SAE2005

Conference

Challenges in Statistics Production for Domains and Small Areas

Scientific Conference Featuring Main Results of, and Issues Raised by, the EURAREA Project

PROCEEDINGS

University of Jyväskylä
28-31 August 2005, Jyväskylä, Finland

SAE2005 Conference
28-31 August 2005, Jyväskylä, Finland

Organizers

University of Jyväskylä
Statistics Finland
The EURAREA Consortium

Scientific Committee

Prof. Ray Chambers, University of Southampton (Chair)
Prof. Risto Lehtonen, University of Jyväskylä (Secretary)
Dr Patrick Heady, Office for National Statistics (EURAREA project coordinator)
Dr Timo Alanko, Statistics Finland
Dr Stefano Falorsi, Italian National Statistical Institute (ISTAT)
Dr Dan Hedlin, Statistics Sweden
Mrs. Montserrat Herrador, National Statistical Institute of Spain (INE)
Prof. Jan Kordos, Warsaw School of Economics
Prof. Aldo Russo, University of Roma Tre
Dr Li-Chun Zhang, Statistics Norway

Organizing Committee

Prof. Risto Lehtonen, University of Jyväskylä (Chair)
Mr. Kari Nissinen, University of Jyväskylä (Secretary)
Dr Timo Alanko, Statistics Finland
Mr. Kari Djerf, Statistics Finland
Mrs. Sari Eronen, University of Jyväskylä
Mr. Mikko Myrskylä, Statistics Finland

Sponsors

University of Jyväskylä
Statistics Finland
SAS Institute
Baltic-Nordic Network in Survey Sampling
City of Jyväskylä
International Association of Survey Statisticians IASS

CONTENTS

Foreword *by Risto Lehtonen* ……...……... 5
EURAREA: an overview of the project and its findings *by Patrick Heady and Martin Ralphs:* ...…….. 9

Keynote Papers

Chris Elbers, Jenny Lanjouw and Peter Lanjouw: Poverty mapping and extensions ….……..……21
Mike Hidiroglou and Marie Cruddas: Developing small area statistics for business surveys ……..……22
Danny Pfeffermann:
Small area estimation using time series models subject to benchmarking constraints ……..……23
Small area estimation under informative sampling . ..……... 24
J.N.K. Rao: Small area estimation: overview, new developments and practical issues ……25
Carl-Erik Särndal: Reliable statistics for subpopulations - some unresolved issues ……..…..26

Invited and Contributed Papers

Michele D'Alò, Loredana Di Consiglio, Stefano Falorsi and Fabrizio Solari: Small area estimation of the Italian poverty rate ……..…..... 29
Esmail Amiri: Bayesian study of small area racial disparities in heart disease mortality in Ghazvin province (Iran) ...……..…30
Julia Aru: Notes on sample covariance matrix under informative sampling ……..... 31
Ray Chambers and Nikos Tzavidis: Bias adjusted distribution estimation for small areas ……... 32
Hukum Chandra and Ray Chambers: Comparing EBLUP and C-EBLUP for small area estimation ……..… 33
Coro Chasco-Yrigoyen: Ecological inference: a new approach based on spatial econometrics ……... 34
Philip Clarke, Fernando Moura and Danny Pfeffermann: Small area estimation with varying area boundaries by low level hierarchical modelling using the synthetic estimator …….…... 35
Emanuela Conza: Small area estimation or simulation by using training images: the advent of multiple-point statistics …….……36
Gauri Datta: Composite estimation of small area means using Fay-Herriot model ……..…... 37
Grazyna Dehnel and Elzbieta Gołata: Attempts to estimate basic information for small business in Poland …….…38
Kari Djerf: EBLUP estimator: comparison of the prediction using true population information with sample level information ……39
Enrico Fabrizi, Maria Rosaria Ferrante and Silvia Pacei: Estimation of poverty indicators at the sub-national level using univariate and multivariate small area models ...…… 40
Piero Demetrio Falorsi, Stefano Falorsi, Paolo Righi and Fabrizio Solari: Sampling designs for small domains estimation through multi-way stratification techniques ..……..……42
Wojciech Gamrot: Estimation of a domain mean under nonresponse using double sampling ..……..…. 43
W. González-Manteiga, M.J. Lombardía, I. Molina, D. Morales and L. Santamaría: Bootstrap approximations of the mean squared error of empirical predictors ..……..…… 44
S.J. Haslett and G. Jones: Small area estimation of poverty and malnutrition in Bangladesh: some practical and statistical issues ……...45
Natalja Jurevitš: The bias-corrected regression estimator …… 46

Jan Kordos: Impact of the EURAREA project on research in small area estimation in Poland ……...47
Danutė Krapavickaitė: Income estimation for small sample size ……..…… 48
Nicholas T. Longford: On standard errors of model-based small-area estimators ……..……..…. 49
A.F. Militino, M.D. Ugarte and T. Goicoa: Combining sampling and model weights in agriculture small area estimation ..……..…… 50
Michal Mladý: Regional labour market statistics at a European level - small number of survey respondents ……..…..……. 51
Mikko Myrskylä: Study on the performance of four variance estimators for logistic GREG estimator for domains ……..…… 52
Kari Nissinen: EBLUP estimation of small area totals for unit-level panel data ……... 53
Haritz Olaeta: Small area estimations in the industrial survey of the Basque country ……54
Erkki Pahkinen: Education of experts in small area statistics …...……..…..…. 55
Jan Paradysz and Tomasz Klimanek: Adaptation of EURAREA experience in business statistics in Poland ..…….…….…….…….…….…….…….…….…… 56
Alessandra Petrucci, Monica Pratesi and Nicola Salvati: Geographic information in small area estimation. Small area models with spatially correlated random area effects. ……..…… 57
Krystyna Pruska: Logistic regression models in small area investigations ……..…..59
Cristina Rueda and José A. Menéndez: A restricted model approach to improve the precision of estimators …….…….…….…….…….…….…… 60
Ayoub Saei, Li-Chun Zhang and Ray Chambers: Generalized structure preserving estimation models for small areas …….…….…….…….……..…… 62
Jorge Saralegui, Montserrat Herrador, Domingo Morales and Agustín Pérez: Small area estimation in the Spanish Labour Force Survey …….…….…….…… 63
Kaja Sõstra: General restriction estimator in small area estimation …….…… 64
Marja Tammilehto-Luode: Register-based statistics and geographic information ……65
Nicola Torelli and Matilde Trevisani: Small area estimation by combining spatially misaligned data ……..…… 66
Ari Veijanen, Risto Lehtonen and Carl-Erik Särndal: The effect of model quality on model-assisted and model-dependent estimators of totals and class frequencies for domains …….…… 67
Tomasz Żądło: On mean square error of EBLU predictors based on the formula of Royall's BLU predictor …….…….…….…….…… 69
Li-Chun Zhang and Ib Thomsen: A prediction approach to sampling design …….……..70

Annex

Conference Program ……..…….…….…….…….…….…… 73
List of Participants …….…….…….…….…….…….…… 76

Foreword

This proceedings publication includes the abstracts of papers submitted for presentation in the SAE2005 Conference – Challenges in Statistics Production for Domains and Small Areas – organized in 28–31 August 2005 at University of Jyväskylä, Finland. The variety of topics and approaches presented in the papers reflects nicely the wide coverage of methods available for statistics production for regional areas and other population subgroups or domains. Over 50 papers were submitted featuring the state-of art of small area and domain estimation methodology and practice and, also importantly, manifesting the increasing demand in the society for regional and domain statistics.

The results of the EURAREA project – Enhancing small area estimation techniques to meet European needs – are discussed in many papers. The project was conducted in 2001–2004 by a consortium of six national statistical agencies and five universities from different European countries and was funded by European Union. The use of data from previous time periods (sometimes called borrowing strength in time) and data from other, neighbouring areas (featuring an attempt to borrow strength in a spatial dimension) are just some of the topics of the project that are highlighted and extended in the presentations. Other important topics are the inclusion of sampling weights in an estimation procedure and the estimation of cross-classifications in the SAE context. In a number of papers, additional developments of high methodological and practical relevance are addressed, such as advances in domain estimation under the design-based framework, the estimation of poverty figures, and special features of small area estimation in the context of business surveys. Authors of SAE2005 conference papers are encouraged to submit manuscripts for publication in a special issue of Statistics in Transition Journal; thanks are due to Editor, Prof. Jan Kordos, for offering pages of the journal for this purpose.

SAE2005 also serves as the final conference of the EURAREA project. Many attendees have been involved in the project as consortium members or in some other role. I am glad to notice that in the group of some 100 participants, many people conduct their SAE and related research under other frameworks, in universities and national statistical agencies and similar institutions for example. This mixture surely increases interaction, exchange of experiences and communication between the various approaches. The conference sessions and the supplement, Short Course on Tools for Small Area Estimation, can be expected to offer a good basis for further extension of the use of the methods and tools in practical applications.

I express thanks to the members of the Scientific Committee and the Organization Committee for their activity during preparatory phases of the conference and in the event itself. Mikko Myrskylä, Hilkka Potila and Riikka Turunen of Statistics Finland edited these proceedings, Kari Nissinen and Sari Eronen, Department of Mathematics and Statistics of the University of Jyväskylä, were the webmasters, and several other department staff members and students assisted in practical arrangements during the conference. Thanks are due to all these people.

Jyväskylä, 22 August 2005

Risto Lehtonen

EURAREA

Overview

EURAREA: An overview of the project and its findings

Patrick Heady and Martin Ralphs[1]

1. A short introduction

One of the purposes of this conference is to present the findings of the EURAREA project, and consider the steps that follow from it. So we need to start by looking at what EURAREA was, what it was attended to achieve, and what was its relation to other applied and theoretical research. In this paper, we would like to consider EURAREA 's contribution under three headings:

1. Empirical evaluation of SAE methods
2. Making SAE "NSI-friendly"
3. Creation of an environment for future empirical research

2. Empirical evaluations and their implications

In the research proposal that we submitted to the European Commission we presented small area estimation as a promising methodology which so far had mostly been applied on the other side of the Atlantic. We proposed to investigate:

1. the potential effectiveness of these methods in the context of European official statistics
2. the scope for using recent theoretical innovations (such as methods involving spatial and temporal autocorrelation) to enhance their effectiveness
3. to make recommendations for their application.

Thus, though the project provided some scope for theoretical innovation (some of which has been published in journals as well as in the EURAREA Report, for example Dehnel et al., 2004 and Zhang and Chambers, 2004), its main focus was on the application and evaluation of existing methods, and of methods that were already being developed elsewhere. And, fortunately, our evaluation of these methods has generally confirmed previously positive results.

2.1. A summary of the conclusions from the EURAREA evaluation

The main conclusions from the evaluation are summarised below.

1. Model-based estimation methods substantially outperform design-based methods for very small areas (NUTS4 / 5), and achieve comparable or slightly better leve1s of precision for medium-sized areas. However, this finding does not always extend to the performance of confidence intervals calculated using model-based methods. Though in some instances they performed well, in others coverage rates were substantially below face value.

2. Model misspecification is a potentia1 source of error. If models are fitted using unit-level covariate data alone, the fixed effect component of the estimators is 1iable to severe bias as a result of the 'ecological' effect. Additionally, misspecification of the distribution of random terms may underlie some of the prob1ems with confidence interva1s.

[1] Office for National Statistics, UK

3. Making use of data from earlier time periods for the area concerned, via either the random or fixed part of the model, substantially enhances the precision of estimates for individual small areas. Interestingly, allowing for the spatial auto-correlation of random area effects was less effective in our simulations. It is possible that greater improvements might be achieved with different spatial autocorrelation structures or distance metrics, but in general we saw a more pronounced gain from incorporating time series data.

4. The enhanced log-linear methodology proved effective in estimating change-since-last-census for cross-classified data, with the use of a generalized linear structural mixed model achieving the best results in most cases. The associated confidence intervals tended to be underestimated for SPREE and GLSM estimators, but were generally too conservative in the case of the GLSMM estimator in our experiments.

5. The standard deviation of the set of estimated area means generated from a single sample tends to either underestimate (in the model-based case) or overestimate (in the design-based case) the standard deviation of the set of actual area means. In principle, model-based estimators can be adjusted to reduce this problem. Such adjustments are not possible with design-based estimators.

6. Effective model-based estimation requires that sample data can be matched to area-level covariates with high explanatory power. If possible, unclustered sample designs are also favourable and increase the success of the estimation models.

2.2. Our results in the European context

Although these results support theoretical expectations and are in that sense unsurprising, they are interesting and new from the point of view of European statistical policy because they show the specific effect of these general findings for the choice of estimators for the kinds of subject matter and spatial unit that are important to European policy makers.

Policy implication 1: Useful estimates for very small areas

A key finding from EURAREA is that useful estimates can be made for very small areas (NUTS4/5) using small area estimation techniques and model-based approaches in particular. The typical gain achieved is illustrated below in Figures 1 and 2. Here, we show Mean Squared Error (MSE) performance for key estimators expressed as a proportion of the MSE that would arise if the National Sample Mean was used as the estimator for each local area. Of course, the National Sample Mean is not actually a sensible small area estimator. But the results do tell us the amount of error that would be incurred by making the false assumption that all areas had the same mean value, which would be the natural default assumption in the absence of any form of small area estimation. They therefore provide a useful benchmark against which to assess the performance of the other techniques.

In Figure 1, we see that at NUTS3 level all of the estimators perform substantially better than the national sample mean, but that model-based estimators are usually (except in the case of income) as good or better than their design-based counterparts.

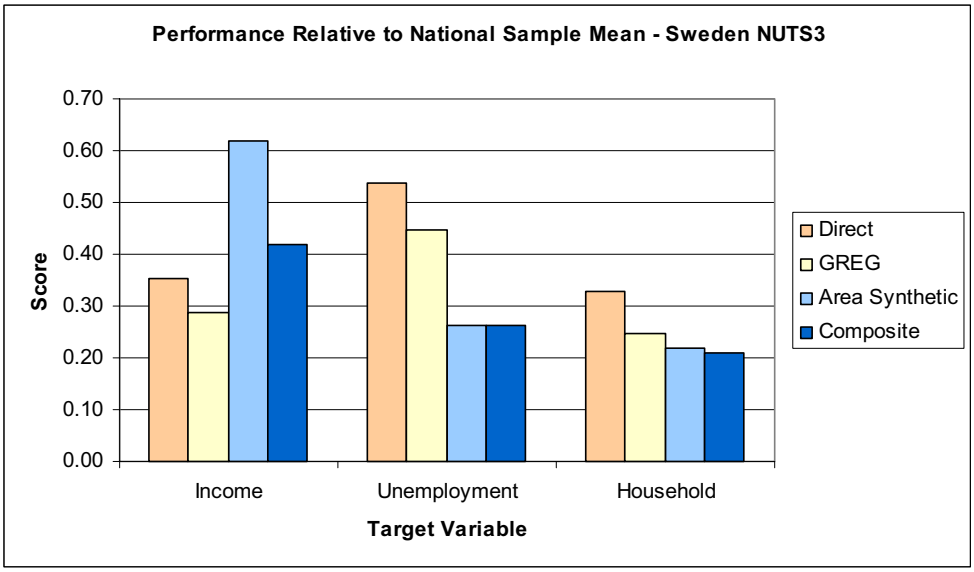

Figure 1 – MSE performance relative to the MSE of the National Sample Mean for three target variables in Sweden at NUTS3 (NSM = 1.0).

