This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Synthetic Estimators for Small Area Estimation

Contents

General section

1. Summary

In surveys conducted by statistical offices one of the main problem is to have reliable estimates for domains for which the sample size is too small or even equal to zero. It is the consequence of the fact that many institutions need more detailed information not only for the whole country but also for some specific subdomains such as geographic areas or other cross-sections. It also concerns business statistics where increasing demand exists for information for different classification of activities (e.g., trade, manufacturing, transport, construction, etc.) including small, medium and large enterprises and many variables (e.g., revenue, operating costs, taxes, etc.). In such situations direct estimates based only on specific domain sample data were insufficient because of high variability and small precision. The remedy could be the methodology of small area estimation (SAE) which plays an important role in the field of modern information provision, which aims to cut survey costs while lowering the respondent burden. Thanks to their properties, SAE methods enable reliable estimation at lower levels of spatial aggregation and with more specific domains, where direct estimation techniques display too much estimator variance. Another advantage over direct estimators is that small area estimation can be used to handle cases with few or no observations for a given domain in the sample. Therefore it is necessary in many situations to use indirect estimates that borrow strength by taking into account values of the variables of interest from related areas and from that point of view increasing the "effective" sample size. Generally speaking there are basically two types of indirect estimators: the synthetic and the composite estimators which can be derived under a design-based approach or taking into account the fact that an explicit area level or unit level model exists. The main aim of module is to provide a set of principles for synthetic estimators. Information about the first group of estimators can be found in the module "Weighting and Estimation – Composite Estimators for Small Area Estimation".

2. General description of the method

One of indirect estimators is the synthetic estimator, which relies on a properly chosen model. Such a model takes into account auxiliary information from different sources, such as sample survey data, census data or administrative records, in other words it "borrows strength" to improve the process of estimation. Modeling in this area involves making use of implicit or explicit statistical models to indirectly estimate small area parameters of interest. The traditional synthetic estimators rely on an implicit linking model. In this case synthetic estimators for small areas are derived from direct estimators for a large area that covers some small areas under the assumption that the small areas have the same characteristics as the large area. In other words, an estimator is called a synthetic estimator if a reliable direct estimator for a large area, covering several small areas, is used to derive an indirect estimator for a small area under the assumption that the small areas have the same characteristics as the large area, see Rao (2003).

Recently explicit linking models have come to play a more important role in the literature on small area estimation and have brought significant improvements in techniques of indirect estimation. Drawing on mixed model methodology, these techniques incorporate random effects into the model. Random effects account for the between-area variation that cannot be explained by including auxiliary variables, see Mukhopadhyay and McDowell (2011). A broader discussion of synthetic estimators derived only fromlinear mixed models can be found later in this module. It is worth noting that model-based estimators can also be derived from linear models without taking into account specific area effects. For more details, see Rao (2003).

It should be mentioned that there is a compromise between direct estimators and synthetic estimators. It relies on the fact that when the sample size for a specific domain is small, direct estimators have large variance and small precision but low or no bias. On the other hand, for the same specific domain, a synthetic estimator is often biased, especially if the above assumption is not fulfilled, but is better than a direct estimator from the point of view of precision, i.e., the variance of this estimator is smaller.This compromise is used by the so-called composite estimators which will be discussed in detailin the module "Weighting and Estimation – Composite Estimators for Small Area Estimation".

As it was stated above synthetic estimators can be considered from design and model-based perspectives. Synthetic design-based estimators make use of survey weights $d_i = 1/\pi_i$, which are based on the probability distribution and depend on the specific sample design, i.e., first order inclusion probabilities are used $\pi_i = P(i \in s)$ but they also take into account information about the domain under study (for instance information about the population area size for domain d or a known total value for the variable X for the d-th small area/domain can be used).Synthetic model-based estimators make use of a properly chosen statistical model that "borrows strength" in making an estimate for one small area from sample survey data collected in other small areas.

