IMPACT OF SOCIAL MEDIA ON THE STOCK MARKET: EVIDENCE FROM TWEETS

Vojtěch Fiala[1], Svatopluk Kapounek[1], Ondřej Veselý[1]

[1] *Mendel University in Brno*

EUROPEAN JOURNAL OF BUSINESS SCIENCE AND TECHNOLOGY

Volume 1 Issue 1
ISSN 2336-6494
www.ejobsat.com

ABSTRACT

The paper deals with the impact of the economic agent sentiment on the return for Apple and Microsoft stocks. We employed text mining procedures to analyze Twitter messages with either negative or positive sentiment towards the chosen stock titles. Those sentiments were identified by developed algorithms which are capable of identifying sentiment towards companies and also counting the numbers of tweets in the same group. This resulted in counts of tweets with positive and negative sentiment. Then we ran analysis in order to find causality between sentiment levels and the stock price of companies. To identify causal effects we applied Granger causality tests. We found bilateral causality between the risk premium and the amount of news distributed by Twitter messages.

KEY WORDS

stock returns, Granger causality, text mining, sentiment analysis, CAPM

JEL CODES

C1, C12, G17, G12

1 INTRODUCTION

The objective of this paper is to identify causal links and their directions between the stock returns and the economic agent sentiment. The main focus will be placed on the social networks, especially messages sent via Twitter. We hypothesize that those messages – tweets, provide quantifiable information about the sentiment. Therefore we apply text mining algorithms to identify positive and negative tweets relating to the analysed companies (Apple and Microsoft).

This paper continues the work of predecessors in this field, mainly Bollen, Mao and Zeng (2011). They have shown that there is a causal link between sentiment on Twitter and the stock market and therefore they were able to predict movements of the Dow Jones index (DJIA) with

FIALA, Vojtěch, KAPOUNEK, Svatopluk, and VESELÝ, Ondřej. 2015. Impact of Social Media on the Stock Market: Evidence from Tweets. *European Journal of Business Science and Technology,* 1 (1): 24–35. ISSN 2336-6494.

87.6% accuracy. Kuleshov (2011) was, on the other hand, not able to reproduce their results with the same procedure and thus questioned the research. No researcher has been able to archive similar results even with different methods, neither with the whole market nor with specific stock titles.

Apart from the majority of its predecessors this paper does not deal with causal links between sentiment and the whole market, which was represented by e.g. the Dow Jones index (Bollen, Mao and Zeng, 2011). It tries to find causality between sentiment and the price of specific stock titles. In order to accomplish that, it employs special algorithms which were not needed in previous research. The purpose of those is on one hand to identify tweets which are in some relationship with the chosen companies and on the other hand, to evaluate the level of sentiment of those tweets and to count them. Algorithms were created to be able to operate with tweets so they also respect some specifics of colloquial language. One of the most important parts is the analysis of causal links. There we employed Granger causality like Bollen, Mao and Zeng (2011).

2 THEORETICAL BACKGROUND

There is a large and rapidly growing literature examining the impact of investors' sentiment on financial markets, especially the predictive power of internet message postings. The empirical studies commonly employ distinct classifier machine learning algorithms to extract sentiment proxies from the huge quantity of text messages published in the news, in social media or on internet message boards (Antweiler and Frank, 2004; Arias, Arratia and Xuriguera, 2013 or Kim and Kim, 2014). These sentiment proxies are associated with specific words or expressions identified by rules or lexicons.

According to the efficient market hypothesis (EFM), the prices of securities are close to fundamental values (Fama, 1970; 1991). Markets are efficient because investors are rational and there are no limits on conducting arbitrage. Any dislocations in asset prices are quickly eliminated by rational investors (Friedman, 1953; Fama 1965), who understand Bayes' law and process all available information when forming expectations. However, empirical observations of capital markets contradict the EFM because the existence of anomalies and excess volatility cannot be explained by changes in fundamentals (LeRoy and Porter, 1981; Shiller, 1981; 2003).

