Pure

Bond University

DOCTORAL THESIS

News analytics for quantitative equity portfolios.

Marty, Thomas

Award date:
2021

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

News Analytics for Quantitative Equity Portfolios

Tom Marty

Submitted in total fulfillment of the requirements for the degree of Doctor of Philosophy

July 2021

Bond Business School

Professor Bruce Vanstone, Dr Tobias Hahn, and Associate Professor Christopher Bilson

This research was supported by an Australian Government Research Training Program Scholarship

Abstract

This thesis consists of three related studies investigating the practical economic utility of news analytics in momentum-style equity portfolios. We employ a dataset of 5.3 million news items relating to historical constituents of the S&P 500 stock index from 2003 to 2018 to examine exploitable relationships between news media and stock returns over one - six month horizons.

The first study inspects the predictive capacity of news media at the stock level through firm-specific regression analyses. Tests include single period regressions subject to a variety of subset and specification robustness tests, extension to multi-period forecast horizons, and VAR cross-dependency analysis. With rates of statistical significance comparable to pure noise covariates and weak effect sizes, we fail to find support for the hypothesis that either news sentiment or news coverage are useful predictors of forward return on a firm-by-firm basis.

The second study examines the individual and joint impacts of news sentiment, news coverage, and stock price momentum on expected return by analysing the performance of decile portfolios formed through univariate, bivariate, and trivariate sorts on these variables. The style of news-enhanced trading strategies identified in the literature do not appear to be profitable over our sample period and investment universe. However, we find some evidence that news sentiment—if used as a screening mechanism in the short leg of momentum strategies—can enhance risk-adjusted returns. Overall, this study provides little evidential support for the claim that news-derived measures provide useful conditioning information for the cross-section of returns in naive implementations.

The third study tests the economic utility of news content using a model-based portfolio procedure in which portfolios are formed using the output of statistical models trained over backward-looking filtrations and tested over out-of-sample periods. We find that the inclusion of news-based variables in the conditioning information set, combined with flexible statistical learning algorithms, offers only modest increase in performance beyond a traditional momentum implementation. Measures of variable importance suggest that news is secondary to size, analyst following, and momentum in relevance for predicting future return.

These findings provide little evidence that news analytics are an economically useful data source in improving quantitative equity strategies operating over multi-month investment horizons. Our results question the robustness and generalisability of previous findings, tested over less-stringent investment investment scopes or in higher-frequency implementations, that find news analytics to be a straight-forward means of generating excess returns.

Keywords: News sentiment, Media coverage, Textual analysis, Investor sentiment, Momentum strategies, Attention, Information diffusion

JEL Codes: G12, G14

Declaration by Author

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy.

This thesis represents my own original work towards this research degree and contains no material that has previously been submitted for a degree or a diploma at this University or any other institution, except where due acknowledgement is made.

Tom Marty

Research Outputs During Candidature

Peer-reviewed Publications

Marty, Tom, Bruce Vanstone, and Tobias Hahn (2020). "News Media Analytics in Finance: A Survey". In: *Accounting & Finance* 60, pp. 1385-1434. https://doi.org/10.1111/acfi.12466

Copyright Declaration

Copyright permission has been granted by the publisher for the following publication, parts of which have been reproduced within Chapter 2 of this thesis:

> Marty, Tom, Bruce Vanstone, and Tobias Hahn (2020). "News Media Analytics in Finance: A Survey". In: *Accounting & Finance* 60, pp. 1385-1434. https://doi.org/10.1111/acfi.12466

Reuse of the content within the above publication was granted as per the relevant Copyright Transfer Agreement between the contributors and the publisher. Reuse rights for the purposes of theses and dissertations are also stated on the publishers website (Wiley, 2021).

Acknowledgements

First, I would like to thank Professor Bruce Vanstone for not only replying to my initial email—which I think had some weird (yet honest) first-line disclaimer that I knew nothing—but agreeing to meet and provide advice on how to proceed with a thesis proposal. At the time, I was a complete stranger and I would not have had this opportunity if not for this act of good faith and generosity. Similarly, I would like to thank Dr Tobias Hahn, who was involved and supportive from the very beginning.

Both Bruce and Tobias have been valuable sources of guidance and wisdom throughout this process, as well as having demonstrated spectacular patience in the face of my insistent and sometimes pugnacious argumentation. Much of what I have learned during the PhD program is the result of their nudges, observations, and advice. This is not to mention all their direct work involved in supervising a thesis, such as proof-reading etc, for which I am also thankful.

I would also like to thank Professor Tom Smith for his two PhD courses. Tom's passion and enthusiasm for finance theory and empiricism were unexpectedly contagious, and his pedagogy truly inspiring. There is a lot more to say about Tom's teaching and its influence, and I would have thanked him earlier, but fear of perceived schmoozing was always enough to prevent my pen hitting the page.

Dr Christopher Bilson joined Bruce and Tobias as a supervisor quite late in the project. Having now had exposure to his intelligence, energy, and creativity, it is clear that not having him on board sooner is only to the detriment of this thesis. Nevertheless, Chris' contributions have certainly shaped the final result, and our discussions have given me much to think about. Thank you Chris.

I am grateful to the Bond Business school and Bond University more generally, for giving me this opportunity and providing all the necessary scaffolding and support services it takes to organise something like this. Further, I could not have attempted this at all without a scholarship, for which I am deeply grateful to have received.

I would also like to thank Dr David Petty and Professor Vince Wheatley, for providing the initial references required for the application to the PhD program in the first place.

I thank Eric Fishkin and Thomson Reuters for providing access to the news analytics archive.

My pursuing this thesis has meant much less of my time being spent with family and friends. To those that wore this cost with support and understanding, thank you.

Finally, I am grateful to my mother for having way more enthusiasm for my small victories along the way than I would ever allow myself.

Contents

List of Tables

List of Figures

Acronyms

ANN artificial neural network. 30

B/M book-to-market. 47

BOW bag of words. 18

CAPM capital asset pricing model. 36

CAR cumulative abnormal returns. 58

CFG context-free grammar. 28

CUSUM cumulative summation. 43

DAX Deutscher Aktienindex. 45

DJIA Dow Jones Industrial Average. 30, 31, 36

DJNS Dow Jones News Service. 53

FI-GARCH fractionally integrated GARCH. 57

FSD first-order stochastic dominance. 76

fsQCA fuzzy-set qualitative comparative analysis. 72

FTSE Financial Times 100. 30

GFC Global Financial Crisis. 51

GI General Inquirer. 23, 31

GJR-GARCH Glosten-Jagannathan-Runkle GARCH. 63

H-IV4 Harvard IV-4. 23, 31

HSI Hang Seng Index. 30

kNN k-nearest neighbours. 27, 30

LDA latent Dirichlet allocation. 26

LEI Conference Board Leading Economic Indicators Index. 42

LM Loughran and McDonald (2011). 23

LPM lower partial moment. 77

LSI Latent Semantic Indexing. 25

MCS model confidence set. 63

MD&A Management Discussion & Analysis. 12

MDH mixture-of-distributions hypothesis. 3

MLE maximum likelihood estimation. 30

MS-EGARCH Markov switching EGARCH. 60

NB naïve Bayes. 26

NER named entity recognition. 21

NLP natural language processing. 27

OLS ordinary least-squares. 36

PCA principal components analysis. 64

PLSI probabilistic LSI. 25

POS part-of-speech. 20

RMSE root-mean-square error. 46

RS-GARCH regime switching GARCH. 57

SPY SPDR S&P 500 ETF Trust. 40

SSD second-order stochastic dominance. 77

SSI Singapore Straights Index. 30

SVM support vector machines. 26, 31

TAIEX Taiwan Capitalisation-weighted Stock Index. 44

TF-IDF term-frequency/inverse-document-frequency. 24

TRMI Thomson Reuters Marketpsych Indices. 60

TRNA Thomson Reuters News Analytics. 33, 34, 36

TSD third-order stochastic dominance. 77

TVIX Taiwan volatility index. 44

UGC user-generated content. 12

VAR vector autoregression. 31

VaR value-at-risk. 60

VIX CBOE Volatility Index. 36

WSJ Wall Street Journal. 34

Chapter 1

Introduction

1.1 Introduction

This thesis arises from challenges in understanding the practical relevance of news analytics for quantitative equity portfolios. It presents a detailed literature survey and three empirical analyses aimed at mitigating these challenges and providing a more complete understanding of the economic utility of news analytics in asset management.

The thrust of the thesis can be summarised as follows:

1. It is conceptually and empirically credible that news media exerts some influence over financial markets.

2. Technological advances have opened a unique avenue of news-based analysis, so-called *news analytics*, still in its relative infancy in financial research.

3. The majority of the news analytics literature has focused on high-frequency applications, limiting its practical relevance to many fund management practitioners.

4. News-based factor-style portfolios, particularly news-informed[1] momentum portfolios, are one of the most promising applications of news analytics in traditional investment environments.

5. Even in this most relevant subset of the field, the practical implications for the use of news analytics in low-frequency investment environments are unclear. This is due in part to both the experimental setting and empirical constraints of the existing literature.

6. This thesis attempts to bridge some of the described literature gap by providing a better understanding of the economic relevance of news analytics in quantitative equity portfolios.

    - The first contribution of this thesis is in its choice of empirical setting. This includes data sources, investment universe, and strategy parameterisations that are directly applicable to asset management.

[1] By *news informed*, we simply mean that we have conditioned on news information in some way.

- The second contribution comprises three empirical analyses. Broadly, these test the firm-level predictive capacity of news analytics, the performance of literature-motivated news-informed portfolios, and the extension of "naive" sorting procedures to statistical learning approaches.
- The third contribution is the synthesis of a wide-spanning literature into a survey of the relevant research.

Points one to five, the motivation for this thesis, will be briefly described in the first section of this chapter—*Motivation*. Point six, the way in which this thesis engages with the stated motivation, is expanded upon in the second section of this chapter—*Objectives and Contributions of Thesis*.

1.2 Motivation

It is conceptually and empirically credible that news media exerts some influence over financial markets.

Efficient valuation of stock prices hinges on the present expected discounted value of future cashflows. Empirically, stock price movements are largely unexplainable in terms of changes in quantitative measures of firms' fundamentals (Malkiel, 1977; Shiller, 1981; Mankiw, Romer, and Shapiro, 1985) and discount rates (Campbell and Shiller, 1988; Hansen and Singleton, 1982) alone, suggesting that other sources of information are also used as conditioning information by investors.

Considering the hard-to-quantify nature of many company and world events, it is not surprising that analysts' forecasts and publicly disclosed accounting variables are potentially incomplete sources of information at any given time. News media, as a high-exposure vehicle for qualitative descriptions of global occurrences and firms current and future production activities, is plausibly a source of value-relevant information with incremental explanatory power for firms' future cashflows above and beyond traditional quantitative variables (Tetlock, Saar-Tsechansky, and Macskassy, 2008).

Even when news does not contain value-relevant information, many may still trade as if it does. Huberman and Regev (2001) document the salient case of EntreMed, in which a *New York Times* story citing potential development of cancer-curing pharmaceuticals saw the firms' stock price rise from $12.063 at the Friday close to open at $85 Monday, and close above $30 over the following three weeks. The enthusiasm generated by the story, which spilled over to other stocks within the sector, was based on information which had been reported in various sources such as *Nature* and the *Times* over five months earlier. While this is a particularly dramatic example, news media has long been recognised as an important force in the movement of financial markets and the relationship between the two has naturally attracted considerable attention within the academic community.

The earliest attempt to quantify the content of news and its impact on stock prices appears to have been Niederhoffer's (1971) analysis of *New York Times* headlines and subsequent movements in the S&P Composite Index. While computational means of content analysis were available to Niederhoffer (Stone, Dunphy, and Smith, 1966), contemporary technology was such that he deemed its use "not a worthwhile undertaking"—a more pragmatic solution was to have three people categorise and score each headline manually (Niederhoffer, 1971). It is only within the last decade, through the widespread digitisation of text, vast expansion of online media resources and the evolution of big data technologies, that quantifying the content of large sets of public information has become feasible. In particular, modern text-mining capabilities have presented a new dimension of inquiry, *sentiment analysis*, through which to examine the financial markets.

The EntreMed example is not only a testament to the role of media in transmitting information in financial markets; the market's response to the "exceptionally optimistic" (Huberman and Regev, 2001, pp. 396) article is also a conspicuous demonstration of the power of language to influence public decision making (empirical evidence for the causal effect of media language on investor behaviour is provided by Engelberg and Parsons, 2011). This apparent relationship between the language of news and investor psychology is the primary driver behind much of the sentiment analysis literature.

As classical asset-pricing models derive demand functions independently of the distribution of equilibrium prices, they have little to say about information arrival or sentiment. Alternatively, the influence of news media on the financial markets may be framed within the context of rational expectations models, or more specifically, in the context of noise trading. Theoretical models in which traders have heterogeneous beliefs and information sets often posit the existence of two types of traders; noise traders who trade on extraneous factors that convey no information about the true value of the asset and rational arbitrageurs who trade as informed Bayesians[2] (DeLong et al., 1990; Subrahmanyam, 1991).

The uninformed demand shocks of noise-traders provides the incentive and liquidity necessary for informed trading, facilitating the price formation process (Kyle, 1985; Foster and Viswanathan, 1993b). This process results in prices impounding the informed traders' information as well as the noise that carried there, a property necessary for trade to occur in the first place (Milgrom and Stokey, 1982; Grossman and Stiglitz, 1976). The bias, or excessive optimism or pessimism of noise-traders' expectations of an asset's value relative to that of a rational and informed investor is referred to in the finance literature as sentiment. Conceptually, this gives rise to an often-cited goal of news-based sentiment analysis; to determine how the content and arrival of news stories impacts investor sentiment about assets and how to ascertain the aggregate direction of resultant noise trading in those assets.

Other theoretical models concerned only with rational- though informationally-incomplete agents link news coverage to returns directly—such as through compensation for imperfect diversification (via investor recognition) (Merton, 1987) or degree of information asymmetry (to the extent that news media reveals private information) (Easley, Hvidkjaer, and O'Hara, 2002; Easley and O'Hara, 2004).

Also relevant to the theoretical framing of news in the financial markets is the mixture-of-distributions hypothesis (MDH), which posits a direct link between public information arrival, trading volume and variance through the price evolution process (Clark, 1973; Epps and Epps, 1976; Tauchen and Pitts, 1983). According to the MDH, the period price change and trading volume are mixtures of independent distributions in which the number of new pieces of information arriving to the market is the mixing variable. It further predicts that the trading volume and associated price variance depend on the extent to which traders disagree in response to the new information. Other models concentrating on the

[2]Noise traders may also be modeled through a noisy supply, see for example Admati (1985) or Easley and O'Hara (2004).

information-variance relationship have further shown that content of news, insofar as it is either "good" or "bad", produces an asymmetric volatility response which is also associated with volatility persistence (see for example Veronesi, 1999; Engle and Ng, 1993). These models motivate a second path of investigation within news-based sentiment analysis; to determine how the arrival and content of news stories impacts the distribution of investor expectations and the resultant volatility and volume response.

Whether news carries genuine but hard-to-quantify information, influences or proxies the biases (optimism/pessimism) of noise traders, or captures risks related to investor recognition and information asymmetry, such causal explanations give credibility to its potential use as a conditioning variable in markets.

Technological advances have opened a unique avenue of news-based analysis, so-called *news analytics*, still in its relative infancy in financial research.

As inefficient, boundedly rational, or otherwise noisy trading in response to news content can result in a price deviation from fundamental values, sentiment analysis provides a potential means of predicting news-induced arbitrage opportunities and heightened holding risks before they occur. It follows then that quantitative proxies for news sentiment may provide important exogenous (Dougal et al., 2012) conditioning information for the investment decisions of portfolio managers. This application of sentiment analysis has not been overlooked.

In industry, mainstream news organisations such as Bloomberg and Thomson Reuters not only distribute stories themselves, but through the use of natural language processing and information retrieval technologies now equip traders with measures of story sentiment, relevance, topic codes, novelty and many other elements of metadata to facilitate trading decisions (Thomson Reuters, 2015; Cui, Lam, and Verma, 2017). It is this quantitative compression of textual news media that we refer to as *news analytics*. The potentially profitable edge provided by the quantification of news information has likewise influenced academia and a vast number of studies have now explored the performance of trading strategies based on signals derived from little more than the occurrence and content of news media. The associated literature spans a diverse range of strategies, assets, markets, news sources, and textual analysis techniques and has yielded an equally diverse range of results.

The majority of the news analytics literature has focused on high-frequency applications, limiting its practical relevance to many fund management practitioners.

Despite its prevalence and methodological diversity, the majority of the news sentiment literature has treated sentiment as either a daily or intraday variable. When measured daily, the market impact has typically been observed to dissipate within five trading days of the measurement or event. When treated as an intraday or per story variable, the measurable market response is on the order of minutes. While the news sentiment literature has robustly demonstrated these measurable and often exploitable news sentiment effects at high frequencies, these intra-week and intraday fluctuations, and the trading strategies presented to exploit them, offer little promise for the use of news sentiment for mid to long-term investment horizons.

The focus of the textual sentiment literature on daily and intraday horizons is surprising given the treatment of sentiment in the wider finance literature. Empirically, investor sentiment is often treated as a low-frequency variable, examined over months and years rather than days and minutes (Schmeling, 2009; Brown and Cliff, 2005; Baker and Wurgler, 2006; Baker and Wurgler, 2007). In some instances this is at least partially due to the availability of survey and accounting data used to proxy sentiment (Brown and Cliff, 2004), however, two reasons have commonly been put forward as to why the relationship between sentiment and price is likely to operate beyond short-term horizons.

First, as suggested by Brown and Cliff (2005) and Schmeling (2007) among others, sentiment is a persistent variable; demand from noise traders is likely to be correlated over time due to positive feedback effects, resulting in cumulative deviations from fundamental values. DeLong et al. (1990) provide a theoretical basis for this phenomenon in which correlation in noise trader sentiment bears upon equilibrium stock returns. Second, limits to arbitrage can allow persistent mispricings to propagate through time for extended periods before correction. These limits to arbitrage include the potential difficulty in measuring sentiment-driven mispricing directly at short horizons (Baker and Wurgler, 2006; Summers, 1986); the risk that noise traders' beliefs may take a long time to revert to the mean, only becoming more extreme in the interim (Black, 1986; DeLong et al., 1990); the interaction of agency costs and capital constraints (Shleifer and Vishny, 1997); and fundamental risk as discussed by Shiller (1984) and Campbell and Kyle (1993).

News-based factor portfolios, particularly news-informed momentum portfolios, are one of the most promising applications of news analytics in low-frequency investment environments.

The view of sentiment as a persistent variable is also consistent with long-established pricing anomalies including short-medium run momentum, long-term reversals, closed-end fund discounts, and post earnings announcement stock price drift, to name a few. Such anomalies, characterised by various forms of over- and under-reaction, have been addressed by a number of behavioural and boundedly rational models of security markets inspired by DeLong et al. (1990). Notable examples from this literature, which include Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998), and Hong and Stein (1999), each explain empirical over- and under-reaction anomalies through different mechanisms of biased or incomplete processing of new information, such as price movements and corporate events.

It is within this theoretical context that the relatively few studies to empirically investigate the relationship between news media and stock returns at beyond-daily horizons are placed. Collectively, they have raised important findings regarding the predictability of returns as it relates to news and have provided promising evidence in support of the development of profitable news-augmented trading systems with investment horizons spanning months and even years.

For instance, it has been shown that the incidence, volume and content of news stories strongly predicts whether past winners and losers will exhibit stock price momentum or experience short-term reversals (Chan, 2003; Hillert, Jacobs, and Mueller, 2014; Sinha, 2016; Huynh and Smith, 2017). Firm-specific news volume has also been shown to have a cross-sectional pricing effect that lasts for at least 12 months after measurement (Fang and Peress, 2009). News sentiment aggregated over weekly intervals has been found to be predictive of returns for the following quarter, while news sentiment aggregated over daily intervals was only found to be predictive of returns for the following one or two days using the same data (Heston and Sinha, 2017).

The practical implications for the use of news analytics in low-frequency investment environments remain unclear. This is due in part to both the experimental setting and empirical constraints of the existing literature.

While the above findings have established a number of promising news-based effects, the utility of news-based variables as conditioning information for explaining future return movements, and the extent to which this information can be exploited in realistic trading environments, is still largely unexplored. Very few of the relevant studies have restricted their test sample to an investable universe comprising only large, highly liquid firms and none have tested their trading strategies using only firms belonging to a prominent market index such as the S&P 500. Such indices describe a clearly defined set of assets with well understood characteristics and are invested in heavily by funds management practitioners. As such, economic performance and effect size within S&P indices represents an important benchmark for the economic relevance of news-based effects which is absent from the extant literature.

Firm-level or longitudinal predictability is not strictly necessary for the type of cross-sectional predictability exploited by factor sorts. However, it is one pathway to the formation of rank-based portfolios that is missed by the assumption of shared factor sensitivities inherent in purely cross-sectional investigations. For example, if firms of one type exhibit continuation in response to news coverage, while firms of another exhibit reversal, a simple sort on news coverage will likely reveal no relationship between news coverage and expected return, even if the firm-level relationships were deterministic. Yet, no studies have performed large-sample firm-level investigations over the horizons of interest, and even preliminary results in this area would be informative.

In the case of cross-sectional approaches, each of the relevant studies have focused on either individual or specific joint effects of news and momentum variables. Methodological differences between these studies make it difficult to infer an overall picture of the contributions and relative importance of each of the conditioning variables in portfolio formation. An investigation in which each of these variables is studied in individual and joint conditioning procedures, using a consistent experimental setting, and with relevant robustness tests, is needed to compare and clarify the news-driven effects documented so far.

As alluded to above, interaction effects between news-based variables and other conditioning information such as formation period return, firm-specific characteristics and market uncertainty constitute some of the most significant findings in the relevant literature regarding the relationship between news information and firms' future stock-price performance (see for example Chan, 2003; Sinha, 2016; Smales, 2016). However, such effects have only been explored in simple monotonic sorting and panel regression frameworks which, while transparent and intuitive, are restrictive and inflexible from a modelling perspective.

These model-based empirical limitations are unique to this subset of the news sentiment literature. For instance, clustering methods, artificial neural networks, support vector machines, decision trees, genetic algorithms and Bayes classifiers are among the statistical learning techniques that have been used to examine the predictive capacity of news at daily and intraday horizons (Wüthrich et al., 1998; Geva and Zahavi, 2014; Schumaker and Chen, 2009; Feuerriegel and Prendinger, 2016; Maragoudakis and Serpanos, 2016). Unlike parametric linear models, non-parametric and semi-parametric statistical models are highly adaptive and are able to exploit complex internal structures within the data with minimal distributional and relational assumptions. The extant literature has shown the relationship between news information and stock price behaviour to be a time-varying, multi-dimensional and nonlinear phenomenon. It is therefore possible that extending the current work beyond least-squares estimators and double sorting to larger, more powerful statistical models may yield new and useful insights into the impact of news in markets.

1.3 Objectives and Contributions of Thesis

1.3.1 Research Aims

The objective of this thesis is to obtain a better understanding of the economic relevance of news analytics in forming realistic equity portfolios. Due to the shortage of literature at the relevant horizons, inconsistencies in experimental choices between existing studies, and the often limited empirical approaches applied, the state of research on this topic is far from conclusive. Clear results on this topic are of interest to both fund management practitioners and academics. In order to engage with the stated objective, this thesis focuses on three broad research questions, which follow from the motivations of the thesis described above:

- Is news data an economically useful conditioning variable in the formation of momentum-style equity portfolios?
- Is the economic utility of news data in the formation of momentum-style equity portfolios enhanced through the use of flexible statistical learning models?
- Is news data a meaningful predictor of excess returns at the firm level?

