The Survey Statistician, 2023, Vol. 88, 23–35.

Historical Overview of Small Area Estimation in the 50$^{\text{th}}$ Birthday of the IASS

Isabel Molina[1] **and J. N. K. Rao**[2]

[1]Institute of Interdisciplinary Mathematics, Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Spain, isabelmolina@ucm.es

[2]School of Mathematics and Statistics, Carleton University, Ottawa, Canada, jrao@math.carleton.ca

Abstract

To celebrate the 50th birthday of the IASS, this paper presents a historical overview of SAE methods, focusing on the main ideas and theories that have had a significant impact in the SAE methodology. Starting from estimators obtained under design-based theory, we describe simple indirect methods, including synthetic and composite estimation procedures. Then we go through model-based SAE methods, starting with area-level models, then going through unit-level models and finally describing the more up-to-date procedures for the estimation of complex non-linear indicators, such as poverty and inequality indicators. Due to the applied nature of SAE, we enhance applications of the methods, describing important government programs that regularly produce SAE estimates.

Keywords: Area effects; Mixed models; Model-based inference; Poverty mapping; Small domain.

1 Introduction

Launched at the 39th ISI conference held in Vienna in August, 1973, the IASS was founded as a section of the ISI by Tore Dalenius, Ivan Fellegi, Morris Hansen, Leslie Kish and P.C. Mahalanobis, so this paper is written to celebrate its 50th birthday.

As Anders Christianson notes in "Aims and history" of the IASS (http://isi-iass.org/home/aims/), apart from being devoted to promote survey sampling, "the most important reason for the creation of the IASS was to address major limitations of sampling theory". The field of small area estimation (SAE) was actually born to address a major limitation of traditional design-based sampling theory, to meet the (public and private) demands of estimates at more disaggregated levels than those for which surveys were originally planned. "Quick and cheap" disaggregated yet reliable statistical information was needed worldwide in policy making, for the formulation of assistance and development programs, or directly for the allocation of government funds in an efficient way.

SAE came out of the shelter of sampling theory in the hands of other disciplines and theories, such as demography and model-based inference, and started growing exponentially by the second half of 20th century, partly thanks to the development and expansion of linear and generalized linear mixed regression models. By the 21st century, this growth has been also stimulated by the pressure that international organizations, like The United Nations, have put on countries to meet the Millennium Development Goals from 2000 to 2015, and the Sustainable Development Goals after 2015. Appropriate monitoring of the progress of these goals relies on timely, accurate disaggregated statistical information.

The IASS has also promoted the expansion of SAE by co-sponsoring several SAE conferences and organizing SAE sessions and short courses on SAE within the IASS meetings. The importance of SAE within the IASS is also witnessed by the many SAE researchers that have been or are currently involved in the IASS committees.

This increasing need of detailed statistical information has lead to the development of a variety of SAE methods that are specific for the type of estimates that need to be produced and the possibilities offered by the information that is available for that. In the US, many of these SAE methods have long been used in official programs to produce regular estimates. For example, the Small Area Income and Poverty Estimates (SAIPE) Programme of the U.S. Census Bureau (https://www.census.gov/programs-surveys/saipe.html), which started back in 1993, produces estimates of school-age children in poverty, regularly for the counties and school districts. The Local Area Unemployment Statistics (LAUS) Program of the Bureau of Labour Statistics produces local monthly and annual employment, unemployment, and labor force statistics. The County Estimates Program of the National Agricultural Statistics Service (NASS) produces county crop yield estimates. The Substance Abuse and Mental Health Services Administration produces estimates of substance abuse in states and metropolitan areas. The US Department of Health and Human Services produces health status, health care access and family income estimates. The latter estimates are used to formulate an energy assistance program for low-income families. For an excellent account of the use of indirect estimators in US Federal Programs, see Schaible (1996).

In Canada, reliable monthly unemployment rates for small areas are used to determine the rules used to administer the employment insurance (EI) program. In Latin America, many countries are currently producing small area estimates of poverty. For example, in Mexico, there is a mandate to produce poverty estimates by municipality every 5 years, and the Mexican National Survey of Household Income and Expenditure (ENIGH in Spanish) alone cannot provide estimates for the municipalities with adequate quality. In Europe, the SAE methodology expanded greatly thanks to projects funded by the European Commission, like EURAREA (https://cros-legacy.ec.europa.eu/content/eurarea_en), SAMPLE (http://www.sample-project.eu/) and AMELI (https://cros-legacy.ec.europa.eu/content/ameli_en). The Italian National Statistical Office uses since 2006 unit-level SAE methods to obtain employment and unemployment indicators for Labour Market Areas(https://www.istat.it/en/archivio/276035). Central and eastern European countries, which moved away from a centralized decision making, have also played a prominent role in the expansion of the SAE methodology, participating in European projects and organizing several conferences related with SAE; for example, the first two international conferences on SAE were held in Warsaw in 1992 (Poland) and in Riga (Latvia) in 1999. Worldwide, the World Bank and the United Nations, specially the Economic Commission for Latin America and the Caribbean (ECLAC), the Economic and Social Commission for Western Asia (ESCWA), UN Statistics Division (UNSD) and UN Population Fund (UNPF), among possibly other, have sponsored multiple activities aimed at building capacities for countries to produce accurate disaggregated socio-economic statistical information, see e. g. the UN Toolkit on SAE (https://unstats.un.org/wiki/display/SAE4SDG).