This work is based on the fact that, according to Dolan (2002), human emotions have a large influence on decision making in general and also, as Gilbert (2010) states, on decision making and behaviour on financial markets. It is expected that with a better or more positive overall mood of investors they will be more prone to buying rather than selling in expectation of following growth of the price and vice versa. Those influences of public mood are then able to explain changes of asset prices which are unexplainable by fundaments. Gilbert and Karahalios (2010) showed that it is possible to abstract those sentiment (emotion) levels from social media. Those levels have aggregate character and expresses public mood. Based on that we suppose that if the sentiment influences decisions on the financial markets it also influences stock prices, which is the same assumption as used by similar research by Bollen, Mao and Zeng (2011), Kuleshov (2011) or Chung and Liu (2011). We also suppose that if this aggregate or general mood influences stock prices in general, mood towards one company influences stock prices of that company, as used by Chung and Liu (2011).

Different social media were used in past research in order to identify both overall sentiment levels and sentiment levels in relationship with one object (e.g. company). The social network Twitter was used with success by Bollen, Mao and Zeng (2011), Zhang, Fuehres and Gloor (2011), especially for its consistency and information value. The desired information is captured in tweets – unique authors' messages which are collectible via Twitter API.

3 METHODOLOGICAL BACKGROUND

In the beginning we had to choose companies with which to run analysis. We chose Microsoft Corporation and Apple Inc., mainly because of two elements. Stocks needed to be publicly traded, so we could run analysis and companies had to be popular. This popularity is for the purpose of research expressed by the combination of elements of market capitalization, average volume of stared stocks, business to consumer character of product and overall popularity of companies. We assumed that popular companies or their products will be mentioned more on Twitter than less popular ones and thus it will boost the importance of sentiment on Twitter in explaining changes in stock price.

The empirical strategy combines text mining algorithms and econometric modelling. First, we used Twitter streaming API in order to extract tweets from Twitter. We extracted tweets during the period from 1.3.2014 to 18.5.2014. We applied a filter to obtain tweets in the English language and keywords filter, which identified tweets with some relationship with the chosen companies. Those words were selected by analysis of companies and their products and with Google Trends which shows the popularity of words in google searches. Words were divided into two groups which identified Apple (M_A) and Microsoft (M_M).

After we created algorithms which were able to identify towards which company tweet carries sentiment, the level of sentiment itself and which also respects some features of colloquial English language. Those algorithms consisted of groups of words with different functions. In order to be counted as negative or positive, a tweet had to contain none or at least one word from each group in the algorithm. Which words the tweet must and must not contain is described by logical math connectors (conjunction $\wedge$, disjunction $\vee$ and negation $\neg$). Formula 1 represents the algorithm for identification of the number of tweets carrying positive sentiment towards Apple, formula 2 negative sentiment towards Apple, formula 3 positive sentiment towards Microsoft and formula 4 negative sentiment towards Microsoft:

$$M_\mathrm{A} \wedge S^{+} \neg V_\mathrm{A}, \quad V_\mathrm{all}, \quad M_\mathrm{M}, \tag{1}$$

$$M_\mathrm{A} \wedge S^{-} \neg V_\mathrm{A}, \quad M_\mathrm{M}, \tag{2}$$

$$M_\mathrm{M} \wedge S^{+} \neg V_\mathrm{A}, \quad V_\mathrm{all}, \quad M_\mathrm{A}, \tag{3}$$

$$M_\mathrm{M} \wedge S^{-} \neg V_\mathrm{A}, \quad M_\mathrm{A}, \tag{4}$$

where M_A and M_M are groups of words identifying Apple and Microsoft, S^{+} and S^{-} are groups of words identifying positive and negative segment, V_A and V_M are groups of words which prevent misinterpretation of M_A and M_M. V_all is a group of words which prevents misinterpretation of both M_A and M_M. Groups of words S^{+} and S^{-} were created from dictionaries and similar sources. The final groups of words were selected on the basis of usage with Google Trends and Google Ngram tools. Groups of words V_A and V_M were selected after thorough analysis of fundamental aspects of companies, products and especially competitive environment. Words in group V_all were selected in order to enable algorithms to identify specifics of colloquial English.

After we established algorithms we standardized data from Twitter. The process of standardization consisted of deletion of tweets of artificial origin (e.g. made by bots or applications) and limiting the number of tweets at 200,000 per day. Then we applied algorithms 5 to 8 which resulted in counts of both positive and negative tweets in relationship to either Apple or Microsoft by days.

