible Sources of Additional Information

Source: adopted from the Australian Bureau of Statistics [2006], *A Guide to Small Area Estimation*, p. 14.

of sampling methods for a finite population [Bracha 1996; Zasępa 1972]. When the sample in a domain is too small to obtain reliable estimates for this domain by using a direct estimator, we must then decide whether to use an alternative procedure to obtain the desired estimates and increase their precision. The considered alternative estimates are those estimates which increase the effective sample size and reduce the variance of their estimates, which increases precision, using data from other domains and periods of time with the help of models, which assume the similarity of domains and periods of time. In extreme situations, in the domain of the sample this may not occur at all, and if the assessment must be obtained, then we will need an alternative estimate. An alternative estimator will be the *indirect estimator* (see the small area modelling framework in Figure 3).

A statistical model is a mathematical representation of the relationship we assume to exist between the variable we are interested in predicting (known as the response or dependent variable) and other associated variables (known as the auxiliary, explanatory or independent variable). A model is then fitted to data that contains observed values for both the dependent variable and the auxiliary variables for each unit. The fitting process produces estimates of the model parameters such as intercepts and slopes. The unit here may be a person, a business or a small area itself, depending upon the level at which we wish to fit the model. The model also includes one or more error terms

to describe the degree of stochastic or random variation with which predicted values for the response variable deviate from the observed values.

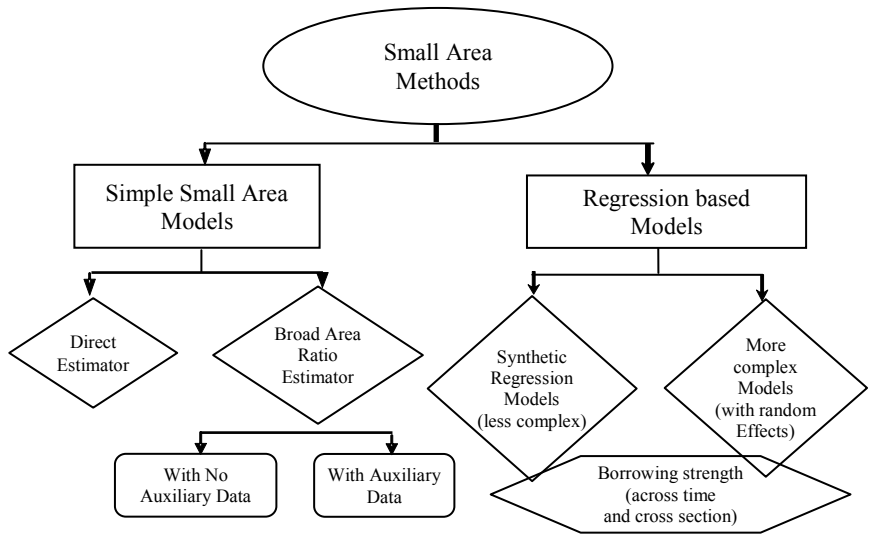

Fig. 3. Small Area Modelling Framework

Source: adopted from the Australian Bureau of Statistics [2006], *A Guide to Small Area Estimation*, p. 29.

Estimators intermediate been characterized in the Bayesian literature as estimators that are *borrowing power* by using the strength of the association studied variable with other variables in the study domain and other domains than the domain one are interested in [Ghosh 2001; Ghosh, Rao 1994; Rao 2003] Thus the indirect estimator uses, depending on the variable studied, also the information from another domain and other period. This shows that indirect estimators depend on the value of the variable domains of study and periods other than the audited. These values are taken into account in the estimation using a model which, in addition to the trivial cases, depends on one or more additional variables known to test the domain and time period. The extent to which these models can be identified, and if there are additional variables, can create indirect estimators to obtain ratings. Please note that the *availability of suitable additional data* and *model combining additional information* with respect to the variable studied *are essential* for the *formation of intermediate estimators*.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

The distinction between direct and indirect estimators reflects the situation at the design stage of the survey. However, the estimates reflect the realities regarding the survey; the distinction becomes a bit less clear, e.g. the lack of response in a survey is a common problem in the data collection process. When there is no response in a survey, even direct estimators have to rely on assumptions, sometimes naive, based on the model regarding information known to the respondents in relation to the unknown information which is not providing answers.

