Area Specific Effects Selection of Small Area Estimation for Constructing the Regional Consumer Price Indices in Indonesia

Novi Hidayat Pusponegoro[1,2]; Anang Kurnia[1]; Khairil Anwar Notodiputro[1]; Agus Mohamad Soleh[1]; Erni Tri Astuti[2]

[1]   Department of Statistics, IPB University, Bogor 16680, Indonesia
[2]   Department of Statistics, STIS Polytechnic of Statistics, Jakarta 13330, Indonesia

Abstract:
Small area estimation (SAE) techniques are now widely employed to produce parameter estimates for smaller domains where sample sizes cannot be used to deliver direct estimation. In the 'Big Data' era, database size and technology has developed rapidly. This leads to computational and statistical challenges since the availability of high data volume. Area specific effects have important roles in attaining reliable parameter estimates in SAE. Thus their exclusion and inclusion in the model should be carefully examined in order to get the optimum rate of convergence and accurate prediction. One 'modern data' type has large observations called as large data scale, the observation perhaps small area. In fact, the number of small areas is relatively small compared to total survey areas. Sparsity on specific area effects can be imposed by assigning zero for the large areas, whereas it preserve the nonzero value for the small areas. It causes the normality assumption of specific area effects can be violated and lead the parameter estimates convergence to be slow. An adaptive model for small area estimation with automatic random effects selection (SARS) using hard-ridge penalty for SAE has been carried out. However, the large scale data challenge an estimation method that not only select the area specific effects but also shrinkage the coefficients at once in order to get the parsimony. In order to fulfill the current challenges, norm1 was proposed as penalty. SAE with norm1 penalty for area specific effects selection empirically deliver minimum mean square error in every condition of auxiliary variables correlation, area effects variance components and percentages of small area. In Indonesia, Consumer Price Indices (CPIs) are calculated on some regencies in urban area and national level. In this decentralization era, CPI statistics are needed to be delivered for other regencies not only in urban area but also in rural areas. Therefore, the aims of this study to investigate the sparsity issue in small area estimation model using the LASSO method. The SAE model using LASSO method applied to estimate the regional CPIs since it important indicator for monetary stability and inflation. The response variables of this paper were the CPIs of January 2018 and the auxiliary information were regard to infrastructure & resources from the village potential census in 2018 (PODES 2018).