With the counts of tweets needed, we applied Arbitrage Pricing Theory (APT) to obtain a basic model describing the relation between the stock returns and other factors (Ross, 1976). To explain stock returns we started with the simple Capital Asset Pricing Model (CAPM):

$$\mathrm{ER}_i = \mathrm{RF} + \beta\,(\mathrm{ER}_m - \mathrm{RF}), \tag{5}$$

where ER_i is the expected return of the specific capital asset i, RF is the risk-free interest rate on the market (usually government bonds), ER_m is the expected return of the market and β_i is the sensitivity of the expected excess of the specific asset return to the expected excess market returns calculated as

$$\beta_i = \frac{\mathrm{Cov}\,(R_i, R_m)}{\mathrm{Var}\,(R_m)}.$$

The formula (5) is extended to the multifactor model as the special case of the APT:

$$R_i - \mathrm{RF} = \delta + \beta_i\,(R_m - \mathrm{RF}) + \gamma_i S_i + \epsilon_i, \quad (6)$$

where R_i and R_m are returns calculated from historical data as a rate of return over a single period on a daily basis, S_i represents sentiment, δ is intercept and ϵ_i is the idiosyncratic component of the stock's return which is minimized by the process of arbitrage. The vector autoregressive model of the order k, VAR(k), can be rewritten in matrix form:

$$\begin{aligned} Y_t &= \delta + B_1 Y_1 + \ldots + B_k Y_k + \epsilon_t = \\ &= \delta + \sum_{j=1}^{k} B_j Y_{t-j} + \epsilon_t, \end{aligned} \quad (7)$$

where all variables have the same lag length of order k. We applied Akaike and Bayesian information criteria to determine the minimal appropriate number of required lags.

To identify the causal relation between the news and the stock returns we applied Granger causality tests (Granger, 1969 and Sims, 1972). The causality is inferred when the lagged values of a variable X_t have explanatory power in a regression of a variable Y_t on lagged values Y_t and X_t. If lagged values of a variable X_t have no explanatory power of any of the variables in a system, then we would view X as weakly exogenous to the system. With respect to the direction of causality we can distinguish two cases: (1) unidirectional causality when X_t is caused Y_t but Y_t does not cause X_t and (2) bilateral causality when variables X_t and Y_t are jointly determined. This causality can be identified using the Granger test (1969) based on the premise that the future cannot cause the present or the past, utilising the concept of the VAR approach. In our analysis we assume a VAR(k) model with three variables X_t, Y_t and Z_t:

$$\begin{aligned} Y_t &= \delta_1 + \sum_{j=1}^{k} \alpha_{1j} X_{t-j} + \sum_{j=1}^{k} \beta_{1j} Y_{t-j} + \\ &\quad + \sum_{j=1}^{k} \gamma_{1j} Z_{t-j} + \epsilon_{1t}, \\ X_t &= \delta_2 + \sum_{j=1}^{k} \alpha_{2j} X_{t-j} + \sum_{j=1}^{k} \beta_{2j} Y_{t-j} + \\ &\quad + \sum_{j=1}^{k} \gamma_{2j} Z_{t-j} + \epsilon_{2t}, \\ Z_t &= \delta_3 + \sum_{j=1}^{k} \alpha_{3j} X_{t-j} + \sum_{j=1}^{k} \beta_{3j} Y_{t-j} + \\ &\quad + \sum_{j=1}^{k} \gamma_{3j} Z_{t-j} + \epsilon_{3t}, \end{aligned} \quad (8)$$

where X_t represents risk premium of the specific stock (R_i − RF), Y_t is risk premium of the specific market (R_m − RF) and Z_t represents sentiment (number of bad and/or good news sent by Twitter on a daily basis). To test causality in the formula (8) we used the Wald test with the defined W statistic (Wald, 1943):

$$\begin{aligned} W &= \frac{(\mathrm{RSS}_r - \mathrm{RSS}_u)/k}{\mathrm{RSS}_u/(n-2k-1)} \\ &\sim F(k, n-2k-1), \end{aligned} \quad (9)$$

and Lagrange multiplier with the defined LM statistic (Aitchison and Silvey, 1958; Silvey, 1959):

$$\begin{aligned} LM &= \frac{(\mathrm{RSS}_r - \mathrm{RSS}_u)/k}{\mathrm{RSS}_r/(n-2k-1)} \\ &\sim F(k, n-2k-1), \end{aligned} \quad (10)$$

where RSS_u is the sum of squared residuals from the unrestricted equation and RSS_r is the sum of squared residuals from the equation under the assumption that a set of variables is redundant (restricted).

