Econometrics and Machine Learning*

Arthur Charpentier, Emmanuel Flachaire and Antoine Ly

Key Question

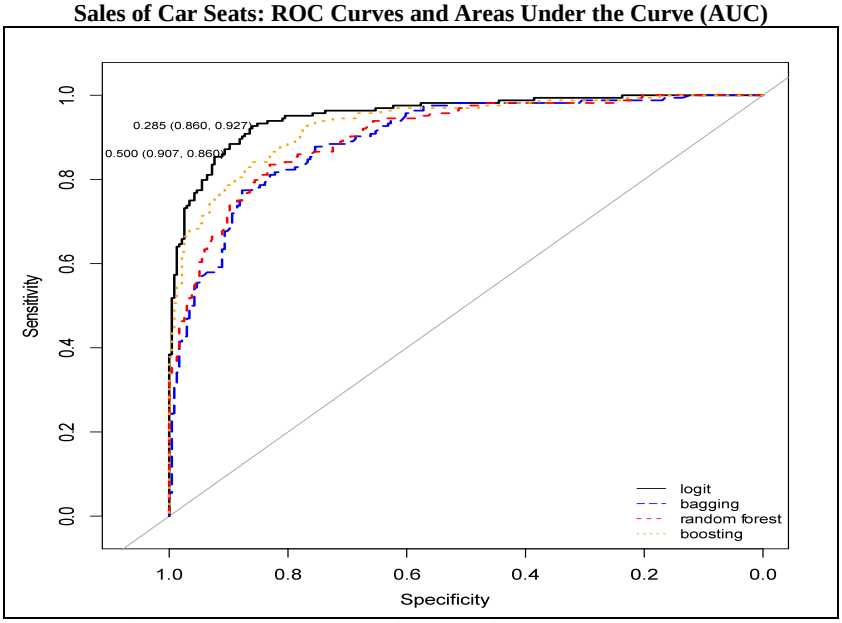
The purpose of this paper is to present machine learning techniques and to compare their performance to traditional econometric methods. The aim is not only to test their performance according to different types of problems, but also to evaluate the extent to which such methods may complement one another In other words, the paper shows how machine learning methods can be used to improve econometric modelling.

Methodology

Standard problems in the machine learning literature are re-assessed by drawing on the databases initially studied. Several different machine learning techniques are applied, as are econometric linear regression methods. The performances of the different methods are then compared.

Main Results

As shown by applications on real data, traditional econometric models, if well-specified, generally perform just as well – if not better – than statistical learning models. However, analysis of the results of learning methods helps to improve the specification of econometric models. Cross-validation provides an alternative to asymptotic developments and bootstrap-aggregation techniques such as bagging provide a means of better capturing nonlinearities. Lastly, penalisation methods provide a simple method for selecting variables, a problem that soon becomes complex in high dimension.



Simulated data from 400 baby car seat points of sale Data from James *et al.* (2013), *carseats* dataset, https://CRAN.R-project.org/package=ISLR

Message

Despite having developed historically in very different contexts, econometric and learning models share an increasing number of common features: algorithmic learning performance currently helps to solve problems traditionally encountered in econometrics; a probabilistic view of learning techniques provides tools for seeing predictions in a new light. While not wanting to challenge their practice, econometricians would benefit from drawing on new machine learning techniques with a view to improving predictions and strengthening certain interpretations.

* Economie et Statistique / Economics and Statistics N° 505-506, 2018. https://www.insee.fr/en/statistiques/3704781