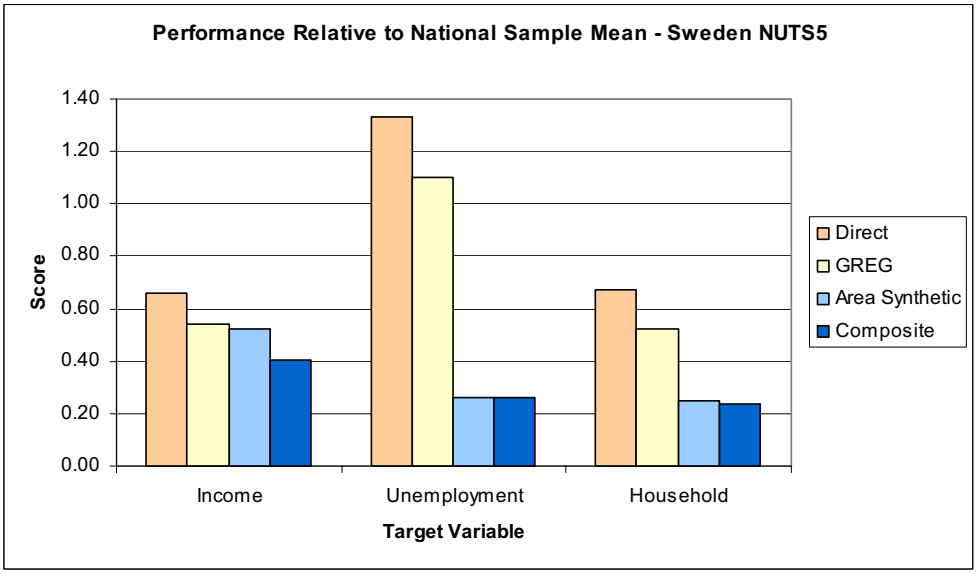

Figure 2 – MSE performance relative to the MSE of the National Sample Mean for three target variables in Sweden at NUTS5 (NSM = 1.0).

In Figure 2, the results are much more clear cut. Direct and GREG estimators actually perform worse than the national sample mean in the case of ILO Unemployment and are always less successful than their model-based counterparts. The composite estimator is usually the best performer.

Policy implication 2: Estimating the distribution of area values: problems of over-shrinkage

The performance of estimators for particular areas is important when resource allocation occurs on an area-specific basis, but other policy applications require estimates that robustly reflect the distribution of area values across the country. This is important if a government wishes to assess the extent of geographic inequality or if applications for funding by some higher-level institution (such as the European Community) are dependent on the number of areas in a country which fall below some specified threshold. From this point of view, a reasonably good set of estimates might be one

for which the empirical standard deviation of the true area values was close to the empirical standard deviation of the estimated area values.

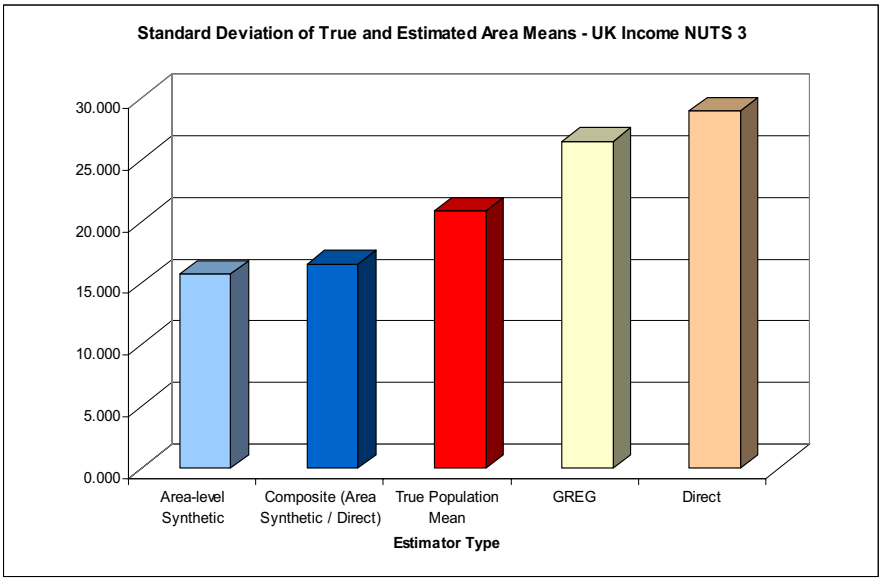

Figure 3 – Comparing the true standard deviation of area means with that produced by different estimation strategies for Income at NUTS3 in Northwest England and North Wales.

In Figure 3, we compare the true standard deviation of area means for NUTS3 areas in the United Kingdom with the standard deviations of estimates of these means produced using the Direct, GREG, area synthetic and composite methods described above. The direct estimator tends to overestimate extremes in the distribution, and as a result the standard deviation of area values is over-inflated. The area level synthetic estimator has the opposite effect, and tends to "shrink" the estimates towards the centre of the distribution. The result is understatement of extreme values, often referred to as "over-shrinkage" in this context, which is equally problematic when our goal is the description of the overall distribution.

There are a number of proposed methods for dealing with over-shrinkage (for example see Spjotvoll and Thomsen, 1987, Rao, 2003 and Zhang, 2004) and this is an area where further empirical work, perhaps using EURAREA datasets, could be valuable.

Policy implication 3: EURAREA findings are consistent

It is important to emphasise that the specific conclusions from the evaluation programme are very much the same for all the European countries in EURAREA despite widely different socio-economic systems and statistical infrastructure.

2.3. Towards the practical implementation of SAE

The findings of the project also point to the remaining work that needs to be done to make SAE operational in European national and EU contexts:

- In all countries, the current design of major national surveys was adequate to support SAE methods that were close in effectiveness to the theoretical optimum;
- The main adaptations that were needed were availability (at least within the NSI) of precisely geo-coded survey data;

- The improved availability of powerful covariates would substantially increase the predictive power of SAE techniques;
- The practical evaluation of alternative approaches to dealing with over-shrinkage was an important area for applied research, particularly in the context of resource allocation within the EU.

Although the methods considered in EURAREA are certainly not exhaustive, the results that have emerged are sufficient to show that, given the political and administrative will to implement them, small area estimation techniques already have the capability to play a major role in resource allocation problems.

3. Making SAE "NSI-friendly"

Important as these findings are, there was more to EURAREA than that. Its wider significance is related to a paradox: the fact that, although European researchers have been prominent in the development and application of SAE and related methods -names such as Särndal, Holt, Goldstein, Pfeffermann and Kordos spring to mind - European statistical offices have been much slower to adopt these methods, and when they have done so, have often applied them in a rather hesitant and marginal way. This is particularly striking when one reflects that most of the key papers on which SAE applications are based are by now anything from 10 to 25 years old. One has to ask whether the statistical offices have simply been waiting for a thorough evaluative study, or whether there are deeper obstacles to the adoption of SAE methods.

We would like to suggest that there are deeper obstacles, and that a second major contribution of the EURAREA project may be the extent to which it helps staff in NSIs to overcome these obstacles. These obstacles can be summed up as follows:

1. The methods are felt to be intellectually inaccessible. The statistical theory that underpins them is quite complex, and the practitioner must grapple with an additional layer of theory to do with computational algorithms in order to implement them efficiently. This becomes increasingly critical as the volume of data increases. The situation is further complicated because the way in which the theory is presented and published means that it is mainly available at researcher rather than practitioner level.

2. The methods are felt to be practically inaccessible, because software requirements, particularly in the case of more advanced models, do not usually fit with extant NSI statistical software systems (in particular the facilities offered by modules such as SAS Proc MIXED or SPSS are rather limited).

NSIs could of course adopt "black-box" solutions: buying in a package that enabled one to specify estimators without fully mastering the underlying theory or the way in which it is implemented. In some ways this makes pragmatic sense, but there is a fundamental problem. NSIs are supposed to be authoritative organisations, taking responsibility for the figures they produce -and this role is hard to reconcile with a "black box" approach.

The EURAREA team started to tackle this problem when we decided to program all our estimators ourselves. It was given in the contract that we would have to write some programs – for those estimators that were not yet implemented in standard packages. However, in the event we resolved, without a great deal of discussion, to program all our estimators ourselves. We believe that it was the wish to fully understand all aspects of the methods that was responsible for this collective decision. The result certainly proved educational for us: there is no better way of testing your understanding of a piece of theory than trying to write an implementation program that actually works!

Of course, if EURAREA is to have a lasting impact on NSI understandings and attitudes, the value of this education must be extended beyond the members of the EURAREA team itself. We have tried to provide for this in two ways: firstly by writing the programs in open code, so that colleagues can play around with them, and so partly replicate our own learning experiences. Secondly, we have tried to structure the EURAREA report in a way that will make the connection between theoretical and implementation issues transparent to readers: whenever possible linking texts on objectives, theory, implementation and actual effectiveness closely together. Before moving on, we hasten to say that the programming work done by the different EURAREA teams was far from being purely educational. Table 1 lists the set of program tools that were developed by the project team, together with the estimators they implement and the groups responsible for developing them. These programs both extend the range of estimators that can be implemented via SAS and greatly improve on the efficiency and speed of some existing programs.

Program	**Implements**	**Authors**
Standard Estimators (SAS v8)	Direct, GREG, Unit-level synthetic, area-level synthetic, composite estimators	SNTL Consulting Office for National Statistics UK
EBLUP_TS (SAS v8)	Composite estimator with area-level time effect	University of Southampton UK Office for National Statistics
EBLUPGREG (SAS v8 / SAS v9)	Unit-level composite estimator with time or spatial effects GREG estimator Synthetic estimator	Statistics Finland University of Jyväskylä University of Southampton
EBLUP_SPACE (SAS v8)	Unit-level composite estimator with spatial effects	ISTAT, Italy University Roma III University of Southampton
FISHERSCORMIX FISHERSCORMIX2 (SAS v8 / C++)	Synthetic estimator with sample weights	INE, Spain University of Miguel Hernandez, Spain
SPREE / GLSM / GLSMM (SAS v8)	Cross-classification estimators for two and three way tables	ISTAT, Italy Statistics Norway

Table 1 – EURAREA programs and functionality

4. Creation of an environment for future empirical research

An equally important part of EURAREA was the research environment that made our evaluation study possible. The considerable investment in datasets and programming that we undertook may make it easier to pursue related research in future.

The simulation strategy that we chose required a considerable investment in database construction and in the construction of programs to run the simulations and implement our chosen performance criteria. The basic simulation setup is illustrated in Figure 4, while the datasets developed for the project are listed in Table 2.

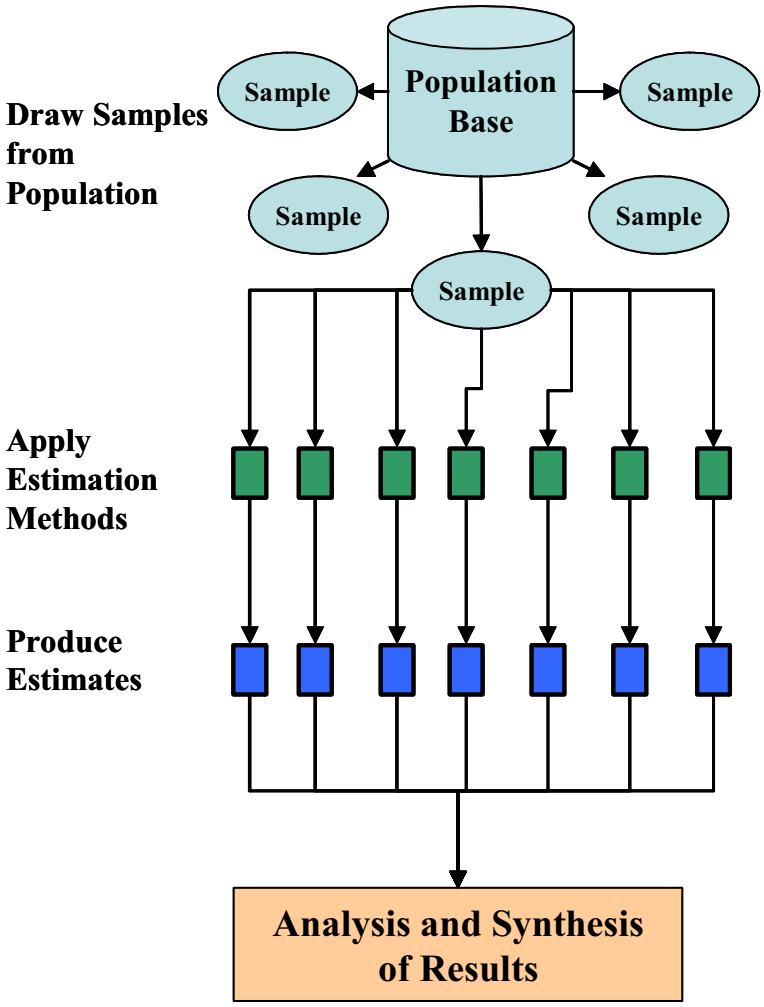

Figure 4 – The EURAREA simulation process. Repeated samples are drawn from a population base and a range of estimation methods are applied to each sample.

Country	Simulation Universe	Total Population	Total Households	NUTS 3 Areas	NUTS 4/5 Areas
Finland	100% of Finland	4.12 million (over 16 years old)	2.25 million	20	85
Italy	25% of Italy			18	151
Poland	5% of population	2.06 million	0.5 million	44	373
Spain	5 Autonomous Communities (Regions)	15.4 million (over 16 years old)	5.92 million	18	215
Sweden	100% of Sweden	5.5 million (16-64 years old)	3.47 million	24	289
United Kingdom	25% of England and Wales	13.9 million	4.9 million	13	2275

Table 2 – Summary information about project databases

One of the main findings of EURAREA is that it is technical1y feasible to simulate statistical procedures and explore their performance on large population databases. This model of experimental design can be taken forward and applied outside the EURAREA project, since the databases and simulation programs remain in existence and can be used to evaluate other statistical techniques. The experience of ONS in making wider use of its EURAREA data resources can serve as an example. Since the completion of the EURAREA research programme, the databases configured for the project have been used for a range of different application, and others are planned for the future:

a) Evaluating further small area estimators. The small area estimation project team at ONS is evaluating a range of small area estimation methods using the EURAREA datasets, focussing in particular on optimal model selection and fitting, deriving consistent estimates for different geographical levels and estimation of change over time.

b) Evaluating area-construction algorithms. ONS has adopted optimisation procedures in the construction of reporting geographies (known as "Super Output Areas") for the 2001 census in order to produce area units of uniform population size and homogeneity in terms of key properties such as tenure mix. We have used the EURAREA population bases to compare the properties of these new, optimised geographical units with existing administrative hierarchies. We also plan to use EURAREA data to evaluate maintenance requirements for the new geography across the inter-censal period.

c) Testing data-recasting methodology and associated confidence intervals. Changing geographical boundaries and consequent problems for temporal comparison between small areas are a particular problem in Britain. ONS has been developing methods to move data between overlapping boundary systems. Again, EURAREA datasets have been used to provide a basis for empirical comparison between different data recasting methods.

