

Correlation 2.0: News and Social Media Predicting Stock Correlations?

Master's Thesis

in

Financial Economics

Department of Banking and Finance

University of Zurich

December 10, 2013

Supervisors:

Prof. Dr. Thorsten Hens

Dr. Sven C. Steude

Author:

Michael Bonifazi

Executive Summary

This master's thesis aims to contribute to the current debate about the ability as well as contribution of news and Social Media flows to predict various economic factors. While most of the current research is focussing on predicting stock returns on behalf of news and Social Media data, this work goes even further by analysing the influence of news and Social Media flows on the stock correlations. In doing so, four possible stock correlation predictors based on sentiment scores and occurrences of news items are defined and tested. To account for the seemingly endless stream of information and growing importance of Social Media within Web 2.0, all correlation predictors are extracted using three different data sources. On the one hand, the worldwide printed news flow is channelled and accessed using two different sources. The first data set contains raw news content including bodies and headlines and is provided by Thomson Reuters Corporation as well as awp Finanznachrichten. The second pillar of the printed news flow is provided by Thomson Reuters News Analytics (TRNA) and includes a lot of metadata such as sentiment scores and relevance of news items. On the other hand, the worldwide internet news and Social Media flow is used to extract the same correlation predictors. This data is prepared and made available by TRNA for internet news and Social Media including the same metadata as TRNA for printed news. Analysing raw news contents allows to calculate sentiment scores on behalf of an own textual analysis approach - which is strongly related to the Harvard's General Inquirer approach and based on the positive and negative word lists described in ? (?).

To show if the correlation predictors defined within the scope of this master's thesis provide any informative value to forecast future changes in stock price correlations, the econometric technique of bivariate Granger Causality (? (?)) is applied. In addition, different signature plots are visualising the relationship of the news and Social Media based correlation predictors and the corresponding shifts in price correlation for different lengths of price data. The result of the Granger Causality analysis does not seem to be strongly related to the lengths of the time series to extract the price correlations. Depending on the data sources, overall sample periods, and time intervals to extract the correlation predictors, all lengths of historical price data to calculate the price correlations seem to yield in similar results. Overall, some evidence

that the correlation predictors extracted from the worldwide news and Social Media flow are providing some information to project future changes in the price correlation of DJIA's stocks is found. Especially occurrence based predictors are showing promising results and further research should take place in this field. Factors like the industry of the firms and historical co-occurrences in news should be of particular interest.

Key words: Co-Occurrence, Relative Occurrence, Correlation Prediction, Linguistic Analysis, Public Mood