

Mathematical Statistics Stockholm University Bachelor Thesis **2021:9** http://www.math.su.se

Evaluating Forecast Accuracy of Random Forests Versus LS_Boost on Simulated Time Series

Adam Goran^{*}

June 2021

Abstract

No machine learning algorithm is perfect, but some are superior at providing accurate predictions and forecasts. In this report, we compare the ability of random forests and LS_Boost to produce onestep forecasts. The study regards mainly the object of finding which algorithm is able to attain the smallest forecast error for unseen data. Different situations are considered to diversify the work and accomplish a fair comparison. Data is obtained from simulations, where three established financial time series models serve as frameworks. In order to understand and apply the algorithms correctly we introduce the theory of regression trees, bagging and boosting. The specific algorithm for random forests and LS_Boost respectively is also presented, along with some general knowledge about the three time series models. We find that LS_Boost achieves lower forecast error in terms of both mean squared error and mean absolute error for all simulations. This is the case regardless of whether the parameters of the algorithms are tuned. Random forests is close in performance though and does actually approach the performance of LS_Boost as the simulated data becomes more sparse.

^{*}Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: adam.goran@hotmail.com. Supervisor: Ola G H Hössjer, Kristofer Lindensjö.