Both test statistics are distributed as χ^2 under the null hypothesis with the same degrees of freedom (same number of restrictions). Berndt and Savin (1977) showed that the both tests are first-order equivalent and asymptotically optimal but they differ in second-order properties, when the null hypothesis is false, but erroneously fails to be rejected.

In order to proceed we obtained daily close stock prices abstracted from splits and dividends and daily Dow Jones Index close prices from Yahoo! Finance. Risk free interest rate is represented by daily treasury real yield curve rates of 5-year bonds provided by U.S. Department of the Treasury. The dataset contains data in the period March 3, 2014 to May 18, 2014.

4 RESULTS

Firstly, we calculated beta from the sample period using the formula

$$\beta_i = \frac{\mathrm{Cov}\,(R_i, R_m)}{\mathrm{Var}\,(R_m)}.$$

The parameter β is 0.6025 for Apple and 1.3721 for Microsoft. Using the adjusted closing price, the parameter β is 0.6093 for Apple and 1.3712 for Microsoft. The results showed that the Microsoft stock returns are much more sensitive to market movements as opposed to idiosyncratic factors. However, the key question is if the CAPM provides appropriate results and estimations of the asset prices and risks. Therefore we apply the simple version of the CAPM (formula 1) to calculate different betas referring to actual stock returns. The identified relations between the systematic risk (beta) and stock returns are presented in Fig. 1.

Obviously, there are too many situations when stocks do not lie on the SML (Security Market Line). Moreover, the results confirmed limitation of the simple CAPM without other idiosyncratic factors.

Regarding the results provided by the simple CAPM we included sentiment of the economic agents. Thus, we assume that economic agents incorporate and reflect all relevant information, including all idiosyncratic factors related to the specific stock returns. We assume that this information is contained in all news related to the companies and its typical products sent by Twitter as well. The results of the Granger causality tests are presented in Tab. 1 to 4. Tab. 1–2 present Wald statistics of variables within the identified VAR(k) models. In the case of Apple (AAPL) we estimated VAR models with k in the range 1–3. The lag was higher in the models of Microsoft stock (MSFT). The maximal lag of the estimated VAR model with significant relationship between the risk premium and news was 8 days.

However, the identified unilateral causality confirmed that risk premium of the specific stocks is affected only by market premium at the 1% and 5% significance level. We also found that the news sent via Twitter is affected by both risk premium of the specific stocks and risk premium of the market. Thus, the news reacts to the capital market movements, capital markets do not react to the news sent via Twitter. Especially in the case of Apple, we identified causal effects of risk premium of the stock and market on the bad news sent via Twitter at 1% significance level and lag of 2 days. On the contrary, changes in risk premium of the stocks and market affect the good news related to the company Microsoft and its products with the lag of 3 days. The only identified causality effect with direction from news to capital markets was identified in the case of Apple stock and the news which combines the names of both analysed companies and their products. This causality was identified in the model VAR(3) at 10% significance level.

Adjusted closing prices of the stocks showed similar results (Table 2). The employed Wald test identified causality only in the direction from capital market to the news sent by Twitter. Especially bad news related to both companies Apple and Microsoft and its products, and bad news related to Apple and its products, react to the changes of the stock and market returns. The causality was identified at 1% significance level and lag of 2 days. On the contrary, good news related to the company Microsoft and its products are sent 3 days after the changes of the appropriate stock returns or market returns.

Bilateral causality between the news and capital markets, as well as the unilateral causality in direction from capital markets to news