Here we make a very limited historical overview of the literature on small area estimation, starting from direct methods based on area-specific survey data, going through simple indirect methods that include synthetic and composite estimators, more advanced indirect methods based on models at the area and unit levels, with the many different variants depending on the target indicators and the available data, and finishing with procedures designed for the estimation of general, possibly non-linear, area parameters. We place emphasis on the ideas and theories that represented breakthroughs in SAE, focusing specially on mainstream model-based SAE methods and mentioning practical applications of many of the methods.

Books on SAE include Mukhopadhyay (1998), Rao (2003), Longford (2005), Chaudhuri (2012), Rao and Molina (2015), and the recent book by Morales et al. (2021). Good accounts of SAE theory are also given in the books by Fuller (2009), Chambers and Clark (2012), Pratesi (2016), Jiang (2017) and Sugasawa and Kubokawa (2022).

Important reviews on SAE are given in Ghosh and Rao (1994), Pfeffermann (2002, 2013), Jiang and Lahiri (2006), Datta (2009), Lehtonen and Veijanen (2009) and Ghosh (2020). Reviews focused on SAE for welfare and poverty are given by Guadarrama, Molina and Rao (2014), Pratesi and Salvati (2016), Rao and Molina (2016), Molina (2019), Molina, Rao and Guadarrama (2019) and, more recently, Molina, Corral and Nguyen (2022).

2 From direct estimation to early indirect methods

The first estimates based on sample surveys that were intended for subpopulations were "direct", in the sense that they used only the survey data from the subpopulation of interest without "borrowing strength". These estimates are developed under the umbrella of sampling theory, which has long history. For nice accounts of this theory, see the books by Cochran (1977), Särndal, Swensson and Wretman (1992), Thompson (1997), Lohr (1999) and Wu and Thompson (2020). Direct estimators have several advantages, when applied to areas with large sample sizes. The usual direct estimators have good design properties (at least design consistency as the area sample size n_d increases) and avoid making distributional assumptions for the study variable. Another important advantage of direct estimators is that they use "all-purpose" expansion weights, in the sense that the same expansion weights are used for the estimation of totals or means of whatever variable of interest, making the production of large amounts of statistical information automatic.

Generalized Regression (GREG) estimators and more general calibration estimators (Deville and Särndal, 1992; Lehtonen, Särndal and Veijanen, 2003) applied to domains were designed to improve the efficiency of direct domain estimators, owing to the knowledge of the domain totals of some auxiliary variables. These procedures adjust the sampling weights, and the adjusted weights can be used similarly to estimate totals or means of other variables of interest. Nowadays, expansion weights are typically calibrated using the known totals of certain auxiliary variables and are also adjusted for non-response. However, the resulting calibration estimators are still inefficient for areas with small sample size n_d. Even if a more efficient allocation of the total survey sample size n among the different areas at the design stage of the survey (which is recommendable if estimates need to be produced for those areas) might ameliorate the SAE problem, "the client will always require more than is specified at the design stage" (Fuller 1999; p. 344).

The way of addressing the scarcity of data within some of the areas is to obtain indirect estimates, which "borrow strength" across areas, by making some homogeneity assumptions that link the areas through common parameters. These common parameters are estimated with a larger sample size, which leads to more efficient small area estimators. The idea of sharing information within a larger area appeared already in the first demographic methods dating back to 1950, such as the Vital Rates (VR) method due to Bogue (1950). This method assumed that the ratios between the birth/death

rates in two time periods in the small area of interest were constant within a larger area covering that small area. These first indirect methods used only census data and demographic information from administrative records, and were absent of sampling. Detailed accounts of the traditional demographic methods are given by Purcell and Kish (1979), National Research Council (1980), Rives, Serow, Lee and Goldsmith (1989), Statistics Canada (1987), Zidek (1982) and Rao (2003).

The VR method is "synthetic", because the change in the birth/death rate between two time periods is assumed to be the same for all the small areas contained in the larger area, without allowing for specific area behaviour. According to Gonzalez (1973), "An estimator is called a synthetic estimator if a reliable direct estimator for a large area, covering several small areas, is used to derive an indirect estimator for a small area under the assumption that the small areas have the same characteristics as the large area". Post-stratified synthetic estimators, which assume that the means of the study variable do not vary within large post-strata and only vary between post-strata, are perhaps the simplest synthetic estimators based on survey data. The US National Center for Health Statistics (1968) pioneered the use of synthetic estimation for developing state estimates of disability and other health characteristics from the National Health Interview Survey (NHIS), because NHIS sample sizes in many states were too small to provide reliable direct state estimates. Synthetic estimators can have very small design variances, but their design bias can be substantial because the assumptions behind synthetic estimators are typically strong and unrealistic. Since their design bias is not negligible, design MSE estimates that account for both bias and variance should be used to accompany the synthetic point estimates. Apart from the potentially large bias, a problem is that obtaining efficient and area-specific design MSE estimates is still a challenge for these estimators.

