

**mgr Małgorzata Krzciuk**

Abstract of the doctoral dissertation entitled  
**“Model-based approach in small area estimation  
and its applications in economic research”**

written under the scientific guidance of dr hab. Tomasz Żądło, prof. UE

In this dissertation the problem of the model-based approach in small area estimation and its application in economic research is considered. The motivation for the choice of the subject of work is the observed increase in the importance of regions and regional policy. This is related to the growing interest in national databases increasingly used in official statistics and the growing demand for information at the lowest possible level of aggregation. It also contributes to the growing demand for local analysis and methods to obtain accurate estimates of the characteristics of the subpopulation without the need to conduct complete studies that do not require large outlays of time and financial resources. These requirements are met by the methods of small area estimation allowing for inference in the case of samples with a small or even zero size in the subpopulation, which have been used in many areas, such as e.g., market analyzes, regional policy, or labor market analyzes. Particular attention in this study is devoted to one of the main approaches in small area estimation —the model-based approach, which enables inference both based on random and non-random samples.

The main goal of this dissertation is to propose methods of predicting the subpopulation characteristics and to analyze the properties of the proposals of predictors, considering the occurrence of correlation between vectors of the random effects. This goal was achieved by: adapting the model-based approach of small area estimation to the longitudinal economic data, proposing some models belonging to the class of general linear mixed models along with the methods of their verification, as well as the presentation and application of proprietary prediction and accuracy estimation methods under the above models. The dissertation also presents the possibilities of using the proposed methods for real economic data in simulation studies conducted with the Monte Carlo method. The achievement of the above goals allowed to answer the research questions: which models, belonging to the class of general linear mixed models, will make it possible to consider the occurrence of correlation between vectors of the random effects for the purposes of prediction based on economic longitudinal data and how to verify the occurrence of these correlation relationships for the proposed class of models. The analyses carried out in the work also allowed answering research questions concerning the

influence of the occurrence of correlation relationships between random effects and the use of information from previous periods on the properties of the considered predictors of the subpopulation characteristics, including the accuracy of prediction, in comparison with methods using information from one period.

The dissertation is divided into five chapters. For the purposes of chapters 1-4, an extensive review of the literature on small area estimation and the model-based approach, has been performed. In the first chapter the main approaches in the small area estimation, with examples of their applications in economic analyzes, are discussed. Particular attention was paid to the model-based approach, including the process of building a superpopulation model and selected prediction methods. The author proposes to generalize the selected predictors for analyzes of longitudinal data. There are also presented author's own proposals of special cases of linear mixed models assuming a correlation between the vectors of random effects. Furthermore, the use of permutation tests of the significance of the parameters of the above class of models and a test based on the parametric bootstrap method enabling the verification of the relationship between random effects are proposed.

The second chapter deals with the issues of single-period and multi-period research. However, the issue of multi-period research, including panel research, is discussed in more detail. Their classification was presented along with examples from Poland and the world. Particular attention was paid to the advantages and disadvantages of these studies, as well as the benefits and limitations of conducting analyzes based on data from multi-period research.

The third chapter presents the BLU and EBLU class of predictors. This subject is considered under the classification of linear mixed models. The author proposes the use of the EBLU class predictor in the small area estimation, assuming the proposal of the linear mixed model with correlated random effects. The problem of the estimation of mean square errors of the these predictors, including a proposed modification of known methods to proprietary EBLU class predictors, is considered. This chapter also includes a review of selected applications of the above prediction methods in economic research.

In chapter four, are discusses the theoretical aspects of prediction using EB and plug-in predictors. The author proposes the use of the above statistics, assuming the proposed class of linear mixed models. Furthermore, in this part of the dissertation modifications of known mean squared error estimators under the assumption of correlation between vectors of random effects are proposed. This chapter also presents selected applications of the EB and plug-in predictors in economic analyzes.

The fifth chapter describes the assumptions and the results of the simulation studies carried out in accordance with the model-based approach. The analyses dealt with the problem of predicting total values and medians in domains, and their main goal was a comparative analysis of the properties of the proposed predictors with appropriate predictors assuming the lack of correlation between random effects and selected estimators. It should be added that three variants of the simulation study were considered, taking into account the different strength of the correlation between the random effect vectors at the stage of generating population data, including the original parameter values obtained based on the considered dataset. The research was carried out with the use of the R language and self-written functions for proposed predictors and codes for conducting the simulation study. It should be added that the obtained results suggest good properties of the considered authors' propositions of the predictors. This is indicated by the low simulation relative values of the prediction standard error and the root of the mean square prediction error. Summary of the results suggests a gain in the accuracy and precision of prediction resulting from the use of the proposed predictors assuming the occurrence of correlation between the vectors of random effects. It should be emphasized that the results obtained in this dissertation may be useful in practice for institutions conducting research as well as users of the data, e.g. state administration at the central and local level.