Tab. 1: Granger causality statistics, Wald test

Bad News			
VAR(3)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (AAPL)	0.0098	4.0895**	2.7522*
Market Premium	0.0064	4.2297**	2.1409
Bad News (AAPL + MSFT)	1.0075	1.1103	0.4162
VAR(2)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL)**
Risk Premium (AAPL)	0.0600	3.9487**	0.2640
Market Premium	0.0039	4.2858**	0.0595
Bad News (AAPL)	15.6588***	14.9487***	0.0714
VAR(8)	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (MSFT)	2.2677	4.7571**	0.5777
Market Premium	2.8260*	5.6873**	0.6698
Bad News (AAPL + MSFT)	3.3559*	3.1535*	2.1429
VAR(3), const.	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (MSFT)**
Risk Premium (MSFT)	0.9478	2.8848*	1.3400
Market Premium	0.6537	2.2899	1.1574
Bad News (MSFT)	0.1198	0.1044	3.7565*
Good News			
VAR(1)	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (AAPL)	0.1581	13.0362***	0.0786
Market Premium	0.3362	15.5776***	0.0584
Good News (AAPL + MSFT)	2.6802	2.7528*	6.8571***
VAR(1)	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL)**
Risk Premium (AAPL)	0.0694	9.2360***	0.0713
Market Premium	0.1605	11.1707***	0.0914
Good News (AAPL)	0.9593	1.0611	9.4526***
VAR(1), const.	**Risk Premium (MSFT)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (MSFT)	0.7724	2.1210	0.0744
Market Premium	0.7899	2.1437	0.0422
Good News (AAPL + MSFT)	3.3267*	3.1550*	7.0371***
VAR(3)	**Risk Premium (MSFT)**	**Market Premium**	**Good News (MSFT)**
Risk Premium (MSFT)	0.0019	0.5253	0.0501
Market Premium	0.0090	0.3240	0.0052
Good News (MSFT)	6.3132**	6.0871**	6.2187**

Notes: *, ** and *** denote significance at the 10, 5 and 1% level.

Tab. 2: Granger causality statistics, Wald test, adjusted closing price

Bad News			
VAR(2)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (AAPL)	0.0297	4.1425**	0.0262
Market Premium	0.0002	4.4257**	0.0009
Bad News (AAPL + MSFT)	13.3267***	12.7574***	0.3279
VAR(2)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL)**
Risk Premium (AAPL)	0.0374	4.0832**	0.2167
Market Premium	0.0007	4.4039**	0.0586
Bad News (AAPL)	15.0801***	14.3688***	0.0824
VAR(8), const.	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (MSFT)	3.6371*	6.1297**	0.4437
Market Premium	4.4618**	7.2933***	0.5599
Bad News (AAPL + MSFT)	3.1972*	3.0871*	2.2757
VAR(8), const.	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (MSFT)**
Risk Premium (MSFT)	3.5802*	6.0588**	0.0033
Market Premium	4.4049	7.2325***	0.0096
Bad News (MSFT)	1.3975	1.3723	0.7956
Good News			
VAR(1)	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (AAPL)	0.3044	14.4829***	0.0770
Market Premium	0.5023	17.3489***	0.0563
Good News (AAPL + MSFT)	2.7703*	2.8362*	6.8545***
VAR(1)	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL)**
Risk Premium (AAPL)	0.1625	10.2406***	0.0540
Market Premium	0.2588	12.1758***	0.0792
Good News (AAPL)	0.9844	1.0847	9.3627
VAR(1)	**Risk Premium (MSFT)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (MSFT)	0.6016	1.9716	0.0531
Market Premium	0.6163	1.9877	0.0217
Good News (AAPL + MSFT)	4.0486**	3.9813**	6.9760***
VAR(3), const.	**Risk Premium (MSFT)**	**Market Premium**	**Good News (MSFT)**
Risk Premium (MSFT)	0.0145	0.5865	0.2118
Market Premium	0.0003	0.3770	0.0759
Good News (MSFT)	6.7689***	6.7192***	7.1182***

Notes: *, ** and *** denote significance at the 10, 5 and 1% level.

