Abstract

In this thesis, we aim to improve forecast accuracy of a heterogenous autoregressive model (HAR) by including market sentiment indicators based on Google search volume and Twitter sentiment. We have analysed 30 companies of the Dow Jones index for a period of 15 months. We have performed out-of-sample forecast and compiled a ranking of the extended models based on their relative performance. We have identified three relevant variables: daily negative tweets, daily Google search volume and weekly Google search volume. These variables improve forecast accuracy of the HAR model separately or in a Twitter-Google combination. Some specifications improve forecast accuracy by up to 22% for particular stocks, others impair forecast accuracy by up to 24%. The combination of daily negative tweets and weekly search volume is a superior model to the basic HAR for 17 stocks according to RMSE and for 16 stocks according to MAE and MASE. The daily negative tweets specification outperforms the basic HAR for 17 and 19 stocks, respectively. And, the combination of daily negative tweets and daily search volume outpaces the basic HAR for 15 and 18 stocks, respectively. Based on the average MASE improvement, the combination of daily negative tweets and weekly search volume is a clear winner as it lowers the average MASE by 0.71%.