*II PhDay-EIO*
*Santiago de Compostela, 23 de xuño do 2021*

# Covariates selection in high-dimensional data and specification tests for the concurrent model

Laura Freijeiro-González[1], Manuel Febrero-Bande[1], Wenceslao González-Manteiga[1]

[1]Department of Statistics, Mathematical Analysis and Optimization. Universidade de Santiago de Compostela.

## ABSTRACT

Nowadays, in many important statistical applications, it is of high relevance to apply a first variable selection step to correctly explain the data and avoid unnecessary noise. Furthermore, it is usual to find that the number of variables $p$ is larger than the number of available samples $n$ ($p > n$). It is in this context of high-dimensional data where the ordinary models fail and, as a result, estimation and prediction in these settings are generally acknowledged as an important challenge in contemporary statistics. The penalized linear regression models face these problems and provide useful solutions, specially those of the LASSO family. Nevertheless, this algorithm exhibits some drawbacks related to the correct selection of important covariates and the exclusion of redundant information. We analyze its performance under different dependence structures and compare this with modifications and alternatives which have showed suitable properties.

Moreover, we can deal with scenarios where the covariates are not multivariate but have a different nature. The concurrent model is a special regression model in the middle of multivariate and functional approaches. The response $Y$ and the $p$ covariates $X_1, \ldots, X_p$ are functions of the same argument $t$, and the influence is concurrent, simultaneous or point-wise in the sense that $X_1, \ldots, X_p$ only influence $Y(t)$ through their values $X_1(t), \ldots, X_p(t)$ at time $t$. As a result, we can face this by means of multivariate techniques, translating previous ideas, or working with functional strategies. Nevertheless, its estimation is quite complex in practice due to its nature, specially for flexible structures. To fill this gap, we work in developing a specification contrast to be able to determine its regression structure.