The Survey Statistician, 2024, Vol. 89, 74–75.

Book and Software Review

A Course on Small Area Estimation and Mixed Models Methods Theory and Applications in R

Caterina Giusti[1]

[1]Department of Economics and Management, University of Pisa, Italy.
caterina.giusti@unipi.it

Abstract

The book *A Course on Small Area Estimation and Mixed Models Methods. Theory and Applications in R* by Domingo Morales, María Dolores Esteban, Agustín Pérez and Tomáš Hobza was published by Springer in 2021 in the series Statistics for Social and Behavioral Sciences. The book will easily become a reference manual for researchers working in Universities and statistical offices and for PhD students who aim at studying small area estimation methodologies and mixed models from both a theoretical and applied perspective. In its 21 chapters the book covers some of the main models of small area estimation with plenty of details in the main mathematical developments, and with applications to synthetic socioeconomic indicators using R code lines.

Keywords: Small area models, random effect models, synthetic data, R coding.

In the last decades, Small Area Estimation (SAE) models have received a growing attention in the scientific literature, with a corresponding increasing number of methodological articles and manuals. However, these materials are often too advanced for researchers approaching the topic for the first time. Moreover, SAE models are relevant not only from a methodological perspective, but also for their application to obtain reliable estimates for unplanned domains in sample surveys. Therefore, training in SAE should focus not only on the mathematical developments of the estimators, but also on the relevant aspects of their application, including open software codes.

The book *A Course on Small Area Estimation and Mixed Models Methods. Theory and Applications in R* by Domingo Morales, María Dolores Esteban, Agustín Pérez and Tomáš Hobza addresses the basic aspects of the theory and practice of SAE for readers who don't need to be expert in sampling, statistical modeling, or programming languages. Specifically, as stated by the authors, the book aims at being useful to researchers from universities and statistical offices, and to doctoral students. For this reason each chapter, dedicated to a specific SAE model, reports the main mathematical developments with plenty of details, and examples of application of the models to synthetic data using R codes. The coding is as simple as possible, so that the reader can easily identify the corresponding mathematical formulas. The synthetic data files and the codes are available by chapter on a dedicated website.

The data consist in two files containing unit-level data from a labor force survey and a from a living conditions survey, together with two corresponding files with aggregated data at the domain level to be used as covariate information in the models.

The content of the book is organized into 21 self-contained chapters. The first six chapters are dedicated to the introduction of the main basic estimators and models, the following nine chapters to models focusing on SAE unit-level models, and the last six chapters to SAE area-level models.

Specifically, Chapter 1 *Small Area Estimation* introduces the basic elements of SAE and mixed models and describes the data files used in the examples with R. Chapters 2 *Design-Based Direct Estimation* and 3 *Design-Based Indirect Estimation* introduce the most popular design-based estimators of domain means and totals and describe some resampling procedures for estimating their mean squared errors. Chapter 4 *Prediction Theory* introduces the prediction theory in finite populations and the Best Linear Unbiased Predictor (BLUP) of a linear parameter and the corresponding prediction variance. Chapter 5 *Linear Models* and Chapter 6 *Linear Mixed Models* present these models in the framework of SAE.

Chapter 7 *Nested Error Regression Models* introduces the basic unit-level model in SAE, the nested error regression model. The EBLUP of domain linear parameters under this model is derived in Chapter 8 *EBLUPs Under Nested Error Regression Models*, while the corresponding MSE in Chapter 9 *Mean Squared Error of EBLUPs*. Chapter 10 *EBPs Under Nested Error Regression Models* refers to the estimation of non linear parameters (e.g. poverty indicators as the poverty incidence), introducing the Empirical Best Predictors (EBPs) and the corresponding MSE parametric bootstrap estimators. Chapter 11 *EBLUPs Under Two-Fold Nested Error Regression Models* and Chapter 12 *EBPs Under Two-Fold Nested Error Regression Models* are dedicated to the two-fold nested error regression model, taking into account the variability between domains and between subdomains inside each domain. Chapter 13 *Random Regression Coefficient Models* introduces random regression coefficient models that can be used when a more flexible model specification is needed. Finally, Chapter 14 *EBPs Under Unit-Level Logit Mixed Models* and Chapter 15 *EBPs Under Unit-Level Two-Fold Logit Mixed Models* introduce two specific unit-level mixed models belonging to the class of generalized linear mixed models.

Chapter 16 *Fay-Herriot Models* introduces the basic area-level model for SAE, with the EBLUPs of domain means and the corresponding MSE estimators. Chapter 17 *Area-Level Temporal Linear Mixed Models* and Chapter 18 *Area-Level Spatio-Temporal Linear Mixed Models* generalize the basic area-level model by considering structures of temporal and/or spatial correlation. Further extensions are introduced in Chapter 19 *Area-Level Bivariate Linear Mixed Models*, Chapter 20 *Area-Level Poisson Mixed Models* and Chapter 21 *Area-Level Temporal Poisson Mixed Models*.

In short, this book is accessible to a wide audience, including students, survey statisticians, practitioners and researchers interested in the use of SAE.