Later in this document the most common synthetic estimators (both design and model-based) for the total value of study variable Y in the d-th domain will be introduced. Various formulas for different synthetic estimators will be shown. All these formulas are based on taking into account the more reliable Horvitz-Thompson direct estimator for the broad area and domain and use it to construct an estimator for the small area/domain. The review of the estimators is based on ESSnet Project on Small Area Estimation (2012b) and Rao (2003).

One of the simplest synthetic estimators is the so-called BARE (Broad Area Ratio Estimator) which takes into account only additional information about the population area size N_d. The formula for the BARE estimator is as follows:

$$\hat{Y}_{d,BARE} = N_d \frac{\sum_{i \in s} d_i y_i}{\sum_{i \in s} d_i} = N_d \frac{\hat{Y}_{BA}}{\hat{N}_{BA}} \ , \tag{1}$$

where N_d is the population area size for domain d, s denotes the sample, d_i is the initial weight associated with the i-th unit in the sample andy_i is the value of the target variable for this unit, $d = 1, \dots, D$. This formula states that the total value for the variable under study y for the large area is proportionally allocated in all small areas according to the population area sizes N_d.

If domain-specific auxiliary information is available in the form of$k$-vector of known totals $\boldsymbol{X}_d^T$, then as estimator for the domain total Y_d the so-called the regression-synthetic estimator can be taken into account. The formula below is appropriate for any model that has been used to derive the parameter$\hat{\boldsymbol{\beta}}$ which is the regression coefficient based on data from a broad area (the whole country or the entire region):

$$\hat{Y}_{d,GRS} = \boldsymbol{X}_d^T \hat{\boldsymbol{\beta}}. \tag{2}$$

The $\widehat{\boldsymbol{\beta}}$ regression coefficient is the solution of the sample weighted least squares equations and its formula can be found in Rao (2003).

The next simple estimator in the class of synthetic estimators is the so-called ratio-synthetic estimator. It takes into account a broad area survey estimate $\hat{Y}_{BA} = \sum_{i\in s} d_i y_i$and can be used when the value of a single auxiliary variable X is available in the form of a total value for each small area from another source. The formula for this estimator is given by:

$$\hat{Y}_{d,RSE} = X_d \frac{\sum_{i\in s} d_i y_i}{\sum_{i\in s} d_i x_i} = X_d \frac{\hat{Y}_{BA}}{\hat{X}_{BA}} \ , \tag{3}$$

where X_d is the known total value for the variable X for the d-th small area/domain and$\hat{X}_{BA} = \sum_{i\in s} d_i x_i$ is the direct survey estimate of the total of the only one auxiliary variable at the broad area.

In some situations “good” direct estimates (acceptable precision) for the broad area are also known for cross-classification of respondents, e.g., sex, age groups, place of residence for social surveys or ownership form in business statistics. In such cases, if population sizes or auxiliary variable totals are known for all cross-classifications for specific small areas, it is possible to construct an appropriate synthetic estimator which is called a post-stratified estimator. In this approach classification counts play the role of poststrata.

When population sizes N_{dg} for cross-classification g in small area d are known it is possible to constructa so-called count-synthetic estimator $\hat{Y}_{d,CSE}$ given by the formula:

$$\hat{Y}_{d,CSE} = \sum_g N_{dg} \frac{\hat{Y}_g}{\hat{N}_g} \ , \tag{4}$$

where $\hat{Y}_g$is the direct survey national estimate of the variable under study for cross-classification cell g,$\hat{N}_g$ is the direct survey national estimate of the national population size for cross-classification cell g, N_{dg} is the known population size for cross-classification g in small area d and g denotes the cross-classifications of poststrata, e.g.,g=1 to 16 can represent firms according to size of the firm (four variants: micro, small, medium and large size) and section (also four variants: trade, manufacturing, transport, construction).