In light of these research questions, the contributions of this thesis are outlined below.

1.3.2 Contributions

All studies employ a proprietary Thomson-Reuters dataset relating to historical constituents of the S&P 500 stock index from 2003 to 2018. The use of an institutional-grade news analytics platform represents a realistic option for asset managers (compared to say, those based on the archive subscriptions of educational institutions) and reduces the idiosyncrasies of custom news-processing techniques often applied in the literature[3].

Similarly, the choice of investment universe is important—none of the most promising news-informed investment strategies cited above have been benchmarked within a well known index. As discussed by Vanstone and Hahn (2017), S&P indices offer a clearly defined and highly investable strata with well-understood characteristics that appeal to fund managers. Testing within such indices therefore provides a useful benchmark for portfolio managers who prioritise liquidity and investability. This also minimises, as far as is reasonably practicable, the documented performance that can be attributed to market frictions, while also eliminating survivorship bias. Since equities within major indices attract a large number of institutional investors (Cao, Han, and Wang, 2017), and

[3]This is not to say custom news processing techniques are without merit. The point is that it contributes significantly to the difficulty in making cross-study comparisons.

institutions tend to trade extremely early in the news cycle (Tetlock, 2011; Hendershott, Livdan, and Schuerhoff, 2015), testing within an index is likely to be a more difficult test for news-based strategies.

Following from this, a major difficulty in drawing practical conclusions from the current literature is that portfolios based on different features of news, and different combinations of news and momentum, are documented over significantly varied empirical settings. The first contribution of this thesis is that it documents the performance of a number of literature-motivated news and momentum portfolios in the consistent experimental setting described above. More specifically, we examine the performance of investment strategies employing nonparametric conditioning on news history and stock price momentum using single, double, and triple-sorted decile portfolios. The results of this analysis offer little evidential support for the utility of news analytics in momentum-style or momentum-enhanced portfolios. Plausibly ex-ante-identifiable strategies, such as those motivated by the literature, failed to generate risk-adjusted excess returns, even after controlling for the GFC.

The second contribution of this thesis is in its extension of the empirical approaches typically employed in the literature to construct news-informed portfolios at the horizons of interest. We move away from the monotonic, sequential factor-sorts employed previously by conditioning on the output of supervised statistical learning models trained over a designated subset of data. Starting with a "classical" OLS-based approach to model-informed portfolio formation before utilising a wider class of algorithms such as gradient-boosted-trees and neural networks, we test the predictive utility of news information in an environment largely uninhibited by prior distributional assumptions. We find that the combined use of news-derived features and flexible statistical learning algorithms offers only a modest increase in theoretical performance beyond a traditional momentum implementation. Measures of variable importance suggest that news is secondary to size, analyst following, and momentum in relevance for predicting future return.

The third contribution of this thesis is an investigation of the predictive capacity of news sentiment at the firm level. As previously stated, firm-level predictability is one pathway to the construction news-informed portfolios that has largely been overlooked by the low-frequency news analytics literature. We test the longitudinal predictability of forward returns through a series of firm-level regressions, including single period, multi-period, and VAR specifications. We find no supporting evidence that either news content or news volume are useful predictors of future return at the firm-level. Rates of statistical significance for predictive model coefficients were comparable to those of random covariates, and the majority of effect sizes were economically irrelevant.

The fourth contribution of this thesis is a literature survey that condenses some of the most important elements of the news analytics literature into one place. Financial news analytics is ultimately a multi-disciplinary field, and we argue that theoretical motivation, natural language processing techniques, approaches to variable construction, and econometric techniques are distinct and separable components of a given news-based analysis. Further, we consider an understanding of each of these topics to be of fundamental importance in drawing conclusions from the literature. The survey contained herein is an attempt to provide an informative summary of each of these topics while also presenting the current state of the literature through a bibliographic analysis.

Overall, we find that while the use of news analytics in finance has sound motivations, its deployment in low-frequency equity portfolios does not appear to offer an economically significant edge—at least not within a highly-investable stock index such as the S&P 500. The evidence in support of this conclusion is of course subject to the limitations and constraints of this work. Greater detail regarding these limitations, as well as suggestions for future research in the area, are provided in Chapter 7.

Chapter 2

Literature Review

2.1 Structure of Literature Review

The first two sections of this literature review can be considered background information to the application of textual analysis in finance. The first section, 2.2, attempts to establish why we may credibly consider public news content to be of empirical interest in markets ex ante.

The second section, 2.3, is intended to provide the reader with some bearing of what is commonly referred to as sentiment in the relevant studies. It is also intended to communicate the large variability in measures of sentiment used between studies and provide an understanding of what type of information they may or may not be capturing and why this might influence findings. The overview of textual analysis techniques is provided upfront so as to allow subsequent discussion of relevant works to focus on experimental design and key findings rather than detailed explanations of the associated measures of sentiment.

Section 2.4 focuses on the different approaches taken to econometric testing of sentiment within the literature. This section is not concerned with the findings of the studies, but how the influence of sentiment and other news-based information has been measured and applied empirically. These topics are important when considering the state of the literature as it relates to the current work, since different variable constructions, econometric techniques and back-testing approaches have varying degrees of applicability to the economic relevance of news sentiment to fund management.

Section 2.5 covers the process used to identify the core collection of publications selected for full-text review. The corpus of review items is a subset of the wider finance literature with particular emphasis on the themes relevant to the current work. In addition to discussing the sourcing and selection of review items, section 2.5 includes a brief bibliographic analysis of news sentiment literature and a discussion of the descriptive statistics of the review corpus.

The subsequent section, 2.6, is concerned with the findings of the relevant studies themselves. As the textual sentiment literature appears to be free from any coherent chronological narrative to guide a discussion of the relevant findings, the studies herein are first grouped according to the time horizons over which the influence of textual sentiment is examined, and then grouped according to whether their findings relate most strongly to expected returns or risk. These partitions were chosen as they naturally align with my focus on the practical use of textual sentiment at mid- to long-term horizons while also correlating with broad groupings in methodological approaches. This section concludes the main body of the literature review.

2.2 Theoretical Background

While this thesis is not motivated by, or intentionally implicated with, any *particular* model or theory governing the role of news media on prices, the theoretical background to be discussed is nevertheless an important and enriching feature of the relevant literature.

Theoretical models provide a basis for understanding the potential causal factors underpinning the themes and findings of the news media analytics literature, and offer researchers a credible starting point for constraining the empirical specifications of their analyses in the face of expansive contemporary data sets. Tetlock (2007), Li (2010a), and Loughran and McDonald (2016) each cite the importance of economic theory in understanding and progressing text-driven research in finance.

Every trade requires willing participants on either side. Speculative markets with only rational and informationally-informed trading result in conditions of adverse selection in which participants are no longer incentivised to trade, and so the market ceases to function (Akerlof, 1970; Milgrom and Stokey, 1982). Noise trading (non-informational trading) supplies the liquidity required for informed trading to exist and thereby carries genuine information into prices, while simultaneously ensuring price is only a noisy reflection of value. This partially revealing property of price is required to incentivise the collection of information in the first place (Grossman and Stiglitz, 1976). Thus, information asymmetry and non-informational trading are inextricable properties of financial markets.

As classical asset-pricing models derive demand functions independently of the distribution of equilibrium prices, they have little to say about these features of markets or how they may interact in response to public information arrival. Alternatively, noisy rational expectations (RE) models (such as the canonical Kyle, 1985 model) have provided a rich theoretical framework in which the mechanics of financial information theory are described.

Admati and Pfleiderer (1988) found that periods of concentrated trading arise as discretionary liquidity traders elect to trade at the same time as each other to protect against privately informed traders. The structure of private information in their model permits the interpretation that privately informed traders are able to process public information faster and more efficiently than others and hence identify profitable trading opportunities.

Foster and Viswanathan (1990) examined an interday model in which one informed trader holds a monopolistic information advantage over several liquidity traders. They found that the informed trader will trade on their information over multiple periods, at an intensity that depends on the quality of public information. With no public information, the informed will trade in such a way that the market maker's price response to new orders is the same and that prices are equally informative each day. With an informative daily public signal, the informed trader's informational advantage depreciates each day and the information released through trading decays over subsequent periods.

Holden and Subrahmanyam (1992) developed a multi-period auction model in which multiple informed traders holding identical information (and being rational agents, identically interpreted) trade aggressively against each other, resulting in the information being incorporated into prices almost immediately. Their result approaches a strong form market efficiency for the common information signal as the interval between auctions approaches zero.

Foster and Viswanathan (1993a) examine variations of the Kyle (1985) model with non normal belief distributions and find that Holden and Subrahmanyam's (1992) result only applies if the informed hold identical information. If the information is only correlated (through non-normal beliefs), the original Kyle (1985) result holds but with conditional heteroskedasticity in price changes and autocorrelation in volume. They further show that when the assumption of normally distributed beliefs is discarded, price variance and trading volume are functions of public information and depend on the degree of unexpectedness of the information.

Although the nature of information asymmetry in RE models is such that they are focused toward the strategic actions of those with a strong information advantage in the face of public news arrival and market depth, they have also been used to more closely examine the actions of other subsets of rational investors. For example, Wang (1994) and Tetlock (2010) examined models in which one group of investors has a private informational advantage and incurs privately observed liquidity shocks (combining the traditional role of the informed and noise traders) and one group that only observes price and public information signals. They showed that the role of public information in resolving information asymmetry induces the uninformed investors to accommodate the liquidity shocks from the informed and generates abnormal trading. He and Wang (1995) draw similar insights using a model in which private information takes the form of differential information—each investor holds some information that other investors do not, but there is no strategic information advantage.

Collectively, these models suggest that public information will generally act to resolve information asymmetry, leading to a convergence of beliefs (Foster and Viswanathan, 1990; Tetlock, 2010; He and Wang, 1995). Commonly observed signals will be rapidly incorporated into prices (Holden and Subrahmanyam, 1992) but those with superior information, or information processing capability can identify profitable trading opportunities (Admati and Pfleiderer, 1988; Foster and Viswanathan, 1993a). Traders with a long-lived private informational advantage will trade on their information over multiple periods and at an intensity proportional to the quality of public signals (Foster and Viswanathan, 1993b). These findings provide a rational basis for considering news-driven trading reactions at various horizons (see for example, Engle, Hansen, and Lunde, 2012).

A traditional view of asset pricing holds that the actions of noise traders (delegated the role of liquidity shocks in the above models) can have no meaningful impact on prices, since they will be traded against aggressively by rational arbitrageurs who will drive prices close to fundamentals in the process. Black (1986) however, argued that noise-induced deviations from fundamentals may persist through time due to the capital constraints and risk aversion of rational investors. DeLong et al. (1990) and Campbell and Kyle (1993), in the spirit of Black (1986), developed equilibriums in which random innovations in noise trader bias (i.e. sentiment) can have a significant bearing on asset return and trading volumes. As demonstrated by Tetlock (2007), such theories of investor sentiment lead to testable predictions if news content is taken to reflect investor sentiment. The theoretical link between noise trader behaviour and public information is made explicit by difference of opinion models.

While RE models are guided by the principle that rational agents will interpret the same information signal identically, investors in difference of opinion (DO) models have heterogeneous priors and interpret information differentially, allowing them to "agree to disagree" (Kandel and Pearson, 1995; Banerjee and Kremer, 2010). In this light, noise trading need not remain a purely exogenous phenomenon unrelated to valid information but may be provoked in certain ways due to the misinterpretation of valid information. This notion is invoked heavily within the news analytics literature, such that measures of 'textual sentiment' are held not only as an empirical proxy for investor sentiment, but as an influencer of it.

Behavioural finance models motivated by specific cognitive biases and bounded rationalities have sought to explain empirical anomalies such as stock price momentum (Jegadeesh and Titman, 1993), event-based return predictability, and long-term reversals (De Bondt and Thaler, 1985) through investors' interaction with different types of public information. For example, the theories of Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998), and Hong and Stein (1999) have featured prominently in the low frequency news analytics literature. These models are relevant to the sequential release of information, the gradual diffusion of information across different investor groups, and the joint effects of endogenous and exogenous information. Additionally, as they characterise different types of stock market under-, and over-reaction, there are implied cross-sectional implications for the way firms' attention profiles and media exposure may affect market expectations, that differ between models.

An alternative perspective sees the role of investor attention and media exposure on stock valuations more directly. Merton (1987) developed a capital market equilibrium where each investor knows only about a subset of the available securities, and invests efficiently in these securities. Market-clearing requires that those invested in firms with small investor bases take large undiversified positions in them. Low-recognition firms must therefore offer their investors higher returns to compensate them for the increased idiosyncratic risk. Assuming media attention can influence investor recognition, Merton's (1987) findings suggest that all else being equal, firms with higher news exposure will generate lower returns.

Easley and O'Hara (2004) showed that the cost of capital increases with the degree of firm information that remains private. Like Merton (1987), their model suggests firms with greater news coverage will provide lower returns, but only insofar as the news conveys genuine information. Behavioural models such as those of Hirshleifer and Teoh (2003) and Peng and Xiong (2006), which emphasise the role of investors limited attention toward different types of news content, also imply cross-sectional effects related to firms' information environments and news exposure.

An often-cited theoretical basis for the influence of news on asset variance is the mixture-of-distributions hypothesis (MDH), which posits a direct link between public information arrival, trading volume and variance through the price evolution process (Clark, 1973; Epps and Epps, 1976; Tauchen and Pitts, 1983). According to the MDH, the period price change and trading volume are mixtures of independent distributions in which the rate of information arrival is the mixing variable. It further predicts that the trading volume and associated price variance depend on the extent to which traders disagree in response to the new information. As discussed by Engle, Hansen, and Lunde (2012), MDH is a general hypothesis capable of subsuming many of the more detailed hypotheses relating information arrival and transmission characteristics to the behaviour of return volatility and trading volume.

2.3 Textual Sentiment Techniques in Finance

Textual analysis distinguishes itself from much of the quantitative research conducted in finance due to the crucial process of transforming a collection of characters to a numeric representation of some aspect of information conveyed by those characters. The nature and complexity of this transformation is necessarily dependent on the type of information one intends to extract from the text. However, even for relatively modest goals, there are numerous methodological degrees of freedom that must be fixed before a quantitative representation of text can be obtained.

Seemingly arbitrary choices from the text source, the earliest stages of document parsing and pre-processing, through to the final representation of the text can have significant impacts on both the accuracy and precision of the output measure. Loughran and McDonald (2016) pay particular attention to this aspect of textual analysis and provide a number of finance-specific examples. Manning, Raghavan, and Schütze (2008, pp. 22-33) provide a detailed discussion of some early document parsing decisions and their consequences within the broader context of information retrieval.

Moreover, many sources of imprecision are sensitive to firm- and industry-specific features of language and reporting, resulting in measures which are not only noisy, but correlated with idiosyncratic and cross-sectional characteristics. For example, words associated with death and disease are recognised as negative by a number of word lists used in language processing, but are pervasive in reportage of pharmaceutical companies. Similarly, company or product names themselves may be inadvertently mapped to certain categories by the processing algorithm.

The sensitivity to process embodied by textual analysis in finance exists alongside the absence of any clear discipline-specific norms or practices; likely a symptom of the field's relative infancy. Consequently, the literature is populated by a diverse spectrum of methodology choices, the results of which are often not easily compared or generalised. These factors do not undermine the importance or veracity of the field, nor are they the focus of the current work. Rather, they are simply a feature of the literature that must be considered; the conclusions drawn from studies using text-based measures must be interpreted in light of the data and processes used to develop those measures. This is particularly true given the myriad of textual analysis techniques employed throughout the financial literature for which the output falls under the umbrella of *sentiment*.

Sentiment analysis or *opinion mining* refers broadly to the process of quantifying opinions, sentiments, emotions, appraisals and attitudes toward an entity or topic as expressed in text (Serrano-Guerrero et al., 2015). Although the terms sentiment analysis and opinion mining tend to be used interchangeably in the wider information sciences[1], this has not been true within accounting and finance, where the two terms present different and important connotations. For this reason it is common for authors to avoid casual use of either of these terms entirely, electing

[1]Sentiment analysis is widely characterised in the context of opinion-dominated applications such as product reviews and customer satisfaction. See for example Miner (2012), ch.4, Pang, Lee, et al. (2008), Liu and Zhang (2012) and Medhat, Hassan, and Korashy (2014).

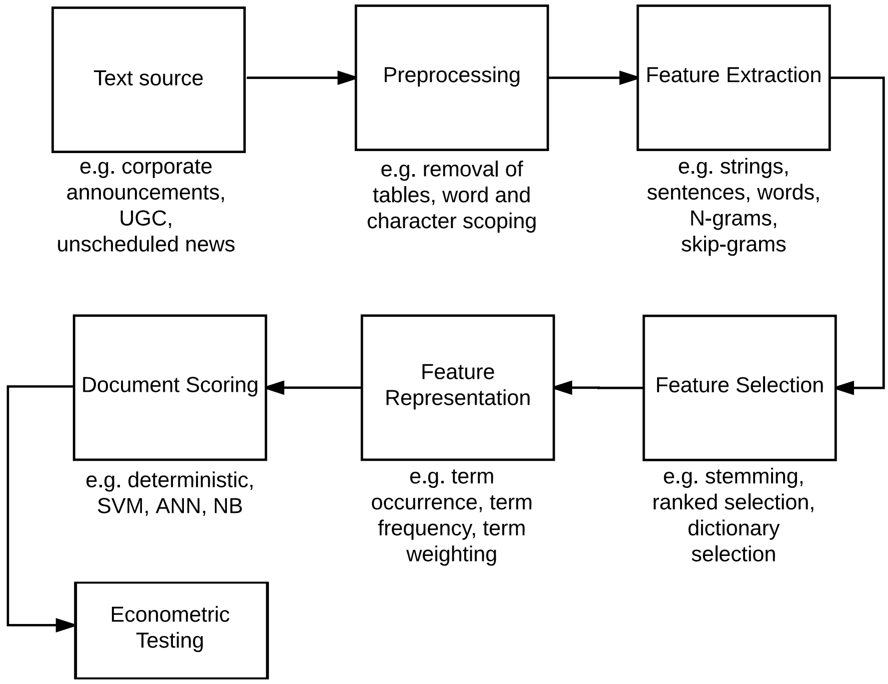

Figure 2.1: Textual analysis workflow

instead for less implicative surrogates such as *media content*, *tone*, *qualitative information* and *linguistic content*. Regardless, *sentiment analysis* remains pervasive and generally accepted terminology in the literature whilst still differentiating between other forms of content analysis such as document readability and topic extraction, and is therefore used broadly from this point forward.

The order of the topics covered in this section follow the order of the textual processing steps required in transforming unstructured text data into numerical values that can be used for econometric analysis and hypothesis testing, as summarised in Figure 2.1.

2.3.1 Sources of Textual Information

Sentiment analysis is applied in the accounting and finance literature not only to reflect the opinions and emotions of the author but also to capture a more objective indication of conditions within firms, institutions and markets (Kearney and Liu, 2014). In this sense, the type of information intended to be captured is in large part reflected by the source of the textual data used for the analysis, with most finance research focusing on one of three main sources: corporate disclosures and filings, unscheduled news and user-generated content (UGC).

These sources vary not only in the information content of the text but also in their frequency, timeliness, breadth, depth, and their perception by market participants.

Corporate disclosures

As a communication vehicle for management to the public, corporate disclosures have been a natural and important source of textual data for researchers, offering a setting in which to understand managements incentives and attain information about the data generating function of long-studied quantitative data (Li, 2010a). The language used by managers to describe their firm's activities and prospects has been shown to be associated with market response (Henry, 2008; Feldman et al., 2010; Kothari, Li, and Short, 2009), future firm performance (Davis, Piger, and Sedor, 2012; Li, 2010b; Loughran and McDonald, 2015) and management fraud (Purda and Skillicorn, 2015).

This area of the literature has predominantly made use of firms' annual and interim reports, earnings releases, and conference call transcripts, with a number of studies focusing on specific sections of these sources such as the Management Discussion & Analysis (MD&A) section of 10-Ks and 10-Qs. These text sources are plausibly rich in information relevant to firm fundamentals and valuation, but as discussed by Kearney and Liu (2014), their quarterly or annual release schedules result in a low-frequency time-series that is not ideal for many types of analyses. Additionally, as corporate disclosures are inherently firm specific and relatively fixed in scope, they do not readily capture broader market conditions or outside perspectives of the same firms. Kothari, Li, and Short (2009) provided evidence that investors do not place much weight on favourable disclosures from management and suggested that investors find alternative sources more credible or timely, but found counter evidence for small firms which tend to have comparatively scarce information environments.

Unscheduled news

Unlike corporate disclosures, media content (or unscheduled news) is produced by a large number of sources, covers a diverse range of entities and events and is released at a relatively high frequency. For these reasons, media content offers a highly flexible medium for sentiment analysis, allowing researchers to study the impacts of sentiment aggregated over varying scales of time and entity. For example, researchers have studied the influence of news at the firm level (Tetlock, Saar-Tsechansky, and Macskassy, 2008; Ferguson et al., 2015), industry level (Smales, 2015[b]; Li, Xie, et al., 2014), market level (Tetlock, 2007; Wei et al., 2017), commodity level (Clements and Todorova, 2016; Smales, 2015[a]), between currencies (Nassirtoussi, Aghabozorgi, Teh, et al., 2015), across countries (Griffin, Hirschey, and Kelly, 2011) and at the global asset-class level (Uhl, Pedersen, and Malitius, 2015). Time horizons vary from intraday (Groß-Klußmann and Hautsch, 2011; Ho, Shi, and Zhang, 2013), daily (Tetlock, 2011; Garcia, 2013), weekly (Sinha, 2016; Lu and Wei, 2013), monthly (Ammann, Frey, and Verhofen, 2014; Cahan et al., 2017) and longer (Hillert, Jacobs, and Mueller, 2014).

The range and depth of unscheduled news allows for versatile channels of examination, but it is also less likely to reflect the same density of value-relevant information captured in corporate disclosures, and often contains information already revealed by the market price. Hendershott, Livdan, and Schuerhoff (2015) showed that institutional order flow predicts the occurrence and sentiment of news announcements and accounts for a significant degree of the price discovery related to news stories. However, not all sources of unscheduled news are absorbed with equal efficiency. The literature tends to differentiate between subcategories of unscheduled news with respect to the information content they offer, with equivalent interpretations of the sentiment derived from these sources.

Tetlock, Saar-Tsechansky, and Macskassy (2008) and Ahmad et al. (2016) found that intraday newswires appear to provide more timely and value-relevant information than stories appearing in lower frequency sources such as newspapers, while Ferguson et al. (2015) found that differences in firm coverage between newspapers resulted in measurable differences in market response. Other sources of financial reporting such as *post mortem* market commentary are considered to provide entertainment more so than genuine news (Tetlock, 2007) and appeal to researchers seeking to investigate the impact of news content in a setting where value-relevant information is less likely to drive results. As Garcia (2013) states, the type of market commentary studied by Garcia (2013) and Tetlock (2007) is almost an *opinion piece* with a primarily *non-informational* impact.

