

Does Social Media Sentiment Trump News?

by Baoqing Gan

Thesis submitted in fulfilment of the requirements for
the degree of

PhD in Finance

under the supervision of Dr Vitali Alexeev

Dr Christina Nikitopoulos Sklibosios

Dr Danny Yeung

University of Technology Sydney
Faculty of Business

July 2020

Does Social Media Sentiment Trump News?

by **Baoqing Gan**

Thesis submitted in fulfilment of the requirements for the degree of

PhD in Finance

under the supervision of

Dr. Vitali Alexeev

Dr. Christina Nikitopoulos Sklibosios

and Dr. Danny Yeung

University of Technology Sydney

Business School

Finance Department

30 July 2020

Certificate of Original Authorship

I, Baoqing Gan, declare that this thesis, is submitted in fulfilment of the requirements for the award of PhD in Finance, in the Business School at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 30 July 2020

Abstract

The importance of investor sentiment and its influence on financial market has been widely documented. The majority of these studies, however, are either US-centred or focus on a single source of sentiment. In this thesis, I contrast the effects of news and social media sentiment and assess their impacts on markets around the world, using both daily and intraday textual analytics sentiment from Thomson Reuters MarketPsych Indices (TRMI).

In the first chapter, I explore the rapidly changing news and social media landscape and its interplay with market returns and volatility. I find that news media activities (buzz) dominate social media before 2013, while social media has become increasingly important especially after 2016. A similar evolution of lead-lag pattern between news and social media sentiment is also uncovered. Moreover, I discover that market variables exert stronger impact on sentiment than the other way around, and the linkage between volatility and sentiment is more persistent than that between returns and sentiment.

The second chapter examines the role of news and social media sentiment in explaining intraday returns. My analysis of the Dow Jones Industrial Average (DJIA) constituents reveals that sentiment during non-trading hours is a strong yet short-lived predictor of opening returns. Specifically, sentiment from social media induces larger changes than news media. Negative sentiment effects work at higher economic magnitudes than positive sentiment. Nonetheless, these phenomena quickly diminish after the first minute of trading. Robustness tests show that these effects are not driven by corporate earnings announcements. This chapter provides a new set of techniques and develops a novel framework for high-frequency sentiment analysis.

The last chapter applies similar intraday analysis into 14 international markets: Australia, Brazil, Canada, the EU, France, Germany, Hong Kong, India, Japan, Singapore, Spain, Switzerland, the UK and the US. I find that the dominant role of social media in US is not representative of other global markets. News media sentiment expounds a greater impact on stock prices in other major financial markets. Robustness tests show that the aggregation of sentiment up to three hours prior to the market opening helps generate an effective signal for predicting the direction of the opening prices. This chapter underscores the importance of avoiding adopting US evidence naively to other markets.

Overall, this thesis contrasts effects of news sentiment with that of social media sentiment. Applying a novel dataset of high-frequency text analytics, this thesis provides an approach to help shed light on the role social media sentiment plays in the dynamics of stock markets.

Acknowledgements

I could not thank enough to my principal supervisor Dr Vitali Alexeev, who is definitely one of the most important persons in my life. His professionalism and integrity in supervision have facilitated to my rapid progress in the last three-and-half years. From the down-to-earth coding skills to a broader scope of strategies on academic career development, his knowledge, experience, guidance and encouragement demonstrate not only an outstanding supervisor, but also a friend and a role model who is always caring, understanding, dedicating and contributing to the best interest of his student. In my most challenging times, his continuous empathy alongside the rational thinking and effective suggestions always pull me out of the negative sentiment in time.

I am also in gratitude for my supervisory panel of their consistent patience and encouragement during my candidature. I thank my co-supervisor Dr Danny Yeung for his effective communication and extraordinary academic writing skills that are crucial to my training process as an early career researcher and instructor.

In addition, I want to thank Dr Christina Nikitopoulos Sklibosios and Professor Ron Bird. The early stage communication with them has helped me so much to successfully obtain my scholarships, without which I cannot finish this research. I thank UTS Graduate Research School and Business School of funding my doctoral study. I thank Thomson Reuters Financial and Risk (now re-structured as part of Refinitiv) for offering MarketPsych Indices (TRMI) as part of my research data, and I am grateful for the assistance of Professor Ron Bird in the TRMI dataset acquisition. I am thankful to my PhD candidate peers and the brilliant staffs at UTS Finance Department, whose hard-working attitudes and outstanding performance pushed me to make my best effort in achieving higher goals.

I am grateful for two examiners, Renee Adams, Shan Chen, Robert Durand, Adrian Fernandez Perez, Sean Foley, Bart Frijns, Brian Lucey, Roberto Pascual, Talis Putnins, Terry Walter, Tianyu Zhang, two anonymous referees and other participants at the 2020 World Finance Conference, the 33rd PhD Conference in Economics and Business (Monash), the WRDS 2019 Advanced Research Scholar Program, the 2019 FIRN Annual Conference PhD Symposium, the 2019 New Zealand Finance Meeting, the 8th SIRCA Young Researchers Workshop and the 2018 New Zealand Finance Meeting PhD Symposium, as well as the UTS Finance PhD Presentations for their helpful comments on the

improvement of the chapters in this thesis.