Potential auxiliary data should be evaluated for their relationship to the variable(s) of interest, both theoretically and statistically as well as the accuracy and reliability with which they have been collected. The theoretical relationship should emanate from tested social or economic theories. A careful examination should be made to understand any major differences between the auxiliary data and the variables of interest.

Consideration should be given to the purpose for which the data was initially collected, how was it was processed and edited, what conceptual definitions were used and what is the scope of the auxiliary data holdings. This will allow appropriate auxiliary information to be chosen to improve the model, in explaining to users what factors are driving the small area estimates and help pinpoint potential sources of error.

It should be noted that the problems of small area and the quality of the results obtained should be considered together, regardless of the source of the results obtained. The quality of the results should be tested even if they come from the complete investigations and different registers. So far on this subject we do not have adequate statistical literature. In the last few years Eurostat has undertaken a special research program on the quality of results from different sources. This is about all kinds of non-sampling errors, such as errors of coverage, completeness, accuracy, timeliness, consistency, comparability and timeliness [Beręsewicz 2016; Bethlehem et al. 2011; Szreder 2015; Szymkowiak 2009].

There are algorithms and routines that allow an estimate for the required parameters, if there is any relevant information from other sources. The development of methods for the estimation of small areas contributed significantly to modern computational techniques which allow the preparation and use of large-scale complicated calculation methods in a relatively short time. The problem is how to use small area estimation. We discuss it below (see also: [Namazi-Rad, Steel 2015]).

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

5. Sources of error in small area data

Small area estimates can be subject to a number of different sources of error depending on the way in which they are produced. There are three broad types of error that may impact upon small area estimates. These are:

1) sampling error,
2) non-sampling error and
3) model error.

The role of these three errors in the small area estimation process is shown schematically in Figure 4. The contribution of these errors to a particular small area application depends upon the method being used. The simpler small area methods such as the broad area ratio estimator and the direct survey estimator will be subject only to:

- ***Sampling and non-sampling error***

These two types of error will be familiar to anyone who has experience of the principles of survey design and estimation.

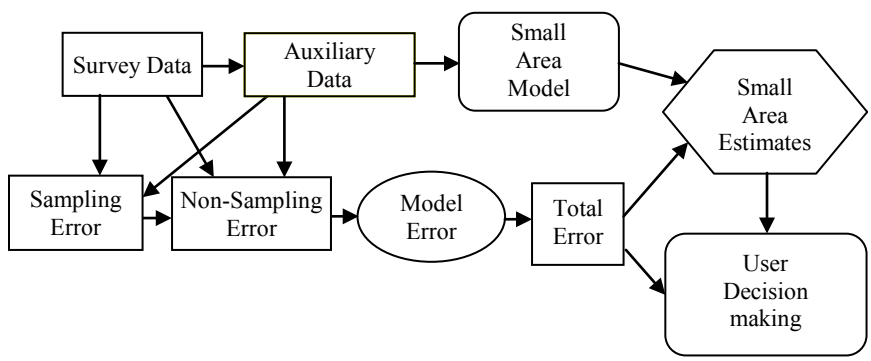

Fig. 4. Sources of Error in Small Area Data

Source: adopted from the Australian Bureau of Statistics [2006].

Model error, on the other hand, will inevitably arise in small area methods that use a statistical model to borrow strength from auxiliary data sources or other relationships in the data. The use of models involves making assumptions about relationships in the data. The suitability of the chosen models for the given data and the validity of the model in describing real world dynamics has a bearing on the nature and magnitude of the errors introduced. In assessing the reliability of small area estimates, it is therefore important to understand the nature

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

of the different sources of model error, and the extent to which they can or cannot be measured adequately, and their likely contribution to overall quality.

6. Evaluation for small areas of a decision of the users

To increase the power of the regression model, there are various functions that can be used in the model. Choosing the right model of small areas is not easy, and its success depends on a reasonable statistical evaluation and experience. Choosing which ones to use depends on the extent to which we are convinced that the resulting gain can be improved. As with any selection, there is a need to decide on certain compromises.