Other studies have focused specifically on this topic and have attempted to completely control for the informational and market-endogenous influence of media sentiment in order to isolate the *causal* effect of linguistic content itself. Dougal et al. (2012) used exogenous scheduling of *Wall Street Journal* columnists as a proxy for their individual writing styles to establish a causal relation between financial reporting and stock performance. Engelberg and Parsons (2011) approached the same problem by comparing the reaction of investors in mutually exclusive trading regions to the same information event but with access to different media outlets. They found that for the same information events, local media coverage strongly predicted local trading and that the timing of local trading was strongly related to the timing of local reporting.

The informational content of sentiment derived from different unscheduled news sources is also important from the perspective of portfolio management, insofar as it affects the ability to explain historical returns and forecast future market behaviour. Tetlock (2011) showed that stock prices respond less to stale news and that the price response which does occur is shortly reversed. Similarly, Leinweber and Sisk (2011) showed that removing redundant or stale news increases the performance of news-based signals in explaining future returns. The unscheduled news sentiment literature is discussed in Section 2.6.

User Generated Content

A relatively recent form of textual information to appear in the finance literature, coinciding with the wide-spread use of the internet, is the content generated by users of web-based message boards, blogs and social media. Whereas corporate disclosures (nominally) reflect the perspectives of inside management, and unscheduled news contains the information and insights of market professionals and financial analysts, user-generated content has often been interpreted as the "mood" (Bollen, Mao, and Zeng, 2011; Li, Wang, et al., 2014) of the retail investor. Web-based forums provide an avenue for individual investors to collaboratively interpret financial press, discuss the fortune of companies, and trade investing ideas in a largely unregulated environment.

The evidence suggests such platforms play an important role in information cascades, rapidly disseminating public information (Antweiler and Frank, 2004) and facilitating interpretation of media releases (Das, Martinez-Jerez, and Tufano, 2005). The first studies to investigate the information content of UGC made use of readily-quantifiable aspects such as message volume (Wysocki, 1999) and user-prescribed ratings of stocks (Tumarkin and Whitelaw, 2001; Dewally, 2003), it was not until the seminal work of Antweiler and Frank (2004) that measures of textual sentiment of UGC were studied. Since then, there has been a plethora of research aimed at extracting the "wisdom of the crowds", and the topic has received much attention from academia and industry alike.

The popularity of social media and stock forums has made UGC a high-frequency and flexible source of data, allowing it to be aggregated in time and over entities in many of the same ways as was discussed with unscheduled news. For example, UGC has been studied at the individual firm level (Das, Martinez-Jerez, and Tufano, 2005; Sabherwal, Sarkar, and Zhang, 2011; Leung and Ton, 2015), sector level (Das and Chen, 2007; Nooijen and Broda, 2016; Souza et al., 2016), market level (Ranco et al., 2015; Nofer and Hinz, 2015; Pineiro-Chousa, Vizcaino-Gonzalez, and Maria Perez-Pico, 2017), and across countries (Garcia, 2016; Siganos, Vagenas-Nanos, and Verwijmeren, 2014).

A common finding throughout the literature is the existence of significant contemporaneous relationships between UGC-derived measures and stock price behaviour, and short-term leading relationships between UGC and volatility and trading volume (see for example Antweiler and Frank, 2004; Das and Chen, 2007; Sabherwal, Sarkar, and Zhang, 2011; Zhang, Swanson, and Prombutr, 2012; Kim and Kim, 2014; Sprenger et al., 2014; Karagozoglu and Fabozzi, 2017). However, findings regarding the extent to which measures of UGC lead returns have been far less consistent. Antweiler and Frank (2004), Das, Martinez-Jerez, and Tufano (2005), Das and Chen (2007), Kim and Kim (2014), Sprenger et al. (2014), and Gabrovšek et al. (2017) find little or no evidence that UGC leads returns, more commonly finding that the returns drive UGC instead.

Some researchers have found that UGC is indicative of future returns in only a narrow range of ex post identifiable assets (Zheludev, Smith, and Aste, 2014), only during certain high-attention events (Ranco et al., 2015), or only among small, illiquid firms with high retail investor ownership (Leung and Ton, 2015). Other studies (Bollen, Mao, and Zeng, 2011; Zhang, Swanson, and Prombutr, 2012; Nguyen, Shirai, and Velcin, 2015; Sun, Lachanski, and Fabozzi, 2016) have collectively found that UGC can be used to explain future return and enhance forecasting accuracy beyond naive or price-only estimates. Some of the research in this area has also made use of the social structure within UGC platforms, where the opinions and expectations of certain individuals exert a greater amount of influence than others (Cha et al., 2010).

Sprenger et al. (2014) found that users of the micro-blogging platform *Twitter* who provide above-average investment advice are paid more attention by other users in terms of re-tweets, mentions and followship. Sabherwal, Sarkar, and Zhang (2011) found that measures of sentiment that were weighted by users' reputation credit were more predictive of subsequent market movements than equally-weighted sentiment measures. Similarly, Nofer and Hinz (2015) found that the sentiment of Twitter posts weighted by number of the followers of the user was significantly related to next-day market returns while a simple aggregation of sentiment was not. Yang, Mo, and Liu (2015) identified financial investment communities on Twitter and found that a weighted sentiment measure which used posts from these social nodes was a stronger predictor of market behaviour than general social sentiment measures.

2.3.2 Document Representation

Once the source of textual information has been selected and the data is available, the unstructured text must be transformed into a machine-readable representation that adequately retains the desired information to be extracted from the document. The collection of steps involved in this transformation may be referred to as *feature processing*. The different approaches to feature processing used in the finance literature can be discussed in terms of three major aspects: document representation, feature selection, and feature representation.

Document representation, or feature extraction, refers to the early parsing procedure in which documents are broken down into the constituents that will be used in subsequent processing steps to distill meaning from the data. This is a fundamental step in the sentiment analysis pipeline as it essentially determines the type and amount of the original information available for subsequent analysis. Relatively simple forms of feature extraction typically involve the mapping of a text d_j into a vector of term weights $\vec{d_j} = \langle w_{1j}, \ldots, w_{|\tau|j} \rangle$, where τ is the set of *terms* (or features) that occur at least once in the document corpus and w_{kj} represents the naive weighting of the term t_k (Sebastiani, 2002). While more complex forms of feature extraction map the text to a tree or network of grammatical structures.

One of the simplest document representations used in the finance and accounting literature is the verbatim occurrence of targeted phrases, since such a representation does not necessarily require analysis beyond the character level. For a small number of phrases or small portions of text, it is efficient enough to simply pattern-match the phrases with very simple regex code. Li (2006) investigated the relationship between the frequency of the words *risk*, *risks*, *risky*, *uncertainty*, *uncertain*, and *uncertainties* in the MD&A section of 10-K filings and firms' subsequent stock price and earnings performance. Similarly, Kravet and Muslu (2013) counted the occurrence of 18 risk related words and their variations in 10-K filings and examined the firms' associated market behaviour and analyst forecast dispersion. Loughran, McDonald, and Yun (2009) counted the frequency of the word *ethic* and its variants and the phrases *corporate responsibility*, *social responsibility* and *socially responsible* in 10-K filings and tested for association with corporate governance measures and lawsuits.

The verbatim matching of words and phrases in running text, however, is highly inflexible and quickly becomes inefficient and difficult to parse outside of a narrow range of problems. Most approaches to textual sentiment therefore require analysis beyond the character level, and thus the running text must first be segmented into meaningful units (or tokens), with words being the common choice of unit. This process of identifying words (or equivalently, their boundaries) is referred to as *tokenisation* or *word segmentation*. In English, a natural choice of word boundary is the whitespace, however this is generally unreliable, for example; words are often attached to punctuation marks which may or may not (as in the case of abbreviations) denote the end of the word, apostrophes can be used to contract to words or alternatively imply ownership, hyphenated expressions may be better understood as single or multiple words, and white spaces may occur within a single word - such as *hot dog* or *a priori*.

During the tokenisation process it is also common to expand clitic contractions (Jurafsky and Martin, 2016) such as *she'll* to the two words *she* and *will*. Simple tokenisation methods include whitespace-based techniques which ignore the type of issues just described or vocabulary-based methods with simple heuristics for unknown words. State-of-the art approaches make use of regular expressions implemented as finite state automaton (Miangah, 2014), or machine learning sequence models (Tomanek, Wermter, and Hahn, 2007). The accuracy required by the tokenisation algorithm is largely dependent on the subsequent choices for feature selection and dimensionality reduction, as certain choices may render essentially all of the complicating tokenisation cases irrelevant.

For many approaches to sentiment analysis, word segmentation is the only form of feature extraction required, and document representation remains at the word level. This is true of the most common method of document representation used in textual analysis - the bag of words (BOW) approach, in which the document corpus is collapsed into a *term-document matrix* of size $n \times d$, where n is the number of documents, d is the size of the vocabulary (or *lexicon*), and the (i, j)th entry is the frequency of the jth word in the lexicon of document i. The critical assumption underlying the BOW approach is term independence, that is, the order of the words as they appear in the document is ignored.

Understandably, the loss of sequencing information removes a large amount of information from the text and limits what kind of analysis can be performed. However, BOW is far simpler from an algorithmic perspective, is more computationally efficient, and has superior statistical qualities (Lewis, 1992; Hagenau, Liebmann, and Neumann, 2013) compared to more complex document representations. Word-based representations are also highly generalisable, robust, and can be combined with more sophisticated representations if required.

A common extension to the BOW representation uses continuous sequences of n words, known as N-grams, as terms in the term-document matrix rather than individual words ('bag of N-grams'). N-grams can serve as valuable feature sets since certain words derive much of their meaning from their collocation with other words and often the surrounding one or two words can provide sentiment-changing context. N-grams have been used in the finance literature by Tetlock (2011), Ranco et al. (2015), Hagenau, Liebmann, and Neumann (2013), Yang, Mo, and Liu (2015), Bollen, Mao, and Zeng (2011), and Butler and Keselj (2009) among others.

A variation of the N-gram approach is often used which allows for a word distance greater than zero between words when defining features. Such N-word combinations are sometimes called *skip-grams*, which specify a maximum of k skips over each set of n words. Word combinations can be used as endogenous statistical discriminators of document sentiment (e.g. Huang, Liao, et al. (2010); Hagenau, Liebmann, and Neumann (2013)), but they can also be useful in identifying targeted phrases without having to match the expression exactly. For instance, Wüthrich et al. (1998) used word combinations and stemming to identify matches for over four hundred key-word phrases, such as *property weak* and *interest rate cut* in online financial news articles. They provide the example of being able to identify the key-word record *stock drop* in a text that contains the phrase *stocks have really dropped*. As N-grams and word combinations retain some sequencing information, they offer a compromise between the generic BOW and higher-level string-based representations.

In moving beyond the word level, a natural requirement is the identification of sentences from running text, known as *sentence tokenisation* or *sentence segmentation*. Sentence segmentation is based largely on punctuation, such as the occurrence of the periods (.), (*?*) or (*!*), and approximately 90% of the time periods do mark the end of a sentence (Riley, 1989). The challenging part for a tokeniser is the remaining 10%, which includes ambiguities such as abbreviations, quotations, and other punctuation-related nuances. Many of these complexities are shared with the problem of word segmentation, and for this reason word segmentation often influences sentence segmentation (Stanford NLP Group, 2017) and the two processes tend to be handled together (Jurafsky and Martin, 2016, ch. 3).

As sentence segmentation can be defined as a sentence boundary binary classification problem, any rule-based or machine learning classification method can be applied to the task, with modern approaches leaning toward the latter. Riley (1989) used decision trees with inputs such as the case and length of the words either side of a period and the probability of the words to occur at the beginning or and of a sentence, Palmer and Hearst (1994) and Palmer and Hearst (1997) use a neural network and the surrounding distribution of part of speech tags, and Reynar and Ratnaparkhi (1997) and Mikheev (1998) employed maximum entropy models. While rule-based approaches require a significant hand coding and tend to be domain-specific (Manning and Schütze, 1999, pp. 135), machine learning models are highly flexible and can be trained on other languages – see for example Silla and Kaestner (2004).

Sentence segmentation is generally used in the textual sentiment literature as a precursor to more advanced feature selection techniques that leverage the sequencing information retained within the individual sentences. However, sentences have also been used as units by which to measure sentiment even when using simple word-based terms. Yu, Duan, and Cao (2013) measured the sentiment of company news articles and UGC using sentiment classification at the sentence level with BOW representation, Feuerriegel and Prendinger (2016) used individual words as features to measure the net-optimism of corporate announcements but inverted the polarity of words following negating terms within each sentence, sentences were also used for negation in this way by Das and Chen (2007). Kravet and Muslu (2013) measured the riskiness of 10-K filings by counting the number of sentences that contained one or more risk-related keywords.

As sentences represent structurally independent grammatical units capable of conveying isolated ideas, and as most grammars apply to sentences (Grefenstette and Tapanainen, 1994), they form the basis for more advanced natural language processing tasks that aim to abstract meaning from text using these grammars. Part-of-speech (POS) tagging is a fundamental step in this abstraction process, and by utilising sentence-level sequencing information, it represents the first major departure from the BOW approaches to feature extraction. Parts of speech are highly relevant to sentiment analysis as they provide a large amount of information about the meaning of a sentence, and certain parts of speech, such as adjectives, are particularly important indicators of opinions (Liu and Zhang, 2012). POS tagging aims to label each word with a tag indicating its appropriate part-of-speech (also known as word class or syntactic category), such as noun, verb, adverb in addition to more fine-grained tags like 'noun plural' or 'phrasal verb'. It is generally performed after, or as part of, the word and sentence tokenisation and tagging process.

An important feature of POS tagging that can lead to a much greater sentiment scoring performance than bag of words approaches is the identification of negation and intensifiers. Negation occurs when one part of a sentence flips the sentiment of the rest. Consider the following phrase: *I was not productive enough to have a good week but I'm staying positive*. A modern POS tagger will identify that the meaning of the word *good* has been reversed and that *but* has essentially split the phrase into to parts. If the words following *but* were also negative, it would be instead be recognised as a modifier. Intensifiers are words such as *very* or *extremely* that strengthen the meaning of a word. The inability of bag of words approaches to correctly identify negation may be one reason positive tone has appeared to offer little explanatory power in financial settings (Tetlock, 2007; Loughran and McDonald, 2011), since negation is used far more often in reference to positive words than negative words in financial text (Loughran and McDonald, 2016).

Although a reasonable tagging accuracy can be achieved based on lexical information (the individual word) alone[2], words are ambiguous and generally have more than one possible part-of-speech; for example, *heat* can be either a noun, as in *the fierce heat of the sun*, or a verb, as in *next, heat the oven*. This is true of 55-67% of word tokens in English running text (Jurafsky and Martin, 2009, ch. 10). POS tagging is therefore a disambiguation task (one of the many in language processing) and accurate tagging cannot rely on lexical information alone. For this reason, modern taggers in some way use both lexical information and syntagmatic (tag sequencing) information to appropriately determine tags (Manning and Schütze, 1999, ch. 10). Such taggers aim to find a stochastic optimal sequence of tags $Y_1^T = t_1, \ldots, t_T$, given a word sequence $X_1^T = x_1, \ldots, x_T$, that maximises $P(Y_1^n, X_1^n)$.

Common state of the art approaches include Hidden Markov Models—a generative sequence technique which maximises the likelihood of the observed words conditioned on tags (the hidden state), and Maximum Entropy Markov Models—a discriminative sequence technique which computes the posterior of each state conditioned on the previous state and observed word. Other statistical approaches to POS tagging include decision trees (Schmid, 1994), neural networks (Benello, Mackie, and Anderson, 1989), and memory-based learning (Daelemans et al., 1996). See Jurafsky and Martin (2016) and Manning and Schütze (1999) for a detailed overview of these POS tagging techniques. POS tagging has been used in the finance literature to extract features such as nouns (Li, Xie, et al., 2014), noun-adjective trigrams (Das, Martinez-Jerez, and Tufano, 2005; Das and Chen, 2007), noun-phrases (Schumaker and Chen, 2009; Hagenau, Liebmann, and Neumann, 2013) and proper-noun phrases (Schumaker, Zhang, et al., 2012), and has been used for noun-based topic detection (Nguyen, Shirai, and Velcin, 2015). POS-tagging is also used in high-level open source software such as Stanford CoreNLP and proprietary commercial systems such as Thomson Reuters News Analytics and RavenPack News Analytics which are common in the literature.

A named entity in a piece of text is a sequence of words that designates a real-world entity, such as a person, organisation or location. Named entity recognition (NER) refers to the combined task of identifying these sequences of words and classifying the entities by type. Named entities are generally anything that can be referred to with a proper name but other types of entity have also been defined

[2]For example, Charniak (1996) showed that a tagger which simply assigned the most common tag to each word achieved an accuracy of 90%.

for use in specific domains, such as biological species and substances in molecular biology (Ohta, Tateisi, and Kim, 2002). There is also general agreement about the inclusion of times and dates and numerical expressions such as monetary amounts and percentages (Nadeau and Sekine, 2007). As with the other language processing systems already discussed, NER systems face ambiguities that make manually defined rule-based approaches unideal for the task.

Most modern NER algorithms treat the task as a word-by-word sequence labeling problem (Jiang, 2012) and so are based on the same types of statistical learning approaches as used for POS tagging, such as Hidden Markov Models (Bikel et al., 1997), Maximum Entropy Markov Models (Bender, Och, and Ney, 2003), and Conditional Random Fields[3] (Finkel, Grenager, and Manning, 2005). Standard features used in modern NER systems include lexical items, parts of speech, surrounding bag of words/bag of N-grams, predictive tokens, occurrence on named entity lists, syntactic chunk labels and shape. Important features for financial news text include entities lists, predictive tokens, which include markers such as *Inc.* and *Corp.*, and shape features, which capture numbers (*3M*), punctuation (*Yahoo!*) and capitalisation (*eBay*). In the finance literature, NER has been used to add the occurrence of entity categories (date, location, money, time etc) as an additional feature set in itself for subsequent classification (Schumaker and Chen, 2009), to keep track of entities for sentiment attribution within articles (Sinha, 2016), and for event identification and information extraction (Boudoukh, Feldman, et al., 2013).

2.3.3 Feature Selection and Dimensionality Reduction

Feature selection refers to the process used to determine the features which retain the most relevant information for the classification task and remove redundant data. It is of particular importance in textual analysis applications due to the high dimensionality and noisiness of word-based representations. Appropriate feature selection is an effective means to enhance computational efficiency and classification accuracy of textual analysis algorithms (Forman, 2003).

Two of the most common forms of feature selection, which are generally used in addition other techniques, are stop word removal and morphological stemming,

[3]CRFs differ from HMMs and MEMMs in that they are undirected graphical models (as opposed to directed graphical models), and labels of current observations can depend on future labels.

or lemmatization. Stop word removal simply removes words such as 'a', 'the' and 'of', which are extremely common but unlikely to provide specific or discriminatory information. This is generally performed using a *stop list*, a predefined list of stop words. Stemming and lemmatization aim to reduce inflectional and derivative forms related to a common base form. Stemming essentially amounts to the truncation of words to remove affixes while lemmatization is a more nuanced effort to correctly identify the base lemma or lexeme in the presence of an inflected form through the use of a vocabulary and morphological analysis. Two common algorithms for stemming English are the Porter stemmer (Porter, 1980) and the older Lovins stemmer (Lovins, 1968).

Dictionary-based

Beyond these standard means of dimensionality reduction, an important and prevalent approach in the finance literature has been the use of dictionary-based feature extraction. This approach typically filters features based on whether they occur within predefined lexicons or word lists which have been manually identified by domain experts to capture particular elements of the text, such as optimism, pessimism and uncertainty. The resulting feature space is a term-document matrix containing the incidence or frequency of the word list items only.

The most frequently used word lists in the accounting and finance literature are the Henry (2008) word list, the General Inquirer (GI) Harvard IV-4 (H-IV4) word lists (Stone, Dunphy, and Smith, 1966), the Diction (Hart, 2000) word lists and the Loughran and McDonald (2011) (LM) word lists. Of these, Diction and H-IV4 were the first word lists available and were used frequently within the earlier literature; particularly H-IV4 after its use in Tetlock's (2007) seminal paper. Diction and H-IV4 are general linguistic word lists premised on sociology and psychology and were not developed for use within any one specific domain. As such, the suitability of their use within accounting and finance has been questioned by a number of researchers. This is because words that are generally deemed to have positive or negative connotations (for example) may have entirely different associations within the context of accounting and finance (Henry and Leone, 2016).

For instance, Loughran and McDonald (2011) found that almost 75% of the negative words in H-IV4 do not have a negative meaning in the context of financial documents, they also note that a number of the H-IV4 terms are likely to proxy for specific industries. Similarly, Loughran and McDonald (2015) report that a large proportion of the terms within the Diction word lists appear to be misclassified when considered in the context of finance. On the other hand, the Henry (2008) and LM word lists were designed specifically for finance based on the content of earnings press releases and 10-Ks respectively. These word lists have been found to be more effective at capturing the targeted attributes of financial texts (Henry and Leone, 2016; Loughran and McDonald, 2011; Price et al., 2012). The Loughran and McDonald (2011) word lists are much more comprehensive than the Henry (2008) word list (85 versus 2,329 negative words, for example) and appear to be the most heavily used word lists in the finance and accounting literature since their publication. A more detailed discussion of these word lists and their usage in accounting and finance is provided by Loughran and McDonald (2016) and Kearney and Liu (2014).

Minimum Occurrence and Rank-Based Selection

A an alternative to using a predefined source such as a dictionary, it is common to perform feature selection based heavily on information occurring within the message corpus. A simple method used by some researchers to avoid unnecessarily large feature spaces is to define a minimum rate of occurrence of each term within the corpus, and remove those terms which occur less-frequently than the cut-off rate. For example, Schumaker and Chen (2009) and Schumaker, Zhang, et al. (2012), following Joachims (1998) remove terms occurring less than three times in the corpus. Alternatively, the rate of occurrence may be used to reduce the feature set to a specified size by including only the n most common features, as in Butler and Keselj (2009).

The goal of feature selection is to reduce the size of the feature space whilst preserving the most information. The intuition behind using term occurrence is that the more often a term appears within a document, the more likely that term is to capture the content of the document. However, if a term appears frequently across all documents it provides little information about the content of the document relative to the others, i.e. it is semantically unfocused. This motivates the use of inverse document frequency as a second measure of informativeness and together with term-frequency these measures inform the common term-frequency/inverse-document-frequency (TF-IDF) weighting scheme. The TF-IDF weighting scheme rewards occurrence within individual documents but penalises occurrence across documents.

A similar principle is applied to statistical weighting schemes that specifically aim to measure discriminatory power across document classes. Given a labelled set of training documents, the occurrence of each term within each of the predefined classes is used to form sample estimates of the relative probability of terms appearing each class. Analogous estimates for the probability of each term to occur in a given document, the relative class frequencies, and the probability of a given document belonging to a specific class conditioned on the occurrence of each term are also formed. These probabilities can be used to measure the informativeness of each term for the document classification task and typically appear through the use of one of five measures: information gain or entropy, Fisher's discriminant (Fisher, 1936), the Gini index, mutual information, and the Chi-square statistic. A brief definition and overview of these measures is provided by Aggarwal and Zhai (2012). These measures and TF-IDF are used for feature selection in the same manner as occurrence, that is, by retaining only the top n terms with the highest weights.

Note that unless the generation of labels for the training corpus were defined using an unsupervised clustering technique, the probabilistic weighting schemes mentioned above are essentially being trained by some market exogenous information (such as manual human labelling) or market endogenous information (such as direction of stock price movement surrounding article release). This is discussed further in Subsection 2.3.5.