Composite estimators, defined as a weighted average of a synthetic estimator and a direct estimator for the same area, were proposed as a compromise between the small design variance but potentially large bias of synthetic estimators and the small design bias but inefficiency of direct estimators. Curiously, averaging different predictors is nowadays one of the main ideas behind modern machine learning procedures.

In the composite SAE estimators, optimal weights are sought from a design-based standpoint. However, the optimal weight depends on the true design MSE estimates of the two estimators involved, encountering again the problem of estimation of the design MSE for synthetic estimators. Griffiths (1996) studied composite estimators and applied them to the estimation of labor force characteristics for US congressional districts.

Purcell and Kish (1979) considered a common weight for all the areas and obtained the optimal weight that minimized the total design MSE for all the D small areas. The resulting composite estimators have good overall efficiency for the D areas, but not necessarily for each small area. In SAE, it is desirable to reduce the largest MSEs, which typically correspond to the areas with the smaller sample sizes, and this is not ensured by these composite estimators.

Composite estimators shrink direct estimators toward the synthetic ones. The idea of shrinking appears already in the James-Stein (JS) method proposed by James and Stein (1961), see also Efron and Morris (1972) and the famous application by Efron (1975) to the estimation of batting averages of major league baseball players in US during 1970 season. In the JS method, direct estimators are shrunk toward a fixed guess of the true quantity for area d, which can be taken as the average across areas of the direct estimators in the absence of auxiliary information, or to the regression-synthetic estimator when auxiliary information is available. This method applies again a constant weight to the two estimators involved, but in SAE, it is much more appealing to consider area-specific weights, with weight attached to the synthetic estimator that grows for the areas with small area sample sizes and decreases for the areas with large sample sizes (giving then more weight to the direct estimator). Following this idea, Drew, Singh and Choudhry (1982) proposed the sample-size dependent

(SSD) estimators, which are composite estimators defined with simple weights that depend on the area sample size. They applied these estimators to produce estimates for Census Divisions from the Canadian Labor Force Survey. In practice, as it happened in the application by Drew, Singh and Choudhry (1982), SSD estimators borrow little or no strength, because the weights attached to the direct estimators often turn out to be either equal or close to one.

The advent of computers produced an explosion in the number and complexity of SAE procedures, most of them based on regression models. The first SAE models actually lead also to composite estimators, but with optimality properties under model assumptions for the study variable. These estimators dominate the above composite estimators by borrowing substantial strength from the other areas. They can achieve large efficiency gains, provided that the model assumptions hold. An important drawback of model-based estimation procedures is that all the modelling and estimation process, including model validation, is specific to each variable of interest, not allowing for automatic production. This might be one of the reasons why there is a delay in the introduction of SAE procedures in the production processes of National Statistical Offices.

3 From the first explicit model to modern area level models

Perhaps the first application of a model for SAE is due to Hansen, Hurwitz and Madow (1953), p. 483, based on the 1945 Radio Listening Survey. The target was to estimate the median number of radio stations heard during the day in the family houses from 500 U.S. counties. They had estimates x_d, $d = 1, \ldots, D = 500$, obtained from a mail survey conducted in the 500 counties, which were biased due to only 20% response rates and incomplete coverage. Unbiased estimates y_d were obtained from an intensive survey conducted in 85 of the counties. A linear regression model for y_d with x_d as auxiliary variable was used, by regarding the y_d as true values for the 85 sample counties. The fitted regression parameters were then applied to predict the number of radio stations heard during the day in the remaining 415 counties, where the mail survey estimates x_d were available. The resulting predicted values do not account for the fact that y_d are subject to sampling error.

The use of linear mixed models (Searle, 1971; Searle, Casella and McCulloch, 1997; Jiang, 2007) that account for unexplained area heterogeneity really represented a breakthrough in the SAE methodology. The best linear unbiased predictor (BLUP) of a mixed effect (a linear combination of fixed and random effects) under a linear mixed model was obtained by Henderson (1950), in a different context from SAE, related with the prediction of the milk yield of dairy cows. On a completely different context, dealing with estimation of mean per capita income in US areas with less than 1,000 inhabitants, Fay and Herriot (1979) also considered a linear regression of the true area means μ_d in terms of certain area-specific covariates x_d (linking model). However, to account for the (important) sampling errors of the direct estimators y_d of μ_d, they considered an additional sampling model for y_d in terms of μ_d, which, together with the linear regression for μ_d, yields a linear mixed model, known popularly as the Fay-Herriot (FH) model. Based on this model, Empirical BLUPs (EBLUPs) of the true area means μ_d were obtained.

