**Workshop on “Small Area Estimation and Mixed Models”**

Instituto Universitario Centro de Investigación Operativa

Universidad Miguel Hernández de Elche

October 31, 2024

**Program**

**THURSDAY October 31th, 2024**

9.30 – 9.45 **Registration**

9.45 – 10.00 **Opening session**

10.00 – 10.30 **Session 1**

Title: M-quantile regression for zero-inflated data and its applications to small area estimation.

Speaker: María Bugallo

Authors: María Bugallo, Domingo Morales, Nicola Salvati, Francesco Schirripa

10.30 – 11.00 **Session 2**

Title: Predicting gender employment discrepancies: Multivariate Fay-Herriot models for compositional data.

Speaker: Esteban Cabello

Authors: Esteban Cabello, Domingo Morales, Agustín Pérez

11.00 – 11.30 **Session 3**

Title: Small area estimation under bivariate Fay-Herriot model with correlated random effects.

Speaker: Domingo Morales

Authors: Esteban Cabello, María Dolores Esteban, Domingo Morales, Agustín Pérez

11.30 – 12.00 **Coffee Break**

12.00 – 13.00 **Session 4**

Title: Estimating Variance of Random Effects to Address Multiple Problems in Small Area Estimation.

Speaker: Partha Lahiri

Authors: Partha Lahiri, Masayo Hirose

13.00 – 16.00 **Lunch**

16.00 – 17.30 **Panel discussion**

17.30 – 17.45 **Closing Ceremony**

**ABSTRACTS**

**Session 1:** M-quantile regression for zero-inflated data and its applications to small area estimation

**Authors:** Maria Bugallo (speaker), Domingo Morales, Nicola Salvati , Francesco Schirripa

**Abstract:** Zero-inflated data are almost inevitably complicated by some form of non-observation or inaccurate measurement. From a probabilistic framework, mixtures of GLMMs for the prediction of zero-inflated outcome-dependent indicators have been extensively investigated, and their results are accurate as long as their strong parametric assumptions hold true. However, the demand for results unaffected by outliers in small areas has encouraged the development of new robust inference techniques in recent years. Prompted by the need to develop robust models for variables with an implausible number of zeros, the definition of M-quantiles and their applications are generalized to small area estimation in this field. The contribution includes the proposal of zero-inflated M-quantiles and M-quantile models, the study of asymptotic properties, the derivation of robust predictors, their optimal bias correction and the analytical calculation of mean squared errors. The new methodology is evaluated by means of model-based simulations, showing the gain that the new proposal brings in the presence of just a few atypical data. An application to the Spanish Living Conditions Survey is concluded with.

**Session 2:** Predicting gender employment discrepancies: Multivariate Fay-Herriot models for compositional data.

**Authors:** Esteban Cabello (speaker), Domingo Morales, Agustín Pérez

**Abstract:** Segregation indices measure the degree of interaction, contact or distance between two groups and are used to quantify occupational discrepancies between genders in a set of occupational sectors. The following talk presents a novel methodology for predicting area-level proportions of employed men and women across various occupational sectors. The challenge arises from the compositional nature of the direct estimators of proportions, which tend to be imprecise when sample sizes are small. To overcome this problem, we propose to use a compositional multivariate Fay-Herriot model. By applying log-ratio transformations to the direct estimators of proportions, we can effectively capture the underlying structure and dependencies within the data. Small area estimators for proportions, entropies, divergences and exposure indexes are derived from the fitted model, and their corresponding root-mean-squared errors are estimated using parametric bootstrap techniques. We conduct two case studies using data from quarters 3 and 4 of the Spanish Labour Force Survey of 2022. The primary objective is to investigate the state of gender occupational segregation in Spanish provinces.

**Session 3**: Small area estimation under bivariate Fay-Herriot model with correlated random effects.

**Authors:** Esteban Cabello, María Dolores Esteban, Domingo Morales (speaker), Agustín Pérez

**Abstract:** This contribution presents an area-level temporal bivariate linear mixed model, incorporating correlated time effects for estimating poverty indicators in small areas. The model is applied through the residual maximum likelihood method, leading to the derivation of empirical best linear unbiased predictors for these indicators. Additionally, it is provided an approximation of the matrix of mean squared errors (MSE) and it is proposed four MSE estimators. The first estimator involves a plug-in approach to the MSE approximation, while the remaining estimators are based on parametric bootstrap procedures. To assess the performance of the fitting algorithm, predictors, and MSE estimators, three simulation experiments are carried out. An application to real data from the 2016 to 2022 Spanish Living Conditions Survey is conducted. The focus is on estimating poverty proportions and gaps for the year 2022, categorized by provinces and sex.

**Session 4:**  Estimating Variance of Random Effects to Address Multiple Problems in Small Area Estimation.

**Authors:** Partha Lahiri (speaker), Masayo Hirose

**Abstract:** For several decades, area-level models have played a critical role in the theory and practice of small area estimation. In this context, we propose a random effects variance estimator that simultaneously (i) improves the estimation of the related shrinkage factors, (ii) protects empirical best linear unbiased predictors (EBLUP) of the random effects from the common over-shrinkage problem, and (iii) avoids complex bias correction in generating strictly positive second-order unbiased mean square error (MSE) estimators, either through the Taylor series or a single parametric bootstrap method. The idea of achieving multiple desirable properties in an EBLUP method through a suitably devised random effects variance estimator is innovative and holds promise for providing robust inferences for random effects within the EBLUP framework. The proposed methodology is evaluated using a Monte Carlo simulation study and real data analysis.