From current practice we may draw conclusions that there are problems with users' communication regarding the quality of the accepted results. There are several propositions to improve this practice, but here we suggested the following Trewin's proposition. Trewin [1999] encouraged NSIs to make greater use of small area estimation methods to generate statistical output. However, in doing so, he emphasised that:

a) the estimates need to be branded differently from other official statistics (the methods and the assumptions should be described in any releases);

b) their validity needs to be assessed to provide user confidence;

c) the underlying models need to be described in terms that users can understand and the validity of the underlying assumptions should be discussed with the key users;

d) their quality should be described in quantitative terms as far as possible; and there should be peer review of the models by an expert as the models are very complex and the choice of methods is considerable.

7. Compromise between quality assessments, cost, time and effort

The basic condition to obtain reliable ratings for small areas is to have good quality of additional data. Quality is of course a relative term and depends mainly on the requirements of the customer decision-maker. In order to effectively set research for small areas that meet the requirements of the user, the user should answer the following questions:

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

What are the key decisions requiring data for small areas?

1. What is the strategic context, objectives and expected results, based on which these decisions are taken and what is the expected geographic distribution?

2. What, according to the user, would be the most relevant data for small areas that meet its requirements?

3. What would be the implications for decision-makers if the data for small areas would not be adequate, say, 5%, 10%, 20% etc.? What assessment would be most suitable for quality?

4. Are there any conceptual models, both social and economic, which is believed to describe the process influenced by variables calculated for small areas?

5. Are there any available administrative data, suitable as additional information that could be used for modeling assessments small areas? How were these data collected, for what purpose and whether it seem equally accurate?

6. Will the data be subject to further disaggregation and according to what category?

7. Was there some previously conducted research to make the right decisions for which there is required assessment for small areas.

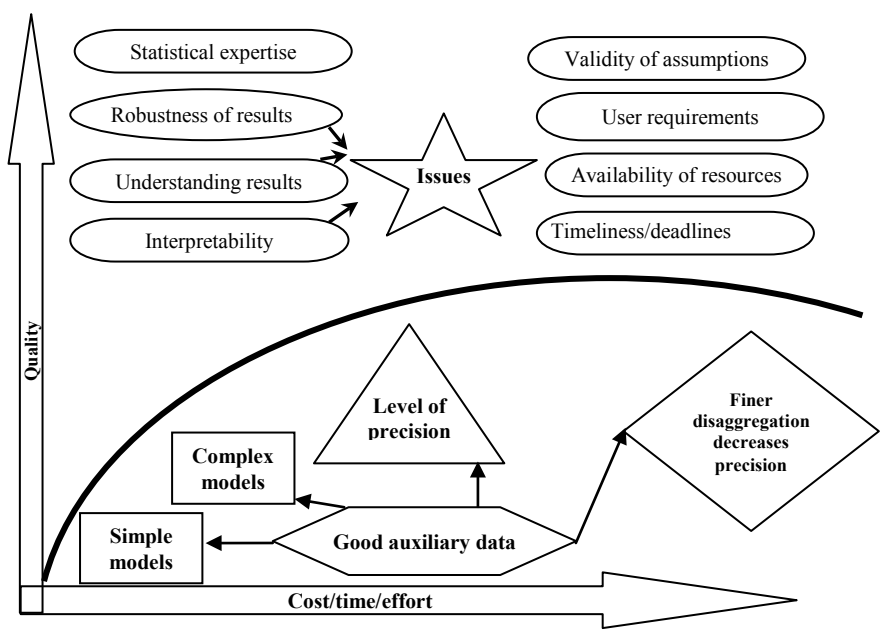

Fig. 5. Trade-off between Quality and Cost/Time/ Effort

Source: adopted from: Australian Bureau of Statistics [2006], A Guide to Small Area Estimation, p. 34.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

Assuming that there are additional data of sufficient quality, then one would expect the use of more complex estimation methods to obtain results of a higher quality. There are also other issues, as shown in Figure 5, which may be important in relation to the relationship quality results of small areas and the cost of obtaining them.

For example, the use of more complex models may require a higher level of knowledge and expertise to assist in the understanding and interpretation of the results of the models. Such knowledge is also important to verify the validity of the assumptions made in the model and check the resistance and sensitivity of the model results.

Finally, there are important points to consider in the relationship issues of quality and costs discussed above. First of all, simplicity is an important aspect of quality in that it helps in the interpretation of the small areas. It should not be inferred from Figure 5 that the simpler methods always lead to lower quality results. More complex methods must be used to obtain significant benefits of quality assessments.