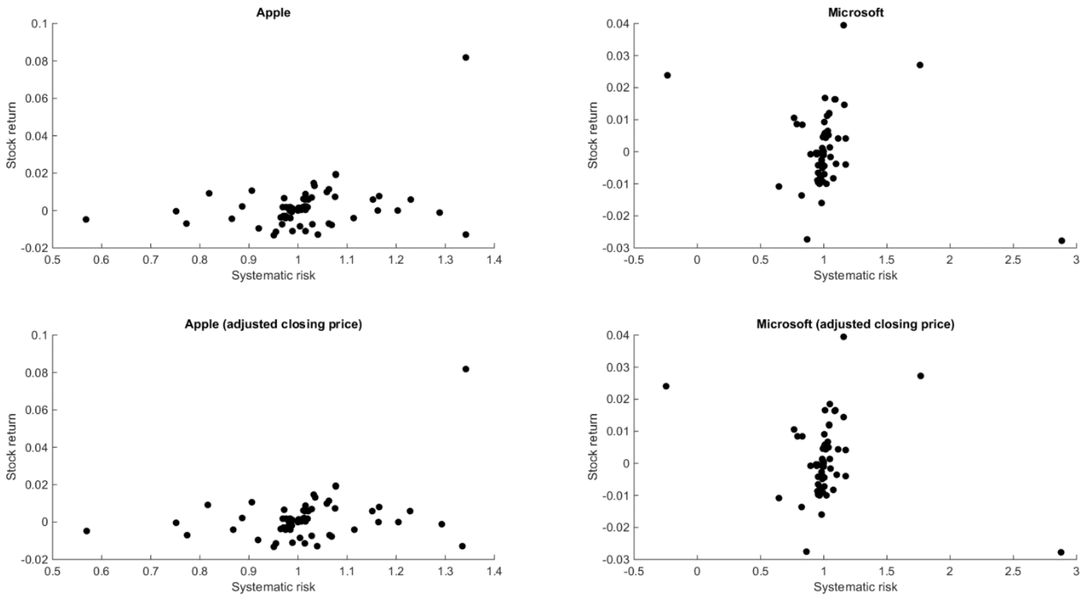

Fig. 1: Relation between the risk and the stock returns

was identified by the Lagrange multiplier test (Tab. 3–4). The results presented in Tab. 3 confirmed that market risk affect stocks in all the identified models at 1% significance level. The effects of the news on the stocks were confirmed only in the model of Microsoft stocks and good news at 1% significance level, and in the case of the other 4 models at 5% significance level and in two models at 10% significance level. The reverse causal effects in direction from the markets to the news was identified in the case of bad and good news related to the company Apple and it products, and in the case of good news related to the company Microsoft and its products. Bilateral causality was identified in the case of good and bad news related to Apple and good news related to both the analysed companies.

Obviously, the results are totally different in comparison with the Wald test. The reason is not only the possible error of the second-order mentioned in the methodological part of this paper but also different VAR models selected for the Granger causality tests. The resulted models, especially the lags, were selected according to the number of significant causal relations (under the assumption of the condition for the minimal required lag given by Akaike and Bayesian information criteria). Summarily, we can assume that the results in Tab. 1–2 may be biased by the error of the second-order. This means that the results may fail to reject a false null hypothesis, thus the results may fail to detect the effects of the news on the capital markets that are present.

Tab. 4 provides results where stock prices are adjusted by any distributions and corporate actions that occurred at any time prior to the next day's opening. These results better record the historical performance and confirmed the effect of the news on the capital markets. However, we can identify much fewer links in comparison with the results presented in Tab. 3. There is significant causal effect of the good news related to the company Microsoft and its products on the stock returns at 1% significance level with the lag of 1 day. In the same model we identified also effects of this news on the market risk premium. The causal effect in the direction from the news to capital markets at 5% significance level was identified in the case of Microsoft stocks and bad news related to both companies.

Tab. 3: Granger causality statistics, Lagrange multiplier test

Bad News			
VAR(2)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (AAPL)	0.6424	9.3083***	0.6597
Market Premium	0.1095	4.9957**	0.0174
Bad News (AAPL + MSFT)	4.6557**	4.6879**	0.3100
VAR(2)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL)**
Risk Premium (AAPL)	0.7648	10.2927***	3.4243*
Market Premium	0.1902	5.1863*	1.4651
Bad News (AAPL)	4.7451**	4.8231**	0.0856
VAR(1)	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (MSFT)	20.8970***	22.9681***	5.2281**
Market Premium	21.9700***	23.4493***	9.1056***
Bad News (AAPL + MSFT)	0.8068	0.7084	6.7864***
VAR(3)	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (MSFT)**
Risk Premium (MSFT)	0.0442	27.3144***	6.5107**
Market Premium	6.4669**	11.3334***	16.7450***
Bad News (MSFT)	1.0849	0.9640	0.0617
Good News			
VAR(1), const.	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (AAPL)	2.3030	14.5587***	2.9538*
Market Premium	0.5477	7.1206***	5.4576**
Good News (AAPL + MSFT)	2.4542	2.3648	9.8139***
VAR(2)	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL)**
Risk Premium (AAPL)	1.0349	29.4166***	4.2049**
Market Premium	1.3066	21.6315***	7.1053***
Good News (AAPL)	5.8221**	5.9787**	12.7513***
VAR(1), const.	**Risk Premium (MSFT)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (MSFT)	20.6646***	24.3159***	4.4195**
Market Premium	24.0339***	25.8299***	1.9896
Good News (AAPL + MSFT)	4.4557**	4.2404**	8.6625***
VAR(1)	**Risk Premium (MSFT)**	**Market Premium**	**Good News (MSFT)**
Risk Premium (MSFT)	18.9652***	22.7299***	21.0364***
Market Premium	21.5313***	23.5420***	10.4915***
Good News (MSFT)	0.0065	0.0149	2.5686