The FH model is still very popular nowadays, because it requires only aggregated data at the area level, accounts for the survey design, through the direct estimators, and accounts for potential unexplained between-area heterogeneity. As a consequence, the resulting EBLUP is a weighted average of the direct and the regression-synthetic estimator, with area-specific weights. Actually, the weight attached to the regression-synthetic estimator is larger for areas where the direct estimator is inefficient (large sampling errors) and smaller for areas where the direct estimator is efficient. The property of approaching the direct estimator as the area sample size grows is appealing, because it ensures design consistency as the area sample size n_d grows. Moreover, if the model parameters were known, EBLUPs based on FH model cannot be less efficient than the direct estimators in

terms of MSE. FH model parameters are estimated by fitting the model to the direct estimators for all the areas (hence borrowing strength). As a consequence, the efficiency of the estimated parameters increases as the number of areas grows. Perhaps the main issue with FH model is that the sampling variances of direct estimators need to be given and are typically deemed as fixed values (without sampling error). Generalized variance function (Vaillant, 1987) is typically applied to smooth these sampling variances, and the smoothed variances are then treated as the true ones. However, when comparing the resulting EBLUPs with direct estimators in applications, it is unclear whether the comparison should be done using the estimated sampling variances or the smoothed versions.

The FH model is regularly used in the US Census Bureau, within the SAIPE project, see Bell (1997). It was also used by Ericksen and Kadane (1985) and Cressie (1989) to estimate the decennial census undercounts in each US state, and Dick (1995) employed the model to estimate Canadian census undercounts. To mention just a few applications of the FH model to estimate welfare indicators, Molina and Morales (2009) estimated poverty rates and gaps in Spanish provinces by gender, Jedrzejczak and Kubacki (2013) estimated income inequality and poverty rates by regions and family type in Poland, and Casas-Cordero Valencia, Encina and Lahiri (2015) estimated poverty rates in Chilean comunas based on the FH model with arcsin transformation of the direct estimators.

Other ways of "borrowing strength" were explored in multiple extensions of the FH model, like the multivariate versions, and models including temporal and/or spatial correlation. Recently, the FH model was extended to include area level covariates obtained from "big data" typically based on non-probability sampling. Marchetti et al. (2015) used big data based on mobility comprised of different car journeys in Italy automatically tracked with a GPS device.

The introduction of Generalized Linear Models (GLMs) by Nelder and Wedderburn (1972) (see also McCullagh and Nelder, 1989), represented a huge step that expanded the use of statistical models in general. After that, two-level GLMs were then applied to estimate mortality or disease rates and obtain corresponding mortality/disease maps. The first proposal, based on a Poisson-Gamma model, was perhaps due to Clayton and Kaldor (1987), who also introduced a model with Conditionally Autoregressive (CAR) area effects. Generalized linear mixed models, or the more general two-level GLMs, have then long been used in many disease mapping and small area applications, with many variants developed, e.g. multivariate versions, or including temporal and/or spatial correlation, etc.

4 Unit level models

The concept of a superpopulation model for two-stage sampling introduced by Scott and Smith (1969) led to important advances in SAE, specially when estimating non-linear area parameters based on unit-level data. The first unit-level model for SAE was proposed by Battese, Harter and Fuller (1988), which was a linear regression model with random area effects, popularly known as the nested error model. They used this model to obtain EBLUPs of county means of crop areas under corn and soybeans, using farm-interview data and auxiliary information obtained from LANDSAT satellite images. Although EBLUPs under a linear mixed model were derived by Henderson under the "infinite" population setup, Royall (1970, 1976) developed EBLUP theory under the finite-population setup without focusing on small areas, see Vaillant, Dorfman and Royall (2001). Current mainstream SAE procedures apply this theory to small areas, by assuming a superpopulation model that links all the areas through common parameters. These common parameters are estimated with the overall survey data from all the areas, which yields substantial increases in the efficiency of model-based estimators compared to direct estimators.

When the area sampling fractions are negligible, the EBLUP of an area mean $\bar{Y}_d$ obtained under the finite population setup with superpopulation model defined by the nested error model, approximates the EBLUP of a mixed effect from the same model under Henderson's infinite population setup, but

this is not the case for non-linear area parameters. To mention just a few other applications of the nested error model, it has been used by Militino et al. (2006) to estimate the area occupied by olive trees in non-irrigated areas at the central region of Navarra in Spain and by Mauro et al. (2015) to estimate means of forest variables of interest by forest regions, based on remote sensing auxiliary data.

Until the first decade of the current century, model-based SAE procedures had focused mainly on means or totals of the variable that is used as model response, since EBLUPs were designed to estimate only linear functions of the model response variables. However, many poverty and inequality indicators cannot be expressed as linear functions of the response variable. Even if the interest was to estimate simple area means of a given variable of interest, once a non-linear transformation (such as log) is taken as response in the model (often done for monetary variables to achieve approximate normality), EBLUPs might not be useful anymore. Note that taking the inverse transformation of EBLUP predictions might lead to severe bias, see Molina and Martín (2018).