The experience of working with these data and simulation systems also contributes to the research environment by bringing certain issues into a clearer focus. Two examples will make the point.

1. Under the hierarchical modelling approach which EURAREA shares with most work on model-based SAE, the areas appear as distinct units with no internal spatial differentiation, and linked at most by spatial auto-correlation of expected area values. Once you start working with databases with individual address data, and use the same databases to construct artificial boundaries within what would otherwise be continuous urban sprawl, it quickly becomes apparent that the usual hierarchical SAE set-up is by no means the only way of posing the estimation problem.

2. A problem that members of the EURAREA team discussed amongst ourselves was the relation between our simulation exercise, based as it was on repeated sample selections from a set of given populations, and the theoretical underpinning of model-based estimation, based (at least in its non-Bayesian versions) on the notion that the model describes the random processes underlying the generation of the observed populations. Some of us felt that this meant that repeated simulations on given populations were not fair tests of the performance of model-based estimators – while others of us felt that contact with the richness of real data, and some acquaintance with the actual processes of community development and boundary construction, exposed the models as merely analytically convenient fictions. We do not want to take sides here! Our point is simply that the process of constructing a shared simulation methodology brought different viewpoints into focus and made possible a meaningful debate in which both theoreticians and practitioners could join.

The final demonstration that EURAREA has succeeded in creating an environment for continuing research in spatial estimation is the fact that many members of the EURAREA team are presenting papers at this conference, based on work that they have continued to do after the formal end of the EURAREA project itself. The ultimate test of the project's value is that it has helped put more European researchers, particularly researchers linked to NSIs, into a position to contribute to, and learn from, wider developments in the field of spatial estimation and modelling.

5. Acknowledgements and Notes

We would like to thank the European Union for funding the EURAREA project, and all of our project partners for working with us in such a constructive manner. We would also like to thank Eurostat for providing Support and organising constructive project reviews.
The estimators referred to in this paper are defined fully in the EURAREA Project Reference Volume. This, and all of the SAS programs developed in EURAREA, are available for download from http://www.statistics.gov.uk/eurarea.

6. References

Dehnel, G., Golata, E. and Klimanek, T ., 2004, Consideration on optimal sample design for small area estimation, Statistics in Transition, Vol. 6, No.5, p. 725–754. Rao, J.N.K., 2003, *Small area estimation,* Wiley, New York.

Spjøtvol1 E. and Thomsen, I. 1987, Application of Some Empirical Bayes methods to small area statistics, *Bulletin of the International Statistical Institute*, Vol. 4, p. 435–450.

Zhang, L., 2004, Simultaneous estimation under nested error regression model, *Statistics in Transition*, Vol. .6, No.5, p. 655–666.

Zhang, L. and Chambers, R.L., 2004, Small area estimates for cross-classifications, *Journal of the Royal Statistical Society, Series B*, Vol. 66, Issue 2, p. 479–496.

KEYNOTE PAPERS

Poverty mapping and extensions

Chris Elbers, Jenny Lanjouw and Peter Lanjouw[1]

'Poverty mapping' is relatively new small area-estimation technique for obtaining high-resolution maps of distributional characteristics of income or expenditure in developing countries.

For many developing countries, high-quality and extensive information on household income or consumption expenditure is collected on a regular basis, but for a relatively small sample of households only. We combine this information with information about covariates avialable from a census to create so-called poverty map. In this lecture we will address a number of issues arising in the construction of maps, give examples of poverty maps from various contexts, and consider some extensions to the basic approach. In particular the lecture will be in three parts.

In part 1 we will give an exposition of poverty mapping: statistical foundation, data requirement, estimation strategy and computational requirements. We will discuss prediction accuracy and compare poverty mapping to other small area-estimation techniques.

In part 2 we give examples of poverty maps and their use and discuss some of the problems and experiences encountered in constructing the maps. Also we will discuss some of the extensions that have been proposed, in particular using the same technique for mapping non-income variables (e.g., health), or mapping along non-geographical dimensions. Also we briefly discuss some related alternative approaches to poverty mapping.

Part 3 discusses how poverty maps can be used in subsequent 'down-stream' research. Essentially small area estimates of income distribution characteristics are synthetic and cannot be used in subsequent analysis without addressing prediction error. Ever since the first map was constructed people have asked for updates without having to wait for a new census. We discuss some of the ideas we currently try out to update poverty maps and how such multiple maps could be used in research. Finally we discuss some research which attempts to verify small-area poverty predictions directly from survey data.

[1] Elbers is with Vrije Universiteit Amsterdam, Jenny Lanjouw is with UC Berkeley, and Peter Lanjouw is with The World Bank. Contact Elbers at: celbers@feweb.vu.nl

Developing small area statistics for business surveys

Mike Hidiroglou and Marie Cruddas[1]

The Office of National Statistics (ONS) has a well-developed programme for estimating for small areas from household surveys. This programme has evolved during the last few years. This is not the case for business surveys, where the majority of statistics are produced at fairly detailed industrial levels within the United Kingdom and its four countries and but not for lower geographical levels. However a recent review of the statistics required for economic policy making in the UK, Allsopp (2004), has identified the need to develop regional economic statistics. In particular, an important aim is to provide good quality Gross Value Added estimates for Government Office Regions (NUTS 1) - London, South East, South West etc - and improved detail at lower levels, as part of an integrated system producing both National and Regional Accounts.

The first step in meeting the requirements of the Allsopp review requires the re-engineering of a large number business surveys as well as the register of businesses that forms the sampling frame for the surveys and provides auxiliary information for estimation. However estimating for small areas will be hampered by known problems with the business surveys (i.e. the impact of outliers; classification in terms of industry, geography and size; and coverage and datedness of the frame). Furthermore, the small area problem needs to be differentiated between the larger and smaller businesses. In business surveys a few large businesses can contribute a large proportion of the variable of interest and these are selected with certainty. However, many will not be able to report at the lower required levels. The problem is how to disaggregate these reported data to these levels, using lower level auxiliary data. For the smaller businesses, the problem will be one of sample allocation and estimation. In terms of allocation, the sample will be allocated to predetermined main domains, bearing in mind that several variables are of interest. In this case the estimation will most likely draw on the methodology for estimating for small areas that has been developed for the household surveys, while for larger businesses this methodology may have applications in the disaggregation problem.

This paper will identify and discuss the problems in developing small area statistics for business surveys.

References

Allsopp, C. (2004) Review of Statistics for Economic Policymaking: Final Report to the Chancellor of the Exchequer, the Governor of the Bank of England and the National Statistician, HMSO: London.

[1] Office for National Statistics

Small Area Estimation Using Times Series Models Subject to Benchmarking Constraints

Danny Pfeffermann[1]

The problem of Small Area Estimation is how to produce reliable estimates of area (domain) characteristics, when the sample sizes within the areas are too small to warrant the use of traditional direct survey estimates. This presentation will focus on the use of time series models as a vehicle for borrowing strength from past surveys. In order to protect against possible model breakdowns and to satisfy arithmetic consistency in publication, it is often required to benchmark the model dependent estimates in the small areas to the corresponding direct survey estimate in a large area for which the survey estimate is sufficiently accurate. This benchmarking process defines implicitly a way of borrowing information across the areas, which can be further enhanced via the model equations.

The presentation will show how the benchmarking can be implemented within state-space time series modelling. The computation of the benchmarked estimators and their variances requires joint modelling of the direct estimators in several areas, which in the case of many areas requires the development of new filtering and smoothing algorithms for state-space models with correlated measurement errors. The application of the proposed procedure is illustrated using U.S. Employment and Unemployment series.

[1] The Hebrew University of Jerusalem

Small Area Estimation Under Informative Sampling

Danny Pfeffermann[1]

The problem of small area estimation (SAE) is how to produce reliable predictors for the true means or proportions in areas with very small or no samples. This can be done by basing the inference on statistical models that permit borrowing information across the areas or over time. In this talk we consider situations where the sampling of areas is with probabilities that are related to the true (unknown) area means, and the sampling of units within the selected areas is with probabilities that are related to the values of the study variable. The problem with this kind of sampling schemes is that the model holding for the sample data differs from the model holding for the population values, giving rise to *informative sampling*. Failure to account for the effects of an informative sampling scheme may result in severe bias of the small area predictors.

We use relationships between the *population distribution*, the *sample distribution* and the *sample-complement distribution* of a study variable in order to derive approximately unbiased predictors of the area means in sampled and nonsampled areas. Appropriate bootstrap MSE estimators of correct order are also developed. The results of a Monte-Carlo study that illustrates the performance of the proposed predictors and their MSE estimators will be shown.

[1] The Hebrew University of Jerusalem

Small Area Estimation: Overview, New Developments and Practical Issues

J. N. K. Rao[1]

Demand for reliable small area statistics has greatly increased in recent years. Traditional area-specific direct estimators may not provide adequate precision because small area sample sizes are seldom large enough or even zero for some areas. This makes it necessary to borrow strength from related areas through linking models based on auxiliary information such as recent census and current administrative data, leading to model-based indirect estimates. Model-based methods based on explicit area level or unit level linking models have been extensively studied including empirical best linear unbiased prediction, empirical best (EB) and hierarchical Bayes (HB). In this talk I will first present a brief overview of those methods with particular attention to measures of variability. I will also present some new results: MSE estimation under the original Fay-Herriot method of estimating the model variance in the basic area level model, use of survey weights under the basic unit level model to ensure automatic benchmarking to reliable large area direct estimates, a new jackknife method of estimating MSE under logistic linear mixed models, choice of matching priors on model parameters in the HB method, efficiency comparison between EB and model-assisted GREG estimators under a two-level model, and effect of measurement errors in the auxiliary variables. I will also address several practical issues including strategies at the design stage, multiple objectives and model diagnostics.

[1] Carleton University, Ottawa, Canada

Reliable Statistics for Subpopulations
Some Unresolved Issues

Carl-Erik Särndal[1]

Two theories currently pave the way for estimation for subdivisions of a sampled finite population: (design-based) domain estimation and (model based) small area estimation. Depending on domain size, it is one or the other of these two theories that will guide our efforts to produce reliable estimates. In one theory as in the other, a weak basis – a lack of sufficient data or other factors – reduces the chances of obtaining reliable estimates.

The presentation will comment on some of the unresolved issues in estimation for subpopulations.

One aspect of importance is that the outlook on estimation for subpopulations varies between countries. They differ in regard to infrastructure, for example in regard to the availability of high quality registers than can provide auxiliary information for the estimation. As a result, the national statistical agencies in different countries adopt different strategies in the choice of methodology for subpopulations.

[1] 2115 Ernbrook #44, Ottawa, Ontario, K1B 4J5, Canada

INVITED AND CONTRIBUTED PAPERS

Small Area Estimation of the Italian Poverty Rate

Michele D'Alò, Loredana Di Consiglio, Stefano Falorsi, Fabrizio Solari[1]

This paper focuses on the application of EURAREA project results to real data provided by Italian households survey. One of the aims of the project was first to assess the performances of some small area standard estimators and then to improve them using a spatial autocorrelation structure based on the Euclidean distance among areas.

On the basis of the methodological aspects developed within EURAREA project, this work intends to compare the performances of standard and enhanced methods to estimate the poverty rate at NUTS3 level. In order to evaluate the properties of the methods under study a simulation study has been carried out using bootstrap techniques.:Two different sets of auxiliary variables has been considered: the first set of variables consists of cross-classification of sex and age, while the second set has been derived clustering the population in homogenous groups with respect to the target variable.

The simulation study has been based on 1000 samples drawn from a pseudo population using two-stages sample design (municipalities-households) and the methods have been compared in terms of relative bias and relative mean square error related to the small area estimates.

The overall evaluation criteria show that the model based estimator implementing the spatial autocorrelation structure performs better than the others estimators.

Bibliography

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

EURAREA Consortium (2003) -*EURAREA Deliverable- WP7*- Final Reference Volume: Vol.1-3
https://www.statistics.gov.uk/eurarea/default.asp

[1] {dalo, diconsig, stfalors, solari}@istat.it
ISTAT - Via Magenta, 2 00185 - Roma - Italy

Bayesian study of small area racial disparities in heart disease mortality in Ghazvin province (Iran)

Esmail Amiri[1]

In a Bayesian approach we study geographical level disparities in heart disease mortality between urban and rural area in Ghazvin province during(1996-2004), for men an women separately. A model involving random effects and auto-correlated errors is proposed for small area estimation, using both time series and cross sectional data. Parameter estimation is performed using Markov chain Monte Carlo methods(MCMC).

Keywords : Bayesian, Auto-regressive, Gibbs sampler, random effect, cross sectional.

[1] Department of Statistics Imam komeini International University,
P.O.Box 288, Ghazvin, Iran.
e_amiri@yahoo.com

Notes on sample covariance matrix under informative sampling

Julia Aru[1]

In this paper I give some notes on the covariance between two variables in the sample compared to the population covariance. The informative sampling design is assumed. As a special case I consider two independent variables in the population and show that independence is preserved in the sample. We also give the general (although not very useful) formula for population covariance that exploits characteristics of sample distribution of considered variables. Some simulation examples will be presented during the presentation.

[1] University of Tartu, Estonia, e-mail: julia_a@ut.ee

Bias Adjusted Distribution Estimation for Small Areas

Ray Chambers[1] and Nikos Tzavidis[1]

Small area estimation techniques are employed when sample data are insufficient for acceptably precise direct estimation in domains of interest. These techniques typically rely on regression models that use both covariates and random effects to explain variation between domains. However, such models also depend on strong distributional assumptions, require a formal specification of the random part of the model and do not easily allow for outlier robust inference.

In a recent paper Chambers and Tzavidis (2005) proposed the use of M-quantile models as an alternative to random effects models for small area estimation. This avoids the problems associated with specification of random effects, allowing inter-domain differences to be characterized by the variation of area-specific M-quantile coefficients. However, they also observed that M-quantile estimates of the small area means are biased, with the magnitude of the bias being related to the presence of outliers in the data. In this paper we propose a bias correction to small area estimates based on the representation of the mean as a functional of the empirical distribution function. The method is then generalized for estimating other quantiles of the small area population distribution of the variable of interest.

Two approaches for small area estimation are considered (a) random effects models and (b) M-quantile models (Chambers and Tzavidis 2005). Distribution estimation for small areas is then performed under these approaches using two estimators of the finite population distribution function (a) a naive estimator and (b) the Chambers-Dunstan (1986) estimator. Variance estimation for the M-quantile small area estimates is discussed. The different approaches are illustrated using both sirnulated and real data.

[1] Southampton Statistical Sciences Research Institute, University of Southampton

Comparing EBLUP and C-EBLUP for Small Area Estimation

Hukum Chandra and Ray Chambers[1]

Several methods for small area estimation have been proposed in the literature. See Rao (2003). However, research is still continuing on the important problem of identifying small area estimation techniques that are efficient and also simple to implement, with estimation of mean squared error an outstanding problem. We describe the C-EBLUP or calibrated approach to small area estimation introduced in Chambers (2005). This approach uses calibrated sample weights for estimation of small area means under linear mixed models, and also includes a simple estimator of the mean squared error of the calibrated estimator. In this paper we present results from a Monte-Carlo study that compares the mean squared error estimates generated under the C-EBLUP approach with those generated under the well know EBLUP-based method of Prasad and Rao (1990). Our results show that the proposed C-EBLUP mean squared error estimator performs well and represents a real alternative to the usual prediction based estimator. We also note that in case of model misspecification, the C-EBLUP approach appears to provide a more robust set of small area estimates.

