STATISTICS IN TRANSITION new series and *SURVEY METHODOLOGY*
Joint Issue: Small Area Estimation 2014
Vol. 16, No. 4, pp. 485–488

FROM THE GUEST EDITORS (PART 1)

The first part of this Joint Issue of Statistics in Transition and Survey Methodology includes eight articles. These two issues have been split according to which guest editors have been looking after the articles. They are not necessarily sequenced according to the themes that appeared in the original conference programme.

The first six contributions in this thematic issue of SIT and SMJ represent articles that are firmly methodological in their perspective. The first paper, by J.N.K. Rao provides a unifying perspective for the remaining five contributions. In this review paper, Rao highlights important new developments in SAE since the publication of his encyclopedic 2003 book. As he notes in his abstract, much of this new methodological development has focused on addressing the practical issues that arise when model-based SAE methods are applied in practice. An important dichotomy in this regard follows from the nature of the available data for SAE. Historically, such data have been area level aggregates of one form or another, typically direct sample-based estimates. Issues addressed in Rao's paper then include the choice of appropriate weights for these aggregates as well as methods for dealing with the not uncommon situation where there is a negligible area level variance component in the basic area-level model (the so-called Fay-Herriot model) used to smooth these aggregates across the areas, or where this smoothing model is necessarily non-linear, reflecting a GLM for the underlying survey variable. Issues associated with estimation of both unconditional as well as conditional MSEs of these model-based estimators are also discussed. In the second half of his paper, Rao switches his attention to SAE where unit level data from the small areas of interest are available. This is a fast-growing set of applications, reflecting new capabilities in data collection. Here, the focus is on sample weighting and benchmarking as important requirements for users interested in design consistency of SAE outputs, together with important new developments in dealing with outliers in the survey data, applications to poverty mapping and dealing with informative sampling methods. Model selection and checking is extremely important in the unit level case, and the paper briefly describes some new developments in this regard.

The next three papers in this issue focus on a new methodology for area level SAE. The first, by Bonnery, Cheng, Ha and Lahiri, notes that users of SAE outputs typically require more than just estimates of area averages, and are often

interested in small area distributions as well as rankings across small areas. In this context, these authors develop a triple goal SAE methodology for US state level unemployment, with estimates structured so that they are simultaneously efficient for estimation of area level average unemployment as well as the empirical distribution of area level unemployment, while also staying as close as possible to the actual ranking of the real small area means. An interesting idea that is discussed in this paper is the fact that in practice it is not just one area average that is of interest, but an "ensemble" of such averages corresponding to the area-level distribution of a characteristic of interest. This immediately leads to a corresponding ensemble of models, which these authors fit using a Bayesian MCMC approach.

The general theme of the usefulness of incorporating time series information in SAE solution is repeated in the paper by van den Brakel and Buelens. Here, though the attention is directed towards appropriate model specification when the estimation must be carried out at regular intervals, using data from repeated surveys and practical considerations rule out survey-specific model optimisation. An approach to covariate selection for small area survey estimates obtained from a repeated survey under a Fay-Herriot specification is defined, with the model specification carried out simultaneously over a number of "editions" of the survey while being constrained to be the same for each edition. The final model is chosen by minimising the average conditional AIC over all the editions, with the small area estimates at each time period computed using a Hierarchical Bayes approach.

The next paper, by Karlberg, switches gears and considers SAE under a unit level model. In particular, in this paper Karlberg addresses two of the difficult issues that arise when the available unit level data are non-negative values drawn from an economic population, as would be the case for a business survey. These conditions often lead to a highly right-skewed distribution of the sample data values, with outliers a not uncommon feature, together with the presence of excess zeros. Both of these data characteristics are not conducive to SAE based on the industry standard linear mixed model for unit level data. Instead, Karlberg combines a log scale linear mixed model for the strictly positive data (to deal with their high skewness) and a logistic model for the presence of zero values (a hurdle model) in order to define a specification for the zero-inflated observed data. Simulation results for SAE based on this approach are promising, but application to a real business survey data set turns out to be disappointing, reflecting the very complex nature of such data. Clearly further research is needed for SAE in business surveys.

The fifth paper, by Franco and Bell, shows how the Fay-Herriot approach can be extended to where the underlying averages are derived from binary survey variables, so that the basic area-level model can be specified as linear on a logit

scale. This model is then combined with time series of aggregates from the small areas, allowing for information to be "borrowed" across both time and space. An application to improving county-level poverty estimates in the SAIPE programme of the US Bureau of the Census is used to demonstrate the efficiency gains of the approach.

The sixth paper, by Luna, Zhang, Whitworth and Piller, represents a fundamental departure from the random area effect-based SAE models that underpin the previous papers. Here, the underlying data consist of historical counts, represented by an out-of-date census (or register)-based cross-tabulation of interest, where one of the dimensions of the tabulation is the area identifier, as well as up-to-date information on margins of the cross-tabulation derived from a current survey. Such data are naturally modelled using a log-linear specification, and the authors consider the use of a generalized SPREE approach to recover the current cross-tabulation. Alternative GSPREE models with increasingly complex interaction structure are investigated and applied to estimation of population counts within ethnic group in small areas in the United Kingdom. Interestingly, these authors report that for these data more complex model specifications do not necessarily lead to improvement in the resulting survey estimates, essentially because the sparse nature of the available data does not allow these more complex models to be adequately fitted.

The last two contributions focus on small area education. Small area estimation is gaining increasing popularity among survey statisticians, economists, sociologists and many others. Unfortunately, small area courses are offered only in a handful of universities and that too just as an elective. However, there is a definite need for small area teaching, and the papers by Burgard and Münnich as well as Golata have addressed this very important issue. The paper by Burgard and Münnich has hit the mark very directly. What the paper emphasizes is that rather than giving a series of lectures on the different small area techniques and the associated theory behind them, it is more important to combine the theory with actual simulations. In this way, students can have hands on experience of the subject as well as are able to make a comparison of the different small area methods which they have learnt. Like Burgard and Münnich, Golata also appreciates very well the need for small area education. To this end, she conducted a survey with participants from both the academics and National Statistical Institutes. Her objective went beyond questions on small area teaching, and enquired several related pertinent questions such as risks encountered in applying SAE as well as important sources on SAE developments. The results of her survey are listed in a series of tables and graphs to provide the reader with a better understanding of the state of the art.

Several persons (in addition to the Editor and Guest Editors) have served as reviewers of papers published in this thematic issue of the journal: we would like to thank all the authors for taking the time to turn their SAE 2014 presentations into the interesting and thought provoking papers published here. We acknowledge the efforts of Giovanna Ranalli, Nicola Salvati, Hukum Chandra and Timo Schmid, who helped review the first six papers: their encouraging and productive comments directly contributed to their obvious quality.

Raymond Chambers and **Malay Ghosh**
Guest Editors