Probably the first SAE procedure that was designed for the estimation of general parameters is that of Elbers, Lanjouw and Lanjouw (2003), known as ELL method. This method was based on the nested error model of Battese, Harter and Fuller (1988), but where the random effects in the model were associated to the sampling clusters (or 1st stage units), and including heteroscedasticity. ELL method was used until 2020 as the default method for mapping poverty or inequality at the World Bank and perhaps was the most extensively used method across the globe for that purpose. This is partly because of the simple point and click software `PovMap` software (Zhao, 2006), which was also extremely computationally fast and efficient in terms of memory.

Banerjee et al. (2006) reviewed the research conducted at the World Bank and did already raise concerns about the ELL method, suggesting that it was not accounting for potential area effects. Actually, as Molina and Rao (2010) showed, even if taking the clusters as the small areas of interest in the ELL method, the ELL estimators of the welfare means under a nested error model for the welfare without any transformation, are synthetic. Banerjee et al. (2006) also raised concerns about the ELL estimated standard errors, which were not accounting for the correlation between the observations in different clusters within the same area. These two problems were solved by the Empirical Best (EB) method and the bootstrap MSE estimation procedure proposed in Molina and Rao (2010), work that was developed under the support of the SAMPLE project.

Similar to the ELL method, EB combines survey data with census (or administrative records) auxiliary data, uses a unit-level model for the welfare variable (or a one-to-one transformation of it) and it is able to estimate very general (and several) indicators that depend on the welfare, based on the same model. Nevertheless, apart from being approximately unbiased, EB estimators are nearly optimal, in the sense of minimum mean squared error under the model. Consequently, EB provides estimators with better efficiency than ELL estimators when the nested error model assumptions hold, and in certain cases the gains in efficiency with respect to ELL may be quite large, as illustrated by Molina and Rao (2010) and later in Corral, Molina and Nguyen (2021). The EB method was implemented within the `sae` R package (Molina and Marhuenda, 2015) in the homoscedastic case, as well as in Stata (Nguyen et al., 2018, https://github.com/pcorralrodas/SAE-Stata-Package). Many SAE methods have been implemented in multiple R packages, as well as in other software packages, but a software review is out of the scope of this paper.

The EB method has been applied to estimate poverty indicators in Spanish provinces by gender (Molina and Rao, 2010), mean income in Mexican municipalities (Molina and Martín, 2018), mean income and (non-extreme) poverty rates for census tracks by gender in Montevideo, Uruguay, and poverty rates and gaps in Palestinian localities by gender (Molina Peralta and García Portugués, 2020).

Corral, Molina and Nguyen (2021) extended the model-based simulation experiment of Molina and Rao (2010) to more realistic scenarios with a much better explanatory power of the model and including also contextual variables, with much larger area population sizes and much smaller sampling fractions, generating errors from a Student's t_5 instead of a normal distribution, and also decreasing the overall sample size and the area sample sizes. Additionally, Corral et al. (2021) performed a design-based validation study, using the Mexican Intracensal Survey as a fixed census, and then drawing from it 500 samples using a realistic sampling method. The superiority, in terms of MSE, of the EB over the traditional ELL in all these experiments lead to a revision of the World Bank methodology for poverty mapping and the corresponding software (https://github.com/pcorralrodas/SAE-Stata-Package). This revision incorporates several variants of the EB estimators of Molina and Rao (2010) and the parametric bootstrap procedure for MSE estimation of González-Manteiga et al. (2008).

The nested error linear regression model has been extended to models with non-parametric mean functions. Opsomer et al. (2008) proposed penalized spline regression models. Recently, Krennmair and Schmid (2022) have used machine learning methods; in particular, mixed-effects random forests, for SAE.

5 Concluding remarks

We have made an overview of SAE methods, going from the basic direct and indirect methods to the modern model-based procedures for SAE, including methods developed for the estimation of non-linear area indicators and variants of the basic methods. Really important topics in SAE like model fitting methods and their properties, methods for MSE estimation or calculation of prediction intervals, have not been covered owing to space-time restrictions, details of those topics can be found in Rao and Molina (2015). Moreover, we have mainly focused on frequentist or empirical Bayes procedures. Descriptions of Hierarchical Bayes (HB) SAE methods can be found in Ghosh and Meeden (1997), Malec et al. (1997), Ghosh et al. (1998) and also Rao and Molina (2015).

Even if the usual SAE models that include area effects are more flexible than the corresponding regression models without the area effects (which lead to synthetic estimators), we cannot forget that properties of all model-based estimators depend on the model assumptions. Hence, the assumed model needs to be carefully checked with the available data, e.g. by using customary residual plots, see Rao and Molina (2015) for model diagnostics in the basic SAE models, although more research is probably needed on this important issue.

