

Tesi di Dottorato in
Economia e Politiche dei Mercati e delle Imprese – curriculum Metodi Statistici

ciclo XXX
A.A 2016/2017

High-dimensional Statistics for complex data

Massimo Pacella

ABSTRACT (ENG)

High dimensional data analysis has become a popular research topic in the recent years, due to the emergence of various new applications in several fields of sciences underscoring the need for analysing massive data sets. One of the main challenge in analysing high dimensional data regards the interpretability of estimated models as well as the computational efficiency of procedures adopted. Such a purpose can be achieved through the identification of relevant variables that really affect the phenomenon of interest, so that effective models can be subsequently constructed and applied to solve practical problems.

The first two chapters of the thesis are devoted in studying high dimensional statistics for variable selection. We firstly introduce a short but exhaustive review on the main developed techniques for the general problem of variable selection using nonparametric statistics. Lastly in chapter 3 we will present our proposal regarding a feature screening approach for non additive models developed by using of conditional information in the estimation procedure.

Differently, the second part of the thesis focuses on the spatio-temporal models in high dimensional contexts. Over the last decade, a particular class of spatio-temporal models has been rapidly developed, the spatial dynamic panel data models (SDPD). Several versions of the SDPD model have been proposed, based on different assumptions on the spatial parameters and different properties of the estimators. The standard version of the model assumes the spatial parameters constant over location, meanwhile another recently proposed version assumes the spatial parameters are adaptive over location. The assumption of different scalar coefficients is motivated by practical situations, in which empirical evidence shows how considering constant effect for each location can be limiting. While chapter 4 is devoted to introduce principal elements of spatio-temporal models in statistical and econometric frameworks, in chapter 5 we propose a strategy for testing the particular structure of SDPD model, by means of a multiple testing procedure that allows choosing between the version of the model with adaptive spatial parameters and some specific versions derived from the general one by imposing particular constraints on the parameters. The multiple test is made in high dimensional setting by the Bonferroni technique and the distribution of the multiple test statistic is derived by a residual bootstrap resampling scheme.