Freie Universität Berlin

Flexible domain prediction of continuous and count outcomes using unit level quantile random effects regression

Timo Schmid*
Freie Universität Berlin, Germany - timo.schmid@fu-berlin.de

Nicola Salvati
University of Pisa, Italy

Nikos Tzavidis
University of Southampton, UK

Having detailed and accurate information on economic and social conditions, summarised by appropriate one-dimensional, multidimensional and composite indicators, is imperative for the efficient implementation of policies. The term detailed is used to signify information that extends beyond aggregate levels into highly disaggregated geographical and other domains (e.g. demographic groups). The term accurate refers to information that is estimated with appropriate level of precision and is comparable over time and space. Simply analysing data from National sample surveys is not enough for achieving the dual target of detailed and accurate information. This is due to the reduction of the sample sizes as the level of detail required increases. For achieving this dual target one must employ appropriate methodology collectively referred to as Small Area Estimation.

In this talk we will present recent work on a new unit-level small area methodology that can be used with continuous and discrete outcomes. The proposed method is based on constructing a model-based estimator of the distribution function by using a nested-error regression model for the quantiles of the target outcome. A general set of domain-specific parameters that extends beyond averages is then estimated by sampling from the estimated distribution function. For fitting the model we exploit the link between the Asymmetric Laplace Distribution and maximum likelihood estimation for quantile regression. The specification of the distribution of the random effects is considered in some detail by exploring the use of parametric and non-parametric alternatives. The use of the proposed methodology with discrete (count) outcomes requires appropriate transformations, in particular jittering. For the case of discrete outcomes the methodology relaxes the restrictive assumptions of the Poisson generalised linear mixed model and allows for what is potentially a more flexible mean-variance relationship. Mean Squared Error estimation is discussed. Extensive model-based simulations are used for comparing the proposed methodology to alternative unit-level methodologies for estimating a broad range of complex parameters.

The aim of this talk is to bring the audience up-to-speed with the state of the art in model-based small area estimation of deprivation indicators. The talk will have an applied focus illustrating the methodological and data challenges and proposing possible solutions.

Keywords: Asymmetric Laplace Distribution; Generalized linear mixed model; Jittering; Non-parametric estimation; Small area estimation.