2.3.4 Feature Transformation and Representation

After feature selection has been used to reduce the original set of attributes to a smaller, information-rich subsample, the remaining features need to be represented by some numeric value in order for them to be processed by the document scoring and classification procedure. This is known as *feature representation*. It is often the case that representation of the reduced set of features of each document is the same as the original set i.e. a vector (standardised by document length) of binary occurrence or frequency of each term within the document. This representation is almost always used when document classification is to be based on statistical learning procedures. For other document scoring and classification techniques it is common for the probabilistic term scores used for feature selection to be retained for feature representation. This may be performed by weighting the occurrence and frequency values in the term-document vector by the relevant score for each term. In the case of dictionary-based feature selection, each term score may also be weighted as positive or negative according to its category.

In some cases feature representation also involves creating a new (and smaller) set of features, i.e. *feature transformation*[4]. An example of such a technique is Latent Semantic Indexing (LSI)—also known as latent semantic analysis— which transforms the text space to a new axis system by applying singular value decomposition to the term-document matrix (i.e. principal component analysis for text). The resulting axes are a linear combination of the original terms and are oriented to capture the greatest amount of variation within the dataset. The resulting textual components can be interpreted as latent semantic features.

For instance, Tetlock (2007) applied LSI to Wall Street Journal columns to create his 'pessimism factor', which was taken as the column's length in the direction of the first principal component (as applied to the previous 12 months of columns). probabilistic LSI (PLSI), proposed by Hofmann (2001), extends LSI by assuming a probabilistic generative process for generating the document terms. Latent Dirichlet allocation (LDA) further extends the PLSI concept by using Dirichlet priors in a Bayesian framework to generate a topic model that best represents the data. For an example of the use of LDA in finance see Huang and Zhou (2017). Linear discriminant analysis is a variant of LSI that aims to capture the features which best discriminate between classes, rather then explain or generate the data. A common method uses Fisher's linear discriminant in an iterative process to identify the axis system which best separates classes. More information on these methods and other feature transformation techniques is provided by Crain et al. (2012) and Aggarwal and Zhai (2012).

2.3.5 Document Scoring and Classification

Once features have been extracted, selected, transformed and represented in a machine-readable format, document classification and scoring is conducted. The term 'document' here simply refers to the collection of terms used to create the score. For instance, rather than scoring the collection of words that occur within a

[4]Whether or not this is considered a feature selection or feature representation process varies from author to author.

single article and then combine the article scores over the formation period, some researchers combine all words from the formation period as a single article.

The are two general ways this has been performed in the literature and they can be separated by whether document scoring is performed analytically or from a machine learning algorithm. For analytical scoring, which is frequently used in conjunction with dictionary-based measures, a commonly used scores consist of one or a combination of: the sum of positively associated terms, the sum of negatively associated terms, and, if the feature representation was not normalised to document length, the total sum of terms. Each of these sums will have slightly different meanings depending on what feature representation was selected (for example, weighted or un-weighted term occurrence or frequency).

For statistical scoring, the same approach can be taken as with any machine learning classification procedure—at this point there is nothing special about the fact that the vector values are representations of some textual attribute. One of the oldest and most common classification technique used throughout the finance literature (and sentiment analysis in general) is the naïve Bayes (NB) classifier. NB classifiers estimate the posterior probability of a class based on the distribution of terms within the document, with the underlying assumption that terms occur independently of one another. The probabilities of each class, term and class-term co-occurrence used for the posterior calculation are informed by the training set.

Another extremely common classification technique is support vector machines (SVM), which appear to be more common in the recent literature than NB classifiers. SVM are maximal margin classifiers, and so they identify the feature surface from the training set that maximises the distance (in whatever the dimension of the feature space) between classes. Documents are then classified depending on what 'side' of the boundary they are positioned on. Other common classification techniques used throughout the literature include neural networks, decision trees, random forests and k-nearest neighbours (kNN). Reviews of these, and other classification methods, and their relative performance in the context of text mining are provided by Sebastiani (2002) and Ravi and Ravi (2015).

An important distinction in the finance literature regarding textual classification is whether the training corpus was labelled from exogenous or endogenous sources. Generally this is the difference between whether documents were labelled manually by a human or based on the market movement surrounding the documents publication (see Hagenau, Liebmann, and Neumann, 2013 for an example of the latter). The latter approach is usually applied such that the market movement used for labelling is the same as the market movement being predicted or explained in econometric testing, i.e. same asset and temporal horizon.

In fact, as discussed in Section 2.4, econometric testing and document classification are often one and the same so that the classification output is a price score (either nominal or continuous) rather than a sentiment score. An example of market-driven labelling it the approach used by Wüthrich et al. (1998), for their training set they labelled online news headlines published prior to market open as either up, down or steady, according to whether the market return for that trading day was above, below or between some threshold. The trained classifier was then used to predict the daily market return using out of sample news headlines published prior to market open.

Training on market movements appears to be much more common in the high-frequency and UGC literature (see Nassirtoussi, Aghabozorgi, Wah, et al., 2014; Nardo, Petracco-Giudici, and Naltsidis, 2016). That said, the system used to generate the proprietary 'Composite Sentiment Score' provided in the frequently used RavenPack News Analytics data sets was trained on the market response to article headlines (Shi, Ho, and Liu, 2016).

2.3.6 Natural Language Processing

The feature selection and transformation methods described above are more easily understood in terms of word or n-gram term-document matrix representations than the more advanced sentence-level representations used by state-of-the-art natural language processing (NLP) systems. These systems utilise the different extraction and scoring techniques already discussed, but in a hierarchical procedure. Following the example in Figure 2.2, the process can be considered in terms of pre-processing, lexical analysis, syntactic analysis and semantic analysis, and a different machine learning procedure or rule-based algorithm may be used for each of these steps. A pre-processing module may perform word and sentence splitting and lemmatisation, a separate lexical analysis module may then identify parts of speech and a syntactic parser could then be applied to this output to identify the sentence phrasal structures. In this example the POS tags have been used to identify a proper noun, a verb and two noun phrases, which is then interpreted as a verb-phrase relating to a single entity.

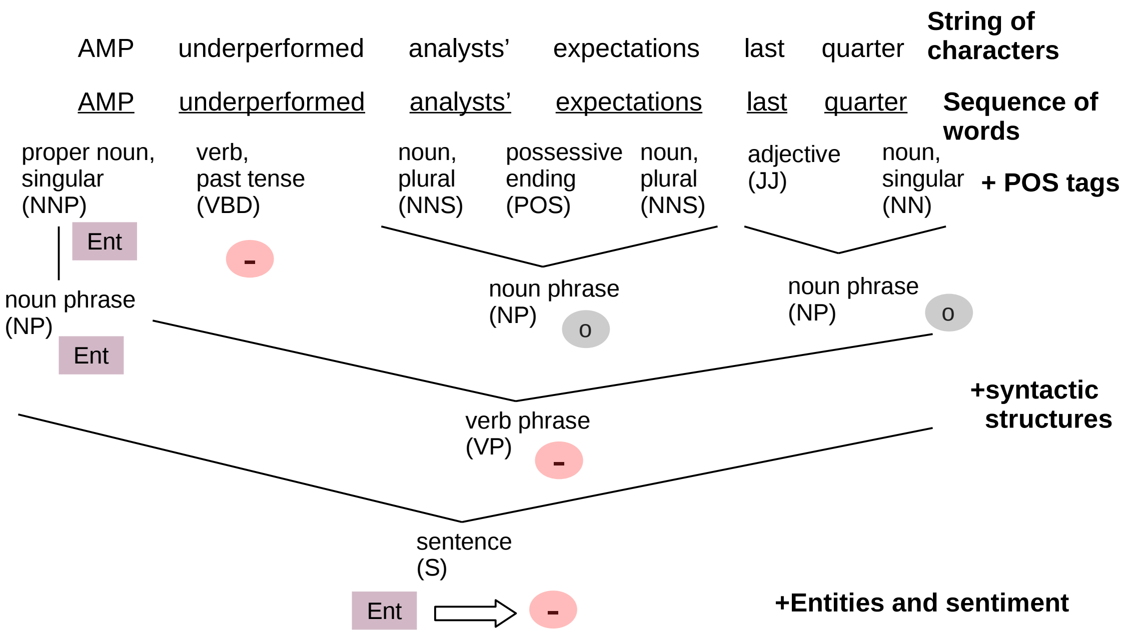

Figure 2.2: Example natural language processing hierarchy. The figure shows the processing steps of an advanced NLP system from a raw character string (top), to a recognised entity and sentiment relationship (bottom).

The most commonly used mathematical system for modelling language structure is the context-free grammar (CFG) (Jurafsky and Martin, 2016), which is graphically depicted by the node labels and tree structure in Figure 2.2. This form of CFG parse tree is generally represented in a format called bracketed notation as shown in Listing 2.1. Another family of grammar formalisms which have become important in contemporary NLP systems are dependency grammars, in which syntactic structure is described only in terms of the words and associated set of directed relations between them (Jurafsky and Martin, 2016). Representation of the example sentence using a dependency grammar is shown in Listing 2.2.

Listing 2.2: Dependency grammar output

```
nsubj(underperformed-2, AMP-1)
root(ROOT-0, underperformed-2)
nmod:poss(expectations-5, analyts-3)
case(analyts-3, '-4)
```

Listing 2.1: Context-free grammar output

```
(ROOT
 (S
   (NP (NNP AMP))
   (VP (VBD underperformed)
     (NP
       (NP (NNS analysts) (POS '))
       (NNS expectations))
     (NP (JJ last) (NN quarter)))
   (. .)))


iobj(underperformed-2, expectations-5)
amod(quarter-7, last-6)
dobj(underperformed-2, quarter-7)
```

After syntactic parsing, semantic analysis modules including NER and a sentiment engine may be used to identify and track known entities (AMP in this case) and map them to semantically relevant text, such as phrases containing sentiment associated words. In the context of financial news, this allows specific portions of text within a story that discusses multiple firms to be meaningfully attributed to the relevant firms. In this example the term *underperformed* has been identified as a term with negative sentiment through the relevant sentiment dictionary. As there are no negations, intensifiers or other sentiment words in the sentence, the negative association gets carried through hierarchically to the sentence level. Finally, these features may be used as input to the sentiment classifier, which may be configured to classify at the paragraph, sentence or entity level depending on the application.

This type of textual processing and sentiment scoring resembles that used within proprietary news analytics feeds such as those provided by Thomson Reuters and RavenPack, and is evidently distinct from bag-of-words approaches. An explanation of the TRNA text-processing engine which includes details of the training corpus and classification accuracy is provided by Sinha (2016), although it is unknown whether the system has since been upgraded; Sinha's description was based on a 2008 *Infonic* white paper. The textual analysis techniques utilised in the works selected for individual review herein are summarised in Tables 2.1 and 2.2.

2.3.7 Word Embeddings

A breakthrough in the field of textual analysis which has recently gained much traction in the development and application of language models is the use of word embeddings. A word embedding is a learned high-dimensional representation of a word (or less commonly, a phrase). Specifically, each word is represented as a real-valued vector of a predefined dimension, often on the order of 100-300.

Word embeddings use the distributional properties of the training corpus to quantify and categorise semantic similarities between terms; words that are used in similar ways will have similar representations, which implies similar meanings Harris (1954). Intuition regarding the representation of words as vectors, and examples of vector-orientated reasoning using such vectors, is provided by Mikolov, Yih, and Zweig (2013).

Two popular approaches for learning word embeddings are the Word2Vec Mikolov, Chen, et al. (2013) algorithm and the Global Vectors for Word Representation (GloVe) algorithm Pennington, Socher, and Manning (2014). Word2Vec uses a shallow, two-layer neural network to reconstruct (predict) the linguistic context of words. The prediction task can be formulated in two ways:

- Predict the current word based on its context (surrounding words).
- Predict the surrounding words given a current word.

The neural network learns the word embedding as a means to achieve these prediction tasks.

The GloVe algorithm is an efficient extension of the Word2Vec algorithm. Rather than using a window to define local context, GloVe constructs a word-context or word co-occurrence matrix. The learning task is then to learn word vectors such that their dot product is a good predictor of their co-occurrence.

Having obtained word embeddings, sentiment classification can then proceed as with other representations. For example, by using the sum or average word embedding as input features into the learning model, or by feeding each of the embedding vectors into an RNN, if information regarding the order of words is to be maintained. In the finance literature, word embeddings are used by Minh et al. (2018), Zhang, Zhang, et al. (2018), and Chen, Liao, and Hsieh (2019), among others.

Table 2.1: Custom textual analysis techniques

Author	Text Source	Feature Extraction	Feature Selection	Feature Representation	Dictionary
Wüthrich et al. (1998)	Wall Street Journal, Financial Times, Reuters, Dow Jones and Bloomberg	Unigrams to 5-grams	Phrase lexicon	Term weighting	Custom
Chan (2003)	Dow Jones Interactive Publications Library - WSJ, NY Times, LA Times, Globe, Chicago Tribune etc.	Item occurrence/ stock direction	NA	NA	NA
Tetlock (2007)	Wall Street Journal	BOW	Dictionary	Term frequency projected onto first principal component	H-IV4
Tetlock, Saar-Tsechansky, and Macskassy (2008)	Wall Street Journal, Dow Jones newswires	BOW	Dictionary	Term frequency	H-IV4
Fang and Peress (2009)	New York Times, USA Today, Wall Street Journal and Washington Post from the LexisNexis database	Item occurrence	NA	NA	NA

Continued on next page

Table 2.1: Custom textual analysis techniques (continued)

Author	Text Source	Feature Extraction	Feature Selection	Feature Representation	Dictionary
Kothari, Li, and Short (2009)	EDGAR, Factiva, Dow Jones Interactive	BOW	Dictionary	Term frequency, term occurrence	Custom, H-IV4
Tetlock (2010)	Dow Jones newswires from Factiva database	BOW	Keyword	Term occurrence	NA
Tetlock (2011)	Dow Jones newswires from Factiva database	Unigrams, bigrams	Not stated	Term occurrence	NA
Dougal et al. (2012)	Wall Street Journal	Journalist, BOW	Dictionary	Term frequency	LM
Garcia (2013)	New York Times	BOW	Dictionary	Term frequency	LM
Xiong and Bharadwaj (2013)	Lydia/TextMap (/*sim* 1000 online news sources)	words, sentences	NER, POS tags, synonym set, dictionary	Term frequency	Lydia/TextMap
Yu, Duan, and Cao (2013)	Google Blogs, BoardReader, Twitter, Google News – ABC News, NYT, USA Today, Fox News, Reuters, WSJ, WP, CNN, The Economist and Forbes.	BOW	Not stated	Term frequency	NA
Ammann, Frey, and Verhofen (2014)	Handelsblatt (German newspaper)	BOW	Dictionary	Term occurrence	LM

Continued on next page

Table 2.1: Custom textual analysis techniques (continued)

Author	Text Source	Feature Extraction	Feature Selection	Feature Representation	Dictionary
Hillert, Jacobs, and Mueller (2014)	New York Times, USA Today, WSJ + 41 local newspapers, from LexisNexis database	BOW	Dictionary	Term frequency	LM
Li, Xie, et al. (2014)	Macroeconomic and company news from FINET archive	BOW	Dictionary	Term weighting (Dictionary)	H-IV4, LM, SenticNet
Bianconi, Hua, and Tan (2015)	New York Times	BOW	Dictionary	Term frequency	Custom - variation of LM
Ferguson et al. (2015)	Financial Times, Guardian, Mirror from LexisNexis UK database	BOW	Dictionary	Term frequency	LM
Wang, Chen, and Wei (2015)	CMoney database	BOW	Dictionary	not reported	not reported
Yang, Mo, and Liu (2015)	Northern Light Single Point online news portal	BOW	Dictionary, lemmatization	Term weighting (Dictionary)	SentiWordNet
Ahmad et al. (2016)	newspapers, industry and trade magazines, newswires, financial blogs, web-based publications from LexisNexis database	BOW	Dictionary	Term frequency	LM
Kroujiline et al. (2016)	Factiva database, online news sources	Not provided	Phrase lexicon	Term occurrence	Custom

Continued on next page

Table 2.1: Custom textual analysis techniques (continued)

Author	Text Source	Feature Extraction	Feature Selection	Feature Representation	Dictionary
Strauß, Vliegenthart, and Verhoeven (2016)	5 leading Netherlands newspapers	BOW	Dictionary	Term frequency	LIWC
Zhang, Härdle, et al. (2016)	Articles and UGC from NASDAQ website	BOW with simple negation	Stemming (Porter), Dictionary	Term frequency	MPQA, LM, BL
Zhang, Song, et al. (2016)	NetEase Financial Channel (Chinese website)	Occurrence only	NA	NA	NA
Chan and Chong (2017)	News articles and blogs from Finet website	Words, sentences	POS tags, term weights, tree topological features, dictionary	Parse tree	H-IV4, SentiWordNet
Kraussl and Mirgorodskaya (2017)	Wall Street Journal, New York Times, Financial Times	BOW	Dictionary	Term occurrence	Custom - variation of LM
Manela and Moreira (2017)	Wall Street Journal	Unigrams, bigrams	Term frequency	NA	NA

Continued on next page

Table 2.1: Custom textual analysis techniques (continued)

Author	Text Source	Feature Extraction	Feature Selection	Feature Representation	Dictionary
Narayan and Bannigidadmath (2017)	New York Times	BOW	Dictionary	Term frequency	LM
Seng and Yang (2017)	KMW database - China Times	BOW	Dictionary, Chi-square	Term frequency	Custom
Wei et al. (2017)	China Times and Commercial Times from InfoTimes database	BOW	Dictionary	Term frequency	Diction (Chinese translation)
Wu and Lin (2017)	Market Observation Post System, and newspapers and websites from Taiwan Economic Journal	BOW	Dictionary	Term occurrence	Lu and Wei (2013) Chinese word list
Yang, Mo, Liu, and Kirilenko (2017)	Northern Light Single Point online business news portal	BOW	Dictionary, lemmatization	Term weighting (Dictionary)	SentiWordNet
Hillert, Jacobs, and Mueller (2018)	New York Times, USA Today, Wall Street Journal and Washington Post from the LexisNexis database	BOW	Dictionary	Term frequency	LM
Johnman, Vanstone, and Gepp (2018)	The Guardian	BOW	Dictionary	Term frequency	LM
Kelly and Ahmad (2018)	Wall Street Journal, Financial Times, Oildrum blogs	BOW	Dictionary	Term frequency	H-IV4, Platts, Oil and Gas UK

Continued on next page

Table 2.1: Custom textual analysis techniques (continued)

Author	Text Source	Feature Extraction	Feature Selection	Feature Representation	Dictionary
Minh et al. (2018)	Bloomberg, Reuters	Truncated article (300 words) / stock direction	Dictionary, stemming (Porter)	Term occurrence, term weighting, word embedding	H-IV4
Myskova, Hajek, and Olej (2018)	Yahoo Finance	BOW	Dictionary	Term frequency	LM
Zhang, Zhang, et al. (2018)	Wind web news	BOW	Verb/gerund dictionary, synonym set	Term occurrence, word embedding	HowNet
Calomiris and Mamaysky (2019)	Thomson Reuters	BOW, 4-grams	Dictonary, stemming	Term occurrence, co-occurrence	LM, Custom
Chen, Liao, and Hsieh (2019)	Online sources including China Times Finance, Yahoo Finance, Google Finance, China Electronics News	Article, BOW, stock direction	Dictionary, POS tags	Term occurrence, word embedding	Jiemba, NTUSD, Custom

Continued on next page

Table 2.1: Custom textual analysis techniques (continued)

Author	Text Source	Feature Extraction	Feature Selection	Feature Representation	Dictionary
Glasserman and Mamaysky (2019)	Thomson Reuters	4-grams	Dictionary	Term occurrence, co-occurrence	LM
Narayan (2019)	100 news sources from Factiva database	BOW	Dictionary	Term frequency	LM
Pyo and Kim (2019)	Online news sources	BOW	Dictionary	Term frequency, term weighting (MI)	Custom
Ahmed, Sriram, and Singh (2020)	Online news sources	BOW	POS tags, dictionary, thesaurus	Term frequency, term weighting (Dictionary)	WordNet, R
Hanna, Turner, and Walker (2020)	Financial Times	BOW	Dictionary	Tern Frequency	LM
Zhou et al. (2020)	EastMoney.com, Hexun.com, Finance.sina.com	BOW	POS tags, dictionary, thesaurus	Term frequency	HowNet, Sogou, custom

Table 2.2: Use of proprietary textual analysis packages in reviewed works

Author	Text Source	Package/ Database	Variables Used
Mitra, Mitra, and Dibartolomeo (2009)	Company news	RavenPack	Relevance, ESS
Nuij et al. (2014)	Company news	ViewerPro	Event classification, event impact score
Uhl (2014)	Market and company news	TRNA	Sentiment classification
Borovkova and Mahakena (2015)	Macroeconomic news (natural gas)	TRNA	Sentiment probability (positive, negative, neutral), novelty
Hendershott, Livdan, and Schuerhoff (2015)	Company news	TRNA	Sentiment probability (positive, negative), relevance, topic codes
Smales (2015[b])	Industry and company news	TRNA	Sentiment probability (positive, negative), novelty, relevance
Smales (2015[a])	Macroeconomic news (gold)	TRNA	Sentiment probability (positive, negative), novelty, relevance
Uhl, Pedersen, and Malitius (2015)	Company and macroeconomic news	TRNA	Sentiment probability (positive, negative)
Khuu, Durand, and Smales (2016)	Market and company news	TRNA	Sentiment probability (positive, negative), novelty, relevance
Nooijen and Broda (2016)	Online company news and social media	TRMI	Social media and news sentiment indices
Shi, Ho, and Liu (2016)	Company news	RavenPack	Relevance, novelty, CSS, ESS
Sinha (2016)	Company news	TRNA	Sentiment probability (positive, negative), novelty, relevance

Continued on next page

Table 2.2: Use of proprietary textual analysis packages in reviewed works (continued)

Author	Text Source	Package/ Database	Variables Used
Smales (2016)	Company news	RavenPack	MCQ sentiment score, relevance, novelty
Allen, McAleer, and Singh (2017)	Company news	TRNA	Sentiment probability (positive, negative)
Cahan et al. (2017)	Company news	TRNA	Sentiment probability (positive, negative)
Heston and Sinha (2017)	Company news	TRNA	Sentiment probability (positive, negative, neutral), novelty
Huynh and Smith (2017)	Company news	TRNA	Sentiment probability (positive, negative), relevance, novelty, headline
Song, Liu, and Yang (2017)	Company news	TRNA	Sentiment probability (positive, negative), relevance
Uhl (2018)	Macroeconomic news	TRNA	Sentiment probability (positive, negative)
Audrino, Sigrist, and Ballinari (2020)	Company news	RavenPack	Relevance, ESS, CSS, NIS
Griffith, Najand, and Shen (2020)	Online news and social media	TRMI	Fear, gloom, joy, stress
Sadik, Date, and Mitra (2019[a])	Macroeconomic news	RavenPack	Event classification, relevance, novelty, sentiment
Sadik, Date, and Mitra (2019[b])	Company news	RavenPack	Relevance, novelty, ESS
Vanstone, Gepp, and Harris (2019)	Company news and social media	Bloomberg	News count, positive news count, negative news count
Al-Maadid et al. (2020)	Business and political news	Bloomberg	Sentiment classification
Coqueret (2020)	Company news and social media	Bloomberg	Sentiment score

Continued on next page

Table 2.2: Use of proprietary textual analysis packages in reviewed works (continued)

Author	Text Source	Package/ Database	Variables Used
Gan et al. (2020)	Company news	TRMI	Social media and news sentiment and buzz indices

Note: MCQ, CSS and ESS represent RavenPack's *multiclassifier for equities*, *composite sentiment score*, and *event sentiment score*, respectively (Mitra and Mitra, 2011).

