

Abstract

The thesis investigates relationship between daily stock return volatility of Dow Jones Industrial Average stocks and data obtained on Twitter, the social media network. The Twitter data set contains a number of tweets, categorized according to their polarity, i.e. positive, negative and neutral sentiment of tweets. We construct two classes of models, GARCH and ARFIMA, where for either of them we research basic model setting and setting with additional Twitter variables. Our goal is to compare, which of them predicts the one day ahead volatility most precisely. Besides, we provide commentary regarding the effects of Twitter volume variables on future stock volatility. The analysis has revealed that the best performing model, given the length and structure of our data set, is the ARFIMA model augmented on Twitter volume residuals. In the context of the thesis, Twitter volume residuals represent unexpected activity on the social media network and are obtained as residuals from Twitter volume autoregression. Plain ARFIMA model was the second best and plain volume augmented ARFIMA was in third place. This means that all three ARFIMA models outperformed all three GARCH models in our research. Regarding the Twitter estimation parameters, we found that higher the activity the higher tomorrow's stock return volatility. This conclusion holds for all Twitter volume variables regardless their polarity.