Key Words: Small area, Calibrated weighting, Prediction approach, MSE estimation, model-based estimation.

[1] Southampton Statistical Sciences Research Institute
University of Southampton
Highfield, Southampton SO17 1BJ
United Kingdom
Email: hchandra@soton.ac.uk and R.Chambers@soton.ac.uk

Ecological inference: A new approach based on spatial econometrics

Coro Chasco-Yrigoyen[1]

In this note we compare the results obtained by the application of two alternative methods of ecological inference. The data is on per capita household disposable income in the 50 provinces and 78 municipalities of Asturias, Spain. The first method is based on Ordinary Least Squares regression model, which assumes coefficient constancy or homogeneity across space. The second method is based on spatial econometrics techniques, which deal with spatial autocorrelation and spatial heterogeneity, assuming spatial exteranlities and some kind of coefficient heterogeneity over geographic space. These second approach implies estimation by Maximum Likelihood or Two-Stage Least Squares to ensure good properties in the estimators.

Key words: Ecological inference, Spatial prediction, Spatial autocorrelation, Spatial heterogeneity, Disposable income.

[1] Department of Applied Economics,Autonoma University of Madrid
Campus de Cantoblanco, 28049 Madrid (Spain)
Tel. ++34914974266
Fax ++34914973943
Email: coro.chasco@uam.es

Small Area estimation with varying area boundaries by low level hierarchical modelling using the synthetic estimator

Philip Clarke[1], Fernando Moura[2], Danny Pfeffermann[3]

This paper investigates the use of hierarchical models for small area estimation with varying area boundaries employing the synthetic estimator. The strategy is to model at the lowest possible area level. The paper shows how the area estimates and corresponding MSE estimates can be obtained at a variety of nested and intersecting boundary systems which build from the low level. The estimates are computed by aggregating from the lowest level and are hence internally consistent. The paper extends the theory of Stukel and Rao (1999) who considered the use of such models together with the EBLUP estimator. The methodology is illustrated by presenting results of a simulation study that uses hierarchical models built at the lowest area level defined by the UK 1991 census, enumeration districts, and producing estimates and MSE estimates for a variety of UK boundaries.

Reference

Stukel D.M., Rao J.N.K., 1999, On small-area estimation under two-fold nested error regression models. Journal of Statistical Planning and Inference 78 131-147.

[1] Methodological Directorate, Office for National Statistics, UK
[2] Federal University of Rio de Janeiro Statistical Department, Brazil
[3] Hebrew University, Jerusalem, Israel and University of Southampton, UK

Small area estimation or simulation by using training images: the advent of multiple-point statistics

E. Conza[1]

In many earth sciences applications, the geological or physical structures to be reproduced are curvilinear, e.g., high permeability sand channels forming preferential flow paths.

Modeling of such curvilinear patterns requires measuring the connectivity in the space of the indicators of such structures; the traditional tool offered by geostatistics is the 2-point statistics covariance/variogram which relates any two points in space, for example establishing the probability that any two locations are in the same facies. Such statistics is largely insufficient to characterize the shape and spatial continuity of the structure under study. The modeling of curvilinear structures requires multiple-point statistics inolving jointly three or more points at a time. The inference of multiple-point statistics needs a vast amount of data on a regular grid, typically not available in the small area or sub-domain (subsurface). In many applications, particularly those related to mapping of categorical variables, facies or rock types distributions, critical structural information can be obtained from training images drawn from prior expertise on similar phenomena. From such training images complex statistics involving jointly values at multiple locations can be extracted.

Such training images depict the expected patterns of geological heterogeneities. The multiple-point statistics inferred from the training images are exported to the reservoir model where they are anchored to the actual subsurface data, both hard and soft, in a sequential simulation mode.

Hence, the objective of this paper/poster is to introduce this new multi-point technique in order to improve estimations/simulations on a small area.

[1] University of Lecce, Italy

Composite estimation of small area means using Fay-Herriot model

Gauri Sankar Datta[1]

Composite estimators are very popular for estimation of small area means. A composite estimator is obtained by taking a weighted average of a model-based synthetic estimator and a traditional survey estimator or direct estimator. One important model in small area estimation based on area level data is the Fay-Herriot model. Empirical best linear unbiased prediction (EBLUP) method is widely used in developing composite estimators of small area means based on suitable models. EBL UP estimator of a small area mean is obtained by estimating the unknown variance parameters in the BL UP estimator of a small area mean. The BL UP estimator is an optimally weighted average of the synthetic estimator and the direct estimator, the weight attached to the latter is proportional to the model error variance.

In this talk, we will consider small area estimation using EBLUPs as well as other composite estimators based on weights either known or obtained from other consideration. For these estimators we will review various mean squared error results. Based on a composite estimator and an estimated measure of uncertainty in estimating a small area mean, we will construct standard approximate confidence interval for a small area mean and study its coverage bias. The coverage bias will be used to calibrate a standard interval to achieve the target coverage probability to a greater degree of accuracy.

[1] Department of Statistics, University of Georgia, Athens GA 30602 USA

Attempts to estimate basic information for small business in Poland

Grazyna Dehnel, Elzbieta Gołata [1]

The paper presents first attempts to use administrative data sources and indirect estimation techniques to estimate basic economic information about small business in the cross-section of Polish Classification of Economic Activities PKD and voivodships.

The study objective, specified as accounting for and applying tax data for a more effective use of a survey of small businesses with up to 9 employees, was understood in a twofold manner. First of all, it was a verification of the hypothesis concerning the possibility of improving estimation precision in studies available to date. Secondly, it was intended as a possible extension of estimation scope by joint distribution by voivodship and economic activity (PKD division). The basic economic information, for the aim of this study, was limited to the paid employment and revenues.

One of the major problems involved in estimating information about economic activity across domains is the small sample size and incompleteness of tax registers rendering integration of data sources difficult. The distribution of small companies by target variables occurs to be considerably skewed to the right, with high variation, high kurtosis and outliers. To tackle the problem, the following solutions were suggested. One involves moving the analysis up from the unit level to the domain level: territorial units, PKD categories or combined domains. Other methods concern application of robust regression or logarithmic transformation in constructing the models.

The Horvitz-Thompson estimates in the joint cross-sections of PKD and voivodships are presented and compared with the results of indirect: ratio synthetic, regression synthetic and composite estimates. The properties of the estimators are discussed from the domain specific point of view and combining all domains. Estimation precision characterizing economic activity of small enterprises is presented and analyzed for different types of domains: PKD sections, regions and joint cross-section of regions and economic activity.

Results obtained in the study entitle to draw the following conclusion. Application of indirect estimation to small business data requires consideration of the heterogeneity of its distribution. Nevertheless the results of the study present the practical possibilities and benefits of adopting the techniques of small area estimation to small business data in Poland.

[1] The Poznan University of Economics

EBLUP estimator: Comparison of the prediction using true population information with sample level information

Kari Djerf[1]

One of the main goals of the EURAREA project was to develop new methods and applications based on empirical linear unbiased predictors (EBLUP) on unit level data. The applications and finally software were based on availability of full population information at least as the tabular data. However, in many situations such data are not completely available and, thus, the most reliable way to predict is to use sample level data only. This presentation shows some preliminary results how the sample level data might be used and how the two approaches will compare with each other based on both simulations and real examples.

[1] Statistical R&D, Statistics Finland

Estimation of poverty indocators at the sub-national level using univariate and multivariate small area models

Enrico Fabrizi[1], Maria Rosaria Ferrante[2], Silvia Pacei[2]
enrico.fabrizi@unibg.it, ferrante@stat.unibo.it, pacei@stat.unibo.it

The assessment and reduction of the large inequalities in the distribution of income and wealth among member countries and regions represents, for the EU, a priority in order to stimulate an equal participation of all regions and members states to the economic life of the Union. Thus availability of reliable estimates of income distribution parameters at a sub-national level is essential for the study of poverty and regional disparities.

The aim of this work is to estimate some of the income inequality parameters suggested in the Laeken European Council (Eurostat, 2003) for the Italian admimistrative regions. In particular we consider the *Per-Capita Income*, the *Poverty Threshold*, the *At-risk-of-poverty rate* based on a regional Poverty Threshold, the *At-risk-of-poverty rate* based on a national Poverty Threshold, the *Gini Index*.

Estimates are based on the 2001 repetition of the European Community Households Panel (ECHP), a survey that was designed to provide reliable estimates for macro-areas (in Italy: North West, North East, Center, South and Islands). Admimistrative regions for which estimates are wanted are smaller.

To obtain reliable regional estimates of the parameters we are interested in , we use data from the ECHP survey. We obtain direct estimates using the survey's weights and derive estimates of their variances by means of a bootstrap resampling method. Then we propose small area estimators based on a univariate area level model and on two different multivariate area level models. Multivariate models are based on the idea of modeling jointly a set of different but correlated indicators. As auxiliary information, since Census related or Administrative data are either not available yearly or not fully reliable, we use the estimates of the average annual unemployment rate, provided by the Italian National Institute of Statistics (ISTAT). Uncertainty associated to these estimates are accounted for in the evaluation of estimators' variability.

As estimation method we use a Hierarchical Bayesian approach implemented by means of MCMC computation methods. Our main results are that model based estimation strategy proposed leads to significant gains in efficiency and that, among the models considered, multivariate models perform better than univariate ones.

We consider also the problem of overshrinkage that may lead to a distribution of estimated indicators across regions less variable than what it should. The problem is faced constraining estimates based on the proposed models. In view of exploiting the panel nature of the ECHP survey, we provide also some preliminary results based on models borrowing strength longitudinally as well as in cross section.

[1] Department of Mathematics, Statistics, Informatics and Applications, University of Bergamo, Italy
[2] Department of Statistics, University of Bologna, Italy

The results obtained show that the model based estimation strategy proposed leads to significant gains in efficiency and that, among the models considered, those borrowing strength from the sampling covariance between estimates of various parameters lead to a major variance reduction respect to the univariate model.

Eurostat (2003), *Laeken' Indicators – Detailed Calculation Methodology*, Working Paper, Working Group "Statistics on Income, Poverty & Social Exclusion", 28-29 April 2003.

Sampling designs for small domains estimation through multi-way stratification techniques

Piero Demetrio Falorsi falorsi@istat.it
Stefano Falorsi stfalors@istat.it
Paolo Righi parighi@istat.it
Fabrizio Solari solari@istat.it
Istituto Nazionale di Statistica – Via C. Balbo 16 – 00184 Roma

The small area problem is usually thought of as one to be dealt with via estimation. However, there are opportunities to be exploited at the survey design stage. In this framework it is crucial to control the sample size for each domain of interest, so that each domain is treated at design stage as planned domain, for which it is possible to produce direct estimate with a prefixed level of precision. In general, this level of precision is useful to keep under control the variance of the direct estimator but it does not guarantee reliable direct estimates. In this paper, the small area problem is dealt with considering the design phase. Some techniques that allow to control the sample sizes for domains of interest which are defined by different partitions of the reference population are presented. Such techniques are useful when the overall sample size is relatively small and by consequence in some of the partitions there are small domains.

When the objectives of the survey is to produce estimates for two or more partitions of the population a standard solution to obtain planned sample sizes for the domains of interest is to use a stratified sample with the strata defined by cross-classification of variables defining the different partitions. In the following this design will be denoted as *cross-classification design*. In many practical situations, the cross-classification design is often unfeasible since it needs the selection of at least a number of sampling units as large as the product between the number of categories of the stratification variables. In order to overcome this problem, an easy strategy is to drop one or more stratifying variables or to group some of the categories and, consequently, small-area estimation problems become more serious since some planned domains become unplanned and some of them can have small or null sample size.

Many methods have been proposed in the literature to keep under control the sample size in all the categories for all the stratifying variables. These approaches may be roughly divided into two main categories or contexts. The first context contains the methods mostly known in the literature as *controlled selection*. They allow to satisfy the sample size planned for each domain of interest without using the cross-classification of the stratifying variables. In the second context there are methods based on *sample coordination*. A separate sample is selected for each partition trying to guarantee the maximum overlap among the different samples. The methods of both contexts avoid to fall into some of the problems previously described.

The aim of this work is to offer a general overview of the techniques allowing to control the selection on the separate stratifying variables and to give account of some recent methodological proposals. The methods described are particularly useful when dealing with small domains or small area problems. In fact, in this situation sampling from the cross-classification of the partition of domains is likely to be unfeasible because the resulting stratification defines a too fine partition of the population.

Estimation of a domain mean under nonresponse using double sampling

Wojciech Gamrot[1]

The well-known two-phase (or double) sampling procedure developped to deal with nonresponse relies on subsampling survey nonrespondents and repeating efforts to collect the data they failed to provide in the initial phase of the survey. If there is complete response in the second phase then unbiased estimates of population parameters may be constructed. Otherwise, if there is incomplete response in the second phase, the introduction of subsample data and construction of estimators utilizing this data allows to reduce the nonresponse bias. In this paper the two-phase sampling procedure is applied to estimation for domains. Estimators of the domain mean are considered and their properties are studied.

Keywords: nonresponse, double sampling, two-phase sampling, domain mean

[1] Department of Statistics, University of Economics, Katowice - Poland

Bootstrap approximations of the mean squared error of empirical predictors

González-Manteiga W.[1], Lombardía M.J.[1], Molina I.[2], Morales D.[3] and Santamaría L. [3]

For linear mixed models with normal distribution, Prasad-Rao approximation of the mean squared error of the EBLUP is currently the most common reference. When dealing with empirical predictors obtained under generalized linear mixed models, the same formula can be applied after a suitable linearization of the model. In alI cases, a conceptually simple, but with high computational cost, are resampling methods. Several bootstrap estimators are introduced, and they are empirically compared with Prasad-Rao formula, under different scenarios for the characteristic of interest including a logistic mixed model.

Key words: Resampling methods, bootstrap, linear mixed models, logistic mixed model, empirical predictor, mean squared error, small area estimation.

AMS Classification(1991): 62DO5, 62JO5.

[1] wenceslao@usc.es and mjoselc@usc.es, Department of Statistics and Operations Research, Universidad de Santiago de Compostela.
[2] imolina@est-econ.uc3m.es, Department of Statistics, Universidad Carlos III de Madrid.
[3] d.morales@umh.es and l.santamaria@umh.es, Operations Research Center, Universidad Miguel Hernandez de Elche

Small area estimation of poverty and malnutrition in Bangladesh: some practical and statistical issues

S.J. Haslett and G. Jones[1]

Working in conjunction with the Bangladesh Bureau of Statistics and the United Nations World Food Programme, we have produce small-area estimates of poverty and malnutrition in Bangladesh at upazila level by combining survey data with auxiliary data derived from a 5% sample of the recent census. A single model is found to be adequate for predicting log average per capita household expenditure, and the poverty measures derived from it at upazila level have on the whole acceptably small standard errors. Small-area estimates are also calculated for food poverty and malnutrition, but these are more tentative as we were unable to find good predictive models for them. The inclusion of GIS variables, especially if these were health related and available at a suitably disaggregated level, might prove useful for these models.