2.4 Econometric Testing

Each finance study making use of news analytics consists of two broad components; attainment of features derived from the arrival and content of news text, and econometric analysis and hypothesis testing. While the textual analysis techniques (such as the dictionary used) and the econometric models employed for a given study should be consistent with the overall hypothesis, econometric testing can generally be considered independently from whatever textual analysis techniques were used. Given this, there are a number of ways presented in the literature in moving from the attainment of news-based features to econometric testing. In the context of this discussion, econometric testing broadly refers to means of predicting or explaining market variables in terms of news variables. It does not capture hypothesis formation or the suitability of a given econometric technique in testing a given hypothesis. The econometric techniques used within the reviewed words are presented in Table 2.3.

2.4.1 Parametric Techniques

In a parameter-focused approach, the testing of hypotheses concerning the relationship between news flow and market dynamics is largely reduced to an estimation problem. Parametric methods first assume a functional form of the relationship between the predictors (or independent variables) and response (dependent variable) being examined. This functional form is described explicitly using a model which is then fitted to the data by estimating the set of model parameters according to some objective function.

The nature of the relationship between the predictors and response can then be judged on the basis of the statistical significance and magnitude of the model parameters and model closeness-of-fit statistics (insofar as the model captures the true or hypothesised relationship between the predictors and response). Simple parametric models are relatively inflexible and are likely to mischaracterise the true functional relationship between the predictors and response. However, they are highly interpretable and therefore suited to hypotheses requiring inference as to the role of individual predictors in the modeled system.

Linear Regression

The most common parametric approach in the news analytics literature is multiple linear regression—a single equation linear model with multiple predictors, fit using ordinary or generalised least squares. Typically, the dependent variable used is raw return or abnormal return calculated from the Fama-French three-factor model or CAPM, and the independent variables include current and lagged values of the relevant news measures and control variables. Autocorrelation and heteroskedasticity in the residuals are commonly corrected using Newey and West (1987) or White (1980) standard errors.

Of the relevant studies considered for this review, multiple linear regression is used as the primary econometric technique by Tetlock, Saar-Tsechansky, and Macskassy (2008), Dougal et al. (2012), Yu, Duan, and Cao (2013), Garcia (2013), Lu and Wei (2013), Ammann, Frey, and Verhofen (2014), Kothari, Li, and Short (2009), Ferguson et al. (2015), Hendershott, Livdan, and Schuerhoff (2015), Smales (2015[b]), Yang, Song, et al. (2015), Bianconi, Hua, and Tan (2015), Strauß, Vliegenthart, and Verhoeven (2016), Khuu, Durand, and Smales (2016), Shi, Ho, and Liu (2016), Cahan et al. (2017), Zhang, Härdle, et al. (2016), Wu and Lin (2017), Narayan and Bannigidadmath (2017), and Seng and Yang (2017). Logistic regression is used when the response variable is categorical, and one wants to model the posterior probabilities of each class as linear functions of the inputs. Hendershott, Livdan, and Schuerhoff (2015) uses logistic regression to estimate the probability of a news event in terms of institutional order volume and the prior day's news incidence.

Vector Autoregression

A variant of the single equation regression model is the vector autoregression (VAR) model, which is used in the literature to capture the linear interdependencies between the separate time series of dependent variables, news measures and control variables. That is, the VAR model explains the time series evolution of each variable in terms of their own lagged values and the lagged values of the other model variables. Of the articles considered, VAR models have been used by Tetlock (2007), Uhl (2014), Ahmad et al. (2016), Hendershott, Livdan, and Schuerhoff (2015), Wang, Chen, and Wei (2015), Strauß, Vliegenthart, and Verhoeven (2016), Smales (2016), Zhang, Härdle, et al. (2016), Wei et al. (2017), and Kraussl and Mirgorodskaya (2017). Xiong and Bharadwaj (2013) used VAR in addition to dynamic panel generalised method of moments (GMM) estimate of the same equation. Properties of the estimated VAR model are often examined through structural analysis techniques such as Granger causality, (and less commonly) impulse responses and forecast error variance decompositions (see Uhl, 2014 for an example of all three). Granger causality is used by Bianconi, Hua, and Tan (2015), Wang, Chen, and Wei (2015), Strauß, Vliegenthart, and Verhoeven (2016), Kraussl and Mirgorodskaya (2017), Checkley, Higon, and Alles (2017), and Chan and Chong (2017).

Volatility Models

Another parametric model commonly used among the studies considered is the GARCH model and its variants. These are used to examine the relationship between news and conditional volatility dynamics whilst allowing for well-known features of financial time series such as volatility persistence, excess kurtosis and distributional asymmetry. Variants of the standard GARCH model used within the literature include EGARCH and GJR-GARCH, which allow for asymmetric response of innovations; Markov-switching GARCH (MS-GARCH), which allows for different volatility regimes, and; FI-GARCH, which allows for long memory persistence. Models of the GARCH family are used by Shi, Ho, and Liu (2016), Nooijen and Broda (2016), and Allen, McAleer, and Singh (2017).

Table 2.3: Use of econometric techniques

Author	Market Variable	Estimation Horizon	Econometric Technique
Wüthrich et al. (1998)	Return	1 day	Rule-based classifier, kNN, ANN, linear regression
Chan (2003)	Return	1-36 Months	Portfolio analysis
Tetlock (2007)	Return, volume	5 days	VAR, event study, trading simulation
Tetlock, Saar-Tsechansky, and Macskassy (2008)	Return, earnings	10 days, 6 quarters	Linear regression, event study, trading simulation
Fang and Peress (2009)	Return	1-12 months	Portfolio analysis

Continued on next page

Table 2.3: Use of econometric techniques (continued)

Author	Market Variable	Estimation Horizon	Econometric Technique
Kothari, Li, and Short (2009)	Return, volatility, earnings	Contemp., 1 quarter	Linear regression
Mitra, Mitra, and Dibartolomeo (2009)	Volatility	1 day	Linear regression, prediction
Tetlock (2010)	Return, volume	10 days	Linear regression
Leinweber and Sisk (2011)	Return	60 days, 20 days	Event study, trading simulation
Tetlock (2011)	Return	5 days	Linear regression
Dougal et al. (2012)	Return	1 day	Linear regression
Garcia (2013)	Return	5 days	Linear regression
Xiong and Bharadwaj (2013)	Return	1 month	Dynamic panel GMM, VAR
Yu, Duan, and Cao (2013)	Return, volatility	1 day	Linear regression
Ammann, Frey, and Verhofen (2014)	Return, industrial production	1 month, 3 months	Linear regression
Hillert, Jacobs, and Mueller (2014)	Return	6-36 months	Portfolio analysis, event study
Li, Xie, et al. (2014)	Return	1 day	SVM
Nuij et al. (2014)	Return	1, 3, 5 days	Genetic algorithm, trading simulation optimisation
Uhl (2014)	Return	1 month	VAR, trading simulation
Bianconi, Hua, and Tan (2015)	Volatility, Co-volatility, VIX	Contemp., 1 day	Linear regression, Granger causality
Borovkova and Mahakena (2015)	Volatility, returns	1 day	GARCH, HEAVY, GJR-GARCH, VAR, Granger causality
Ferguson et al. (2015)	Return	1 day	Linear regression, trading simulation

Continued on next page

Table 2.3: Use of econometric techniques (continued)

Author	Market Variable	Estimation Horizon	Econometric Technique
Hendershott, Livdan, and Schuerhoff (2015)	Return, institutional order flow	10 days	Logit regression, event study, linear regression, VAR, impulse response
Smales (2015[b])	Return	Contemp.	Linear regression
Smales (2015[a])	Volatility	1 day	Linear regression, TGARCH
Uhl, Pedersen, and Malitius (2015)	Return	Dynamic, 1 week	Trading simulation, linear regression
Wang, Chen, and Wei (2015)	Return	5 days	VAR, Granger causality, impulse response
Yang, Song, et al. (2015)	Return, volume, volatility	11 days	Linear regression
Ahmad et al. (2016)	Return	up to 250 days	VAR
Khuu, Durand, and Smales (2016)	Return	1 day contemp.	Linear regression
Kroujiline et al. (2016)	Return	$\approx$ 45 days	Trading simulation, model fitting
Nooijen and Broda (2016)	Volatility, return	1 day	EGARCH, MS-GARCH
Shi, Ho, and Liu (2016)	Volatility, return	1 month	EGARCH, linear regression
Sinha (2016)	Return	13 weeks	Portfolio analysis, trading simulation
Smales (2016)	VIX, return	Contemp., 1-250 days	VAR, linear regression
Strauß, Vliegenthart, and Verhoeven (2016)	Return	1 day	VAR, Granger causality
Zhang, Härdle, et al. (2016)	Volume, volatility, return	1 day	VAR
Zhang, Song, et al. (2016)	Return, volume	50 days	Event study, linear regression

Continued on next page

Table 2.3: Use of econometric techniques (continued)

Author	Market Variable	Estimation Horizon	Econometric Technique
Allen, McAleer, and Singh (2017)	Returns	1 day	Linear regression, EGARCH, GJR-GARCH, entropy, MI
Cahan et al. (2017)	Volatility, liquidity volatility	1 month	Linear regression
Chan and Chong (2017)	Return	5 days	Granger causality
Heston and Sinha (2017)	Return	13 weeks	Portfolio analysis, event study
Huynh and Smith (2017)	Return	1 year	Portfolio analysis
Kraussl and Mirgorodskaya (2017)	Return, volatility	1 month	VAR, Granger causality
Manela and Moreira (2017)	Return, volatility, VIX	1-24 months	SVR, linear regression
Narayan and Bannigidadmath (2017)	Return	1 day	Linear regression, trading simulation
Seng and Yang (2017)	Volatility	1 year	Linear regression forecasts
Song, Liu, and Yang (2017)	Return	1 week	Neural network, trading simulation
Wei et al. (2017)	Return, trading value, turnover ratio, volatility	1 week, 1 month	VAR, portfolio analysis
Wu and Lin (2017)	Return, institutional order flow	1 month contemp.	Linear regression

Continued on next page

Table 2.3: Use of econometric techniques (continued)

Author	Market Variable	Estimation Horizon	Econometric Technique
Yang, Mo, Liu, and Kirilenko (2017)	Return	$/approx$ 1 day	Genetic algorithm, trading simulation optimisation
Hillert, Jacobs, and Mueller (2018)	Return	1 day	Linear regression, trading simulation
Johnman, Vanstone, and Gepp (2018)	Return, volatility	1 day	Linear regression, trading simulation
Kelly and Ahmad (2018)	Return	1 day	VAR, trading simulation
Minh et al. (2018)	Price	1-10 days	ANN (Two-stream GRU RNN)
Myskova, Hajek, and Olej (2018)	Volatility	3 days	Tree-based meta learners / ensembles
Uhl (2018)	Implied volatility (1m – 3y), Implied skew	1 day	Linear regression
Zhang, Zhang, et al. (2018)	Price	1 day	Coupled matrix and tensor factorisation, SVM, TeSIA
Audrino, Sigrist, and Ballinari (2020)	Volatility, VaR	1-22 days	HAR, prediction
Calomiris and Mamaysky (2019)	Return, volatility, drawdown	1 month, 1 year	Linear regression, penalised linear regression,
Chen, Liao, and Hsieh (2019)	Return	1 day	ANN (LSTM)
Glasserman and Mamaysky (2019)	Volatility, implied volatility	1 month	VAR, linear regression
Griffith, Najand, and Shen (2020)	Return, volatility	1 day	VAR, TGARCH, impulse response
Narayan (2019)	Return	1 month	VAR, GARCH

Continued on next page

Table 2.3: Use of econometric techniques (continued)

Author	Market Variable	Estimation Horizon	Econometric Technique
Pyo and Kim (2019)	Return, implied volatility, volatility	1 day	VAR
Sadik, Date, and Mitra (2019[a])	Price	1 day	VAR, Kalman filter, prediction
Sadik, Date, and Mitra (2019[b])	Volatility	1 day	GARCH, EGARCH, news augmented GARCH
Vanstone, Gepp, and Harris (2019)	Price	1 day	ANN-AR, prediction
Ahmed, Sriram, and Singh (2020)	Price	1 day	Single layer perceptron, multi-layer perceptron, ANN, SVM, DT, RF, LVQ, GBM, NB
Al-Maadid et al. (2020)	Return, volatility	1 week	HMM, linear regression
Coqueret (2020)	Return	0,4,9,20 days	Linear regression
Gan et al. (2020)	Return, volatility, volume	1-20 days	VAR, impulse response, Granger causality
Hanna, Turner, and Walker (2020)	Return, volume	1 day	Linear regression
Zhou et al. (2020)	Return	1-3 days	SVM

Note: Estimation horizons listed with approximate values reflect studies in which the estimation horizon was not fixed but could be estimated, such as dynamic trading signals with a cited average holding period.

Analytic Models

Although infrequent in the review sample, authors may describe the relationship between news information and market variables explicitly through an analytic model. In addition to the clear role of such models in hypothesis formation (see Tetlock, 2010), some models may be capable of being fit to the data directly. In this case, the model may be used for forecasting, or for empirically describing news-market dynamics in ways that weren't explicit predictions of the model equations (see Kroujiline et al., 2016).

2.4.2 Nonparametric, Semiparametric and Model-free Techniques

Cross-sectional Portfolios

The most prominent nonparametric econometric technique within the reviewed literature is the cross-sectional portfolio approach, in which groups of assets (portfolios) reflecting particular portions of cross-sectional input space are compared in terms of a given response variable. The input space being conditioned on is a cross-sectional location in that it represents a particular quantile of the input distribution at a given point in time, for which the actual values of the underlying variables are not pre-specified and are likely to vary through time. In practice, the conditioning process amounts to ranking all assets based on an input variable and selecting particular groups for portfolio construction (often via a long-short portfolio) and comparison. If more than one variable is being conditioned upon, the ranking and grouping procedure is usually conducted sequentially, with ranking of subsequent variables occurring within each group of the previous variable–this is case for all portfolio analyses in the review sample conditioning on multiple variables, reflected by “Sorting quantiles" column of Table 2.4.

Portfolios are formed based on an input space consisting of news and market data for a specified lookback window (formation period), with post-formation return (the response variable) then measured over a given holding period. The time, if any, between the formation event and the holding period is termed the skip period, and may be included to avoid capturing microstructure effects or specific market activity such as short-term return reversal. The specification of a portfolio procedure employing a given set of predictor and response variables can be summarised using the Jegadeesh and Titman (1993) J/S/K notation, whereby J,S and K represent the portfolio formation, skip and holding periods, respectively.

In the standard portfolio analysis approach, portfolios are formed at every time interval (such as monthly), such that the holding periods overlap and each point in calendar time is associated with the returns of K portfolios. The average return of the overlapping portfolios results in a time series of returns with useful properties, Fama (1998) writes:

> The time-series variation of the monthly abnormal return on this portfolio accurately captures the effects of the correlation of returns across event stocks missed by the model for expected returns. The mean and variance of the time series of abnormal portfolio returns can be used to test the average monthly response of the prices of event stocks for [K periods]... following the event.

As the output is a time series of returns, the exposure of the portfolio to common risk factors, such as the Fama-French three factors and the Carhart four factors, can be estimated by regressing the return series against the contemporaneous value of the relevant portfolios (market, size, book-to-market and UMD portfolios, for example). To account for the possibility that the documented strategy has nonlinear risk exposures which are not captured by the linear factor models, various authors also report benchmark adjustments. This involves constructing portfolios matched on the basis of the specified risk factors and subtracting the returns contemporaneously generated by the risk-matched portfolio from the strategy portfolio. Portfolio intrinsic measures such as volatility and average return per unit of variance are also be used to indicate risk.

The cross-sectional portfolio approach is used primarily within the low-frequency news analytics literature and is a particularly useful way of examining the economic significance and exploitability of news-driven pricing effects. The time series of returns produced by the analysis can be interpreted as the returns of a trading strategy implementing the portfolio rules over the test period, while the strategy itself implicitly offers regular and well-defined rebalancing characteristics that are not offered by signal-based strategies. These properties make the technique particularly amenable to portfolio management research. Of the reviewed studies, overlapping portfolio analysis is used by Chan (2003), Fang and Peress (2009), Hillert, Jacobs, and Mueller (2014), Sinha (2016), Heston and Sinha (2017), Wei et al. (2017), and Huynh and Smith (2017). Song, Liu, and Yang (2017) generate cross-sectional portfolios each week but holding periods are non-overlapping. Table 2.4 lists the specifications of the cross-sectional portfolios examined in each of these studies.

Event Study

Event studies are another common model-free econometric method used in the literature and are generally applied in addition to other forms of analysis. Event studies typically attempt to measure the impact of different news events on a firm's stock price or the comparative post-formation performance of different portfolios. Most often, an estimate of the expected or 'normal' value of the variable of interest is determined using an average of recent values or a market model such as the CAPM–in the case of returns. The difference between the normal value of the market variable and its actual value is calculated and plotted throughout an event window featuring the event. Of the review sample, Heston and Sinha (2017), Zhang, Song, et al. (2016), Hendershott, Livdan, and Schuerhoff (2015), Hillert, Jacobs, and Mueller (2014), Leinweber and Sisk (2011), and Tetlock, Saar-Tsechansky, and Macskassy (2008) employed event studies in their analyses.

Table 2.4: Selected cross-sectional portfolio analyses

Author	J/S/K	Sorting quantiles	Sorting variables
Song, Liu, and Yang (2017)	1W/0W/1W	decile	ANN output
Wei et al. (2017)	1M/0M/1M	quintile, quintile	tone, [PB,TO,MV]
Heston and Sinha (2017)	1W/[0,1W]/13W	decile	tone
Huynh and Smith (2017)	1W/[0,4W]/52W	tercile, tercile, tercile	attention, tone, return
Sinha (2016)	1W/0W/13W	decile	tone
Hillert, Jacobs, and Mueller (2014)	6M/1M/[6-36M]	quintile, tercile, median	coverage, return, tone
Fang and Peress (2009)	[1-6M]/0M/[1-12M]	binary, median	incidence, coverage
Chan (2003)	1M/1W/[1-36M]	binary, tercile	incidence, return

Note: 'J', 'S', and 'K' represent formation, skip, and holding period, respectively. 'M' and 'W' reflect months and weeks repsectively. 'PB', 'TO', and 'MV' represent price-to-book ratio, turnover, and market value, respectively, as defined in Wei et al. (2017).

Nonparametric Models

The parametric econometric techniques applied in the literature are generally linear, in that the response can be expressed in terms of a linear function of the predictors. This may impose restrictions on the relationship between the response and the dependent variables, as well as the interactions between dependent variables, that do not capture the true nature of the data in which the statistical relationships are more complex.

In contrast, nonparametric models do not imply explicit a priori assumptions about the nature of the functional relationship between the predictors and response. While nonparametric models may contain parameters, the number and nature of the parameters are flexible and determined by the data. Due to their flexibility, nonparametric models are able to estimate a much wider range of news-market functional forms than their commonly used parametric counterparts and are therefore more capable of exploiting inherent features of the data. However, this added flexibility comes at the cost of interpretability, as complicated functional forms obscure the relationship between any individual predictor and the response. For this reason, nonparametric models are best suited to hypotheses reducible to prediction tasks for which interpretability is not important.

Wüthrich et al. (1998) trial K-nearest neighbours (KNN) and an artificial neural network (ANN) for their daily market prediction system, but decide on a probabilistic rule set.

Song, Liu, and Yang (2017) use an ANN trained with two different learning-to-rank loss functions for their week-by-week trading strategy. Li, Xie, et al. (2014) use support vector machines (SVM) for their news calssification and stock prediction system. The use of nonparametric models within the review sample is limited, suggesting that most hypotheses are primarily associated with understanding the impact of different characteristics of news flow rather than the potential for news analytics to optimise forecasting and trading performance.

Performance-focused analyses appear to be more common in the intraday literature, for example Fung, Yu, and Lam (2003) and Schumaker and Chen (2009) use SVM; Geva and Zahavi (2014) use an ANN, a genetic algorithm (GA) augmented decision tree and stepwise logistic regression; Hagenau, Liebmann, and Neumann (2013) and Li, Wang, et al. (2014) use support vector regression (SVR); Feuerriegel and Prendinger (2016) use a decision tree; Maragoudakis and Serpanos (2016) use Markov-chain Monte Carlo (MCMC) tree-augmented naïve Bayesian classifier, random forest (RF), SVM and an ANN.

Forecast Comparison

A number of the reviewed works use the forecasting performance of different (parametric or nonparametric) models, either directly or in a trading strategy, to provide evidence as to the predictive capacity of news content. For example, GARCH models with and without news-based parameters may be compared on the basis of one-step-ahead value-at-risk (VaR) forecasts. If the model containing the news-based parameters out-performs the news-free model, this may be interpreted as a demonstration of the predictive content of the news information used. Sometimes the forecasting problem can be interpreted as a classification task (such as directional price predictions), in this case, the relative accuracy of the news-driven model may be compared to the accuracy expected by chance (Wüthrich et al., 1998). These types of tests are considered nonparametric because the function relating the news variable(s) to the out-of-sample forecasting performance of models with different specifications is not made explicit. Forecast comparison is used by Uhl (2014), Ammann, Frey, and Verhofen (2014), Nooijen and Broda (2016), and Bianconi, Hua, and Tan (2015). This testing philosophy is also the basis for the use of metaheuristics.

Metaheuristics

Metaheuristic procedures are sometimes employed to algorithmically select the combination of model type(s), model specifications, inputs and parameters to converge on a combination that delivers a high-level of performance in terms a predefined fitness function. As they make few assumptions about the optimisation problem, they are often suitable for complex search spaces that are not well understood, relatively unstructured, and too large to be completely sampled (De Jong, 1988, pp. 122).

Metaheuristics are used in the news analytics literature as a means to test whether news data, including news-driven events and signals, is a competitive source of conditioning information for predicting stock movements in the context of a trading system. Nuij et al. (2014) and Seng and Yang (2017) both employ genetic algorithms (GAs), a metaheuristic belonging to the larger class of evolutionary algorithms, to find optimum trading rules among a set of news-driven and market-driven indicators.

Trading Simulations

It is well understood that statistically significant econometric relationships are not necessarily economically meaningful. In order to investigate the economic significance of news-market relationships and the potential relevance to financial practitioners, many researchers have conducted portfolio analyses and trading simulations based on the news information being investigated. While such analyses do shed light on the economic size of an effect and help to frame the investigated relationship in terms of systematic trading rules, there are number of reasons why they may still fail to reflect the performance of a realistic investment strategy. The availability of the assets used for testing, proposed rebalancing frequency, the inclusion of transaction fees, survivorship bias, strategy market depth and reported evaluation metrics are among the experimental parameters that help determine the economic *relevance* of a documented strategy. This discussion focuses on portfolio analyses and trading simulations with a rebalancing frequency of at least one week, as the current work is specifically concerned with the economic relevance of low frequency news-market dynamics.