In the case when the total value of a single auxiliary variable X is known at cross-classifications for each small area and is measured in the survey, a so-called combined ratio-synthetic estimator $\hat{Y}_{d,CRSE}$ can be constructed. Its formula is given by:

$$\hat{Y}_{d,CSE} = \sum_g X_{dg} \frac{\hat{Y}_g}{\hat{X}_g} \ , \tag{5}$$

where $\hat{X}_g$is the direct survey national estimate of the auxiliary variable for cross-classification cell g and X_{dg} is the known value of the auxiliary variable for cross classification cell g of the small area d.

Estimators discussed above were derived under a design-based approach and make use of design weights during the process of estimation.What follows below is a discussion of synthetic estimators obtained assuming that an explicit area level or unit level model exists.

One very important class of synthetic estimators are those which are based on linear mixed models, see Inglese, Russo and Russo (2008), Eurarea (2004). The first synthetic estimator, called Synth A in the EURAREA project, is given by Eurarea (2004):

$$\hat{Y}_d^{Synth\,A} = \mathbf{X}_d^T \hat{\boldsymbol{\beta}}^{unit}, \tag{6}$$

where $\mathbf{X}_d^T$ is a vector of area level k covariates of known population totals and which is based on a unit level mixed model:

$$\mathrm{y}_{di} = \mathbf{x}_{di}^T \boldsymbol{\beta} + \mathrm{u}_d + \mathrm{e}_{di}, \tag{7}$$

where $\mathrm{u}_d \sim iid\ N(0, \sigma_u^2)$, $\mathrm{e}_{di} \sim iid\ N(0, \sigma_e^2)$, $\mathbf{x}_{di}$ is the vector of k covariates relates to the i-th unit within area d, $\boldsymbol{\beta}$is the $(k \times 1)$vector of the model coefficients, u_d is the random area effect associated with small area d, and e_{di} is the unit level random error. In most cases, the variances σ_u^2 and $\sigma_e^2$are unknown so they have to be estimated using for example the restricted maximum likelihood method (REML). Other possibilities of estimation also exist. For example, Proc Mixed in SAS can be used to calculate ML or REML estimates of σ_u^2 and σ_e^2. For details, see Rao (2003). The weighted least squares estimator for the model coefficients $\boldsymbol{\beta}$of size $(k \times 1)$is given by:

$$\hat{\boldsymbol{\beta}}^{unit} = \left(\sum_{d=1}^{D} \mathbf{x}_d^T \hat{\mathbf{V}}_d^{-1} \mathbf{x}_d\right)^{-1} \left(\sum_{d=1}^{D} \mathbf{x}_d^T \hat{\mathbf{V}}_d^{-1} \mathbf{y}_d\right) \tag{8}$$

where $\mathbf{x}_d$ is an $(n_d \times k)$ matrix of values of the k covariates related to area d, $\mathbf{y}_d$ is the $(n_d \times 1)$ vector of the target variable and $\hat{\mathbf{V}}_d$ is the estimated variance-covariance matrix of the vector $\mathbf{y}_d$ given by the formula:

$$\hat{\mathbf{V}}_d = \hat{\sigma}_e^2 \mathbf{I}_{n_d} + \hat{\sigma}_u^2 \mathbf{1}_{n_d} \mathbf{1}_{n_d}^T, \tag{9}$$

where$\mathbf{I}_{n_d}$ and $\mathbf{1}_{n_d}$ denote an identity matrix of dimension n_d and an n_d-dimensional vector of 1s respectively and $\hat{\sigma}_e^2$, $\hat{\sigma}_u^2$ are estimates of the variance components σ_e^2 and σ_u^2 respectively.