8. Final remarks

It seems reasonable to give some recommendations and suggestions compiled from different papers, conferences and projects related to SAE methods (see: [Kordos 2016]):

1. Good auxiliary information related to the variables of interest plays a vital role in model-based estimation. Expanded access to auxiliary data, such as census and administrative data, through coordination and cooperation among federal agencies is needed.
2. Preventive measures at the design stage may significantly reduce the need for indirect estimators.
3. Model selection and checking plays an important role. External evaluations are also desirable whenever possible.
4. Area-level models have wider scope because area-level data are more readily available. But the assumption of known sampling variance is restrictive.
5. Model-based estimates of area totals and means are not suitable if the objective is to identify areas with extreme population values or to identify areas that fall below or above some pre-specified level.
6. Suitable benchmarking is desirable.
7. Model-based estimates should be distinguished clearly from direct estimates. Errors in small area estimates may be more transparent to users than errors in large area estimates.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

8. A proper criterion for assessing the quality of model-based estimates is whether they are sufficiently accurate for the intended uses. Even if they are better than direct estimates, they may not be sufficiently accurate to be acceptable.

9. An overall program should be developed that covers issues related to sample design and data development, organization and dissemination, in addition to those pertaining to methods of estimation for small areas.

International practice shows that statistical data for small areas may be more widely used by users, provided that:

a) the assessment is presented in a different way than other statistics (the methods and assumptions to be described);

b) their accuracy is evaluated and explained to inspire user confidence;

c) the used models are described in an understandable way, and the validity of the assumptions made understandable for the user;

d) if possible, the quality of the evaluation should be described in terms of quantity.

Small area estimates are often used by program administrators to determine or benchmark their funding allocations. Without small area information, the administrators have difficulty in assessing the actual need for goods and services in each area. This can result in undesirable scenarios such as "the squeaky wheel gets the grease", whereby interest groups or areas which are most vocal receive a greater share of the funding allocations. Small area estimates provide detailed information on each area allowing for objective and informed decision making.

Local government demand for small area data has also increased as they become increasingly aware and interested in the role statistics can play in informing them about what is happening in their own jurisdictions.

Small area estimates should only be produced when there is strong and justified user demand as well as no alternate data at the small area level that will serve the required purpose. In addition there needs to be adequate survey and auxiliary data to ensure that the outputs produced will be of sufficient quality to fit their intended purpose.

We have to agree with Baesens [2007; 2014], that the best way to increase the efficiency of the analytical model is not to seek fantastic tools and techniques, but we must first improve data quality. In many cases simple analytical models are useful, depending mainly on the data.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

Acknowledgements

The author would like to thank Dr. M.E. Beręsewicz, from the Poznan University of Economics and Business, for his comments, and the Australian Bureau of Statistics, for permission to use *A Guide to Small Area Estimation* in this paper.

References

ABS (Australian Bureau of Statistics), 2006, *A Guide to Small Area Estimation* – Version 1.1. Internal ABS document.

Baesens B., 2007, *It's the Data, You Stupid!,* Data News.

Baesens B., 2014, *Analytics in a Big Data World: The Essential Guide to Data Science and its Applications*, Wiley and SAS Business Series.

Beresewicz M.E., 2015, *On representativeness of internet data sources for real estate market in Poland*, Aust. J. of Stat., 44, 2, pp. 45-57.

Beręsewicz M., 2016, *Internet data sources for real estate market statistics*, https://berenz.github.io/assets/phd/Beresewicz_Maciej_dissertation.pdf.

Beręsewicz M., Klimanek T., 2013, *Wykorzystanie estymacji pośredniej uwzględniającej korelację przestrzenną w badaniach cen mieszkań*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 279, pp. 281-290.

Bethlehem J., Cobben S., Schouten B., 2011, Handbook of Nonresponse in Household Surveys, Wiley.

Box G.E.P., 1976, Science and Statistics, Journal of the American Statistical Association, vol. 71, no. 356, pp. 791-799.

Bracha Cz., 1994, *Metodologiczne aspekty badania małych obszarów*, Z prac Zakładu Badań Statystyczno-Ekonomicznych, z. 43.

Bracha Cz., 1996, *Teoretyczne podstawy badań reprezentacyjnych,* PWN, Warszawa.