One of the most important experimental parameters governing the practical value of a simulated portfolio is the universe of investable assets on which the test is conducted. The selected market itself may be undesirable to portfolio managers due to liquidity, regulatory issues, governmental interference or home bias, but as most tests are conducted with US securities, this is generally not considered a problem. A more relevant issue is the particular subset of securities used for testing. In particular, the inclusion of small stocks or otherwise thinly traded securities may introduce spurious results due to microstructure issues such as bid-ask bounce, and may severely limit the scalability of reported strategies due to the price impact of large trades.

Sinha (2016) and Heston and Sinha (2017) perform their analyses using all US common stocks on the CRSP database for which they have news data. Given such a broad universe of assets, there is a strong possibility that many of the stocks contributing to their results represent negligible investment opportunities due to availability (short positions) and trading frictions. For instance, Chan (2003), who used a random 25% of US common stocks on the CRSP database, noted that most of his observed news/momentum effect was on the downside of smaller, less liquid stocks and found that the news-related return patterns diminished (but remained significant) when accounting for liquidity. To account for such effects, Fang and Peress (2009) used all stocks listed on the NYSE which "contains mainly large stocks" and 500 randomly selected NASDAQ stocks, and excluded stocks with a price below $5. Even within this relatively restricted sample of assets, Fang and Peress cited bid-ask bounce as a having a potentially significant impact on their results and conducted additional robustness checks to ensure this wasn't the case. Hillert, Jacobs, and Mueller (2014) used all stocks in the CRSP database (with news) listed on NYSE, AMEX, or NASDAQ with prices above $5, and further excluded stocks below the first (lowest) NYSE size decile.

Uhl (2014), Yang, Mo, and Liu (2015) and Kroujiline et al. (2016), Uhl, Pedersen, and Malitius (2015) and Smales (2016) presented news-based market-timing strategies based on the DJIA, SPY, the MSCI WEI, and SPX, respectively. While instruments for these indices (and shares for the case of SPY) are readily accessible and highly liquid, single-asset market-timing strategies are a seemingly risky proposition for many portfolio managers and investors; at any one time, all invested wealth fluctuates with the rise and fall of a single security and an assessment of the portfolio manager's skill is easily distilled down to in-or-out decision-making ability. Wei et al.'s (2017) analysis included all stocks listed on the Taiwan Stock Exchange and is therefore likely to contain many small and infrequently traded assets. Additionally, a strategy's performance as documented in the Taiwanese stock market may yield very different results when employed in London or the US.

The inclusion or exclusion of transaction costs and slippage, whether by fees or price impact, may also affect the realism of a strategy's reported performance. For example, Tetlock (2007) and Tetlock, Saar-Tsechansky, and Macskassy (2008) found that their strategies would no longer be profitable once round-trip transaction costs exceeded 4.4 and 8 bp respectively. These are both daily strategies, and all else-equal, transaction costs become less important as rebalancing frequency decreases. However, transaction costs may still have an appreciable impact on the performance of strategies which rebalance less frequently, particularly those holding a large number of securities and rebalancing all positions upon each investment (such as equal-weighted portfolios). Chan (2003) noted that his equal-weighted strategy which rebalanced monthly would be economically costly in practice and presented the results of a more feasible strategy based on buy-and-hold returns. Fang and Peress (2009) and Hillert, Jacobs, and Mueller (2014) also tested equal-weighted strategies rebalanced monthly, but did not present buy-and-hold alternatives.

Since the approximate number of positions taken each month is not stated within these studies, it is difficult to estimate the relative economic impact of rebalancing costs between them.

Sinha (2016) and Heston and Sinha (2017) employ equal-weighted strategies with weekly rebalancing and a 13 week holding period. Considering their large and indiscriminate sample, this rate of rebalancing would likely result in significant transaction costs for the documented strategy. The strategy proposed by Uhl, Pedersen, and Malitius (2015) did not use a predefined trade frequency, but averaged eight trades per year over the sample period and included a per-trade transaction cost of 0.75%. Kroujiline et al. (2016) tested a forecast-based strategy with an average holding period of 45 and 8 business days over the two test periods included, and included a transaction cost drag of 0.8% per annum. Yang, Mo, and Liu (2015) and Smales (2016) also tested strategies with on-condition trade frequencies, which averaged 20 and 22 trades per year over the sample period, respectively. Yang, Mo, and Liu included a per-trade transaction cost based on the historical bid-ask spread and a brokerage fee, but did not provide any details as to the magnitude of these costs. Uhl (2014) and Wei et al. (2017) tested strategies with a one-month holding period between trades and did not include transaction costs. As the strategies tested by Uhl (2014) and Smales (2016) only relied on a single long or short position each trade, transaction costs represent much less of a potential issue compared to the other strategies discussed.

An important measure of the true economic size of an effect is the maximum dollar capacity of a strategy designed to exploit it. An upper bound for this value can be determined by the total traded volume of the securities held by the strategy. This can be reported as an average daily volume or the returns generated by a rolling portfolio that is assumed to have invested in the historical traded volume at each rebalance (see Vanstone and Hahn, 2017). Another approach is to estimate the dollar value trade volume that would lead to a price impact equal to the strategy's estimated alpha. This is the approach taken by Fang and Peress (2009), who provide the only estimate of strategy depth out of the studies examined. They found that the average daily trading volume of the stocks in their media portfolio had an average trading volume of about \$2 million and using the Amihud (2002) illiquidity ratio, estimated that it would take a trade of approximately \$0.61 million over a single day to eliminate the strategy's alpha.

The performance of the strategies presented in the examined literature have been reported in a number of different ways but have generally considered measures of average return (often in addition to those in excess of some benchmark) and risk exposure. Fang and Peress (2009) and Hillert, Jacobs, and Mueller (2014) reported average monthly returns using the overlapping portfolio approach of Jegadeesh and Titman (1993) with rolling monthly windows, which shows the average return in calendar time of implementing such a strategy. Chan (2003) also used the Jegadeesh and Titman portfolio approach but reports summed, instead of average returns, which shows how a strategy performs over time. Sinha (2016) used weekly average returns for overlapping and non-overlapping positions.

The latter corresponds to the average weekly return of each trade, for each week in the holding period, and is the primary reporting method used by Heston and Sinha (2017). Sinha (2016) also presented the cumulative portfolio return over the sample period, i.e. the strategy's equity curve. Smales (2016) reported the average forward return of different holding periods following the strategy entry signal, using bootstrapping with 10,000 resamples. Wei et al. (2017) presented average monthly portfolio return, which, given the way the strategy is described, also corresponds to the average return of each trade. Uhl (2014) and Uhl, Pedersen, and Malitius (2015) reported total return and average per annum return, while Yang, Mo, and Liu (2015) and Kroujiline et al. (2016) reported average per annum return.

Risk exposure is generally reported by either regressing strategy returns against common risk factors, by subtracting the return of portfolios matched on the basis of those risk factors (benchmark adjustments), or through portfolio intrinsic measures such as volatility and average return per unit of variance. Chan (2003), Fang and Peress (2009), Hillert, Jacobs, and Mueller (2014), and Sinha (2016) regressed portfolio time series returns on the market (CAPM) model, Fama and French (1993) three-factor model, and Carhart (1997) four-factor models. Hillert, Jacobs, and Mueller (2014) also regressed returns on a six-factor model which included the Carhart four-factor model with additional factors for short-term and long-term reversal, while Fang and Peress (2009) regressed against the Pástor and Stambaugh (2003) liquidity factor model. Kroujiline et al. (2016) regressed on the CAPM.

To account for the possibility that the documented strategy has nonlinear risk exposures which aren't captured by the linear factor models, various authors also report benchmark adjustments. This involves constructing portfolios matched on the basis of the specified risk factors and subtracting the returns contemporaneously generated by the risk-matched portfolio from the strategy portfolio. Chan (2003) used size and B/M benchmarks, Heston and Sinha (2017) used 26-week momentum and size benchmarks, Fang and Peress (2009) used Daniel, Grinblatt, et al. (1997) (DGTW) benchmarks, and Hillert, Jacobs, and Mueller (2014) used size, B/M, turnover, industry and DGTW benchmarks.

Uhl (2014), Uhl, Pedersen, and Malitius (2015), Yang, Mo, and Liu (2015), Smales (2016), and Kroujiline et al. (2016) rely on binary (in-out or long-short) strategies with market-based instruments and so rely more on portfolio metrics than factor models. Uhl (2014) reported market-excess return and Sharpe ratio; Uhl, Pedersen, and Malitius (2015) reported market excess return, volatility, maximum drawdown, tracking error and Sharpe ratio; Kroujiline et al. (2016) reported Sharpe and Sortino ratios, volatility, 5% monthly VaR, and maximum monthly drawdown, and Yang, Mo, and Liu (2015) reported volatility and Sharpe ratio. Smales (2016) reported the coefficient of variation associated with the bootstrapped mean forward return. Wei et al. (2017) did not report on risk exposure.

In addition to the liquidity and frequency of trade issues already discussed, another consideration concerning position availability is whether certain simulated portfolio positions could have been taken at all. An obvious case of this kind is the short sale constraints enforced during the global financial crisis (discussed by Boulton and Braga-Alves, 2010; Yerkes, 2011, and many others). Of the strategies discussed, those of Hillert, Jacobs, and Mueller (2014), Sinha (2016), and Heston and Sinha (2017) relied on the short-sale of individual stocks during the GFC. Sinha (2016) addressed this by issue by presenting additional results with the GFC dates removed from the sample. Hillert, Jacobs, and Mueller (2014) did not address the GFC directly, but did document portfolio performance individually for each year in the sample. Heston and Sinha (2017) did not include subperiod results or corrections.

Out of the surveyed literature, the portfolio analysis results of Hillert, Jacobs, and Mueller (2014) are likely to represent the most realistic indication of how a fund manager following the strategy may have performed over the sample period. This is largely due to the investable universe being the strictest in terms of the size and price cut-offs, which suggests that the small cap effect is not as significant as may be the case in the other studies. However, there are still other factors that obscure the realisable performance of the strategy and limit the applicability to fund managers.

Hillert, Jacobs, and Mueller's (2014) strategy assumes rebalancing of one sixth of all stocks within the portfolio (one of six overlapping portfolios) each month. For the baseline analysis, this is 12% of all stocks within their investable universe at the time of rebalancing. It is not clear how many stocks on average this equates to each month, but given their relatively large sample of stocks for the period (7,815 firms between 1989 and 2010) trading costs could potentially have an appreciable effect on realisable returns. And, as the emphasis of the study was on the determinants of momentum and not necessarily on strategy performance, common portfolio evaluation metrics such as those presented in the more practitioner-focused studies (see Uhl, 2014; Uhl, Pedersen, and Malitius, 2015; Kroujiline et al., 2016) were not presented.

Another potential source of imprecision arises from delistment returns. Shumway (1997) documented an upward delisting return bias in CRSP's NYSE and AMEX data due to missing returns for many stocks delisted as a result of poor performance. Shumway and Warther (1999) further reported that the bias in the CRSP data is 4.7 times larger for NASDAQ stocks than that documented by Shumway (1997). Chan (2003) included only delisted firms for which the dataset included delistment returns, while Fang and Peress (2009) explicitly accounted for delistment bias in their robustness tests. Hillert, Jacobs, and Mueller (2014) did not comment on delistment returns or corrections, so it is unclear what effect it may have had on results. However, as their test sample excluded stocks below certain size break points, and delistment bias in CRSP data is most significant within the smallest NASDAQ stocks (Shumway and Warther, 1999), the effect of delistment bias may have been significantly mitigated. Neither Sinha (2016) or Heston and Sinha (2017) comment on delistment returns or corrections.

Importantly, none of the multi-investment strategies discussed above (i.e. those not trading a single index) have been benchmarked within a well known index such as the S&P 500. As discussed by Vanstone and Hahn (2017), S&P indices offer a clearly defined and highly investable strata with well-understood characteristics that appeal to fund managers. Testing within such indices therefore provides a useful benchmark for portfolio managers who prioritise liquidity and investability. This also minimises, as far as is reasonably practicable, the documented performance that can be attributed to market frictions and eliminates survivorship bias. Since equities within major indices attract a large number of institutional investors (Cao, Han, and Wang, 2017), and institutions tend to trade extremely early in the news cycle (Tetlock, 2011; Hendershott, Livdan, and Schuerhoff, 2015), testing within an index is likely to be a more difficult test for news-based strategies. Tables 2.5 and 2.6 summarise the features of selected news-based trading strategies.

Table 2.5: Summary of Selected News-Based Trading Strategies

Author	Investable Universe	Per-Trade Investment	Testing Period
Wei et al. (2017)	All firms on Taiwan stock exchange	Quintile of quintile	Jan2003 - Dec2012
Heston and Sinha (2017)	All CRSP US common stocks	Top and bottom decile	Jan2003 - Jan2011
Kroujiline et al. (2016)	SPY	SPY	2003 - 2016

Continued on next page

Table 2.5: Summary of Selected News-Based Trading Strategies (continued)

Author	Investable Universe	Per-Trade Investment	Testing Period
Smales (2016)	SPX	SPX	Jan2010 - Dec2012
Sinha (2016)	All CRSP US common stocks	Top and bottom decile	Jan2003 - Jan2011
Uhl et al. (2015)	MSCI WEI, 2-5yr US Government bonds	Either MSCI WEI or Government bonds	Aug2004 - Dec2003
Yang et al. (2015)	SPY	SPY	Jul2012 - Oct2014
Uhl (2014)	DJIA	DJIA	Jan2010 - Dec2010
Hillert et al. (2014)	NYSE, AMEX and NASDAQ common stocks above bottom NYSE size decile and stock price > \$5 at the end of the formation period.	Top (bottom) media coverage decile and top (bottom) return tercile	Jan1989 - Dec2010
Fang and Peress (2009)	All NYSE common stocks and 500 randomly selected NASDAQ stocks, with prices > \$5	Top and bottom media coverage terciles	Jan1993 - Dec2000
Chan (2003)	Random 25% of all CRSP US common stocks, additional analysis only prices > \$5	Top and bottom return tercile of all stocks with news (that month)	Jan1989 - Dec2000

Table 2.6: Summary of Selected News-Based Trading Strategies

Author	Rebalancing Frequency	Holding Period	Performance Metric	Risk Measure
Wei et al. (2017)	1 month	1 month	Average trade return	—
Heston and Sinha (2017)	1 week	13 weeks	Average weekly portfolio return	Momentum- and size-corrected
Kroujiline et al. (2016)	Variable	Variable: averaged 45 and 8 trading days in two tests	Mean yearly return, alpha, Sharpe ratio, Sortino ratio	CAPM, volatility, 5% monthly VaR, maximum drawdown
Smales (2016)	Variable: 15 to 28 time/year in test	1, 5, 10, 60, 250 days	Bootstrapped average trade return	Coefficient of variation
Sinha (2016)	1 week	13 weeks	Average weekly trade return, average weekly portfolio return, total portfolio return	CAPM, FF3, FFC4 regressions
Uhl et al. (2015)	Variable: averaged 8 times/year in test	Variable	Total return, average annual return, Sharpe ratio, tracking error, information ratio	Volatility, maximum drawdown
Yang et al. (2015)	Variable: averaged 20 times/year in test	Variable	Average annual return, Sharpe ratio	Volatility

Continued on next page

Table 2.6: Summary of Selected News-Based Trading Strategies (continued)

Author	Rebalancing Frequency	Holding Period	Performance Metric	Risk Measure
Uhl (2014)	Monthly	1 month	Total return, total excess return, average annual return, Sharpe ratio	—
Hillert et al. (2014)	1 month	6 months	Average monthly return	CAPM, FF3, FFC4, 6F regressions; Size, B/M, turnover, industry and DGTW benchmark adjustment
Fang and Peress (2009)	1 month	1, 3, 6, 9 and 12 months	Average monthly return	CAPM, FF3, FFC4, FFC4+PSL regressions; DGTW benchmark adjustment
Chan (2003)	1 month	1, 3, 6, 9, 12, 24 and 36 months	Average monthly return, average cumulative monthly trade return	CAPM, FF3, FFC4 regressions; size, B/M benchmark adjustments

Note: *FF3* refers to the Fama and French (1993) model, *FFC4* refers to the Carhart (1997) model, *PSL* refers to the Pastor and Stambough (2003) liquidity factor, *DGTW* refers to the original adjustment in Daniel et al. (1997) and *6F* refers to a model augmented with factors for short-term and long-term reversal.

2.4.3 Aggregate Variable Construction

Most approaches to econometric testing in the news analytics literature make use of variables constructed from the information contained within multiple news items. This step is important to consider, as it embeds potentially significant assumptions about the way investors respond to news. Most aggregate variables used in the literature make use of some measure of news tone aggregated over the relevant formation period. If individual items (documents, words or sentences) are scored prior to variable formation, typical measures take the form

$$\begin{aligned} Pos &= \frac{\sum_i^N (I^{\text{pos}} \times S^{\text{pos}})}{N} \\ Neg &= \frac{\sum_i^N (I^{\text{neg}} \times S^{\text{neg}})}{N} \\ Polarity &= Pos - Neg \\ &\text{or} \\ &= Neg - Pos \end{aligned}$$

Where I^{pos} (I^{neg}) is an indicator variable which takes a value of 1 if the item is classified as positive (negative) and zero otherwise, and S^{pos} (S^{neg}) is the positive (negative) score of each document or word.

It is often the case that items are classified categorically as either positive or negative rather than scored, so that the above measures simply reduce to the fraction of positive items, the fraction of negative items, and the difference between them, respectively. For formation periods exceeding one day, these measures are often computed daily before being averaged or summed over the entire formation period. It is also common for each document to be weighted by some measure of relative importance such as story relevance (Shi, Ho, and Liu, 2016), firm market capitalisation (Smales, 2015[b]), or document size (Calomiris and Mamaysky, 2019).

The commonly used measures described above are invariant to any temporal aspects of the news flow over the aggregation period, as information pertaining to the order, dispersion, and recency of the news stories is lost through averaging. This is likely to be inconsequential for the daily formation periods typically studied, but may carry more weight over longer formation periods. In this sense, news aggregation techniques may be thought of in terms of a bias-variance trade-off. Daily measures of news flow tend to be extremely noisy with respect to low-frequency formation periods, temporal aggregation can therefore go a long way in extracting the underlying signal of interest (Uhl, Pedersen, and Malitius, 2015). However, for formation periods spanning a number of months, it would be surprising if the temporal characteristics of the news flow, such as the relative recency of periods of particularly good or bad news exposure, had no bearing on market behaviour.

For certain analyses, vector autoregressive (VAR) models with multiple lags and joint significance tests can go some way to alleviate aggregation bias, in that they reduce the temporal invariance from the entire formation period to the length of each lag and capture the linear dynamics between lags. Another approach is to weight sentiment scores by some decaying function of time when aggregating (Sadik, Date, and Mitra, 2019[b]). Other studies explicitly use changes in temporal characteristics, such as sentiment spikes, momentum, and trend changes, as news events.

The common sentiment or tone measures presented above do not account for the total number of stories being considered in each aggregate, which has been shown to be important for both return, volume and volatility estimates (Ferguson et al., 2015; Mitchell and Mulherin, 1994; Tetlock, 2010). This information (referred to as news volume, attention, or coverage) can be incorporated into analyses a variety of ways, and is usually applied as a separate variable or by weighting the sentiment measures defined above. When measured as a separate variable, the joint effect of news volume with other components of news (such as tone) can be examined explicitly (with the exception of nonparametric models). This is usually performed by including interaction terms in the relevant regressions or by conditioning assets on their relative news volume either before or after conditioning on other news variables.

In the case of incorporating news volume into measures of sentiment, different approaches will provide more or less weight to the item count. Antweiler and Frank (2004) discuss this issue when constructing their measure of 'bullishness' from online message boards and present three alternative specifications. The first takes into account the ratio of positive messages minus the ratio of negative messages, but disregards the total number of messages. This is equivalent to *Polarity* as defined above. Their second measure considers the log of the ratio of positive to negative messages and is approximately equal to *Polarity* scaled by the log of the total number of messages. The third measure is the number of positive messages minus the number of messages. This is equal to *Polarity* scaled by the total number of messages and is more sensitive to the total number of messages than the second measure. Antweiler and Frank found that although their key results could be produced using each of the specifications, the performance of each measure was different.

An additional variable construction practice which has been found to be effective in the small number of studies in which it is applied is the use of residual news scores. This is similar to the more common practice of standardizing the observed news variable with respect to some rolling window (such as previous 12 months), but also accounts for firm specific characteristics that may explain the average news tone and coverage of the firm. This ensures that the news-based variables used for analysis are not proxies for firm size, sector, index membership or other determinants of news exposure.

2.5 Literature Scoping and Bibliometric Analysis

To establish a baseline citation corpus, a systematic search was performed on the Clarivate Analytics Social Sciences Citation Index (SSCI) from 1900 to August 2020 using the Web of Science platform[5]. (Note that all citation counts, unless specified otherwise, are as of August 2020). The SSCI was queried for all peer-reviewed articles with titles, abstracts and keywords fulfilling any one, or combination of, the following search criteria:

> TS = (((sentiment OR tone OR analytics) AND (news OR media)) AND ((stock OR market) AND (volatility OR risk)))
> TS = ((((language OR linguistic) AND (content OR tone OR positive OR negative)) AND (news OR media)) AND ((stock OR market) AND (volatility OR risk)))
> TS = ((((language OR linguistic) AND (content OR tone OR positive OR negative)) AND (news OR media)) AND (“stock market" OR “stock return*" OR “stock price" OR volatility))
> TS = (((sentiment OR tone OR analytics) AND (news OR media)) AND (“stock market" OR “stock return*" OR “stock price" OR volatility))
> TS = ((news OR media) AND (sentiment OR tone OR analytics) AND “trading strategy"))

This returned 542 items, to which an additional 10 items (Leinweber and Sisk, 2011; Chan, 2003; Ferguson et al., 2015; Wüthrich et al., 1998; Sinha, 2016; Mitra, Mitra, and Dibartolomeo, 2009; Kothari, Li, and Short, 2009; Yang, Mo, and Liu, 2015; Tetlock, 2010; Manela and Moreira, 2017) were added from other sources, such as Google Scholar. Before discussing the filtering process used to narrow down the resulting corpus to a subset of items for individual full-text review, it is instructive to briefly examine some aspects of the unrefined sample.

[5]The citation search and review was conducted at two different points in time—the primary search from 1900-2018, coinciding with confirmation of candidature and publication submission, and an update from 2018 to August 2020. Due to minor changes in database export formatting and major changes in the bibliometrix software used, updating certain bibliometric graphs was found to be a non-trivial exercise. As a result, some bibliometric figures are based on the primary citation analysis (late 2017 - early 2018) and are labelled as such.

2.5.1 Literature Growth and Historiographic Analysis

With the increasing pervasiveness of internet platforms, extensive digitisation of text media and rapid improvements in the technologies used to access, store and process unstructured data, text analytics and the subfield of sentiment analysis have experienced increasing growth and attention over the last decade. A simple reflection of this is captured by the search rate of the term 'sentiment analysis' since 2004, illustrated in Figure 2.3.