The second synthetic estimator, called Synth B in EURAREA project, is given by Eurarea (2004):

$$\hat{Y}_d^{Synth\,B} = \mathbf{X}_d^T \hat{\boldsymbol{\beta}}^{area} \tag{10}$$

and it is based on an area level mixed model with auxiliary variables available at area level and random area-specific effects and errors independently normally distributed:

$$\mathrm{y}_d = \mathbf{X}_d^T \boldsymbol{\beta} + \mathrm{u}_d + \mathrm{e}_d, \tag{11}$$

where$\mathrm{u}_d \sim iid\ N(0, \sigma_u^2)$, $\mathrm{e}_d \sim iid\ N(0, \sigma_e^2)$, y_d is the value for the target variable in d-th area and e_d is the area level random error. The weighted least squares estimator for the model coefficients $\boldsymbol{\beta}$of size $(k \times 1)$is given by:

$$\hat{\boldsymbol{\beta}}^{area} = \left(\sum_{d=1}^{D} \mathbf{X}_d \mathbf{X}_d^T / (\hat{\sigma}_u^2 + \hat{\sigma}_e^2)\right)^{-1} \left(\sum_{d=1}^{D} \mathbf{X}_d \hat{Y}_d / (\hat{\sigma}_u^2 + \hat{\sigma}_e^2)\right). \tag{12}$$

Estimation of variance and MSE of synthetic estimators is described in details in Rao (2003). Generally speaking, the estimation process is different for variance and MSE in the design-based and model-based situation. The variance of the design-based synthetic estimators will be small compared with the variance of a direct estimator because of the fact that it depends only on the precision of direct estimators at a larger area level. From this point of view the variance of design-based synthetic estimators is estimated using standard design-based methods. For example the variance of the ratio-synthetic estimator or of the count-synthetic estimator can be estimated using the Taylor linearisation method. Similarly the variance of the regression-synthetic estimator can be estimated using resampling methods such as the jackknife. A broad discussion of the variance and the MSE of estimators under the design-based approach can be found inSärndal et al. (1992) and Rao (2003).

For synthetic model-based estimators the problem of constructing an MSE is more complicated than in the case of design-based estimators because it depends on the underlying model which can be very complex. A discussion devoted to MSE estimation in the model-based approach can be found in a report prepared during Essnet Project on Small Area Estimation (2012c). More details about MSE estimation in the model-based approach can also be found in the monograph by Rao (2003) and in Jiang and Lahiri (2006). Formulas for the MSE for Synth A and Synth B estimators can also be found in Inglese, Russo and Russo (2008) and Eurarea (2004).

3. Preparatory phase

4. Examples – not tool specific

We refer to Rao (2003) for many examples of applying synthetic estimators in real surveys including health variables, county estimates of wheat production in the state of Kansas with evaluation, etc.The article written by Marker (1999) contains also a broad discussion of synthetic estimators with many examples presented in detail.

5. Examples – tool specific

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Boonstra, H. J. andBuelens, B. (2011), *Model-based estimation*.Contribution to the Methods Series, Statistics Netherlands, The Hague.

Essnet Project on Small Area Estimation (2012a), *Report on Workpackage 3 – Quality Assessment*, Final Version, March 2012.

Essnet Project on Small Area Estimation (2012b), *Report on Workpackage 4 – Software Tools*, Final Version.

Essnet Project on Small Area Estimation (2012c), *Report on Workpackage 6 – Guidelines*, Final Version, March 2012.

EURAREA (2004),*Project Reference Volume*.http://www.statistics.gov.uk/eurarea

Fabrizi, E., Rosaria, M., andPacei, S. (2007), Small Area Estimation of Average Household Income Based on Unit Level Models for Panel Data.*Survey Methodology***33**, 187–198.

Ghosh, M. andRao, J.N.K. (1994), Small Area Estimation: An Appraisal.*Statistical Science***9**, 55–76.

Gonzalez, J.F., Placek, P.J., and Scott, C. (1996), Synthetic Estimation in Followback Surveys at the National Center for Health Statistics.In: W.L. Schaible (ed.),*Indirect Estimators in U.S. Federal Programs*.

