Editorial Introduction

Raymond Chambers
University of Wollongong, Australia

This issue of Survey Research Methods contains four papers that were presented at the ITACOSM11 conference held in Pisa, Italy, June 27–29, 2011. ITACOSM is an Italian scientific conference on the development, testing, and application of survey sampling methodologies in the fields of economics, social and demographic sciences and official statistics and in studies of biological and environmental phenomena. It is organized by the Permanent Working Group on Sample Survey Methodology of the Italian Statistical Society.

The four papers published here represent a good cross-section of current Italian research on sample survey design and estimation. They describe new methodologies for survey design based on multivariate auxiliary information, using imputation to account for survey nonresponse, semiparametric methods for small area estimation and outlier robust survey design.

Three of the papers focus on business surveys. The paper "Multivariate Boundaries of a Self-Representing Stratum of Large Units in Agricultural Survey Design" by Benedetti and Piersimoni describes an innovative approach to defining stratum boundaries when a multivariate size measure is available. In particular, this paper considers the problem of defining the boundary between a 'take all' stratum consisting of large population units and a 'take some' (i.e. sampled) stratum consisting of the remaining population units for this case. The authors use an optimization algorithm based on simulated annealing to simultaneously determine the smallest overall sample size and the corresponding optimal partition of the population in order to achieve a minimum specified level of precision for each of a set of target variables.

The paper "Robust Lavallée-Hidiroglou Stratified Sampling Strategy" by Bramati considers a similar sample design problem for a business survey, but in the context of a population containing both outliers in the survey variables as well as influential values of a single auxiliary size variable. This author takes a model-based approach, assuming that the regression relationship between a survey variable and the auxiliary variable can be modelled multiplicatively (or additively on a log scale). She then modifies the well-known Lavallée-Hidiroglou algorithm for defining a set of optimal size stratum boundaries and corresponding stratum allocations (including a 'take all' stratum made up of the largest population units) to make it robust to extreme size values in the population.

Missing data due to sample refusals or non-contacts are a pervasive problem for sample surveys, including business surveys, and a common strategy for dealing with them is to adopt some form of reweighting of the responding sample units. However, depending on the response propensity model used for this purpose, this strategy can lead to extreme final sample weights for some of these units. The paper "Handling Nonresponse in Business Surveys" by Borgoni, Marasini and Quatto considers this problem and advocates a weight-trimming strategy. Since such trimmed weights no longer define unbiased estimates under the assumed propensity model, these authors develop a bootstrap-based approach to determing an optimal trimming threshold that minimises the estimated mean squared error.

The last of the four papers "Robust Small Area Estimation and Oversampling in the Estimation of Poverty Indicators" by Giusti, Marchetti, Pratesi and Salvati is somewhat different. The focus here is on estimation, and the increasingly important application of model-based methods to small area (or to be more precise, small sample) estimation. Since the quality of the small area estimates produced via these methods is dependent on the appropriateness of the assumed model, there is always a question about their adequacy. Here the authors consider the special situation where small area estimates of household poverty for the ten provinces making up the Tuscany region of Italy were required, but increased funding meant that a much larger sample could be taken in just one of these ten provinces (Pisa). The authors compare provincial estimates derived using a semi-parametric M-quantile modelling approach based on data from the original Tuscany region sample with direct estimates based on the same sample, as well as with the direct estimates for Pisa obtained using the enhanced Pisa sample. Their results indicate that the model-based estimates stand up well to this comparison.

There is always a big gap between the version of a paper presented at a conference and the final version that is published. Bridging this gap requires considerable effort from the authors and constructive input from anonymous referees. The high quality of these four papers reflects both sources. It was a privilege to be able to be part of this process.