In the case of clear model departures, the model should be changed to accommodate to data features or the final estimates should be taken with a lot of caution. This is related to another important issue, which is the estimation of area parameters in non-sampled areas. Note that the model assumptions cannot be checked for non-sampled areas and, unless additional information is available, we cannot be sure that these areas satisfy the assumed model. Moreover, as already discussed, synthetic estimators used for those areas are inefficient if area effects are significant. Hence, unless legally bound, a general recommendation is not producing estimates for non-sampled areas.

Once the sample is drawn from the population, the model for the sample part $\mathbf{y}_s$ of the population vector $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_c')'$ (for which a superpopulation model is assumed) is simply obtained by marginalization; that is, integrating out with respect to the sample complement part $\mathbf{y}_c$. The sample model for $\mathbf{y}_s$ then has the same shape as the superpopulation model when sampling is ignorable, but this does not hold for non-ignorable (informative) sampling. Similarly, the model for the respondents might be different from the model for the sample units under non-ignorable non-response. Methods for SAE accounting for the sampling design have been discussed already. However, concerning model checking, skeptical survey samplers might raise the concern that the superpopulation model cannot be checked under informative selection and/or non-ignorable non-response, because population data

are not available. In this regard, it is important to point out that only the sample/respondents model needs to be checked with the available sample/respondents data.

Another important point is that, when estimating non-linear area parameters based on unit-level models, the values of the auxiliary variables are required for each population unit. This microdata is typically obtained from the most recent census or from administrative records, which are usually protected for privacy reasons, and this protection limits the practical applicability of these methods. Another important issue is that outdated information in the census file for inter-censal years might yield severely biased small area estimators. Corral et al. (2021) analyzed the empirical properties of the common approaches for that case, but further research is probably needed on this important issue.

Finally, conventional MSE estimates of model-based estimators are obtained assuming that the corresponding model assumptions hold, even if we know that "All models are wrong, but some are useful". Hence, these MSE estimators might be understating the real uncertainty. Molina and Strzalkowska-Kominiak (2020) and others proposed to use the same idea of "borrowing strength" behind SAE, for the estimation of the design MSE of small area means, which accounts for model uncertainty. Design MSE estimation for general non-linear indicators is an interesting topic that also deserves further research.

References

Banerjee, A. V., Deaton, A., Lustig, N., Rogoff, K., and Hsu, E. (2006) An evaluation of World Bank research, 1998-2005. Available at SSRN 2950327.

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988) An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.

Bell, W. (1997) Models for county and state poverty estimates. Preprint, Statistical Research Division, U.S. Census Bureau.

Bogue, D. J. (1950) A technique for making extensive postcensal estimates. *Journal of the American Statistical Association*, **45**, 149–163.

Casas-Cordero Valencia, C., Encina, J., and Lahiri, P. (2016) Poverty mapping for the Chilean comunas. In: *Analysis of Poverty Data by Small Area Estimation*, (ed. M. Pratesi), Wiley, New York.

Chambers, R., and Clark, R. (2012) *An Introduction to Model-Based Survey Sampling with Applications*. Oxford University Press, Oxford

Chaudhuri, A. (2012) *Developing Small Domain Statistics: Modelling in Survey Sampling*. LAP LAMBERT Academic Publishing GMbH & Co. KG, Saarbrücken.

Clayton, D., and Kaldor, J. (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671–681.

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

Corral, P., Himelein, K., McGee, K., and Molina, I. (2021) A map of the poor or a poor map? *Mathematics*, **9**(21), 2780; https://doi.org/10.3390/math9212780

Corral, P., Molina, I., and Nguyen, M. (2021) Pull your small area estimates up by the bootstraps. *Journal of Statistical Computation and Simulation*, https://doi.org/10.1080/00949655.2021.1926460

Corral, P., Molina, I., Cojonaru, A., and Segovia, S. (2022) Guidelines to small area estimation for poverty mapping. The World Bank.

Cressie, N. (1989) Empirical Bayes estimation of undercount in the decennial census, *Journal of the American Statistical Association*, **84**, 1033–1044.

Datta, G. S. (2009) Model-based approach to small area estimation. In: *Sample Surveys: Inference and Analysis*, (eds. D. Pfeffermann and C. R. Rao). *Handbook of Statistics*, Volume 29B, North-Holland, Amsterdam, 251–288.

Deville, J. C., and Särndal, C. E. (1992) Calibration estimation in survey sampling. *Journal of the American Statistical Association*, **87**, 376–382.

Dick, P. (1995) Modelling net undercoverage in the 1991 Canadian census. *Survey Methodology*, **21**, 45–54. Drew, D., Singh, M. P., and Choudhry, G. H. (1982) Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methodology*, **8**, 17–47.

Efron, B. (1975) Biased versus unbiased estimation, *Advances in Mathematics*, **16**, 259–277.

Efron, B., and Morris, C. E. (1972) Limiting the risk of Bayes and empirical Bayes estimators, Part II: The Empirical Bayes Case. *Journal of the American Statistical Association*, **67**, 130–139.

Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2003) Micro-level estimation of poverty and inequality. *Econometrica*, **71**, 355–364.

Ericksen, E. P., and Kadane, J. B. (1985) Estimating the population in census year: 1980 and beyond (with discussion). *Journal of the American Statistical Association*, **80**, 98–131.

Fay, R. E., and 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*, **85**, 398–409.

Fuller, W. A. (1999) Environmental surveys over time. *Journal of Agricultural, Biological and Environmental Statistics*, **4**, 331–345.

Fuller, W. A. (2009) *Sampling Statistics*. Wiley, New York.

Ghosh, M. (2020) Small area estimation: its evolution in five decades (with discussion). *Statistics in Transition*, **21**(4), 40–44.

Ghosh, M., and Meeden, G. (1997) *Bayesian Methods for Finite Population Sampling*. Springer, New York.

Ghosh, M., and Rao, J. N. K. (1994) Small area estimation: an appraisal (with discussion). *Statistical Science*, **9**, 55–93.

Gonzalez, M. E. (1973) Use and evaluation of synthetic estimates. In: *Proceedings of the Social Statistics Section*, American Statistical Association, 33–36.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008) Bootstrap mean squared error of a small-area eblup. *Journal of Statistical Computation and Simulation*, **78**(5), 443–462.

Griffiths, R. (1996) Current population survey small area estimations for congressional districts. In: *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 314–319.

Guadarrama, M., Molina, I., and Rao, J. N. K. (2014) A comparison of small area estimation methods for poverty mapping. *Statistics in Transition*, **1**(17), 41–66.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953) *Sample Survey Methods and Theory I*, Wiley, New York.

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

James, W., and Stein, C. (1961) Estimation with quadratic loss. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 361–379.

Jedrzejczak, A., and Kubacki, J. (2013) Estimation of income inequality and the poverty rate in Poland, by region and family Type. *Statistics in Transition-New Series*, **14**(3), 359–378.

Jiang, J. (2007) *Linear and generalized linear mixed models and their applications*. Springer-Verlag, New York.

Jiang, J. (2017) *Asymptotic Analysis of Mixed Effect Models: Theory, Applications and Open Problems*. CRC Press, Boca Raton, FL.

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

Krennmair, P., and Schmid, T. (2022) Flexible domain prediction using mixed effects random forests. *Journal of the Royal Statistical Society, Series C*, **71**(5), 1865–1894.

Lehtonen, R., Särndal, C.-E., and Veijanen, A. (2003) The effect of model choice in estimation for domains including small domains. *Survey Methodology*, **29**, 33–44.

Lehtonen, R., and Veijanen, A. (2009) Design-based methods of estimation for domains and small areas. In: *Sample Surveys: Inference and Analysis*, (eds. D. Pfeffermann and C. R. Rao). *Handbook of Statistics*, Volume 29B, North-Holland, Amsterdam, 219–249.

Longford, N. T. (2005) *Missing Data and Small-Area Estimation*. Springer, New York.

Lohr, S. L. (2010) *Sampling: Design and Analysis*. Duxbury, Pacific Grove, CA.

Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., Pappalardo, L., and Gabrielli, L. (2015) Small area model-based estimators using big data sources. *Journal of Official Statistics*, **31**, 263–281.

Mauro, F., Molina, I., García-Abril, A., Valbuena, R., and Ayuga-Télleza, E. (2015) Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. *Environmetrics*, **27**, 225–238.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. Chapman & Hall, Cambridge.

Militino, A. F., Ugarte, M. D., Goicoa, T., and González-Aud'icana, M. (2006) Using small area models to estimate the total area occupied by olive trees. *Journal of Agricultural, Biological, and Environmental Statistics*, **11**, 450–461.

Molina, I. (2019) Desagregación de datos en encuestas de hogares: metodologías de estimación en áreas pequeñas. Series of the Economic Commission for Latin America and the Caribbean (ECLAC) from United Nations, Estudios Estadísticos LC/TS.2018/ 82/Rev.1, CEPAL.

Molina, I. (2020) Discussion on "Small area estimation: its evolution in five decades", by M. Ghosh. *Statistics in Transition*, **21**(4), 40–44.

Molina, I., Corral, P., and Nguyen, M. (2022) Poverty mapping methods: a review. Test, DOI: 10.1007/s11749-022-00822-1

Molina, I., and Marhuenda, Y. (2015) sae: An R package for small area estimation. *The R Journal*, **7**, 81–98.

Molina, I., and Morales, D. (2009) Small area estimation of poverty indicators. *Boletín de Estadística e Investigación Operativa*, **25**, 218–225.

Molina, I., and Rao, J. N. K. (2010) Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, **38**, 369–385.

Molina, I., Rao, J. N. K., and Guadarrama, M. (2019) Small area estimation methods for poverty mapping: a selective review. *Statistics and Applications*, **17**, 11–22.