Notes: *, ** and *** denote significance at the 10, 5 and 1% level.

Tab. 4: Granger causality statistics, Lagrange multiplier test, adjusted closing price

Bad News			
VAR(1)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (AAPL)	8.5346***	12.4047***	1.8510
Market Premium	2.7211*	9.7663***	7.0824***
Bad News (AAPL + MSFT)	0.5500	0.6024	9.3682
VAR(1)	**Risk Premium (AAPL)**	**Market Premium**	**Bad News (AAPL)**
Risk Premium (AAPL)	8.4035***	11.0118***	2.7150*
Market Premium	1.9990	8.1766	7.8238***
Bad News (AAPL)	0.3145	0.0000	7.3322***
VAR(1)	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (AAPL + MSFT)**
Risk Premium (MSFT)	22.6093***	25.3058***	5.8036**
Market Premium	22.8656***	25.1621***	9.5345***
Bad News (AAPL + MSFT)	0.7169	0.6215	6.8462***
VAR(1)	**Risk Premium (MSFT)**	**Market Premium**	**Bad News (MSFT)**
Risk Premium (MSFT)	18.6683***	24.0590***	0.2101
Market Premium	23.1452***	26.1626***	0.2435
Bad News (MSFT)	1.4336	1.2405	10.2091***
Good News			
VAR(1), const.	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (AAPL)	4.5873**	15.0571***	2.8407*
Market Premium	0.6713	7.4638***	5.1118**
Good News (AAPL + MSFT)	2.6622	2.5635	9.8050***
VAR(1)	**Risk Premium (AAPL)**	**Market Premium**	**Good News (AAPL)**
Risk Premium (AAPL)	4.8234**	30.0526***	1.9303
Market Premium	2.6874	23.8666***	3.6264*
Good News (AAPL)	6.2321**	6.8226***	12.4235***
VAR(1)	**Risk Premium (MSFT)**	**Market Premium**	**Good News (AAPL + MSFT)**
Risk Premium (MSFT)	19.2966***	23.9271***	2.7963*
Market Premium	22.6996***	25.3400***	0.9935
Good News (AAPL + MSFT)	4.5975**	4.6390**	8.4946***
VAR(1)	**Risk Premium (MSFT)**	**Market Premium**	**Good News (MSFT)**
Risk Premium (MSFT)	19.4082***	24.1781***	20.3466***
Market Premium	21.7214***	24.6692***	10.2522***
Good News (MSFT)	0.0024	0.0082	2.5621

Notes: *, ** and *** denote significance at the 10, 5 and 1% level.

On the contrary, capital markets affected news sent via Twitter in the four models. Three of these four models are VAR(2). Thus, the effects of the news on the capital markets may be slightly faster in several cases than the effects in the direction from the capital markets to the news sent via Twitter.

The section should contain an evaluation and exact description of the achieved results. If the nature of a paper allows it, also state the statistical significance of the results.

5 DISCUSSION AND CONCLUSIONS

In this paper we employed Granger causality to identify causal links between users' content on the social network Twitter – tweets and price of stocks of Apple Inc. and Microsoft Corporation on the New York Stock Exchange. The Wald test which was used proved causality from the direction of the market to the sentiment on Twitter. LM statistics on the other hand showed the existence of both one directional and two directional causal links. The causality of stock markets on the sentiment of tweets was proven mostly in models with positive sentiment tweets, which is a similar result as in the research of Chung and Liu (2011). In both tests causality of sentiment on the premium of Apple and Microsoft was proven, which may for example indicate that some Twitter users are owners of the stocks in question and opinion leaders such as news agencies are informing about the performance of markets. There were also identified causal links from the direction of the whole market (in this case the DJIA) to Twitter.