Finally, I am full of gratitude to my family members and friends, especially my “teammate”—my husband Kai Zhang for keeping telling me that I am an outstanding lady who deserves the opportunity to pursue whatever career I wanted to take on, with his full support, for being together on this road less travelled, for comforting my blue days and for cooking fancy dishes during my ups-and-downs. Thanks my parents for staying strong in their difficult times when mum spent 11 days in ICU and slowly recovered from a cerebral hemorrhage without their single child alongside. Thanks Kay and Matt for organising various events on weekends and holidays, and led us quickly adapt into lives in beautiful Sydney. Thanks my "neighbours" in the office: Dr Kristoffer Glover, Prof Jianxin Wang, and Dr Deborah Cotton whose exemplary working attitudes and encouragement around me everyday in the past four years have boosted my confidence in each step of my progress. Lastly, thanks Sergei Rachmaninoff and Pyotr Ilyich Tchaikovsky, whose immortal productions of piano concertos kept me companion, calmed my nerves and sparkled with my coding enthusiasm while I was working at my workstation alone.

Contents

1	Introduction	1
2	Sensitivity to Sentiment: News vs Social Media	7
2.1	Introduction	9
2.2	Literature Review	11
2.3	Data and Methodology	16
2.3.1	Sentiment Data	16
2.3.2	Stock Market Data	18
2.3.3	Data Aggregation Process	19
2.3.4	Econometric Framework	19
2.4	News vs Social Media: Dominating Causality Pattern	20
2.5	Media vs Market: Sub-sampling Period Comparison	25
2.5.1	Sentiment vs Return	25
2.5.2	Sentiment vs Volatility	31
2.6	Conclusion	34
3	Investor Sentiment under the Microscope	37
3.1	Introduction	39
3.2	Data and Methodology	44
3.2.1	Sentiment Data	44
3.2.2	Stock Price Data	45
3.2.3	Data Aggregation	45
3.3	Findings	49
3.3.1	Opening Return Patterns	52
3.3.2	After-Hours Media Sentiment Patterns	67
3.4	Robustness Checks and Discussion	71
3.4.1	Tackling the Causality Loop	71
3.4.2	Investor Sentiment and Earnings Announcements	74
3.4.3	Event window choice	76
3.5	Conclusion	79
4	Do Emotions Trump Facts? Evidence from around the World	81
4.1	Introduction	83
4.2	Data and Methodology	87
4.2.1	Data Source	87

4.2.2	Model Specifications	91
4.2.3	Descriptive Statistics	94
4.2.4	Model Validity	95
4.3	Sensitivity to Overnight Sentiment in Global Markets	97
4.4	Robustness Checks	109
4.5	Conclusion	113
5	Conclusion	115
A	Main Appendix	127
A.1	Appendix for Chapter 2	128
A.1.1	List of acronyms and notation	128
A.1.2	Testing for Unit Roots	129
A.1.3	ACF and PACF for main TRMI series	130
A.2	Appendix for Chapter 3	133
A.3	Appendix for Chapter 4	138
B	Supplementary Appendix	143
B.1	Tried-and-true vs Bold-and-New: on commonality between Baker & Wurgler and MarketPsych Indices	144
B.2	Robustness check: model selection	144
B.2.1	Optimal Lag Length: Information Criterion	146
B.2.2	Optimal Lag Length: Why VAR(1) is Parsimonious Form of VAR (7)	147
B.2.3	Optimal Lag Length: Likelihood Ratio test	149
B.2.4	Model specification: VAR subsystems and omitted variable bias	151
B.3	TRMI data and variables	154
B.4	Correlation Schema-Balls	161
B.5	One day lag cross correlations between social and news.	162
B.6	Principal Component Analysis	163
B.6.1	Scree Plots for Social and News Series	163
B.6.2	Biplots of the first two principal component coefficients	164
B.7	Number of TRMI Observations: Raw	165
B.8	Data pre-processing and Distribution of Event-days	167
B.9	Days with Excessive Sentiment and the Overlap with Earnings Announcements	168
B.10	Examples of other stocks	171