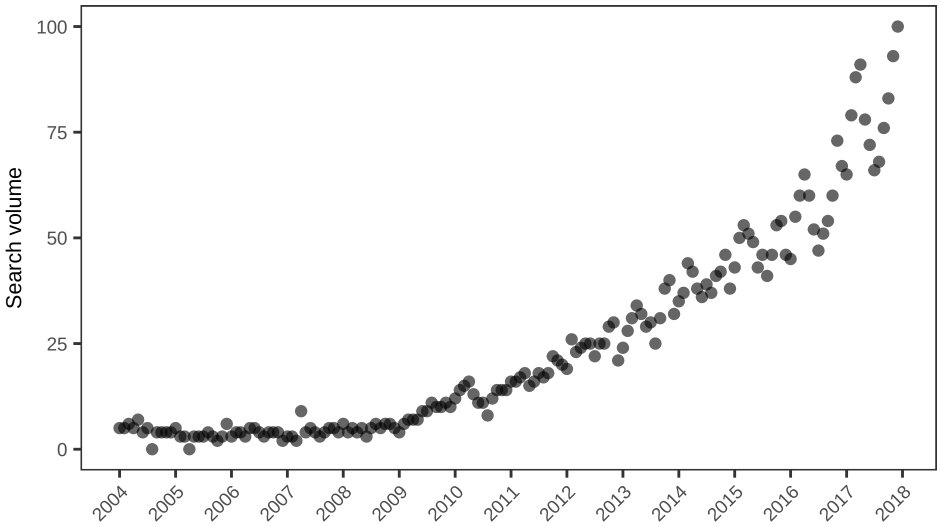

Figure 2.3: Search interest of the term 'sentiment analysis' through from January 2004 to December 2017, relative to the peak popularity of the term over the given date range. Data source: Google Trends (www.google.com/trends).

Interestingly, there has been comparable growth in academic interest regarding the domain-specific use of news analytics and news sentiment in finance. The annual production of peer-reviewed articles returned by the above citation search is shown in Figure 2.4.

While several works exist in the late 90's and early 2000's, it can be seen that academic interest starts to pick up from around 2008-2009 onwards, which is consistent with Tetlock's *Journal of Finance* paper released in 2007 being the first influential work in the field, followed the year after by Tetlock, Saar-Tsechansky, and Macskassy (2008). Starting in approximately 2003, Thomson Reuters undertook a technological overhaul of their news process, resulting in a significant increase in the depth, breadth and meta-data content of their news feed commencing from operational deployment of the system in 2006 onward (Leinweber and Sisk, 2011). TRNA was subsequently released by 2009 and data began to emerge in the academic literature in 2011 with Groß-Klußmann and Hautsch (2011) and Leinweber and Sisk (2011). This coincided with the emergence of other proprietary news analytics data streams such as RavenPack and their appearance in the literature Mitra,

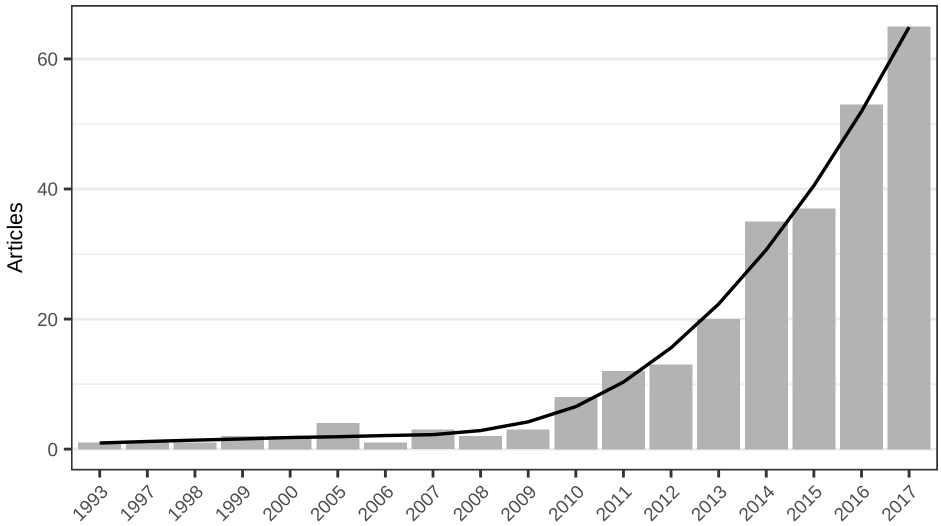

Figure 2.4: Number of peer-reviewed articles in topics related to the study of news analytics in equity markets, from January 1993 to December 2017. Data source: Social Science Citation Index via Web of Science (www.webofknowledge.com).

Mitra, and Dibartolomeo (2009). As demonstrated in Table 2.2, these systems only appear to have grown in popularity with the academic community since their emergence, and form the basis for a large portion of the extant news analytics literature.

Although this database search was designed to capture news-based studies, there is evidently spillover from the UGC literature; not only in the results returned but for interest in the field more generally. Early works using data obtained from Yahoo! Finance message boards (Antweiler and Frank, 2004; Das and Chen, 2007) garnered significant interest in the use of automated content analysis and set the stage for the use of data from more mainstream UGC platforms as they grew in popularity. Twitter, having been launched in July 2006 became one such platform, and reached the finance literature with the work of (Bollen, Mao, and Zeng, 2011) (819 citations at the time of writing). Twitter has now become the primary data source for UGC-based sentiment analytics in finance and is the basis for a number of proprietary data streams such as PsychSignal and MarketPsych.

This brief retrospective assessment is loosely supported by an algorithmic historiographic analysis (Garfield, Pudovkin, and Istomin, 2003; Garfield, 2004). The historiograph illustrated in Figure 2.5 shows a chronological mapping of the 15 most internally-cited articles occurring within the unfiltered database result, the size of each node is proportional to the number of internal citations for that article. The key to the historiograph is shown in Table 2.7, along with local and total citation counts.

The historiograph includes one article based on sentiment extraction from message board postings (Das and Chen, 2007) and one using Twitter (Sprenger et al., 2014), indicating the

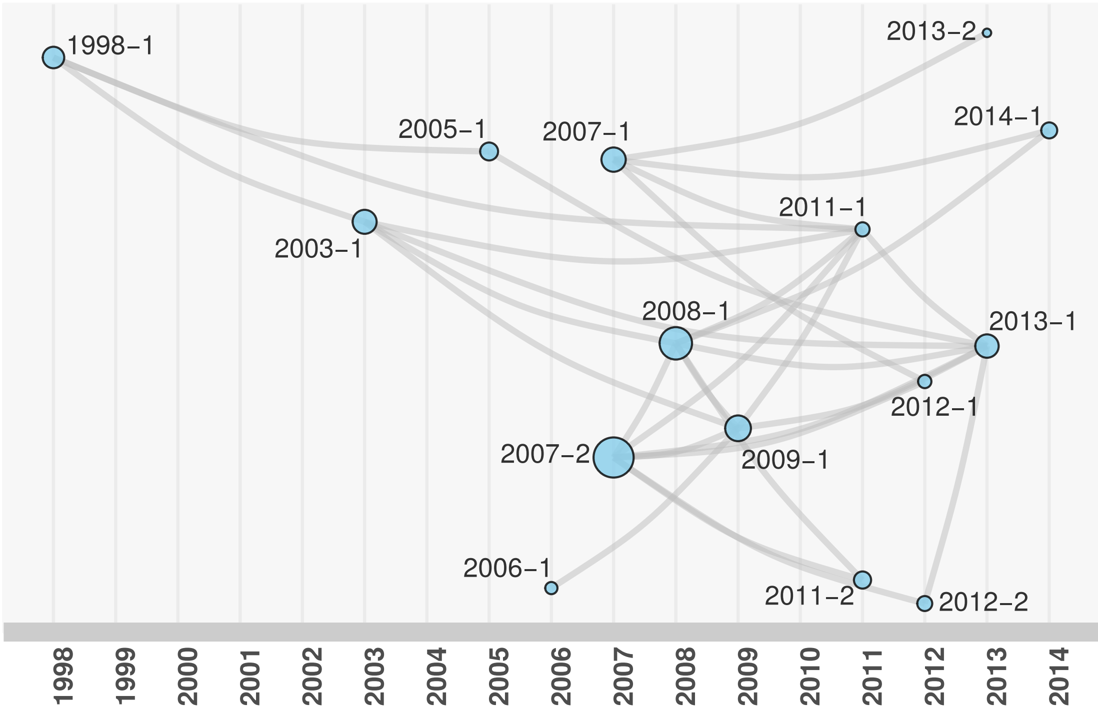

Figure 2.5: Historiograph of the 15 most internally-cited articles in citation corpus. Figure generated in R using the Bibliometrix package (Aria and Cuccurullo, 2017)

Table 2.7: Citation Key for Figure 2.5

key	Label	Year	LCS	GCS
1998-1	Barberis, Shleifer, and Vishny (1998)	1998	62	1521
2003-1	Chan (2003)	2003	77	334
2005-1	Brown and Cliff (2005)	2005	45	365
2006-1	Kumar and Lee (2006)	2006	29	418
2007-1	Das and Chen (2007)	2007	79	418
2007-2	Tetlock (2007)	2007	237	1091
2008-1	Tetlock, Saar-Tsechansky, and Macskassy (2008)	2008	147	651
2009-1	Fang and Peress (2009)	2009	89	476
2011-1	Tetlock (2011)	2011	33	148
2011-2	Groß-Klußmann and Hautsch (2011)	2011	42	107
2012-1	Schumaker, Zhang, et al. (2012)	2012	31	121
2012-2	Dougal et al. (2012)	2012	35	127
2013-1	Garcia (2013)	2013	73	217
2013-2	Yu, Duan, and Cao (2013)	2013	26	167
2014-1	Sprenger et al. (2014)	2014	38	136

spillover of UGC alluded to above. The historiograph also includes two papers (Brown and Cliff, 2005; Kumar and Lee, 2006) from the subset of empirical investor sentiment research that has made use of non-textual proxies for sentiment. This branch of the literature arose only prominently just before the use of textual analysis in finance gained traction, and is itself motivated in large part by theoretical treatments of market efficiency in the context of behavioural finance (such as Barberis, Shleifer, and Vishny, 1998, shown). Although most such citations (correctly) weren't captured by the news-focused search, we can test their influence on the evolution of news sentiment literature by including a few of the major theoretical behavioural finance/noise trader articles (such as those cited in the motivation section of this document) to the chronological mapping of the citation list.

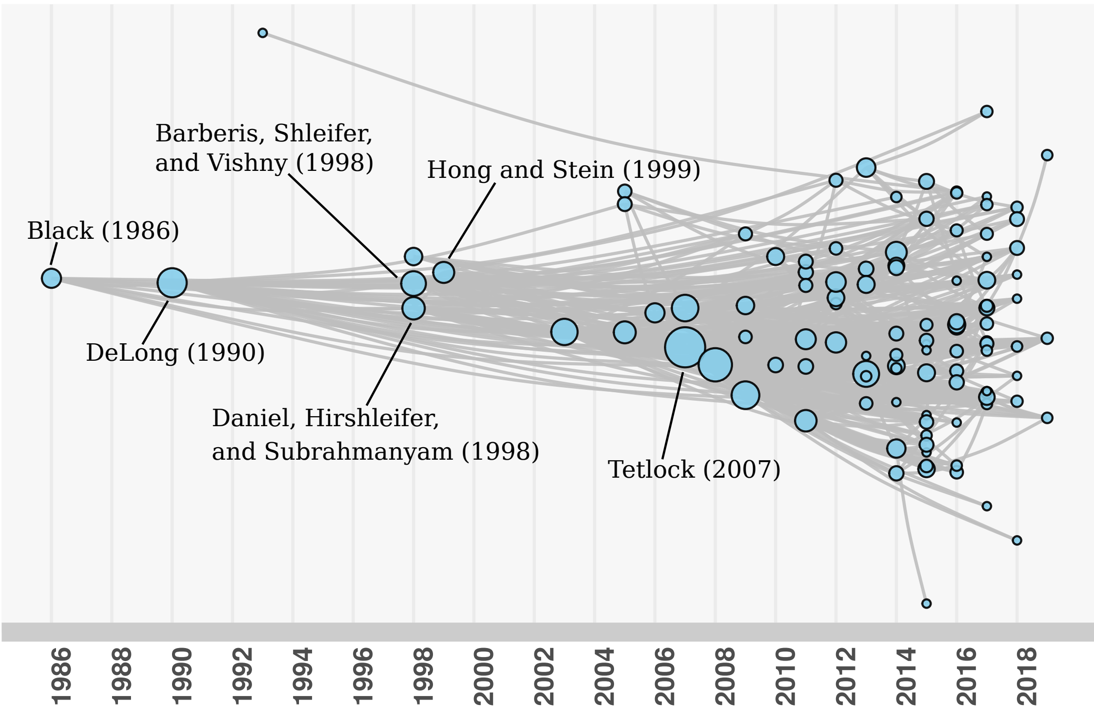

Figure 2.6: Historiograph of the 100 most internally-cited articles in citation corpus with addition of selected behavioural finance/noise trader works. Figure generated in R using the Bibliometrix package (Aria and Cuccurullo, 2017)

Figure 2.6 shows a historiograph of the top 100 internally-cited articles returned from the database search, with the addition of DeLong et al. (1990), Daniel, Hirshleifer, and Subrahmanyam (1998), and Hong and Stein (1999). From Figure 2.6 we can see that the news-sentiment literature has evolved out of these earlier components of the finance literature and was not born entirely from technological development and the 'big data' explosion.

The themes attributed to the growth of the literature also align with the keyword occurrences of the citation list. We calculated the frequency of keywords specified by the authors and the *KeyWords Plus* terms attributed to each study across all items in the citation list. KeyWords Plus are association terms assigned by Web of Science, they are the result of algorithmic subject indexing of citations and reflect the most-commonly recurring words and phrases appearing in the titles of cited papers. Table 2.8 includes the 20 highest occurring keywords and their respective frequencies. Here we can see that social media and Twitter are commonly mentioned themes across the literature, highlighting their role in energising text-based sentiment studies in finance. Similarly, terms relating to predictive analytics, such as machine learning, big data and prediction, also feature regularly within the corpus. The prominence of investor sentiment, behavioural finance, information and noise related topic matter again points to the broader fields of finance from which the news sentiment literature has emerged.

Table 2.8: Keyword frequency across citation search publications

Author Keywords	Frequency	Keywords Plus	Frequency
Investor sentiment	62	Investor sentiment	153
Sentiment	53	News	151
Sentiment analysis	51	Sentiment	120
Social media	48	Returns	113
Stock market	38	Risk	113
News sentiment	32	Information	104
Stock returns	27	Media	98
Twitter	25	Volatility	89
Text mining	21	Market	84
News	20	Cross-section	83
Textual analysis	19	Impact	83
Volatility	18	Prices	56
Analysis	17	Noise	53
Event study	16	Stock returns	52
Big data	15	Model	49
Predictability	14	Information-content	47
Prediction	14	Behavior	43
Stock	14	Attention	41
Machine learning	13	Markets	41
Media coverage	12	Stock	39

Core Features of Citation Corpus

Table 2.9 includes the top ten most cited publications from the corpus with their total citations and citations per year. Unsurprisingly, there is complete overlap with the most internally cited publications included in Figure 2.5. The top ten most relevant journals for the field, in terms of the number of publications in the corpus, are provided in Table

Table 2.9: Most-Cited Articles - Full Sample

Paper	Citations	Citations/Year
Barberis N, 1998, J Financ Econ	1521	66.1
Tetlock Pc, 2007, J Financ	1091	77.9
Tetlock Pc, 2008, J Financ	651	50.1
Fang L, 2009, J Financ	476	39.7
Das Sr, 2007, Manage Sci	418	29.9
Kumar A, 2006, J Financ	418	27.9
Brown Gw, 2005, J Bus	365	22.8
Barberis N, 2005, J Financ Econ	335	20.9
Chan Ws, 2003, J Financ Econ	334	18.6
Kothari Sp, 2009, Account Rev	295	24.6

2.10, while the most productive countries are shown in Figure 2.7 with their associated number of single-country and multiple-country publications. Based on these statistics, Australia is seen to be a surprisingly relevant source of research for the field. This is supported by *Pacific Basin Finance Journal* and *Accounting and Finance* appearing among the most relevant sources.

Table 2.10: Most Relevant Sources - Full Sample

Source	Article
Journal of Behavioral Finance	19
Journal of Banking & Finance	17
Journal of Financial Economics	17
International Review of Financial Analysis	14
Pacific-Basin Finance Journal	13
Decision Support Systems	12
Emerging Markets Finance and Trade	12
Finance Research Letters	12
Physica A-Statistical Mechanics and its Applications	12
Accounting and Finance	9

Figure 2.8 shows a thematic map of the unfiltered citation corpus, based on co-occurrence of WoS keywords (Callon, Courtial, and Laville, 1991; Börner, Chen, and Boyack, 2003). Each theme reflects a keyword cluster derived from application of Louvain's algorithm (Blondel et al., 2008) to the co-occurrence network. Each cluster is positioned by *centrality* and *density* as per Cobo et al. (2011). The centrality of a cluster measures the degree of interaction of a network with other networks and can be understood as the importance of a theme in the development of the entire research field. Degree measures the internal

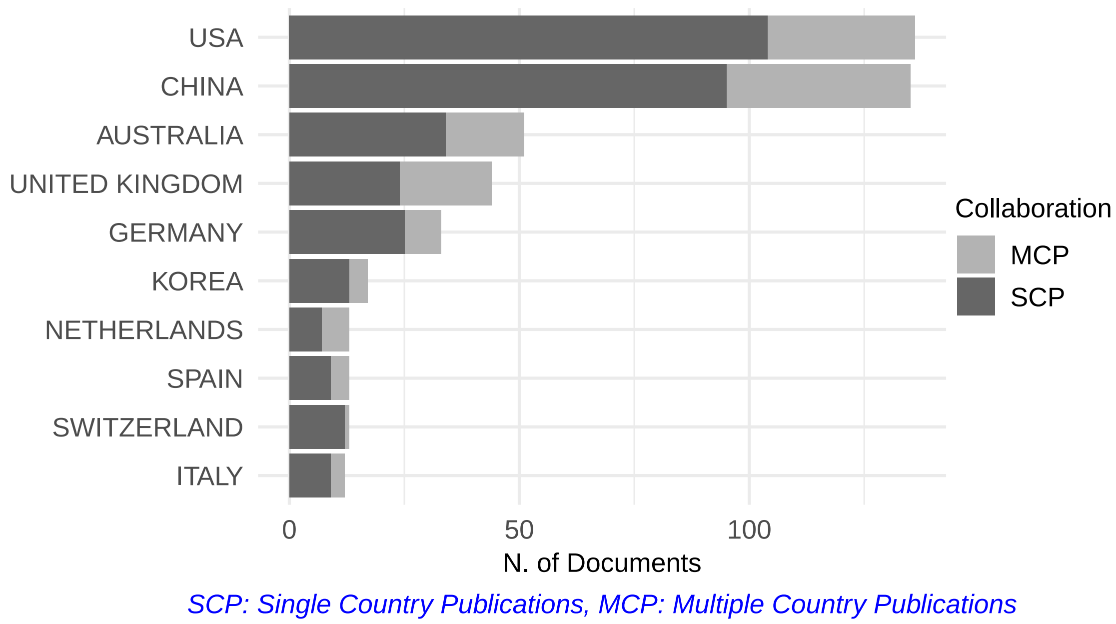

Figure 2.7: Bar chart of the 10 most productive countries in citation corpus by number of publications

strength of the network, and can be understood as a measure of the theme's development.

Representing themes in centrality-density space allows categorisation according to the quadrant in which they are placed (Cahlik, 2000; Coulter, Monarch, and Konda, 1998), to paraphrase Cobo et al. (2011):

- *Motor themes* (upper-right): Well developed and important for the structuring of a research field.
- *Basic and transversal themes* (lower-right): Important for the research field but relatively undeveloped.
- *Emerging or declining themes* (lower-left): Low-density and low-centrality may represent either emerging or disappearing themes.
- *Highly developed and isolated themes* (upper-left): Well developed internal ties but weak external ties, reflecting themes that are specialised and peripheral in character.

We used a simple rule to estimate theme membership of corpus items—if an item has three or more keywords in common with the keyword cluster representing the theme,

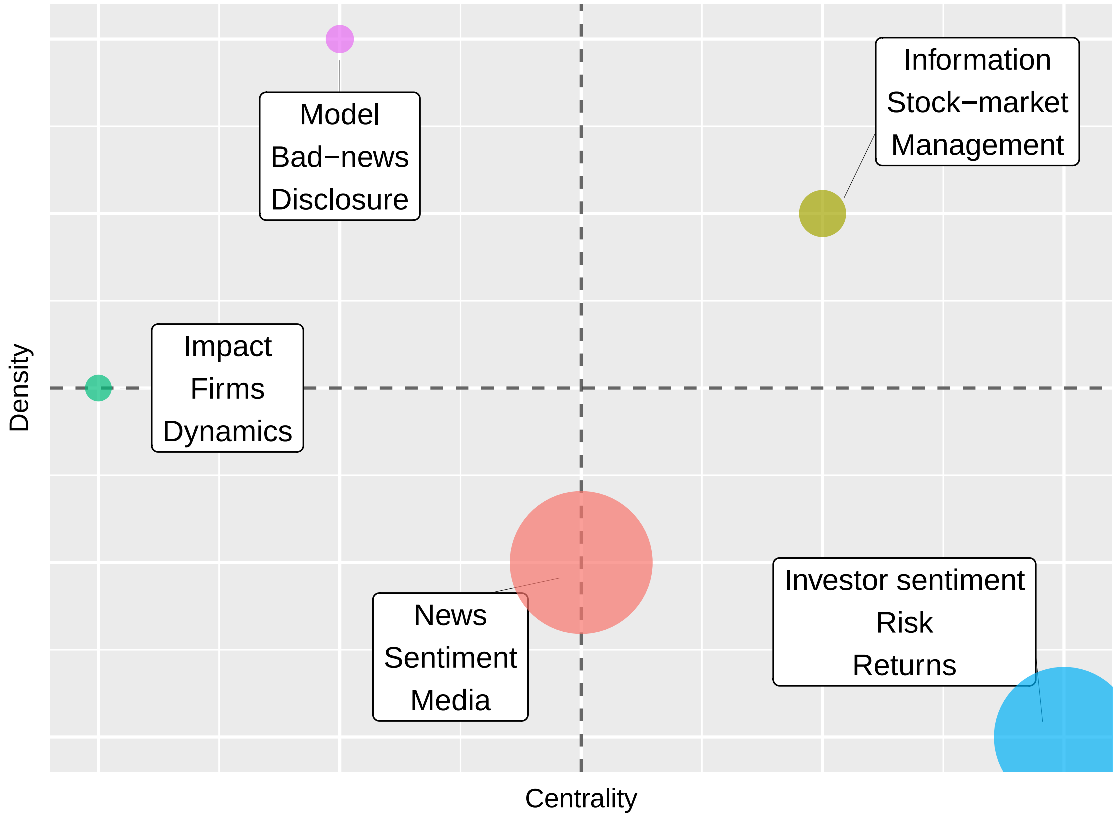

Figure 2.8: Thematic map of citation corpus via WoS KeyWord Plus co-occurrence networks.

we say it is linked to that theme[6]. By looking at the titles and abstracts of the 20 or so most heavily cited papers belonging to each theme and the corresponding keywords used, we were able to get some intuition into the relevant features of the field.