These small-area estimates are derived by combining expenditure and food consumption data from the 2000 Household Income and Expenditure Survey with predictor variables common to both the survey and the 2001 Population Census. To do this we adapted the World Bank's procedure, which has been used successfully in a number of other countries.

In this paper we discuss our adaptation of the standard procedure and touch on a number of general methodological issues, including 'matching' variables between survey and census, the use of robust regression procedures, design-based versus model-based adjustments, appropriate selection of regression predictors, 'multiple' versus 'single' models, use of GIS data, and use of a sample from the Census rather than the full Census itself. We also discuss some practical limitations on how fine a partition can be achieved with this method. Maps of the small-area estimates will be presented.

[1] Dr Stephen Haslett is Professor and Dr Geoff Jones is Senior Lecturer at the Statistics Research and Consulting Centre / Institute of Information Sciences and Technology, Massey University, Private Bag 11222, Palmerston North, New Zealand. Email of first author: s.j.haslett@massey.ac.nz

The bias-corrected regression estimator

Natalja Jurevitš[1]

It is well known that the regression estimator is biased for the total it has to estimate. The bias is small for big samples. Nevertheless, beside the sample size, the bias depends on the auxiliary variables, on their relation to the study variables and on the sampling design. It is important to know sources of the bias and in some cases to use the bias-corrected regression estimator.

The bias of the regression estimator is developed from its Taylor expansion and its main term has a general form (Musting, 2004):

$$\begin{aligned} B \approx & -vec'\left(Cov\left(\hat{\mathbf{t}}_{\mathbf{xy}}, \hat{\mathbf{t}}_{\mathbf{x}}\right)\right) vec\mathbf{T}^{-1} \\ & + vec'\left(Cov\left(vec\hat{\mathbf{T}}, \hat{\mathbf{t}}_{\mathbf{x}}\right)\right) vec\left(\mathbf{T}^{-1} \otimes \mathbf{t}'_{\mathbf{xy}} \mathbf{T}^{-1}\right), \end{aligned} \tag{1}$$

where the involved totals are:

$$\mathbf{T} = \sum_U \frac{\mathbf{x}_i \mathbf{x}'_i}{\sigma_i^2},\ \mathbf{t}_{xy} = \sum_U \frac{\mathbf{x}_i y_i}{\sigma_i^2},\ \mathbf{t}_x = \sum_U \mathbf{x}'_i,\ t_y = \sum_U y_i, \tag{2}$$

and $\hat{\mathbf{T}}, \hat{\mathbf{t}}_{xy}, \hat{\mathbf{t}}_x, \hat{t}_y$ are their design-unbiased estimators.

In the presentation we will give the important special cases of the bias. We assume the model with one auxiliary variable both with and without intercept and the group mean model. We will consider different designs, e.g. SI and multinomial.

We construct the bias-corrected regression estimator:

$$\hat{t}_{y,corr} = \hat{t}_y - \hat{B}, \tag{3}$$

and study its properties.

In a simulation study we use the data taken from (Knottnerus, 2003). In addition to the existent real variables we simulate some different study variables, so that correlation between *y* and *x* is large, small or negative.

Practical study shows that in cases of small correlation the bias-corrected regression estimator (3) is more accurate than ordinary regression estimator. In most cases the variability of the corrected estimator is the same as of the ordinary regression estimator.

References

Musting, K. (2004) Study of the bias of generalized regression estimator. *Workshop on Survey Sampling Theory and Methodology*, June 18-22, 2004, Tartu, Estonia, p. 78-81

Knottnerus, P. (2003) *Sample Survey Theory. Some Pythagorean Perspectives.* Springer-Verlag, New York, p. 300.

[1] University of Tartu, Estonia

Impact of the EURAREA project on research in small area estimation in Poland

Jan Kordos[1]

The EURAREA project was preceded by two international conferences devoted to small area estimation (SAE) (the Warsaw Conference in 1992 and the Riga Conference in 1999), where Polish statisticians presented their first contributions in this field. The author starts with synthetic description of these contributions, emphasizing involvement in the EUREAREA project.
The above mentioned international conferences and the EARAREA project have had significant impact on the following statistical activities in Poland: (i) attempts of application of SAE methods in several fields; (ii) yearly country statistical conferences; (iii) international conferences where Polish statisticians presented their contributions.

The author distinguishes here the following topics in which SAE methods were used: a) estimation of some employment and unemployment characteristics by region and poviat (county) using the 2002 Population Census data; b) estimation of some characteristics of the smallest enterprises by region and poviat; c) application of Hierarchical Bayes method in estimation of unemployment by region and poviat; d) estimation of some agricultural characteristics by region and poviat using agricultural sample surveys and agricultural census data. The author also discusses some aspects of data quality of small area statistics obtained from different sources of statistical data, and the Polish experience in this field.

The author pays special attention to two of the above mentioned topics: (i) estimation of unemployment characteristics by poviat using the 2002 Population Census, and (ii) improving small area estimates by region and poviat using agricultural census data, using area-level and unit-level approaches (considering ecological effect). The author briefly outlines the methodology used and summarises some of the empirical findings.

[1] Warsaw School of Economics, Warsaw, Poland

Income estimation for small sample size

Danutė Krapavickaitė [1]

Keywords: income model, small area estimation, calibrated estimator

1. Introduction. Household budget survey (HBS) is one of the most important sample surveys in official statistics of every country. It estimates income in cash and kind and expenditure per capita of the population of the country and in various parts of the population. The Lithuanian HBS uses a stratified two stage sampling design. Estimator ofincome per capita is investigated like estimator of the ratio of two totals. A sample size and the accuracy of the estimates in the rural area of districts are low. Small area estimation methods using auxiliary information from the neighboring areas are applied in order to improve the accuracy of the estimates of the income per capita in the rural area of Lithuania.

2. Methods used. The data of the HBS survey of the fourth quarter of 2002 are used here for estimation. Two kinds of small area estimators -James-Stein estimator ([2]) and empirical best linear unbiased predictor (EBLUP) ([3]) -are applied. The linear regression model of income per capita id build. Demographical data and agricultural production data are used are used to it. The direct estimates, based on the sample design, and currently used calibrated estimates ([1]) are also presented to compare. The modelling results are aimed at choosing the most suitable estimators in HBS ([4]

3. Estimation results. The James-Stein estimator performs better than the direct one. The MSE of the composite estimator EBLUP performs equally along the areas, improving a very low accuracy of the direct estimator in some areas, however, without any improvement in the average accuracy. It can be explained by the low correlation between the direct estimates in the areas and auxiliary variables. The currently used calibrated estimator has the smallest average estimated mean square error over the small areas.

References

[1] J. Deville, C.-E. Särndal, and O. Sautory. Generalized raking procedures in survey sampling. *JASA*, 88,1013-1020 (1993).

[2] R. E. Fay, and R. A. Herriot. Estimates of income for small places: an application of James-Stein procedures to census data. *JASA*,. 74.269-277 (1998).

[3] M. Ghosh, and J. N. K. Rao. Small area estimation: an appraisal. *Statistical Science*, 9(1),55-93 (1994).

[4] D. Krapavickaite. An example of small area estimation in finite population sampling. *Lietuvos matematikos rinkiny*s, 43, spec. nr. (2003) (86-89).

[1] Institute of Mathematics and Informatics, Akademijos 4, 08663 Vilnius, Lithuania
e-mail: krapav@ktl.mii.lt

On standard errors of model-based small-area estimators

Nicholas T. Longford[1]

The EURAREA project confirmed the superiority of model-based estimators of small area means and proportions, with several qualifications, but reported rather disappointing results regarding estimators of their standard errors. We trace this problem to the contradiction between (replicate) sampling from a population, its division to small areas and values of alI variables that were fixed, and application of random-effects models which assume that a different population of subjects and a set of small areas are drawn from a hypothetical superpopulation in each replication. The two corresponding perspectives, design-based and model-based, are related by an averaging applied in deriving the standard errors ofshrinkage estimators. We regard the design-based perspective as appropriate, but dismiss the standard design-based estimators because they fail to draw on the auxiliary information available in the form of data from other areas, related variables and other surveys.

We show that the model-based estimator of the sampling variance of a small-area estimator is approximately unbiased only when the small-area target is in the typical distance from the national mean or its regression adjustment. Based on this diagnosis, we derive an estimator of the mean squared error (MSE) of the empirical Bayes and composite estimator of the local-area mean in the standard small-area setting. The MSE estimator is a composition of the established estimator based on the conditional expectation of the random deviation associated with the area and a naive estimator ofthe design-based MSE. Its performance is assessed by simulations in settings that range from the congenial (in close agreement with the assumption) to distinctly uncongenial, exploring the sensitivity of the estimator with respect to some of the model assumptions. Variants of this MSE estimator are explored and some extensions outlined.

Key phrases: Composite estimation, empirical Bayes estimation, shrinkage, small-area estimation.

[1] Address for correspondence: N. T. Longford, SNTL, 23 Fairstone Hill, Oadby, Leicester LE2 5RL, England. Email: NTL(@SNTL.co.uk

Combining sampling and model weights in agriculture small area estimation

Militino, A. F.[1,2], Ugarte, M. D.[1,2], and Goicoa, T.[3]
E-mail: militino@unavarra.es

This work is focussed on agriculture small area models for predicting minor crops. In the application considered here, the study domain is often poor in the crop of interest leading to irregular and sparsely distributed plots where the sampled quadrats or segments do not need to be completely included in the domain. Hence, the variability among the sampled units becomes large in those areas with a high number of segments. To date, models including weights to account for heteroscedasticity, as well as models considering sampling weights to achieve design-consistency have been proposed to derive estimators of small area means or totals. In this work, we discuss extensions of these models and the convenience of using both types of weighting. The models performance is illustrated for predicting the total area occupied by olive trees in a region of Navarra, Spain.

References

[1] Battese GE, Harter RM, Fuller WA. 1988. An Error-Components Model for Prediction of Country Crop Areas Using Survey and Satellite Data. Joumal of the American Statistical Association, 83, 28-36.

[2] Prasad NGN, Rao JNK. 1999. On Robust Small Area Estimation Using a Simple Random Effects Model. Survey Methodology, 25,67-72.

[3] Rao JNK. 2003. Small Area Estimation. Wiley Series in Survey Methodology.

[4] Stukel DM, Rao JNK. 1997. Estimation of regression models with nested error structure and unequal error variances under two and three stage cluster sampling .Statistics and Probability Letters, 35, 401-407.

[1] Universidad Publica de Navarra
[2] Centro Asociado de la UNED en Pamplona
[3] Campus de Arrosadia, 31006 Pamplona, Spain

Regional labour market statistics at a European level – small number of survey respondents

Michal Mladý[1]

The regional labour market statistics which Eurostat provides for EU-25, EFTA (Norway, Island) and candidate countries (Bulgaria and Romania) could be a very rewarding field for applying small area estimation techniques. The speaker will give an overview of the available data set at Eurostat. He will briefly explain the EU Labour Force Survey (LFS), then talk about the various regional levels of the labour market data and, finally, clarify the LFS publishing limits and their harmonisation.

1. EU Labour Force Survey (LFS)

LFS represents the main source of regional labour market data provided by Eurostat. The LFS is a quarterly household sample survey and its target population is made up of all persons in private households aged 15 and over.

All regional labour market data provided by Eurostat can be found on the Eurostat web-site http://europa.eu.int/comm/eurostat/ according to the following categories: Regional economically active population, employment, unemployment, socio-demographic labour force statistics and labour market data based on pre-2003 methodology (data up to 2001).

2. Regional levels of the labour market data

Down to NUTS level 2, LFS represents the only source of the regional labour market data. For NUTS level 3, LFS NUTS level 2 data are apportioned to level 3 according to the distribution of either LFS NUTS-3 data or NUTS-3 register data (if the LFS results at NUTS level 3 are considered unreliable). At NUTS level 3, Eurostat publishes the following statistics: economically active population, unemployed persons and unemployment rates by sex and age (15-24, 25 and over). Unemployment figures in many NUTS level 3 regions often represent only a small number of survey respondents – especially in the age group 15-24. These figures are considered to be unreliable and are not published. Regional unemployment statistics is thus an area in which the application of SAE methods could be studied.

3. LFS publishing limits (thresholds)

In order to avoid the publication of figures which are statistically unreliable, Eurostat implemented LFS publishing guidelines introducing two limits (thresholds) based on the sample size and sample design in the various Member States:

A limit – figures below this limit are considered to be unreliable, are not published and are replaced by a colon (:).

B limit – figures between A and B limit are published with a warning concerning their reliability.

As relative standard errors of the limits set by National Statistical Institutes vary significantly, countries were asked to provide Eurostat with harmonised limits corresponding to different level of relative standard errors (10 %, 15 %, 20 %, 25 % and 30 %).

[1] Eurostat

Study on the performance of four variance estimators for logistic GREG estimator for domains

Mikko Myrskylä[1]

Let $U = \{1, 2, \ldots, N\}$ be the population. We estimate frequencies of class A in domains $U^{(d)}, d = 1, 2, \ldots, D$; these are $\sum_{U^{(d)}} I_{\{i \in A\}} \equiv T^{(d)}$, where $I_{\{i \in A\}} = 1$ if i belongs to A and 0 other-wise. Sampling vector $\underline{\mathbf{I}} = (I_1, I_2, \ldots, I_N)$ has distribution $p(\mathbf{I})$, and realisation $\mathbf{I} = (k_1, k_2, \ldots, k_N)$ of $\underline{\mathbf{I}}$ is the sample so that unit i is sampled k_i times. Sample set and sample set in domain are $s = \{i : k_i > 0\}$ and $s \cap U^{(d)} \equiv s^{(d)}$, respectively. Sampling weights are $w_i = k_i / E(I_i)$. The generalised regression (GREG) estimator for $T^{(d)}$ is $\hat{T}^{(d)} = \sum_{U^{(d)}} \hat{y}_i - \sum_{s^{(d)}} w_i \hat{e}_i$, where $\hat{e}_i = I_{\{i \in A\}} - \hat{y}_i$ and $\hat{y}_i$ is prediction from a statistical model. If the model is linear with fixed effects, we call this estimator GREG-lin, if logistic, GREG-log, respectively.

The accuracy of GREG estimator with respect to functional form of the model has been studied in [4], [5], [2] and [6]. Results indicate that for class frequencies, GREG-log is more accurate than GREG-lin. However, the Sen-Yates-Grundy (SYG) variance estimator $\sum\sum_{U^{(d)}} w_i \hat{e}_i w_k \hat{e}_k$, which is often used for GREG-lin, seems to underestimate the variance of GREG-log, and especially if i) domains are minor[4], ii) if the assisting model is complex in the sense that it has a large number of covariates[6], and iii) if auxiliary information is very strong[6]. In addition, variance of SYG often becomes unbearably large in these cases[6].

Using Monte Carlo simulation, I study the performance of four variance estimators for GREG-log for domains under simple random sampling without replacement (SRSWOR). The baseline estimator is SYG, which under the SRSWOR design is $N^2 n^{-1}(1 - n/N)\hat{S}^2_{e^{(d)}}$. The reason for this estimator failing as well as the performance of three alternative variance estimators are studied. The alternative estimators are standard iid bootstrap[7], bootstrap without replacement[1],[3], and delete-one jackknife[8]. These estimators are externally scaled in a standard way so that the linear condition (unbiasedness in the case of linear estimator) is fulfilled. Performance of the variance estimators is then compared by means of bias, MSE, and coverage rate.