List of Figures

2.1	Rolling Window VAR(1) Off-Diagonal Elements - daily <i>Buzz</i>	22
2.2	Rolling Window VAR(1) Off-Diagonal Elements - daily <i>Sentiment</i>	25
2.3	<i>Sentiment</i> vs <i>Return</i> Sub-sample Comparison	28
2.4	<i>Sentiment</i> ² vs <i>VIX</i> Sub-sample Comparison	33
3.1	Visual representation of sentiment and return data for Apple Inc	47
3.2	Social Media: Overnight Sentiment and Opening Returns (CSCO.OQ)	50
3.3	News Media: Overnight Sentiment and Opening Returns (CSCO.OQ)	51
3.4	Returns on long-short strategies conditional on sentiment	60
3.5	Cumulative Returns and Ensuing Sentiment (CSCO.OQ)	69
3.6	Contrasting Betas	75
4.1	Heatmap of News Media sentiment	89
4.2	Heatmap of Social Media Sentiment	90
4.3	Proportion of the overnight return variation explained	103
4.4	Market sensitivities to overnight sentiment	107
4.5	Market reactions to positive and negative sentiment	108
A.1	Time-series Analysis of raw <i>Buzz</i> data	130
A.2	Winsorized and De-Seasoned <i>Buzz</i> Time Series Check	130
A.3	Raw <i>Sentiment</i> Time Series Check	131
A.4	De-Seasoned and Market Merged <i>Sentiment</i> Time Series Check	131
A.5	Maximum likelihood estimate of the Box-Cox transformation	132
A.6	Contrasting estimated of controlled and uncontrolled models: The case of News Media	135
A.7	Alternative Event Window lengths: the case of social media for CSCO.OQ	136
A.8	Alternative Event Window lengths: the case of news media for CSCO.OQ	137
B.1	Likelihood ratio test results	150
B.2	Rolling Window VAR(1) Off-Diagonal Elements - daily <i>Buzz</i>	152
B.3	Rolling Window VAR(1) Off-Diagonal Elements - daily <i>Sentiment</i>	153
B.4	Contemporaneous Correlation Dynamics between Key Social and News Indices	156
B.5	Timeline of textual content analysed for the social and news media TRMI	158
B.6	Correlation coefficients between various emotional scores for the Company Group	161
B.7	One Day Lag Cross-Correlation between Key Social and News scores	162
B.8	Scree Plots from Principal Component Analysis of emotional scores for the Company Group	163

B.9 Principal Component Analysis of the Sentiment Indices	164
B.10 <i>Buzz</i> Data Availability: News vs Social media	166
B.11 <i>Sentiment</i> Data Availability: News vs Social media	166
B.12 Event Clustering and Earnings Announcements Overlap (AAPL.OQ)	168
B.13 Distribution of TRMI Social and News Media Sentiment by Time of Day for AAPL.OQ	171
B.14 Social Media: Overnight Sentiment and Opening Returns (AAPL.OQ)	172
B.15 News Media: Overnight Sentiment and Opening Returns (AAPL.OQ)	173
B.16 Cumulative Returns and Ensuing Sentiment (AAPL.OQ)	174

List of Tables

2.1	Descriptive Statistics for the company group	18
2.2	Before vs After Transition Period VAR Slope Coefficients: Social vs News	23
2.3	Before vs After Transition Period VAR Slope Coefficients: Sentiment Vs Market	27
3.1	Daily CARs conditional on the top and bottom overnight sentiment	58
3.2	The first half-hour CARs conditional on the top and bottom overnight sentiment	61
3.3	The first hour CARs conditional on the top and bottom overnight sentiment	63
3.4	Morning session’s CARs conditional on the top and bottom overnight sentiment	65
3.5	Cumulative Excess Return and Overnight Sentiment	70
3.6	Sentiment as a Predictor for Returns	74
3.7	Coincidence between earnings announcement days and strong sentiment days	77
4.1	Data Source	87
4.2	Descriptive Statistics	95
4.3	Pairwise correlation coefficients: DJIA and FTSE100 close-to-open	96
4.4	Coincidence rates in extreme sentiment	97
4.5	Sentiment Sensitivity of the DJIA.	100
4.6	Sentiment Sensitivity of the FTSE100.	101
4.7	Global market sensitivity to overnight sentiment	106
4.8	DJIA three-hour cumulative sentiment regressions	111
4.9	FTSE100 three-hour cumulative sentiment regressions	112
A.1	List of acronyms, data sources and variable names in Chapter 2	128
A.2	Unit Root Test	129
A.3	List of acronyms, data sources, and variable definitions in Chapter 3	133
A.4	Sentiment Data Availability	134
A.5	Acronyms, Variables and their Definitions	138
A.6	TRMI Descriptive Statistics	139
A.7	Oxford-Man Descriptive Statistics	140
A.8	Variance Inflation Factors	141
B.1	Correlation Between BW and TRMI Sentiment Indices	144
B.2	Optimal Lag Selection using Information Criteria	146
B.3	VAR(7) Parsimonious Form Examination (A)	147
B.4	VAR(7) Parsimonious Form Examination (B)	148
B.5	Descriptive Statistics for TRMI MPTRXUS500 Company Groups based Social Media	159

B.6 Descriptive Statistics for TRMI MPTRXUS500 Company Groups based News Media . 160
B.7 TRMI DJIA Constituents Sample Number of Observations 165
B.8 Overnight Sentiment Events Sample Distribution 169