Inglese, F., Russo, M., and Russo, A. (2008),Different approaches for evaluation precision Small Area Model-Based Estimators.*Proceedings of Q2008, European Conference on Quality in Official Statistics*.

Jiang, J. andLahiri, P. (2006), Mixed model prediction and small area estimation.*TEST***15**, 1–96.

Marker, D.A. (1999), Organization of Small Area Estimators Using a Generalized Linear Regression Framework.*Journal of Official Statistics***15**, 1–24.

MEETS (2011), *Use of Administrative Data for Business Statistics*, Final Report, Poznań (Poland) 2011.

Mukhopadhyay, P. K. andMcDowell, A. (2011), *Small Area Estimation for Survey Data Analysis Using SASSoftware*. SAS Institute Inc., Cary, NC.

Rao, J.N. K. (2003),*Small area estimation*. John Wiley & Sons, Hoboken, New Jersey.

SAE package developers (2007), *Introduction to Small Area Estimation*. www.bias-project.org.uk/software/SAE.pdf

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

Stasny, E., Goel, P.K., and Rumsey, D.J. (1991), County Estimates of Wheat Production.*Survey Methodology***17**, 211–225.

Specific section

8. Purpose of the method

The method is used for small area estimation and comprises some techniques (including the case with no and auxiliary information) used for estimation when the sampling size in the domain of interest is too small to obtain reliable estimates using a direct estimator. The purpose of this method is to provide acceptable estimates for small areas when using direct estimators is impossible (no units in the sample for specific domain) or there are only some units in particular small areas. Synthetic estimation is based on the concept of "borrowing strength" and uses both survey and auxiliary data from outside as well as within the domain/small area of interest. As a consequence using additional sources of information in synthetic estimators generally leads to higher precision and, if the key assumption of homogeneity within the larger domain is fulfilled, to reduction of bias.

9. Recommended use of the method

1. These estimators can be applied for estimation when sample data are not available for the domain of interest, since the only required informationis local covariate means or totals and the value of $\widehat{\boldsymbol{\beta}}$, which is based on data from the entire region, or country, covered by the survey.
2. Synthetic estimators can be used even when sampling was not involved. It is especially important in business statistics where units are taken into a sample not always according to an appropriate sampling scheme. For example, we can use the synthetic approach when information about the mean or total value of y is known from some administrative source and means or totals of covariates are also known for the domain of interest and at the level of the population.
3. This class of indirect estimators should be recommended in all surveys when information for small areas/domains is needed mainly because of its simplicity, applicability to general sampling designs or surveys where a sample design is not present, and potential of increased accuracy in estimation by borrowing "strength" from similar small areas.

10. Possible disadvantagesof the method

1. If the assumption that small areas have the same characteristics as the large area is not fulfilled, then estimates may not be appropriate. Such an assumption is quite strong, and in fact for some areas or domains, synthetic estimators can be heavily biased in the design-based framework, see Ghosh and Rao (1994).
2. When one wants to use synthetic estimators for small areas, it is very important that good auxiliary information is available.
3. When one wants to use synthetic estimators for population totals or means in small areas, it is very important to take possible selection effects into consideration as far as possible. Selection effects may cause systematic differences in the target variable between sample and population. In synthetic estimators based on a model and used for the entire population it may be less

useful to predict the non-observed part of the population and this may lead to huge bias, see Boonstra and Buelens (2011).

4. For some synthetic estimators, the estimates $\hat{Y}_d$ for small areas do not add up to the direct large area estimate $\hat{Y}$. In such cases adjustment is needed in order to ensure coherence of estimates at different levels. A potential solution is to use the following formula:

$$\hat{Y}_{d,adj} = \frac{\hat{Y}_d}{\sum_d \hat{Y}_d} \hat{Y}. \tag{13}$$

A detailed discussion of several adjusting methods can also be found in the topic "Macro-Integration".