References

[1] Bickel, P. J. – Freedman, D. A. (1984): Asymptotic Normality and the Bootstrap in Stratified Sampling. The Annals of Statistics, Vol. 12, No. 2, 470-482.
[2] Duchesne, P. (2003). Estimation of a Proportion with Survey Data. J. of Stat. Educ. Vol. 11, No. 3.
[3] Gross, S.T. (1980). Median estimation in sample surveys, ASA Proc. of Surv. Res. Met. Sect.: 181-4.
[4] Lehtonen R. and Veijanen A. (1998). Logistic generalized regression estimators. SMJ 24, 51-55.
[5] Lehtonen, R., Särndal, C.-E. and Veijanen, A. (2003). The effect of model choice in estimation for domains, including small domains. SMJ 29, 33-44
[6] Myrskylä, M. (2004). Estimation of class frequencies with micro level auxiliary information. Master's thesis (unpublished), University of Jyväskylä.
[7] Sitter, R.R. (1992). A Resampling Procedure For Complex Survey Data, JASA, 87, 755-765.
[8] Wolter, K. M.(1985). Introduction to Variance Estimation. New York: Springer-Verlag.

[1] Statistics Finland, Box 5V FI-00022, email: mikko.myrskyla@stat.fi

EBLUP estimation of small area totals for unit-level panel data

Kari Nissinen[1]

The EBLUP estimation (see e.g. Rao, 2003) is one of the most common approaches to estimation of small area totals or means. In the EURAREA research project (The EURAREA consortium, 2004) an EBLUP estimator as well as estimator of its mean squared error was derived for the case, where unit-level data from previous time points were available for each area to be utilized in the estimation of current small area totals by appropriate mixed models. However, these models, with possibly autocorrelated area-level random effects, were defined only for the non-panel case, where different units were observed at different time points.

The purpose of this work is to adjust the EBLUP estimator for the case of panel or rotated panel survey data, where there are short time series available for each unit. In the underlying model the unit-level error terms are assumed to follow the AR(1) covariance structure. The performance of the estimator is illustrated with an application to Finnish survey data.

References

Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New York.

The EURAREA consortium (2004). *Project Reference Volume D7.1.4.*

[1]Department of Mathematics and Statistics
University of Jyväskylä, Finland
e-mail: knissine@maths.jyu.fi

Small Area Estimations in the Industrial Survey of the Basque Country

Ana Fernandez Militino[2], Patxi Garriddo[1], Haritz Olaeta[1] and Lola Ugarte[2]

The increasing demand of small area estimations has forced Eustat (the Statistical Institute of the Basque Country in Spain) to work on small area estimation techniques for different surveys. In this work we briefly describe the Fixed Effects Model and the Linear Mixed Model that are being used in the Industrial Survey and show the first estimates for comarcas (i.e. administrative clusters of municipalities) that will be released in September. Several issues concerning the estimate production and release policies of special relevance in official statistics, specially for statistical offices dealing with small population sizes will be addressed.

[1] Eustat, Vitoria-Gasteiz, Spain
[2] Universidad Publica de Navarra, Pamplona, Spain
Address for correspondence: Haritz Olaeta, Eustat, Donostia 1, 01010 Vitoria-Gasteiz, Spain.
Email: H_Olaeta@eustat.es

Education of experts in small area statistics

Erkki Pahkinen[1]

Achieving expertise in order to produce small area statistics is considered. A short review of commonly used analysis strategies and statistical tools in small area research shows that skills of different academic disciplines are needed. A small area specialist should master topics in mathematical statistics, survey methodology and widely in experimental mathematics for performing simulation and disclosure tasks. In addition, professional skills are needed for empirical research in some appropriate substance field as for example in economics, social statistics or epidemiology. This knowledge helps to collect relevant auxiliary information and to interpret analysis results. Thus the learning process of a student can not be completed by some academic courses only. Good knowledge can be reached by advanced studies of different disciplines and by participating in practical research at some organization, which produces small area statistics. Timing and content of this kind of training program is sketched.

Key words: Multidisciplinary know-how, team work, political relevance.

[1] Emeritus professor, University of Jyväskylä, Finland
E_mail pahkinen@maths.jyu.fi

Adaptation of EURAREA experience in business statistics in Poland

Jan Paradysz, Tomasz Klimanek[1]

Based on the experiences of the EURAREA project an attempt was made to use small area statistics methods to improve estimation precision with respect to basic information about economic activities of small businesses.

At first available data sources and the possibility of integrating them are discussed. The SP3 survey is presented in terms of sample size, sample design and the range of information estimated. Then the characteristics is enlarged by using additional information from other data sources like: Database of Statistical Units - BJS, Register of Economic Entities REGON and Tax Register POLTAX. The data sources consistence is discussed. We assess them in terms of how useful they are to provide information across type of economic activity (PKD sections) and regions combined.

Application of indirect estimation methods in compliance with the standards developed within the EURAREA project and necessary modifications is presented. An analysis of indirect estimation precision in comparison with traditional estimation techniques is provided.

Another problem considered is the sample allocation for the SP3 survey in terms of optimisation criteria used in small area estimation. We take into account the optimal sample allocation in terms of direct and composite estimators and compare them with the examined sample size across domains and the estimation precision obtained.

Our concerns about the lack of population homogeneity were confirmed. It turns out that it can affect the use of estimators involving values of auxiliary variables at the unit level since GREG Synth_A and EBLUP_A estimators do not comply with direct estimator results. The study confirmed our hypothesis that one method of coping with the high level of variation in distributions of estimated variables is to construct applicable models at the domain level rather than at the unit level. Estimates obtained by Synth_B and EBLUP_B estimators show more compliance with those produced by direct estimators.

[1] The Poznan University of Economics

Geographic information in Small Area Estimation. Small area models with spatially correlated random area effects.

Alessandra Petrucci[1], Monica Pratesi[2], Nicola Salvati[2]
alex@ds.unifi.it, m.pratesi@ec.unipi.it, salvati@ec.unipi.it

Small area indirect estimators are often based on area level random effects models. Under this class of models, when only aggregate specific covariates are available. the Best Linear Unbiased Predictor (BLUP) is obtained under the assumption of uncorrelated random area effects (Fay and Herriot, 1979). The EBLUP takes advantage of the between small area-variation. The evidence is that the EBLUP estimator is significantly better than the sample-size dependent estimators, especially when the between small area-variation is not large relative to the within small area variation (Rao and Choudhry , 1995). This suggests that the location of the small areas may also be relevant in modelling the small area parameters and that further improvement in the EBLUP estimator can be gained by including eventual spatial interaction among random area effects (Petrucci, Salvati. 2004a; Pratesi. Salvati, 2005). Spatially correlated effects can also have a pragmatic role (Cressie, 1991). Ideally all relevant variables are chosen in the model. or proxies for them appear in the regression relation. These variables -and the dependent variable -often all vary spatially, so the benefit obtained from including spatial dependence is presumed to be considerable. In addition, it should be noted that small area boundaries are generally defined according to administrative criteria without considering the eventual spatial interaction of the variable of interest. As a result, there is no reason to exclude the assumption that the random effects between the neighbouring areas are correlated and that the couelation decays to zero as distance increases.

This work deals an extension of the Fay-Herriot model with spatial couelation between the random small area effects modelled through the Simultaneously Autoregressive (SAR) process (Petrucci, Salvati, 2004a; Pratesi, Salvati, 2005). The best linear unbiased predictor under this model is called Spatial BLUP. Its empirical version (EBLUP) is obtained and an estimator of its MSE is proposed. Relative performances of the Spatial EBLUP are evaluated through a Monte Carlo experiment (Pratesi, Salvati, 2005).

Moreover, in some study it happens that, some small areas are not represented in the sample. This problem can be addressed specifying a nested error unit level regression model with dependent area level random effects (Petrucci, Salvati, 2004b). Allowing area random effects to be spatially correlated, the Empirical Best Linear Unbiased Predictions for the area parameters can be computed, taking into account also the contribution of the random part of the model, for sampled areas as well as out of sample areas (Saei, Chambers, 2005).

The properties of various estimators are evaluated applying the proposed estimator to two environmental case studies.

[1] Dipartimento di Statistica "G. Parenti, Universita degli Studi di Firenze. viale Morgagni. 59- 50134 Firenze
[2] Dipartimento di Statistica e Matematica Applicata all.Economia. Universita degli Studi di Pisa. via Ridolfi. 10- 56124 Pisa

References

Cressie, N. (1991): Small-Area Prediction of Undercount Using the General Linear Model. *Proceedings of the Statistic Symposium 90: Measurement and Improvement of Data Quality,* Ottawa: Statistics Canada, 93-105.

Fay, R.E., Herriot, R.A. (1979): Estimates of income for small places: an application of James-Stein procedures to census data, *Journal of the American Statistical Association*, 74, 269-277.

Petrucci Alessandra, Salvati Nicola (2004a): Small Area Estimation Using Spatial Information. The Rathbun Lake Watershed Case Study, *Working Paper n° 2004/02, Dipartimento di Statistica "G. Parenti", Firenze.*

Petrucci, A., Salvati, N. (2004b): Small Area Estimation considering spatially correlated errors: the unit level random effects model, *Working Paper n° 2004/10, Dipartimento di Statistica "G. Parenti", Firenze*

Pratesi, M., Salvati, N. (2005): Small Area Estimation: the EBLUP estimator with autoregressive random area effects, *Report n° 261, Dipartimento di Statistica e Matematica Applicata all ' Economia, Pisa.*

Rao, J.N.K., Choudhry, G.H. (1995): Small Area Estimation: Overview and Empirical Study in Business Survey Method, Edited by Cox, Binder, Chinnappa, Christianson, Colledge, Kott, John Wiley & Sons, 38,527-540.

Saei, A, Chambers, R. (2005): Empirical Best Linear Unbiased Prediction for Out of Sample Areas, *Working Paper M05/03, Southampton Statistical Sciences Research Institute, University of Southampton.*

Logistic regression models in small area investigations

Krystyna Pruska[1]

Different qualitative variables are considered in statistical investigations. Some of them have only two variants of values (categories). We can assume that these variables have Bernoulli distribution. Logistic regression models are applied to the analysis of such variables. It means these models are used to the analysis of binary data.

If we consider a population analysed with respect to Bernoulli variable and some auxiliary variables then we can construct logistic regression models for this population.

In this paper we consider a population divided into M small areas: $A_1,...,A_M$. We assume that Y is an investigated variable and x_i is a vector of auxiliary variables for i-th small area. Moreover, a distribution of Y is given by the function:

$$P(Y=1|A_i) = \theta_i \text{ and } P(Y=0|A_i) = 1 - \theta_i \quad \text{for } i = 1,...,M, \tag{1}$$

where θ_i is unknown.

We construct the following logistic regression model:

$$L^{-1}(p_i) = x_i'\alpha + \varepsilon_i \qquad \text{for } i = 1,...,M, \tag{2}$$

where

$$L^{-1}(p_i) = \ln \frac{p_i}{1-p_i} \tag{3}$$

and p_i is an estimator of parameter θ_i, α is the model parameter, ε_i is a random error, $E(\varepsilon_i) = 0$.

We draw m small areas from M small areas. We determine the values of estimators p_i for drawn areas on the basis of small area sample ($i = 1,...,m$).We estimate the parameters of model (2) on the basis of data for drawn areas. Next we determine the estimates of θ_i for undrawn areas on the basis of model (2).

We can consider different sampling methods and different estimators of θ_i for drawn small areas.

In this paper an estimation of variance of estimator p_i ($i = 1,...,M$) is considered, too. The simulation experiments are conducted for this purpose. The estimator p_i is determined on the basis of small area sample for drawn m areas and on the basis of model (2) for other areas.

[1] Chair of Statistical Methods
University of Lodz
Poland
kpruska@uni.lodz.pl

A restricted model approach to improve the precision of estimators

Cristina Rueda and José A. Menéndez[1]

In this paper we propose a new approach to small area estimation that uses the methodology of constrained statistical inference (CSI) to improve the precision of direct estimates. The idea is to formulate a linking model for related domains using prior knowledge that is incorporated as restrictions on the model parameters.

The proposed estimators are indirect domain estimators and could be developed using explicit or implicit models, which will be called restricted models.

We focus on the estimation of a domain mean. Consider m domains and their means, $Y = (Y_1,..., Y_m)'$, as the parameters of interest and also consider the total mean $\overline{Y}$. The corresponding sample means based on a sample on each domain give the direct estimators, $\hat{Y} = (\hat{Y}_1,...,\hat{Y}_m)'$ and $\hat{\overline{Y}}$ respectively.

An example of a simple restricted model is given when the information $\overline{Y} \geq 0$ is included in the model. The corresponding restricted estimator is then $\hat{Y}^r = p(\hat{Y} / C_+)$, where $C_+ = \{ v \in R^m : \sum_{i=1}^{m} v_i \geq 0 \}$ and $p(y/C)$ is the projection of y onto a cone C. We introduce a more complex model when supplementary information in the form of an auxiliary variable $X = (X_1,....,X_m)'$ is available and a monotone relationship between X and Y exists, which can be formulated as follows: $X_i \leq X_j \Rightarrow Y_i \leq Y_j$. Statistically, this information is incorporated in the estimation process through the order relationship $\leq_x$ induced by X, the order cone $C_X = \{ v \in R^m : v_i \leq v_j \text{ if } i \leq_X j \}$ and the restriction $Y \in C_X$. The corresponding estimator is then $\hat{Y}^r = p(\hat{Y} / C_X)$. Intuitively, we would expect to do better by incorporating such additional information than by ignoring them.

In a similar way a restricted model could be defined using an explicit linear mixed constrained model given by $\hat{Y}_i = \mu_i + v_i + e_i,\ \mu \in C_X$. In this case a new restricted estimator would be obtained that have not been referenced before in the literature. In these and other similar models the properties of restricted and related estimators must be compared with classical alternatives to small area estimation.

The properties of the restricted estimator $\hat{Y}^r = p(\hat{Y} / C_X)$ have been extensively studied in the CSI literature, but as far as we know there have been no applications to small area problems. The more relevant result for the small area context is that using the criterion of the Mean Square Error (MSE),

[1] Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Valladolid. C/ Prado de la Magdalena s/n, 47005 Valladolid, Spain
E-Mail: crueda@eio.uva.es

$\sum_{i=1}^{m} E(\hat{Y}_i^r - Y_i)^2$, the restricted estimator performs much better than the direct estimator when the hypothesis $Y \in C_X$ is true.

Taking into account constraints in the models produces a reduction of the parameter or sample space. We think that it would be possible to design methods that properly incorporate the constraints in the models, producing efficient estimators for small area applications.

In this first attempt we have considered the simplest model where only the information $\overline{Y} \geq 0$ is available.

