
Variable selection in spatial models, application to income analysis in Uruguay rural areas

Liliane Bel^{*1}, Mathias Bourel², and Romina Gonella Furtado²

¹Mathématiques et Informatique Appliquées – Université Paris Saclay, INRAe, AgroParisTech, UMR MIA-Paris, 75005, Paris, France – France

²Universidad de la República – Uruguay

Résumé

This work focuses on variable selection for spatial regression models, with locations on irregular lattices and errors modeled as CAR or SAR. The strategy is to whiten the residuals by estimating their spatial covariance matrix and then proceed by performing the standard LASSO for independent data on the transformed model. A result is stated that proves the consistency in sign for general dependent errors provided that the transformed design matrix fulfills standard assumptions for the LASSO procedure and that the estimate of the residual covariance matrix is consistent. Then sufficient conditions on the adjacency matrix of the SAR or CAR model are given that ensure those assumptions are fulfilled. An extensive simulation study is driven that shows this method gives good result in terms of variables selection, while some underestimation of the coefficients is noted. It is compared to a strategy that estimates both the regression and the covariance parameters in a LARS procedure. Coefficient are better estimated with the LARS procedure but it gives in some cases much more false positive in the variable selection. The application is on the regression of income data in rural area of Uruguay on a set of covariables describing various characteristics of the households. The selection variables procedure after whitening results in a set of selected variables much less than the LASSO procedure for independent observations with coefficient values much more interpretable, showing the interest of the method.

*Intervenant