Bracha Cz., 2003, *Estymacja danych z Badania Aktywności Ekonomicznej Ludności na poziomie powiatów dla lat 1995-2002*, GUS, Warszawa.

Bracha Cz., Lednicki B., Wieczorkowski R., 2003, *Estimation of Data from the Polish Labour Force Surveys by poviats (counties) in 1995—2002* (in Polish), Central Statistical Office of Poland, Warsaw.

Bracha Cz., Lednicki B., Wieczorkowski R., 2004, *Wykorzystanie złożonych metod estymacji do dezagregacji danych z Badania Aktywności Ekonomicznej Ludności w roku 2003*, Studia i Prace – Z Prac Zakładu Badań Statystyczno-Ekonomicznych GUS i PAN, z. 299, Warszawa.

Brakel J.A. van den Bethlehem J., 2008, *Model-Based Estimation for Official Statistics*, Statistics Netherlands, Voorburg/Heerlen.

Cochran W.G., 1977, *Sampling Techniques*, 3rd ed., Wiley, New York.

Dehnel G., 2010, *The development of micro-entrepreneurship in Poland in the light of the estimation for small areas,* Publ. University of Economics in Poznan, Poznan.

Dehnel G., Golata E., Klimanek T., 2004, *Consideration on Optimal Design for Small Area Estimation*, Statistics in Transition, vol. 6, no. 5, pp. 725-754.

Domański Cz., Pruska K., 1996, *Reprezentatywność próby w statystyce małych obszarów,* Wiadomości Statystyczne, nr 5, pp.11-16.

Domański Cz., Pruska K., 1997, *Prognozowanie w przedsiębiorstwie z wykorzystaniem statystyki małych obszarów*, [in:] M. Cieślak (red.), *Prognozowanie w zarządzaniu firmą*, materiały konferencyjne, Akademia Ekonomiczna, Wrocław.

Domański Cz., Pruska K., 2001, *Metody statystyki małych obszarów*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Eurostat, 2007, *Handbook on Data Quality Assessment: Methods and Tools*, Luxembourg.

Ghosh M., 2001, *Model-Dependent Small Area Estimation – Theory and Practice*, [in:] *Lectures Notes on Estimation for Population Domains and Small Areas*, eds. R. Lehtonen, K. Djerf, „Reviews" no. 5, Statistics Finland, University of Jyväskylä.

Ghosh M., Rao J.N.K.,1994, *Small area estimation: An appraisal*, Statistical Science, 9, pp. 55-93.

Gołata E., 1996, *Statystyka małych obszarów w analizie rynku*, Wiadomości Statystyczne, nr 3, pp. 45-59.

Gołata E., 2004a, *Estymacja pośrednia bezrobocia na lokalnym rynku pracy,* Wydawnictwo Akademii Ekonomicznej, Poznań (praca habilitacyjna).

Gołata E., 2004b, *Problems of estimate unemployment for small domains in Poland*, Statistics in Transition, 6, 5, pp. 755-776.

Gołata E., 2012, *Data integration and small domain estimation in Poland – experiences and problems*, Statistics in Transition – New Series, 13(1), pp.107-142.

Gołata E., 2015, *SAE education challenges to academic and NSI,* Statistics in Transition New Series and Survey Methodology*,* vol. 16, no. 4, s. 611-630.

Heady P., Hennell S., 2001, *Enhancing small area estimation techniques to meet European needs*, Statistics in Transition*,* 5, 2, pp. 195-203.

Hidiroglou M.A., 2014, *Small-Area Estimation: Theory and Practi*ce, Section on Survey Research Methods, Statistics Canada,

Kalton G., Kordos J., Platek R., 1993, *Small Area Statistics and Survey Designs,* vol. I: *Invited Papers*, vol. II: *Contributed Papers and Panel Discussion,* Central Statistical Office, Warsaw.

Kish L., 1965, *Survey Sampling*, New York.

Kish L., 1987, *Statistical Design for Research,* John Wiley and Sons, New York.

Klimanek T., 2012, *Using indirect estimation with spatial autocorrelation in social surveys in Poland*, Przegląd Statystyczny, 59 (numer specjalny 1), pp.155-172.

Kordos J., 1959, *Szacunek rozkładu ludności według grup zamożności*, Wiadomości Statystyczne, nr 3.