This research has also proven that simple CAMP is not enough to describe stock price creation and that the factor of feelings and emotions plays its role as is described by behavioural economics.

Possible limitations of our results originate in two causes. It is possible that the methodology of finding the causality between Twitter and specific stocks was creating limitations and research of causality between Twitter and whole market would resulted in proving even more significant causality (e.g. in all the models) as happened in the case of Bollen, Mao and Zeng (2011). The second limitation comes from identifying the sentiment of the tweets. The algorithms that we used were more complex than in previous research, which means that we were able to recognize sentiment and companies with more precision. On the other hand, specifics of colloquial language are far more complicated than the algorithms used could capture.

6 REFERENCES

Aitchison, J. and Silvey, S. D. 1958. Maximum Likelihood Estimation of Parameters Subject to Restraints. *Annals of Mathematical Statistics*, 29, 813–828.

Antweiler, W. and Frank, M. Z. 2004. Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance*, 59, 1259–1294.

Arias, M., Arratia, A. and Xuriguera, R. 2013. Forecasting with Twitter Data. *ACM Transactions on Intelligent Systems and Technology*, 5 (1).

Berndt, E. R., Savin, N. E. 1977. Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model. *Econometrica*, 45 (5), 1263–1278.

Bollen, J., Mao, H. and Zeng, X. J. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 1–8.

Chung, S. and Liu, S. 2011. *Predicting Stock Market Fluctuations from Twitter: An analysis of the predictive powers of real-time social media*. [online]. Available at: `http://www.stat.berkeley.edu/~aldous/157/Old_Projects/Sang_Chung_Sandy_Liu.pdf`. [Accessed 2015, February 16].

Dolan, R. J. 2002. Emotion, cognition, and behavior. *Science*, 298, 1191–1194.

Friedman, M. 1953. The case for flexible exchange rates. *Essays in Positive Economics*. University of Chicago Press, Chicago.

Fama, E. F. 1965. The Behavior of Stock-Market Prices. *The Journal of Business*, 38 (1), 34–105.

Fama, E. F. 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25 (2), 383–417.

Fama, E. F. 1991. Efficient capital markets II. *Journal of Finance*, 46 (5), 1575–1617.

Gilbert, E. and Karahalios, K. 2010. Widespread Worry and the Stock Market. In *Proceedings of the international conference on weblogs and social media (ICWSM 10)*.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral models. *Econometrica*, 37, 424–438.

Kim, S. H. and Kim, D. K. 2014. Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*, 107, 708–729.

Kuleshov, V. 2011. *Can Twitter predict the stock market?* [online]. Available at: `http://cs229.stanford.edu/proj2011/Kuleshov-CanTwitterPredictTheStockMarket.pdf`. [Accessed 2015, February 16].

LeRoy, S. and Porter, R. 1981. The present-value relation: tests based on implied variance bounds. *Econometrica*, 49, 97—113.

Ross, S. A., 1976. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13, 341–360.

Shiller, J. R. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71, 421—436.

Shiller, J. R. 2003. From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, 17 (1), 83–104.

Silvey, S. D. 1959. The Lagrange Multiplier Test. *Annals of Mathematical Statistics*, 30, 389–407.

Sims, C. 1972. Money, income, and causality. *American Economic Review*, 62, 540–552.

Wald, A. 1943. Tests of Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, 54, 426–482.

Zhang, X., Fuehres, H. and Gloor, P. 2011. Predicting Stock Market Indicators Through Twitter: "I hope it is not as bad as I fear". *Procedia – Social and Behavioral Sciences*, 26, 55–62.

AUTHOR'S ADDRESS

Ondřej Fiala, Department of Finance, Faculty of Business and Economics, Mendel university in Brno, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: fialavojtech1@gmail.com

Svatopluk Kapounek, Department of Finance, Faculty of Business and Economics, Mendel university in Brno, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: kapounek@mendelu.cz

Ondřej Veselý, Faculty of Business and Economics, Mendel university in Brno, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: xvesely5@mendelu.cz