Molina, I., and Strzalkowska-Kominiak, E. (2020) Estimation of proportions in small areas: application to the labour force using the Swiss Census Structural Survey. *Journal of the Royal Statistical Society, Series A*, **183**, 281—310.

Molina Peralta, I., and García Portugués, E. (2020) Short guide for small-area estimation using household survey data: illustration to poverty mapping in Palestine with expenditure survey and census data. UN Economic and Social Commission for Western Asia (ESCWA), E/ESCWA/SD/2019/TP.4

Morales, D., Esteban, M. D., Perez, A., and Hobza, T. (2021) *A Course on Small Area Estimation and Mixed Models: Methods, Theory and Applications in R*. Springer, Cham, Switzerland.

Mukhopadhyay, P. (1998) *Small Area Estimation in Survey Sampling*. Narosa Publishing House, New Delhi.

National Center for Health Statistics (1968), *Synthetic State Estimates of disability*, P.H.S. Publications 1759, Government Printing Office, Washington DC, U.S.

National Research Council (1980) *Panel on Small-Area Estimates of Population and Income. Estimating Income and Population of Small Areas*. National Academy Press, Washington DC, U.S.

Nelder, J. A., and Wedderburn, R. W. M. (1972) Generalized linear models. *Journal of the Royal Statistical Society, Series A*, **135**, 370–384.

Nguyen, M. C., Corral, P., Azevedo, J. P., and Zhao, Q. (2018) Sae: A stata package for unit level small area estimation. World Bank Policy Research Working Paper 8630.

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G., and Breidt, F. J. (2008) Nonparametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society, Series B*, **70**, 265–286.

Pfeffermann, D. (2002) Small area estimation-new developments and directions. *International Statistical Review*, **70**(1), 125–143.

Pfeffermann, D. (2013) New important developments in small area estimation. *Statistical Science*, **28**(1), 40–68.

Pratesi, M. (2016) *Analysis of Poverty Data by Small Area Estimation*, Wiley, New York.

Pratesi, M., and Salvati, N. (2016). Introduction on measuring poverty at local level using small area estimation methods. In: *Analysis of Poverty Data by Small Area Estimation* (ed. M. Pratesi), Wiley, New York, 1–18.

Purcell, N. J., and Kish, L. (1979) Estimates for small domain. *Biometrics*, **35**, 365–384.

Rao, J. N. K. (2003) *Small Area Estimation*, 1st. Ed. Wiley, Hoboken, NJ.

Rao, J. N. K., and Molina, I. (2015) *Small Area Estimation*, 2nd. Ed. Wiley, Hoboken, NJ.

Rao, J. N. K., and Molina, I. (2016) Empirical Bayes and hierarchical Bayes estimation of poverty measures for small areas. In: *Analysis of Poverty Data by Small Area Estimation* (ed. M. Pratesi), Wiley, New York, 315–324.

Rives, N. W., Serow, W. J., Lee, A. S., and Goldsmith, H. F. (Eds.) (1989) *Small Area Analysis: Estimating Total Population*, National Institute of Mental Health, Rockville, MD.

Royall, R. M. (1970) On finite population sampling theory under certain linear regression. *Biometrika*, **57**, 377–387.

Royall, R. M. (1976) The linear least-squares prediction approach to two-stage sampling. *Journal of the American Statistical Association*, **71**, 657–664.

Särdnal, C.-E., Swensson, B., and Wretman, J. H. (1989) The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, **76**, 527–537.

Scott, A., and Smith, T. M. F. (1969) Estimation in multi-stage surveys. *Journal of the American Statistical Association*, **64**, 830–840.

Searle, S. R. (1971) *Linear Models*. Wiley, New York.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992) *Variance Components*. Wiley, New York.

Schaible, W. A. (1978) Choosing weights for composite estimators for small area statistics. In: *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 741–746.

Statistics Canada (1987) *Population Estimation Methods in Canada*, Catalogue 91–528E, Statistics Canada, Ottawa.

Sugasawa, S., and Kubokawa, T. (2023) *Mixed-Effects Models and Small Area Estimation*. Springer, Singapore.

Thompson, M. E. (1997) *Theory of Sample Surveys*. Chapman & Hall, London.

Valliant, R. L. (1987) Generalized variance functions in stratified two-stage sampling. *Journal of the American Statistical Association*, **82**(398), 499–508.

Valliant, R., Dorfman, A. H., and Royall, R. M. (2001) *Finite Population Sampling and Inference: a Prediction Approach*. Wiley, New York.

Wu, C. and Thompson, M. E. (2020) *Sampling Theory and Practice*. Springer Nature, Switzerland.

Zhao, Q. (2006) User manual for povmap. World Bank.
http://siteresources.worldbank.org/INTPGI/Resources/342674-1092157888460/Zhao_ManualPovMap.pdf.

Zidek, J. V. (1982) *A Review of Methods for Estimating Population of Local Areas*. Technical Report 82–4, University of British Columbia, Vancouver, Canada.