Kordos J., 1960, *Próba określenia dokładności szacunków*, Wiadomości Statystyczne, nr 3.

Kordos J., 1963, *Rozkład ludności pozarolniczej według wysokości dochodów na osobę w 1960 r.,* Biuletyn Komitetu Przestrzennego Zagospodarowania Kraju PAN, nr 8.

Kordos J., 1988, *Jakość danych statystycznych*, PWE, Warszawa.

Kordos J., 1991, *Statystyka małych obszarów a badania reprezentacyjne*, Wiadomości Statystyczne, nr 4, pp. 1-5.

Kordos J., 1992, *Podejście do statystyki małych obszarów w Polsce*, Wiadomości Statystyczne, nr 10, pp. 1-5.

Kordos J., 1997, *Efektywne wykorzystanie statystyki małych obszarów*, Wiadomości Statystyczne, nr 1, pp. 11-19.

Kordos J., 1999, *Problemy estymacji danych dla małych obszarów*, Wiadomości Statystyczne, nr 1, pp. 85-101.

Kordos J., 2001, *Nowy projekt zastosowania estymacji dla małych obszarów*, Wiadomości Statystyczne, nr 8, pp.1-10.

Kordos J., 2004, *Metody estymacji dla małych obszarów w badaniach procesów społeczno-ekonomicznych*, Roczniki Kolegium Analiz Ekonomicznych, Szkoła Główna Handlowa, 13, Warszawa, pp. 11- 29.

Kordos J., 2005, *Some Aspects of Small Area Estimation and Data Quality*, Statistics in Transition, 7, pp. 63-83.

Kordos J., 2016a, *Small Area Statistics and Quality Management*, Zarządzanie. Teoria i Praktyka, 15(1), pp. 25-34.

Kordos J., 2016b, *Development of Small Area Estimation in Official Statistics*, Statistics in Transition New Series and Survey Methodology, 17, 1, pp. 105-132.

Kubacki J., 1997, *Ważniejsze metody estymacji w statystyce małych obszarów*, Wiadomości Statystyczne, nr 5, pp. 13-21.

Kubacki J., 2004, *Application of the Hierarchical Bayes Estimation to the Polish Labour Force Survey*, Statistics in Transition, 6, 5, pp. 785-796.

Kubacki J., 2006, *Remarks on using the Polish LFS data for unemployment estimation by County*, Statistics in Transition, 7, 4, pp. 901-916.

Longford N., 2005, *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician*, Springer.

Marker D.A., 2001, *Producing small area estimates from national surveys: Methods for minimizing use of indirect estimators*, Survey Methodology, 27, pp. 183-188.

Namazi-Rad M.-R., Steel D., 2015, *What level of statistical model should we use in small area estimation?,* Australian & New Zealand Journal of Statistics, 57, pp. 275-298. http://onlinelibrary.wiley.com/doi/10.1111/anzs.12115/abstract.

Paradysz J., 1998*, Small Area Statistics in Poland – First Experiences and Application Possibilities*, Statistics in Transition, 3, 5, pp. 1003-1015.

Paradysz J., 2012, *Statystyka regionalna: stan, problemy i kierunki rozwoju*, Przegląd Statystyczny, nr 2, pp. 191-204.

Platek R., Rao J.N.K., Särndal C.E., Singh M.P. (eds), 1987, *Small Area Statistics,* John Wiley & Sons, New York.

Rao J.N.K., 2003, *Small Area Estimation*, John Wiley & Sons, New Jersey.

Szarkowski A., Witkowski J., 1994, *The Polish labour force survey,* Statistics in Transition, 1, 4, pp. 467-483.

Szreder M., 2015, *Big data wyzwaniem dla człowieka i statystyki*, Wiadomości Statystyczne, nr 8, pp. 1-11.

Szymkowiak M., 2009*, Estymatory kalibracyjne w badaniu budżetów gospodarstw domowych,* rozprawa doktorska, http://www.wbc.poznan.pl/dlibra/docmetadata?id=115816.

Szymkowiak M., 2010, *ESSnet on Small Area Estimation. Report on the analysis of questionnaires used in WP 2*, October 2010.

Szymkowiak M., Klimanek T., 2012, *Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 242, pp. 601-609.