11. Variants of the method

1. Variants of the method depend on the availability (or not) of auxiliary information. For example in the situation where the only available additional information is the population area sizes, the broad area ratio estimator can be used. If domain-specific auxiliary information is available in the form of known totals then the regression-synthetic estimator is a good solution.
2. Variants of the method depend also on how synthetic estimators were derived: under a design-based approach or taking into account the fact that an explicit area or a unit level model exists.

12. Input data

1. Input data set can be classified according to the type of synthetic estimator. For example, for the BARE, ratio-synthetic estimator, count-synthetic estimator and combined ratio-synthetic estimator, information about the design weights d_i for all units in the sample s is required as well as about the values of the target variable y_i and the auxiliary variable x_i (e.g., for the ratio-synthetic estimator). Depending on the synthetic estimator, as mentioned above, information is required about known population sizes N_d in the domain d, population sizes N_{dg} for cross-classification g in small area d, known total value X_d for the variable X for d-th small area or known value X_{dg} of the auxiliary variable for cross classification cell g of the small area d. This information can come from a census or administrative registers. In the case of the model-based synthetic estimators, information about known totals $\boldsymbol{X}_d^T$ of auxiliary variables for all small areas is needed.

13. Logical preconditions

1. Missing values
   1. When an area contains no data in the sample, synthetic estimators may be used. This is one very important advantage of synthetic estimators compared especially to direct estimators.
2. Erroneous values
   1. Standard small area methods do not take into consideration errors in auxiliary variables. A possible misspecification of the area level variables or correction in the variables is not taken into account.

3. Other quality related preconditions

   1.

4. Other types of preconditions

   1.

14. Tuning parameters

1. The tuning parameters of synthetic estimators should be specified only for synthetic model-based estimators. Parameters for the convergence of the iterative method for such estimators may be:the maximum number of iterations,convergence criterion. Details of macros in SAS for unit and area level synthetic estimators are described in the Eurarea documentation (2004). Some functions are also available in R and put on the SAE page ate the Cross portal, see Essnet Project on Small Area Estimation (2012b).

15. Recommended use of the individual variants of the method

1. When a domain in the sample is not represented at all or there are only a few sampled units in specific domains (using direct estimators is doubtful due to the large variance), synthetic design-based estimators should be taken into consideration at first because they are easy to implement and easy to understand by the recipient of statistical information.
2. In some cases basic synthetic estimators based on the design-based approach may give unacceptable results (e.g., when there are too many empty domains). In such situations, the model-based approach may be taken into account especially when additional covariates and/or correlations can be included in a model. This approach can be used both when the target variable y is quantitative (linear regression can be used) or categorical (logistic regression can be used).
3. If the auxiliary information used for synthetic estimators is not very predictive for the target variable, then predicted area means are pulled too much towards the general sample average. In this situation small area methods based on models with random area effects are more suitable, see Boonstra and Buelens (2011).

16. Output data

1. In many examples devoted to synthetic estimators, which can be found in the literature when the true value is known (simulation studies), an output dataset usually contains a table with the following information: estimates for a small area, variance of the estimator, MSE, confidence intervals and bias. In real applications, when the true value is not known, the output data set usually is poorer and consists of estimates for all small areas, the variance of the estimator or model-based MSE.

17. Properties of the output data

1. The user should check the quality of estimates based on their knowledge of the investigated phenomenon and the variance of the estimators. In simulation studies also MSE, bias of estimates and confidence intervals may be checked, see ESSnet Project on Small Area Estimation (2012a).

18. Unit of input data suitable for the method

Processing unit level data and domain level variables for computations of the synthetic estimator (area level Synth B) and its variance.

19. User interaction - not tool specific

1. Select the model (no models, unit-level model, area-level model), choose auxiliary variables to be included into the model.
2. Establish the level of aggregation.
3. In the case of synth A and synth B establish tuning parameters (convergence criteria, starting point, stopping point).
4. After the use of synthetic estimators quality indicators should be checked and verified in order to evaluate the final results (variance, MSE, interval confidence).