We propose a family of estimators defined by $\hat{Y}^w = P(\hat{Y} / C_w)$ where C_w is a circular cone, $C_w = \left\{ v \in R^m : \langle c, v \rangle \leq \cos(w) \| v \| \right\}$, $\| c \| = 1$ and $w \in [0, \pi/2]$. Particular cases are the synthetic estimator, $\hat{Y}^{w=0} = \hat{\overline{Y}}$, and the restricted estimator, $\hat{Y}^{w=\pi/2} = p(\hat{Y} / C_+) = \hat{Y}^r$. In other cases $\hat{Y}^w$ could be considered as a kind of composite estimator. An *"empirical restricted"* estimator is selected from de above family in two steps. In the first step an optimum angle is defined by $\hat{w}_{opt} = \arg\min_w E(\hat{Y}^w - Y)^2$ and in the second step a plug-in estimator is obtained from the sample as $\hat{Y}^{\hat{w}_{opt}}$. In this paper we study some properties of this estimator and compare it with other classical counterparts as the positive part James-Stein estimator.

Generalized Structure Preserving Estimation Models for SmaII Areas

Ayoub Saei[1], Li-Chun Zhang[2], Ray Chambers[3]

The structure preserving estimation (SPREE) method improves the small area estimates when no auxiliary information other than from past census is available. In this paper we generalise the SPREE in two ways. The first model adds coefficients in association with census values to allow for possible changes in the association structure. Area random effect is included to account for the variation that is not explained by auxiliary information in the second model. The random effects are allowed to vary with response category levels

Estimates of the parameters in the models are obtained by using maximum likelihood and residual/restricted maximum likelihood methods. The small area estimates are called empirical best linear unbiased-type (EBLUP-type) estimates. The approach is applied to Italian Labour Force Survey (LFS) and Italian household composition at NUTS3 level. We report results from simulation study of the performance of the new method. In this study the area random effect is a normal variable with a general variance-covariance structure between response category levels.

Key words: EBLUP , Labour Force Survey, REML, Structure Preserving Estimation

[1] Southampton Statistical Sciences Research Institute, University of Southampton, Highfield, Southampton SO 17 1 BJ, UK, axs96@,soton.ac.uk
[2] Statistics Norway, PostBoks 1831, Dep, 0033 Oslo, li.chun.zhang@.ssb.no
[3] Southampton Statistical Sciences Research Institute, University of Southampton, Highfield, Southampton SO 17 1 BJ, UK, rc6@yoton.ac.uk

Small area estimation in the Spanish Labour Force Survey

Jorge Saralegui[1], Montserrat Herrador[1], Domingo Morales[2], Agustín Pérez[2].

Once the EURAREA project was finished, the activities of the research group of experts which had participated in the EURAREA project focused on assimilating and applying project results to real data provided by surveys from the National Statistical System. In addition to an overview of such activities within the framework of the Spanish Labour Force Survey, some small area estimators of ILO unemployment totals and rates are presented. Survey and aggregated data are taken from the autonomous community of Catalonia in the second trimester of 2003. Practical problems appearing when applying small area estimation techniques are described and, from the analysis of the obtained results, some recommendations are given. In addition a naïve two-stage bootstrap method is proposed to introduce performance measures to compare estimators.

Key words and phrases: Labour force survey, small area estimation, linear models, mean square error, bootstrap, unemployment totals, unemployment rates.

AMS subject classification: 62E30, 62J12.

[1] Instituto Nacional de Estadística
[2] Centro de Investigación Operativa, Universidad Miguel Hernández de Elche

General restriction estimator in small area estimation

Kaja Sõstra[1]

Several techniques have been introduced for small area estimation. The performance of small area estimators depends on sample size: model-assisted estimators perform better in large areas and model-dependent in relatively smaller areas. Using different estimators for small and large areas can cause problem that estimated totals of areas do not sum up to population total. The focus of this paper is to investigate possibilities to use general restriction estimator in small area estimation to solve this problem.

The performance of two small area estimators are compared in the paper: 1) Generalised Regression estimator (GREG); 2) Empirical Best Linear Unbiased Predictor (EBLUP). Simulation study showed that quality of GREG and EBLUP estimators depends significantly on the sample size of area. GREG-estimator performs better in relatively large areas according to mean square error (MSE). EBLUP estimator tends to overestimate large areas especially with informative sampling where units with large value of study variable have higher inclusion probability.

For better results it is appropriate to use different estimates for smaller areas and large sub-populations. Obtained estimates do not satisfy the criteria that the sum of estimated small area totals is equal to estimated population total or estimated totals of large domains. One solution of the problem is general restriction estimator developed by Knottnerus (2003). My simulation study showed that general restriction estimator is good procedure for calibration the small area estimators to meet certain conditions. In addition restriction estimators perform slightly better than EBLUP estimator.

References

Knottnerus, P. (2003) *Sample Survey Theory*. Springer, New York:

Rao, J.N.K. (2003) *Small Area Estimation*. Wiley, New York.

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

The EURAREA consortium (2004). *Project Reference Volume D7.1.4*

[1] Statistical Office of Estonia, Estonia
University of Tartu, Estonia
e-mail: kaja.sostra@stat.ee

Register-based statistics and geographic information

Marja Tammilehto-Luode[1]

Register-based statistics are compiled from data that usually cover domains exhaustively, for example, all inhabitants, houses or businesses. The total coverage of the data makes it possible to aggregate statistical units depending on the accuracy of their location. All buildings in Finland have map co-ordinates, which make it possible to locate residents and businesses into them. So at a first sight there appear to be no special challenges in Finland to the compilation of statistics on small areas from a register-based statistical system. In theory, data secrecy is the only restriction to the production of small area statistics.

However, there are also other challenges to the production of good quality small area statistics from a register-based statistical system. In my presentation I will discuss these challenges from two perspectives: the register-based statistical system itself and the use of geographic information in statistics production.

A register-based statistical system has many pros and cons. It has been said that the results from a register-based census are at least as reliable as the results from a conventional census made by interviewers or questionnaires. However, there is certain controversy about how the quality of a register-based system should be described and documented. There is no existing theory for assessing the accuracy of statistics based on administrative registers (Platek and Särndal 2001). As a matter of fact, there is lack of methodological knowledge about the quality of register information, e.g. nature and meaning of errors and missing information. In my presentation I will compare the quality factors of a sample survey and a statistical register and discuss some special characteristics of statistical registers and register-based studies.

When geographic information is used in statistics production, the location of statistical units and /or statistical regions does not only help in the classification of data. Accurate locational identifiers of statistical units, such as building co-ordinates in Finland, make it possible to delimit statistical areas flexibly but also to calculate distances between objects and formulate comparable density indicators between different areas as examples. However, mappable statistics also present new challenges. Geographical information possesses quality factors that cannot be measured or managed by statistics. Instead of uncertain information one could talk about imprecise information (Niskanen 1998). In my presentation I will consider why geographic information is an essential part of a register-based statistical system and discuss major challenges to the use of geographic information in statistics production.

[1] Statistics Finland, marja.tammilehto-luode@stat.fi

Small area estimation by combining spatially misaligned data

Nicola Torelli, Matilde Trevisani[1]

The paper addresses the problem of small area estimation by using data defined on different partitions of the relevant territory. Namely, we will show how appropriate (area level) models, within the hierarchical Bayesian setting, can allow to use data on covariates available on non nested areal partitions to provide small area estimates.

As a motivating example we consider the estimation of the number of unemployed for Local Labour Markets (LLMs) in Italy by using two misaligned source data, i.e. design based estimates for LLMs from the Italian Labour Force Survey and auxiliary information about the number of enrolled in Labour Exchange Offices available for an administrative partition incompatible with LLMs subdivision.

Keywords: spatial misalignment; hierarchical Bayesian methods; atombased models.

[1] Dipartimento di Scienze Economiche e Statistiche Universita di Trieste
mailing address: Dipartimento di Scienze Economiche e Statistiche, Facolta di Economia, Universita di Trieste, Piazzale Europa 1 -34127 Triestej
e-mail:matildet@econ.units.it

The effect of model quality on model-assisted and model-dependent estimators of totals and class frequencies for domains

Ari Veijanen[1], Risto Lehtonen[2] and Carl-Erik Särndal[3]

We compare the effects of model quality on model-assisted and model-dependent estimators of totals and class frequencies for population subgroups or domains. As an example of a model-assisted method, we consider the generalized regression (GREG) estimator. The GREG estimator is known to retain good properties even when the model is incorrect; it is always nearly design-unbiased, for example. We compare GREG with a synthetic estimator, defined as the sum of fitted values over each domain. A synthetic estimator is known to have small variance, but it may suffer from considerable design bias. If the bias is large, a confidence interval is misleadingly narrow and does not cover the true value with the desired degree of confidence. Synthetic estimators can be expected to depend heavily on model quality. The paper draws on results in Lehtonen, Särndal and Veijanen (2003, 2004) and on more recent research (Lehtonen, Veijanen and Särndal 2005).

We study four aspects of model quality: (1) The mathematical form of the model. This aspect is likely to be particularly important for binary variables, for which logistic models are preferred to linear ones. (2) The kind of auxiliary information included in the model. We assume that the auxiliary variables values are known for all population elements and that domain membership is known for all population units. We can expect that inclusion of domain indicators in the model is important. (3) Should we formulate a fixed domain effects model or a mixed model with random effects for each domain? (4) How sensitive are GREG and synthetic estimators to outlying domains or outliers in the data?

Our simulation experiments concerning quality aspects (1)-(3) are based on repeated sampling from a fixed population. Our study of aspect (4) involves simulation of populations from a superpopulation model. The sampling weights were constant throughout the population, so as not to create an unfair advantage for the GREG methods (considering that in synthetic and other model-dependent methods, the sampling weights are usually ignored).

In the experiments, model improvement has a distinct impact on the accuracy of synthetic estimators, especially in large domains. For the synthetic estimators, the inclusion of domain indicators in the model was important. Without them, the synthetic estimators are highly inaccurate. A mixed model was found preferable to a model with fixed domain indicators. Model improvement was a much less important factor for the GREG estimators. As expected, they were nearly unbiased regardless of the model, whereas the bias of the synthetic estimators was sometimes large.

Nevertheless, the synthetic estimators usually had smaller mean squared error than GREG estimators. An exception to this was found in a robustness study with a single outlier domain. The presence of the outlier domain reduced the benefits of synthetic estimation. For a distinctly deviating domain, the GREG estimator assisted by mixed model was clearly better than corresponding

[1] Statistics Finland
[2] University of Jyväskylä
[3] Canada

synthetic estimator. The synthetic estimator is affected by the problem of estimated random effects that are biased towards zero, whereas the bias correction in the GREG estimator provides robustness. A general conclusion is that GREG estimators are little affected by the quality of model and they may in many cases be preferable to synthetic estimators, especially when the underlying model is of questionable quality.

References

Lehtonen R., Särndal C.-E. and Veijanen A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, 24, 51-55.

Lehtonen R., Särndal C.-E. and Veijanen A. (2004). Does the model matter? A comparison of the accuracy of model-assisted and model-dependent estimators of domain totals. (Under review for Survey Methodology).

Lehtonen R., Särndal C.-E. and Veijanen A. (2005). Robustness of GREG estimators and synthetic estimators of domain totals to outlying domains. (Draft manuscript).

On mean square error of EBLU predictors based on the formula of Royall's BLU predictor

Tomasz Żądło[1]

Key words: model approach in survey sampling, general linear model, general mixed linear model, BLUP and EBLUP

In the paper we consider the problem of the Best Linear Unbiased and the Empirical Best Linear Unbiased Predictors under the general mixed linear model. The BLU predictor was proposed by Henderson (1950) (following Rao (2003)). Formula of the BLU predictor includes unknown elements of the variance-covariance matrix of random variables. If the elements in the formula of the BLU predictor proposed by Henderson (1950) are replaced by some type of estimators, we will obtain the two-stage predictor called the EBLU predictor which is model-unbiased (Kackar and Harville (1981)). Kackar and Harville (1984) gave an approximation to the MSE of the predictor and proposed an estimator of the MSE. The MSE and estimators of the MSE were also studied by Prasad and Rao (1990), Datta and Lahiri (2000), Das, Jiang and Rao (2004).

In the paper we consider the BLU predictor proposed by Royall (1976). Żądło (2004) showed that the BLU predictor proposed by Royall (1976) may be treated as a generalisation of the BLU predictor proposed by Henderson (1950) and proved model unbiasedness of the EBLU predictor based on the formula of the BLU predictor proposed by Royall (1976) under some assumptions. In the paper we derive the formula of approximate MSE of the EBLU predictor and its estimators. We prove that the approximation of the MSE is accurate to terms $o(D^{-1})$ and the estimator of the MSE is approximate unbiased in the sense that its bias is $o(D^{-1})$ under some assumptions, where D is the number of domains. The proof may be treated as a generalization of the results received by Datta and Lahiri (2000). Using our results we present some BLU and EBLU predictors based on special cases of the general linear model and formulae of their MSEs and estimators of their MSEs.

References

[1] Das K., Jiang J., Rao J.N.K (2004), Mean squared error of empirical predictor, *The Annals of Statistics,* Vol. 32, No.2, 818-840.

[2] Datta G. S. and Lahiri P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613-627.

[3] Henderson C.R. (1950). Estimation of genetic parameters (Abstract). *Annals of Mathematical Statistics*, 21, 309-310.

[4] Kackar R.N. and Harville D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics*, Series A, 10, 1249-1261.

[5] Kackar R.N. and Harville D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association,*

[6] Prasad N.G.N and Rao J.N.K. (1990). *The estimation of mean the mean squared error of small area estimators*. Journal of the American Statistical Association, 85, 163-171.

[7] Rao J.N.K. (2003). *Small area estimation*. John Wiley & Sons, New York.

[8] Royall R.M. (1976). The linear least squares prediction approach to two-stage sampling. *Journal of the American Statistical Association*, 71, 657-473.

[9] Żądło T. (2004) On unbiasedness of some EBLU predictor. In: *Proceedings in Computational statistics 2004*, eds. Antoch J., Physica-Verlag, Heidelberg-New York, 2019-2026

[1] University of Economics in Katowice, Department of Statistics, Bogucicka 14, 40-226 Katowice, Poland. E-mail: zadlo@ae.katowice.pl

A prediction approach to sampling design

Li-Chun Zhang and Ib Thomsen[1]

In standard approach to sampling, whether model-based or not, the main aim is to estimate one or several finite population totals, or some predefined sub-totals. Within the model-based prediction approach the implications on sampling design can be extreme as e.g. in the case of purposive selection for ratio regression populations. This is mainly because a purposive sample is unlikely to be suitable for other uses than the inference of population totals (or means). For instance, we may want to use the data for micro simulations in an econometric model. Or we may need the data for small area (or domain) estimation.

An alternative point of departure is individual prediction. Formally, consider the class of functions $\sum_{i \in U} \lambda_i y_i$, where $U = \{1,...,N\}$ denotes the finite population, and the λ_i's are fixed constants such that $\sum_{i \in U} |\lambda_i| = 1$, and y_i is the variable of interest for the *i*th unit. The population mean is given by setting $\lambda_i = 1/N$. Whereas in the prediction of any individual we set $\lambda_i = 1$ for that chosen unit and let $\lambda_i = 0$ otherwise, which is in some sense the linear function within the above class that differs most from the one implied by the prediction of population total. Moreover, for general database-like uses of the survey data, a natural criterion for sampling design is to make the unconditional individual prediction MSE equal for all the units in the population.

We derive the equal prediction designs for a number of populations. The models we consider here are primarily suitable for continuous variables. It turns out that balancing between equal prediction of the individuals and optimal prediction of the population totals provides a useful model-based approach to sampling design. In this talk we will concentrate on some implications on small area estimation by this prediction approach to sampling design.