Trewin D., 1999, *Small area statistics conference*, Survey Statistician, 41, pp. 8-9.

Trewin D., 2002, *The importance of a quality culture*, Survey Methodology, 28, 2, pp. 125-133.

UN, 2011, *Using Administrative and Secondary Sources for Official Statistics – A Handbook of Principles and Practices*, United Nations Commission for Europe.

Wallgren A., Wallgren B., 2014, *Register-based Statistics*. 2th ed, Wiley, New York.

Wawrowski Ł., 2014, *Wykorzystanie metod statystyki małych obszarów do tworzenia map ubóstwa w Polsce*, Wiadomości Statystyczne, nr 9, pp. 46-56.

Witkowski J., 1992, *Szacowanie bezrobocia dla małych obszarów*, Wiadomości Statystyczne, nr 11, pp. 1-5.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

Yates F., 1980, *Sampling Methods for Censuses and Surveys*,4th ed., London.

You Y., Zhou, Q.M., 2011, *Hierarchical Bayes small area estimation under a spatial model with application to health survey data,* Survey Methodology, 37, 1, pp. 25-37.

Zarkovich S.S., 1966, *Quality of Statistical Data*, FAO, Rome.

Zasępa R., 1972, *Metoda reprezentacyjna*, PWE, Warszawa.

Żądło T., 2004, *On Unbiasedness of Some EBLU Predictor*, [in:] Antoch J., ed. *Proceedings in Computational Satistics 2004*, Physica-Verlag, Hidelberg.

ŻądłoT., 2006, *On prediction of total value in incompletely specified domains*, Aust. NZ. J. Stat., 48, pp. 269-283.

Żądło T., 2009, *On MSE of EBLUP*, Stat. Papers, 50, pp. 101-118.

Żądło T., 2012, *O predykcji wartości globalnej w domenie z wykorzystaniem informacji o zmiennych dodatkowych przy założeniu modelu Faya-Herriota*, Acta Iniversitis Lodziensis, Folia Oeconomica, 271, pp. 243-256.

Żądło T., 2014, *On the prediction of the subpopulation total based on spatially correlated longitudinal data*, Mathematical Population Studies, special issue: Survey Sampling Methods, eds M. Ghosh, T. Żądło, 21, 1, pp. 30-44.

Żądło T., 2015a, *On longitudinal moving average model for prediction of subpopulation total*, Statistical Papers, 56 (3), pp.749-771.

Żądło T., 2015b, *On prediction for correlated domains in longitudinal surveys*, Communications in Statistics – Theory and Methods, 44(4), pp. 683-697.

Żądło T., 2015c, *Statystyka małych obszarów w badaniach ekonomicznych – podejście modelowe i mieszane*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.

STATYSTYKA MAŁYCH OBSZARÓW I JAKOŚĆ ZARZĄDZANIA – POLSKA PERSPEKTYWA

Streszczenie: Autor rozpoczyna od omówienia głównych czynników wpływających na efektywność procesu decyzyjnego, a przede wszystkim na jakość użytych informacji, zastosowanych modeli oraz wiedzę decydentów w zakresie wdrażania. Po pierwsze, GIGO jest używany w informatyce jako akronim („śmieci włożone, śmieci wyjdą"), wyrażający wpływ relacji między jakością informacji wykorzystywanej w modelu a ich wynikiem, a następnie aforyzm Boksa na temat przydatności modeli („Wszystkie modele są złe, ale niektóre są przydatne"). Szczególną uwagę zwraca się na statystykę małych obszarów, omawiając ich źródło, wielkość, a przede wszystkim ich jakość, podając aktualną definicję jakości danych. Następnie autor przedstawia rozwój metod estymacji małych obszarów, pomijając formuły matematyczne, a ogranicza się do prezentacji niektórych schematów blokowych. Omawia też rolę statystyki małych obszarów w zarządzaniu jakością, koncentrując się na ocenie ich jakości, co ma szczególne znaczenie przy podejmowaniu decyzji. Opracowanie przygotowano z polskiej perspektywy, odnosząc się głównie do polskich statystów i niektórych istotnych publikacji w języku angielskim.

Słowa kluczowe: efektywność podejmowania decyzji, jakość informacji, model, akronim GIGPO, statystyka małych obszarów, jakość zarządzania, źródła danych statystycznych.