20. Logging indicators

1. The specific logging indicators depend on the type of synthetic estimator. It can include run time of the application and /or number of iterations to reach convergence in the estimation process and characteristics of the input data.

21. Quality indicators of the output data

1. Variance of the estimator (both in real and simulation studies).
2. MSE, bias and confidence intervals – usually in simulation studies when the real value of the parameter is known, see ESSnet Project on Small Area Estimation (2012a).

22. Actual use of the method

1. The method is applied in a wide range in the U.S. Federal Statistical System. It should be mentioned that in fact some of these applications had an experimental nature and were not disseminated in official statistics. For example synthetic estimators are used by The National Center for Health Statistics in United States and in agricultural surveys, see Gonzalez, Placek and Scott (1996), Stasny, Goel and Rumsey (1991). Synthetic estimators are also used by Statistics Canada to estimate some characteristics of the labour market. This technique is also used to estimate average household income, see Fabrizi, Rosaria and Pacei (2007). In business statistics synthetic estimators are rather used in a very limited range. Some applications can be found in the MEETS project, of which the main goal was to highlight possibilities of using administrative data resources for purposes of estimating enterprise indicators and the resulting benefits. In this project, small area estimators, including synthetic, were implemented to estimate some characteristics (revenue, number of employees, wages) according to short-term and annual statistics of medium and large sized enterprises. For details, see MEETS (2011).In the literature, however, it was pointed out that synthetic estimators may be applied to replace design-based methods to estimate population totals when a known random sample design is not present. It may, for instance, concern estimation based on incomplete registers, of which the VAT turnover register is an example, see Boonstra and Buelens (2011).

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Weighting and Estimation – Small Area Estimation
2. Macro-Integration – Main Module

24. Related methods described in other modules

1. Weighting and Estimation – Composite Estimators for Small Area Estimation
2. Weighting and Estimation – EBLUP Area Level for Small Area Estimation (Fay-Herriot)
3. Weighting and Estimation – EBLUP Unit Level for Small Area Estimation
4. Weighting and Estimation – Small Area Estimation Methods for Time Series Data

25. Mathematical techniques used by the method described in this module

1. For design-based synthetic estimators basic knowledge of linear algebra is needed. For model-based synthetic estimators the knowledge of iterative methods is required.

26. GSBPM phases where the method described in this module is used

1. 5.6 Calculate weights
2. 5.7 Calculate aggregates

27. Tools that implement the method described in this module

1. A review of the available small area estimators routines and software can be found in Essnet Project on Small Area Estimation (2012b, c). It covers such statistical programs as R (packages sae2, arm, mass lme4, MCMCglmm, INLA and many others are mentioned), SAS (including proc MIXED and proc IML), STATA, SPSSS, MLwiN and WinBUGS. Some specific information devoted to synthetic estimators, their application and implementation can also be found in some other documents.

   - Eurarea project http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/spatial-analysis-and-modelling/eurarea/index.html
   - R package sae2 which is not available on CRAN and can be downloaded from the website of the BIAS project: http://www.bias-project.org.uk/, see SAE package developers (2007). Information about R packages can also be found in Essnet Project on Small Area Estimation (2012b).

28. Process step performed by the method

Estimation of parameters in disaggregated domains

Administrative section

29. Module code

Weighting and Estimation-M-Synthetic Estimators for SAE

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	10-02-2012	first version	Marcin Szymkowiak, Tomasz Józefowski	GUS (Poland)
0.2	14-01-2013	second version	Marcin Szymkowiak, Tomasz Józefowski	GUS (Poland)
0.3	31-01-2014	third version	Marcin Szymkowiak, Tomasz Józefowski	GUS (Poland)
0.4	14-03-2014	fourth version	Marcin Szymkowiak, Tomasz Józefowski	GUS (Poland)
0.4.1	19-03-2014	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

31. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	26-3-2014 13:34