University of Zurich[UZH]

Department of Banking and Finance
Chair Prof. Dr. Marc Paolella

MASTER THESIS

Stock market volatility:

Identification of risk drivers and forecasting using random forest

Supervisor:
DR. BORIS WÄLCHLI

Author:
FABIAN SMITS

Field of Study: Banking and Finance

Executive Summary

To this day, volatility is the most prevalent and most intensively researched financial risk measure and is used in a vast array of applications. Most of the academic literature regarding stock market volatility is concerned with models which condition volatility on historical return data. Despite a number of studies which investigate the influence of alternative variables on stock market volatility, there exists no consensus to which degree such data is able to improve forecasting. Neither is there a consensus on which variables the research should be focused on. This thesis investigates both these questions by using the random forest algorithm, a method which has been successfully employed for the identification of predictor variables in various academic fields, but has never been applied to investigate stock market volatility.

Following the approach of Mittnik et al. (2015), I use a large set of financial and macroeconomic predictor variables to produce volatility forecasts of the S&P 500 index for a 1 to 6-month horizon. The forecast accuracy compares favourably with standard benchmark models (GARCH, eGARCH). For short time horizons, random forest achieves a significant outperformance of its benchmarks.

I identify relevant predictors through the comparison of forecasting results of differently specified random forests and by analysing random forest's importance measures. It appears that only a small number of predictors can benefit forecasting, the most important of which is the volatility index VIX. Both the VIX and historical return data display a stable importance ranking, which (plausibly) implies a stable interdependence with stock market volatility. Other predictors occasionally receive a high importance score, but simultaneously display high fluctuations in their importance ranking, suggesting an unstable, time-varying association to volatility. The dividend price ratio is the only one of these variables which achieves a high overall importance score, putting forward this variable as a candidate for future research. Furthermore, the employed time trend variable receives one of the highest overall importance scores, confirming the finding that the dependencies within the examined dataset are unstable and time-varying.

Finally, I investigate the properties of random forest when applied to financial market data. While confirming certain acclaimed benefits of random forest, like the high degree of robustness against the misspecification of tuning parameters, I identify several pitfalls of the method in the domain of time series data. I find that serial correlation causes a bias in out-of-bag errors and identify difficulties of random forest in handling seasonal effects and switching regimes. I analyse the consequences of these deficiencies and propose potential solutions.