[1] Statistics Norway

ANNEX

Sunday 28th August

11:00–18:00	**Registration** Building MaA, 1st floor lobby
12:00–18:00	**Short Course on Tools for Small Area Estimation** Featuring SAS macro programs developed for small area estimation Organizer *Kari Djerf* Statistics Finland
18:30–23:30	**Sauna Party** **Ladun Maja Recreation District** (transportation will be arranged)

Monday 29th August

8:00–15:30	**Registration** Building MaA 1st floor lobby
9:00– 9:15	**Opening** *Risto Lehtonen* University of Jyväskylä **Room MaA102**
9:15– 9:45	**Plenary Session** *Patrick Heady* Office for National Statistics (ONS), UK **Room MaA102** Chair *Montserrat Herrador*
9:45–10:30	**Plenary Session** *Danny Pfeffermann* Hebrew University and University of Southampton **Room MaA102** Chair *Montserrat Herrador*
10:30–11:00	Coffee
11:00–12:30	**Session 1** Temporal and spatial models and GIS **Room MaA102** Chair *Domingo Morales*
11:00-11:30	*Marja Tammilehto-Luode*
11:30-11:50	*Esmail Amiri*
11:50-12:10	*Coro Chasco-Yrigoyen*
12:10-12:30	*Nicola Torelli & Matilde Trevisani*
11:00–12:30	**Session 2** Applications **Room MaA211** Chair *Jan Kordos*
11:00-11:30	*Michele D'Alò, Loredana Di Consiglio, Stefano Falorsi & Fabrizio Solari*
11:30-11:50	*Enrico Fabrizi, Maria Rosaria Ferrante & Silvia Pacei*
11:50-12:10	*Danute Krapavickaite*
12:10-12:30	*Michal Mlady*
12:30–13:30	Lunch
13:30–15:00	**Plenary Session** *J.N.K. Rao* Carleton University **Room MaA102** Chair *Li-Chun Zhang*
15:00–15:30	Coffee
15:30–17:00	**Session 3** Estimation of uncertainty **Room MaA102** Chair *Lola Ugarte*
15:30-16:00	*Nicholas Longford*
16:00-16:20	*González-Manteiga W., Lombardía M.J., Molina I., Morales D. & Santamaría L.*
16:20-16:40	*Mikko Myrskylä*
16:40-17:00	*Tomasz Żądło*
15:30–17:00	**Session 4** Applications **Room MaA211** Chair *Paavo Väisänen*
15:30-16:00	*Jorge Saralegui, Montserrat Herrador, Domingo Morales & Agustín Pérez*
16:00-16:20	*Grazyna Dehnel & Elzbieta Golata*
16:20-16:40	*Wojciech Gamrot*
16:40-17:00	*Jan Paradysz & Tomasz Klimanek*

17:10–18:00 **Poster Session**
MaA, 1st floor lobby
Organizer *Kari Nissinen*
Poster presentations *Emanuela Conza / Ana Militino, Patxi Garrido, Haritz Olaeta & Lola Ugarte / Kari Nissinen / Cristina Rueda & José A. Menéndez*

19:00–20:30 **Welcoming Reception**
City Hall Vapaudenkatu 32

Tuesday 30th August

8:00–15:30 **Conference office** Building MaA 1st floor lobby

8:00– 8:50 **Early Bird Session 5** New SAE developments
Room MaA102
Chair *Marie Cruddas*
8:00- 8:30 *Gauri Datta*
8:30- 8:50 *Ray Chambers & Nikos Tzavidis*

9:00–10:30 **Plenary Session** *Chris Elbers* Vrije Universiteit Amsterdam, *Jenny Lanjouw* University of California, Berkeley & *Peter Lanjouw* The World Bank
Room MaA102
Chair *Patrick Heady*

10:30–11:00 Coffee

11:00–12:30 **Session 6** Temporal and spatial models and GIS
Room MaA102
Chair *Ulrich Rendtel*
11:00-11:30 *Alessandra Petrucci, Monica Pratesi & Nicola Salvati*
11:30-12:00 *Philip Clarke, Fernando Moura & Danny Pfeffermann*
12:00-12:30 *Stephen Haslett & Geoff Jones*

11:00–12:20 **Session 7** Design and weighting issues
Room MaA211
Chair *Imbi Traat*
11:00-11:30 *Piero Demetrio Falorsi, Stefano Falorsi, Paolo Righi & Fabrizio Solari*
11:30-12:00 *A.F. Militino, M.D. Ugarte & T. Goicoa*
12:00-12:20 *Krystyna Pruska*

12:30–13:30 Lunch

13:30–15:00 **Plenary Session** *Carl-Erik Särndal* University of Montreal
Room MaA102
Chair *Timo Alanko*

15:00–15:30 Coffee

15:30–17:00 **Session 8** Evaluation of SAE methods
Room MaA102
Chair *Martin Ralphs*
15:30-16:00 *Ari Veijanen, Risto Lehtonen & Carl-Erik Särndal*
16:00-16:20 *Hukum Chandra & Ray Chambers*
16:20-16:40 *Kari Djerf*
16:40-17:00 *Kaja Sõstra*

15:30–17:00 **Session 9** Applications and miscellaneous
Room MaA211
Chair *Dan Hedlin*
15:30-16:00 *Jan Kordos*
16:00-16:20 *Julia Aru*
16:20-16:40 *Natalja Jurevitš*
16:40-17:00 *Erkki Pahkinen*

18:45–23:30 **Conference Dinner**
Varjola Farm Restaurant (transportation will be arranged)

Wednesday 31st August

8:00–13:00	**Conference office** Building MaA 1st floor lobby
8:00– 9:00	**Early Bird Session 10** New SAE developments **Room MaA102** Chair *Seppo Laaksonen*
8:00- 8:30	*Ayoub Saei, Li-Chun Zhang & Ray Chambers*
8:30- 9:00	*Li-Chun Zhang & Ib Thomsen*
9:00– 9:45	**Plenary Session** *Michel Hidiroglou & Marie Cruddas* Office for National Statistics, UK **Room MaA102** Chair *Nick Longford*
9:45–10:30	**Plenary Session** *Danny Pfeffermann* Hebrew University and University of Southampton **Room MaA102** Chair *Nick Longford*
10:30–11:00	Coffee
11:00–12:45	**Panel Discussion** Future Challenges of Small Area Estimation **Room MaA102** Organizer and chair *Ray Chambers* University of Southampton Discussants *Jan van den Brakel, Dan Hedlin, Michel Hidiroglou/Marie Cruddas, Risto Lehtonen, Imbi Traat, Li-Chun Zhang*
12:45–13:00	**Closing**

List of Participants

LAST NAME	FIRST NAME	ORGANIZATION OR CONTACT	EMAIL
Alanko	Timo	Statistics Finland	timo.alanko@stat.fi
Amiri	Esmail	Imam Khomeini International University	e_amiri@yahoo.com
Aru	Julia	University of Tartu	julia_a@ut.ee
Balea	Cornelia	National Institute of Statistics	cecilias@insse.ro
Bihler	Wolf	Statistisches Bundesamt	wolf.bihler@destatis.de
Chambers	Raymond	University of Southampton	rc6@soton.ac.uk
Chandra	Hukum	University of Southampton	hchandra@soton.ac.uk
Chasco-Yrigoyen	Coro	Universidad Autónoma de Madrid	coro.chasco@uam.es
Clarke	Philip	Office for National Statistics	philip.clarke@ons.gov.uk
Colistru	Elena	Department for Statistics and Sociology	statistic_md@mail.ru
Conza	Emanuela	Universita' degli Studi di Lecce	emanuelaconza@libero.it
Cruddas	Marie	Office for National Statistics	marie.cruddas@ons.gov.uk
Culliford	David	University of Southampton	djc202@soton.ac.uk
D'Alò	Michele	ISTAT	dalo@istat.it
Datta	Gauri	University of Georgia	gauri@stat.uga.edu
Dehnel	Grazyna	The Poznan university of Economics	g.dehnel@ae.poznan.pl
Djerf	Kari	Statistics Finland	kari.djerf@stat.fi
Elbers	Chris	Vrije Universiteit Amsterdam	celbers@feweb.vu.nl
Fabrizi	Enrico	University of Bergamo	enrico.fabrizi@unibg.it
Gamrot	Wojciech	University of Economics	gamrot@ae.katowice.pl
Golata	Elzbieta	The Poznan university of Economics	elzbieta.golata@ae.poznan.pl
Gómez Rubio	Virgilio	Imperial College London	v.gomezrubio@imperial.ac.uk
Halmeenmäki	Tuomo	The Local Government Pensions Institution	tuomo.halmeenmaki@keva.fi
Haranen	Michael	University of Jyväskylä	miharan@cc.jyu.fi
Haslett	Stephen	Massey University	s.j.haslett@massey.ac.nz
Haslinger	Alois	Statistik Austria	alois.haslinger@statistik.gv.at
Heady	Patrick	Office for National Statistics	patrick.heady@ons.gov.uk
Hedlin	Dan	Statistics Sweden	dan.hedlin@scb.se
Hernandez Jimenez	Francisco	National Institute of Spain	fhernan@ine.es
Herrador	Montserrat	Instituto Nacional de EStadistica (INE)	herrador@ine.es
Hidiroglou	Michel	Office for National Statistics	mike.hidiroglou@ons.gsi.gov.uk
Högmander	Harri	University of Jyväskylä	hogmande@maths.jyu.fi
Jurevits	Natalja	University of Tartu	junat@math.ut.ee
Kankainen	Annaliisa	University of Jyväskylä	kankaine@maths.jyu.fi
Karhunen	Jukka	University of Jyväskylä	jikarhun@cc.jyu.fi
Klimanek	Tomasz	The Poznan university of Economics	t.klimanek@ae.poznan.pl
Kokkonen	Elina	University of Jyväskylä	elmakokk@cc.jyu.fi
Konnu	Janika	University of Jyväskylä	jakonnu@cc.jyu.fi
Kordos	Jan	Warsaw School of Economics	j.kordos@stat.gov.pl
Krapavickaite	Danute	Institute of Mathematics and Informatics	krapav@ktl.mii.lt
Kärkkäinen	Salme	University of Jyväskylä	samk@maths.jyu.fi
Laaksonen	Seppo	Statistics Finland	seppo.laaksonen@helsinki.fi
Laiho	Johanna	Statistics Finland	johanna.laiho@stat.fi
Lavikainen	Piia	Tilastokeskus	piia.lavikainen@stat.fi
Lehto	Kristi	Statistical Office of Estonia	kristi.lehto@stat.ee
Lehtonen	Risto	University of Jyväskylä	risto.lehtonen@maths.jyu.fi
Lensu	Anssi	University of Jyväskylä	anssi@maths.jyu.fi
Liberts	Martins	Central Statistical Bureau of Latvia	pm90015@lu.lv
Lombardía Cortiña	María-José	Universidad de Santiago de Compostela	mjoselc@usc.es
Longford	Nicholas	SNTL	ntl@sntl.co.uk

LAST NAME	FIRST NAME	ORGANIZATION OR CONTACT	EMAIL
Masiulaityte	Inga	Statistics Lithuania	inga.masiulaityte@std.lt
Menendez	Jose Antonio	University of Valladolid	josan@eio.uva.es
Mielityinen	Markku	University of Jyväskylä	mmmm@cc.jyu.fi
Militino	Ana	Universidad Publica de Navarra	militino@unavarra.es
Mlady	Michal	EUROSTAT	michal.mlady@cec.eu.int
Molina	Isabel	Universidad Carlos III de Madrid	isabel.molina@uc3m.es
Morales	Domingo	Universidad Miguel Hernandez	d.morales@umh.es
Musat	Sofica	National Institute of Statistics	smusat@insse.ro
Myrskylä	Mikko	Statistics Finland	mikko.myrskyla@stat.fi
Nedeljkovic	Ranko	Statistical Office of Serbia and Montenegro	ranko@szs.sv.gov.yu
Nikic	Boro	Statistical Office of the Republic of Slovenia	boro.nikic@gov.si
Nissinen	Kari	University of Jyväskylä	knissine@maths.jyu.fi
Ogbozor	John Paul	Nigerian Institute of Statistics and Demography	mosesolawuyi@yahoo.com
Olaeta	Haritz	Eustat (Basque Statistical Institute)	h_olaeta@eustat.es
Olawuyi	Moses	Institute of Statistics and Demography	mosesolawuyi@yahoo.com
Ollila	Pauli	Statistics Finland	pauli.ollila@stat.fi
Pahkinen	Erkki	University of Jyväskylä	pahkinen@maths.jyu.fi
Paradysz	Jan	The Poznan university of Economics	jan.paradysz@ae.poznan.pl
Petrucci	Alessandra	Univeristà di Firenze	alex@ds.unifi.it
Pfeffermann	Danny	Hebrew University	msdanny@huji.ac.il
Piela	Pasi	Statistics Finland	pasi.piela@stat.fi
Pratesi	Monica	Universitá de Pisa	m.pratesi@ec.unipi.it
Pruska	Krystyna	University of Lodz	kpruska@uni.lodz.pl
Putcha	Venkata	Thames Cancer Registry	venkata.putcha@kcl.ac.uk
Ralphs	Martin	Office for National Statistics	martin.ralphs@ons.gov.uk
Ranalli	Maria Giovanna	Universita' degli Studi di Perugia	giovanna@stat.unipg.it
Rao	Jon N. K.	Carleton University	jrao@math.carleton.ca
Rendtel	Ulrich	Freie Universität Berlin	rendtel@wiwiss.fu-berlin.de
Righi	Paolo	ISTAT	parighi@istat.it
Rueda	Cristina	University of Valladolid	crueda@eio.uva.es
Räikkönen	Eija	Statistics Finland	eija.raikkonen@stat.fi
Saei	Ayoub	Southampton University	axs96@soton.ac.uk
Salonen	Riku	Statistics Finland	riku.salonen@stat.fi
Sirkiä	Seija	University of Jyväskylä	ssirkia@maths.jyu.fi
Solari	Fabrizio	ISTAT	solari@istat.it
Särndal	Carl-Erik	University of Montreal	carl.sarndal@rogers.cOM
Sõstra	Kaja	Statistical Office of Estonia	kaja.sostra@stat.ee
Tammilehto-Luode	Marja	Statistics Finland	marja.tammilehto-luode@stat.fi
Taskinen	Sara	University of Jyväskylä	slahola@maths.jyu.fi
Traat	Imbi	University of Tartu	imbi.traat@ut.ee
Turunen	Reijo	Jyväskylän yliopisto	returune@cc.jyu.fi
Tzavidis	Nikolaos	University of Southampton	ntzav1@socsci.soton.ac.uk
Tångdahl	Sara	Statistics Sweden	sara.tangdahl@scb.se
Ugarte	Lola	Universidad Publica de Navarra	lola@unavarra.es
Veijanen	Ari	Statistics Finland	ari.veijanen@stat.fi
Verkasalo	Pia	National Public Health Institute	pia.verkasalo@ktl.fi
Väisänen	Paavo	Statistics Finland	paavo.vaisanen@stat.fi
Zadlo	Tomasz	University of Economics	zadlo@ae.katowice.pl
Zhang	Li-Chun	Statistics Norway	lcz@ssb.no