Approximately 45% and 76% of the corpus respectively are associated with clusters one ("Investor sentiment") and two ("News"); the largest themes. Cluster one, the most central and least-dense theme, mainly reflects broad aspects of empirical and behavioural finance, while cluster two, with medium centrality and relatively low density, captures associations relevant to particular sources of information and their processing. Cluster 3 ("Information") has both high centrality and high density, captured here are the survey and review items, specific methods for textual analysis, and papers associated with corporate governance. It is related to 10% of corpus items. Cluster 4 ("Impact") captures studies concerned with the impacts of social media and communication networks not directly associated with stock returns, such as business and branding, sales, product-level performance, and consumer response. 7% of the corpus are associated with this theme. Cluster 5 ("Model"), linked to 7% of the corpus, reflects themes related to vari-

[6]Three was chosen after trying several different values, as it provides a reasonable balance of distinguishing between themes while retaining members within the smallest themes.

ance, volatility forecasting, implied volatility, market risk and investor uncertainty. The top 20 keywords within each them are listed in Table 2.11 with their respective frequencies.

Table 2.11: Keyword Clusters

Cluster				
5. Impact	3. Information	1. Investor Sentiment	4. Model	2. News
Word (freq.)	Word (freq.)	Word (freq.)	Word (freq.)	Word (freq.)
Impact (83)	Information (104)	Investor-Sentiment (153)	Model (49)	News (151)
Firms (14)	Stock-Market (22)	Risk (114)	Bad-News (18)	Sentiment (120)
Dynamics (14)	Management (17)	Returns (113)	Disclosure (18)	Media (98)
Word-of-Mouth (13)	Financial-Markets (14)	Market (85)	News-Sentiment (13)	Volatility (89)
Reviews (9)	Media-Coverage (12)	Cross-Section (83)	Time-Series (12)	Noise (53)
Sales (9)	Language (10)	Prices (56)	Announcements (12)	Information-Content (47)
Decision (9)	Trading-Volume (10)	Stock-Returns (52)	Heteroskedasticity (9)	Behavior (42)
Event (8)	Policy (9)	Markets (41)	Variance (9)	Attention (41)
Communication (7)	Contagion (8)	Liquidity (33)	Gold (8)	Stock (39)
Online (7)	US (8)	Performance (32)	Overconfidence (8)	Talk (35)

This table includes the top 10 most frequently occurring keywords for each of the thematic clusters shown in Figure 2.8. Cluster labels correspond to the top keyword (first line in Figure 2.8) of each cluster.

2.5.2 Filtering and Article Selection

To remove irrelevant items and establish a refined corpus subject to individual written review, the 552 initial items were filtered by removing those that:

1. Did not make use of measures derived from financial news content.

2. Did not examine forward market movements at time horizons of at least one day.

3. Did not empirically examine movement of market-based variables.
4. Only examined market responses surrounding earnings announcements.
5. Studied only a particular news event (case studies) or events (e.g. airline crashes) or a particular type of stock (e.g. Islamic stocks).

As discussed in Section 2.3.1, the different categories of unstructured text sources used in the finance literature to extract measures of sentiment and other text-based analytics differ from one another in terms of information content, authorship, availability, versatility and audience. As such, there is no reason to believe that findings relating to one particular text source are generalisable to the others. As the focus of the current work is the use of financial news analytics in the context of fund-style investment decisions, this review is only concerned with those studies making use of financial news text, either by itself or in combination with other sources; hence the first selection rule above. A similar line of reasoning was also taken for the different temporal horizons used for analysis; intraday responses to news items and the strategies used to exploit them, do not confidently tell us anything about the behaviour of low-frequency dynamics operating over months and years, if they exist at all. While this may also be true for studies at within-week horizons, many of the seminal works and stylised facts regarding financial news sentiment emanate from intraweek analyses. The temporal cut-off was therefore taken to be one trading day; hence the second selection rule above.

Two exceptions were made to rule number one, for publications in which content-derived measures were not used–Chan (2003) and Fang and Peress (2009)–as these studies have been particularly influential in the low-frequency literature. Note that our database query was designed to capture studies relating to the use of news media analytics in equity markets, opposed to commodity or FX markets. While we have included non-equity publications that pass the filtering criteria, our publication corpus is unlikely to be comprehensive with respect to such studies. These filtering criteria excluded a total of 234 items, leaving 75 items, exceptions included, for review.

2.5.3 Descriptive Statistics of Review Items

Annual Production

Figure 2.9 displays the number of review items published by year. In line with the scientific production of the unfiltered corpus, the majority of review items were published in the last five years (2015 - 2020). Figure 2.10 illustrates the annual productivity of the review corpus by the number of total citations attributed to items published in each year.

Despite the increase in publications in later years, the most significant publication years remain 2007-2009.

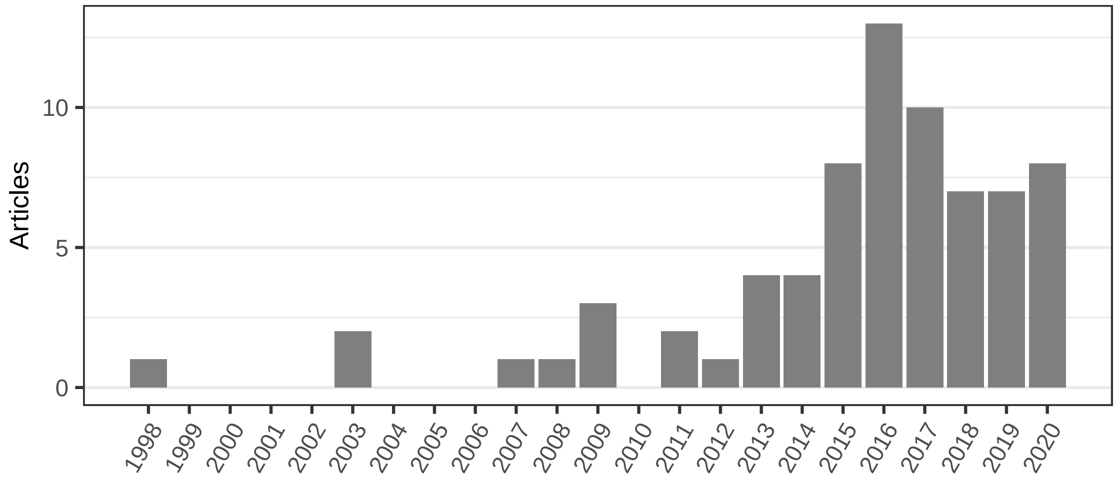

Figure 2.9: Number of review items published, by year.

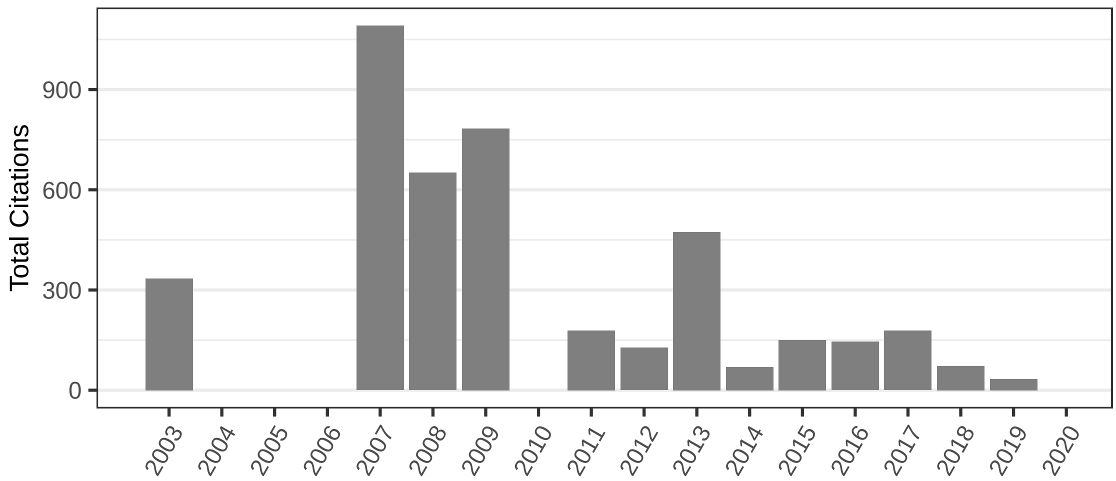

Figure 2.10: Number of total citations of review items, by year

Historiograph

The historiograph illustrated in Figure 2.11 shows a chronological mapping of the ten most internally-cited review items, the size of each node is again proportional to the number of internal citations for that article. The key to the historiograph is shown in Table 2.12, along with local and total citation counts.

From the historiograph we can see that a number of the most prominent works in the raw database have been retained, but the spillover from UGC-based studies, behavioural finance, and non-textual investor sentiment has been removed. The earliest work of the

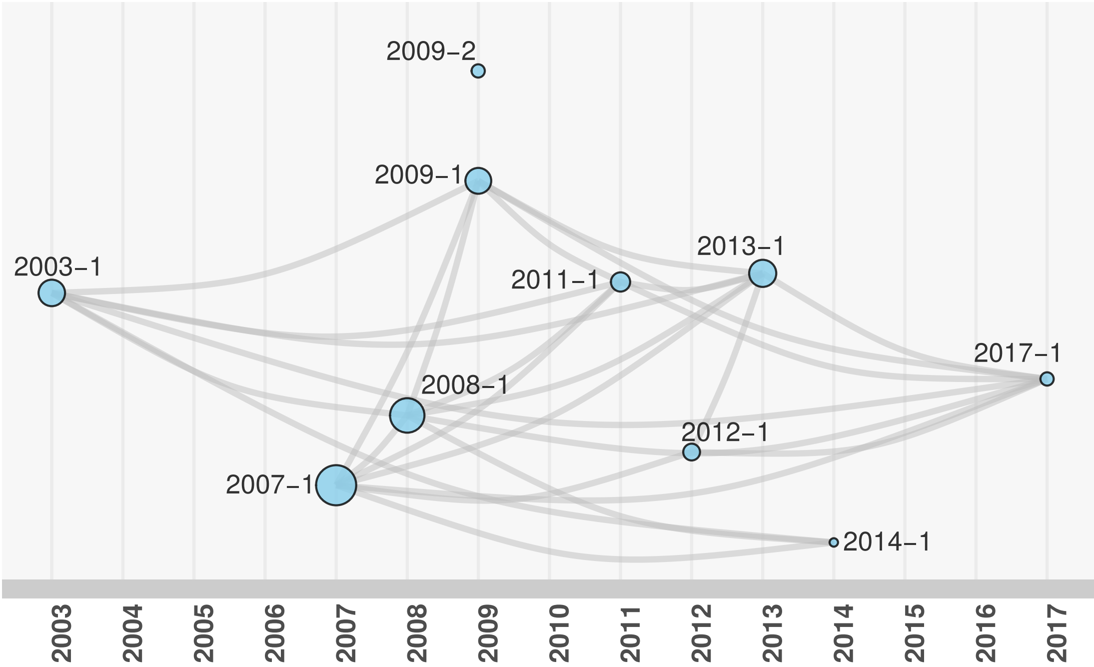

Figure 2.11: Historiograph of the ten most internally-cited articles selected for review. Figure generated in R using the Bibliometrix package (Aria and Cuccurullo, 2017).

Table 2.12: Citation Key for Figure 2.11

key	Label	Year	LCS	GCS
2003-1	Chan (2003)	2003	19	334
2007-1	Tetlock (2007)	2007	48	1091
2008-1	Tetlock, Saar-Tsechansky, and Macskassy (2008)	2008	34	651
2009-1	Fang and Peress (2009)	2009	18	476
2009-2	Kothari, Li, and Short (2009)	2009	6	295
2011-1	Tetlock (2011)	2011	10	148
2012-1	Dougal et al. (2012)	2012	8	127
2013-1	Garcia (2013)	2013	20	217
2014-1	Uhl (2014)	2014	5	30
2017-1	Heston and Sinha (2017)	2017	6	29

review items with significant impact is Chan's (2003) study of news headlines and momentum. Chan's analysis did not make use of textual sentiment but the incidence of news – the sign of the news article was assumed from the sign of the return over the look back period. It was still not until Tetlock (2007) that textual sentiment was introduced into the wider finance literature. Tetlock's (2007) paper most directly influenced Tetlock, Saar-Tsechansky, and Macskassy (2008), Dougal et al. (2012), and Garcia (2013), although it is referenced by all subsequent historiograph items with the exception of Kothari, Li, and Short (2009), whose paper does not reference any other review items. As the citation linkages suggest, Tetlock (2011) is an amalgamation and extension of the major themes within (Chan, 2003; Tetlock, 2007; Tetlock, Saar-Tsechansky, and Macskassy, 2008), in that it is concerned with investor *overreaction* to *uninformative* news content at the *firm-level*. Fang and Peress (2009) descends from the same three articles, but is most thematically relevant to (Chan, 2003) in its focus on news volume (rather than content) and methodology. Uhl (2014) was one of the earlier papers to aggregate firm-level TRNA sentiment to forecast market returns, while Heston and Sinha (2017), extending the theme of Chan's (2003) study, apply TRNA at the firm-level with conditioning on weekly price momentum. Each of these papers, along with the rest of the review items, are covered in more detail in Section 2.6 and Tables 2.1 through 2.16.

Core Publications

Table 2.13: Most-Cited Articles - Filtered Sample

Paper	Citations	Citations/Year
Tetlock Pc, 2007, J Financ	1091	77.9
Tetlock Pc, 2008, J Financ	651	50.1
Fang L, 2009, J Financ	476	39.7
Chan Ws, 2003, J Financ Econ	334	18.6
Kothari Sp, 2009, Account Rev	295	24.6
Garcia D, 2013, J Financ	217	27.1
Yu Y, 2013, Decis Support Syst	167	20.9
Tetlock Pc, 2011, Rev Financ Stud	148	14.8
Dougal C, 2012, Rev Financ Stud	127	14.1
Hendershott T, 2015, J Financ Econ	68	11.3

Table 2.13 includes the ten most-cited items within the review corpus. There is a 50% overlap with the unfiltered corpus, with the less relevant items being replaced. This overlap between the raw and filtered corpus suggests that the initial citation parameters were well calibrated for the current work, while the high overlap between Tables 2.13 (total citations) and 2.12 (local citations) points to the thematic consistency among the review items being appropriately directed.

Most Productive Countries

Figure 2.12 illustrates the top ten most productive countries for the review corpus, in terms of number of publications produced, including single-country and multiple country publications. Here, Australia and USA have the highest number of publications, followed by China. Figure 2.13 depicts the corresponding country collaboration network,

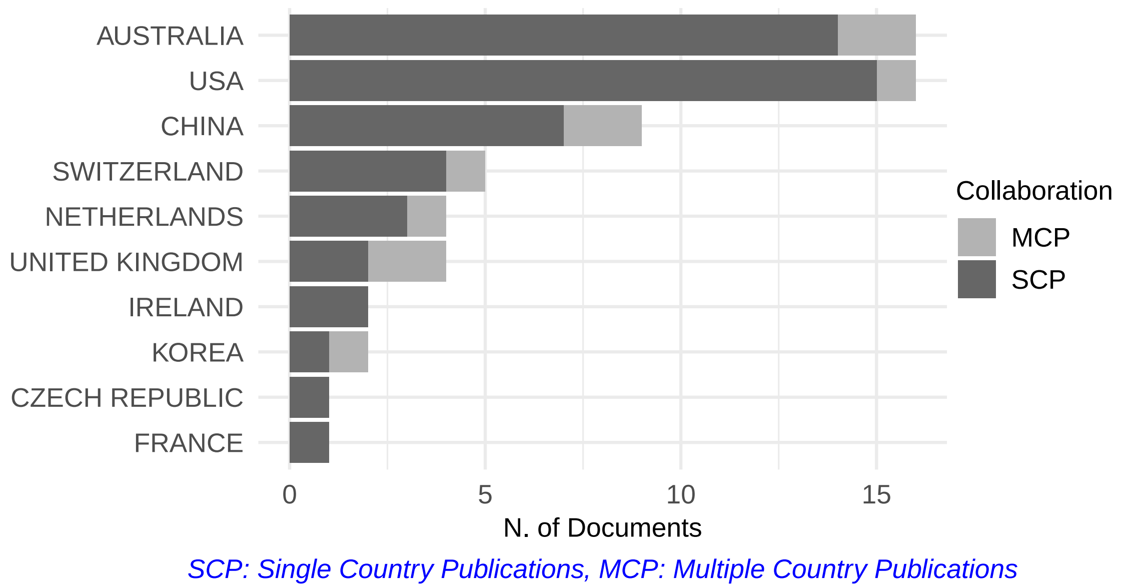

Figure 2.12: Bar chart of the 10 most relevant countries in review corpus by number of publications

in which edges represent collaborations and node size is proportional to the number of publications.

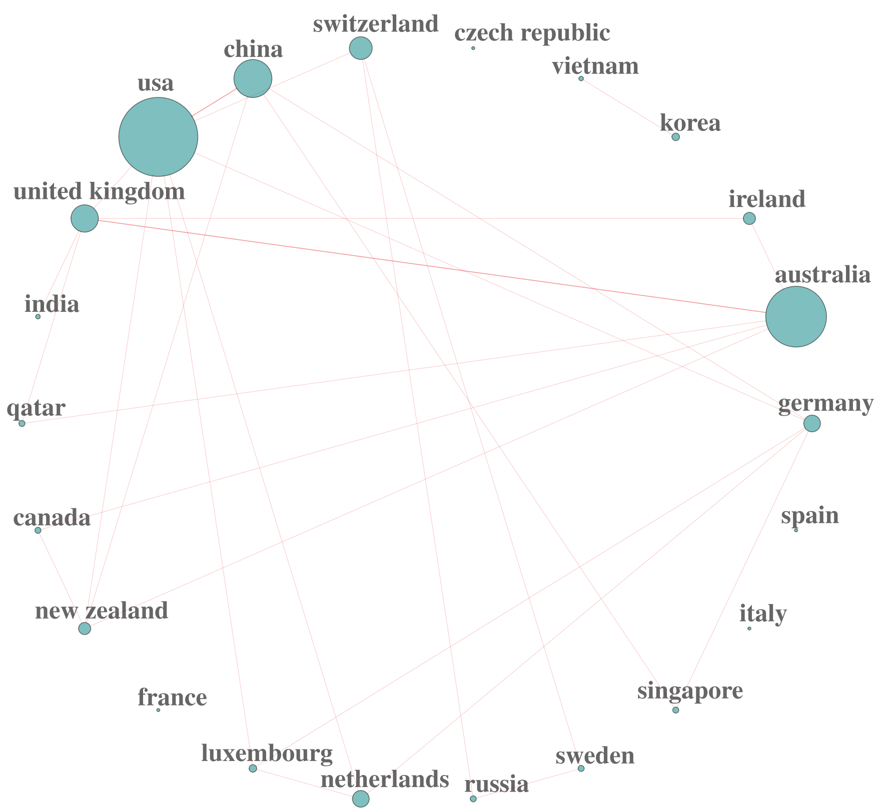

Figure 2.13: Country collaboration network of review publications.

Most Productive Authors

Figure 2.14 includes a timeline of the most productive authors, by number of publications, within the review corpus. Each point in the chart represents a publication year for a given author, the size and transparency of which reflect the number of publications and the citation rate of those publications, respectively. L.A. Smales (Australia) is the most prolific author with a total of six publications between 2014 and 2016 within the sample. P.C. Tetlock has produced three very highly cited publications in the field but has not been active in this area since 2011. G. Mitra's contributions span the longest period among these authors, from 2009 to 2020.

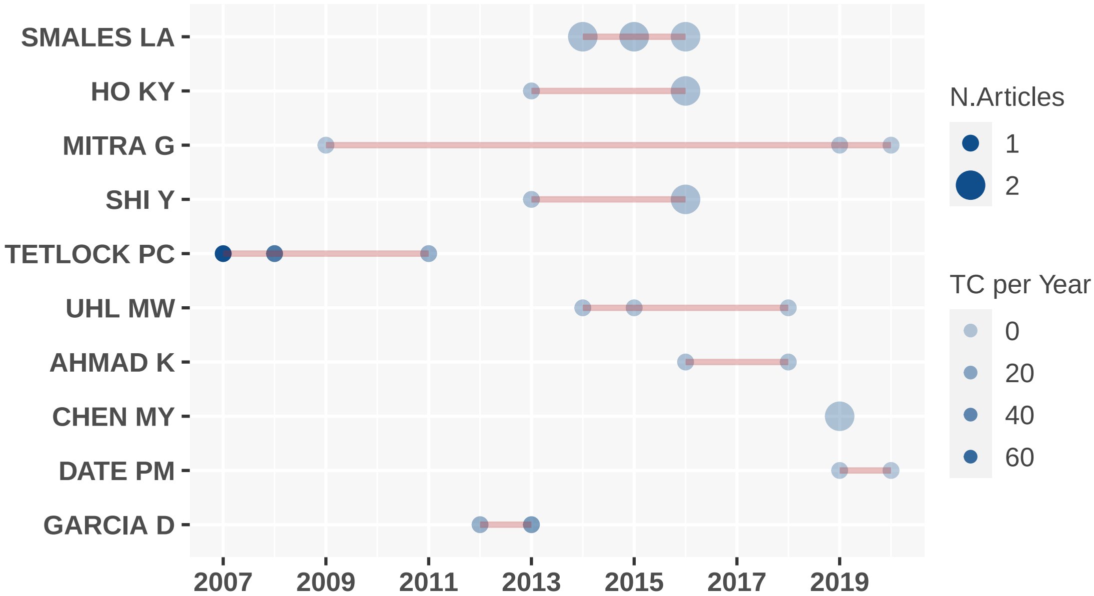

Figure 2.14: Publication timeline of the ten most productive authors in review corpus.

Core Journals of Publications

The 75 review items were recorded by 50 different journals. Of these, 36 have only one publication in the review corpus. Table 2.14 lists the top 10 journals of the review items ranked by total citation count.

Keywords of Publications

The selection of review articles contained a total of 22 unique author keywords, of which 34 occurred more than once, and a total of 179 KeyWord Plus terms, of which 61 occurred more than once. Other terms to capture the themes within the corpus were extracted manually from the titles and abstracts of the publications. Whitespace was used to define the boundaries of words occurring in the title and abstract of each publication and

Table 2.14: Core Journals of Review Publications

Journal	Articles	Citations
Journal of Finance	4	2435
Journal of Financial Economics	4	478
Accounting Review	1	295
Decision Support Systems	1	167
Journal of Behavioral Finance	4	50
Journal of Portfolio Management	2	50
Journal of Marketing Research	1	49
Pacific-Basin Finance Journal	3	46
North American Journal of Economics and Finance	2	45
Quantitative Finance	4	33

stop words and numbers were removed. Table 2.15 lists the top 20 author keywords, KeyWords Plus, and manually extracted bigrams, ranked by the number of publications they were associated with. Bigrams from the title and abstracts as, upon inspection, many of the most common unigrams were parts of a two-word phrase in the original text.

Table 2.15: Keyword frequency of review publications

Author Keywords	Freq.	Keywords Plus	Freq.	TI/AB	Freq.
News sentiment	14	Investor sentiment	20	News sentiment	24
Sentiment	8	Returns	18	Stock market	24
News	6	Market	15	Stock returns	14
Textual analysis	6	Sentiment	15	Financial news	8
Investor sentiment	5	Media	14	Financial markets	7
Sentiment analysis	5	Risk	14	News articles	7
Stock market	5	Cross-section	13	Positive negative	7
Stock returns	5	Impact	13	Predictive power	7
Social media	4	Model	12	Implied volatility	6
Trading strategy	4	News	12	Market returns	6
G14	3	Prices	10	Sentiment analysis	6
Garch	3	Volatility	10	Social media	6
Implied volatility	3	Noise	9	Stock prices	6
News analytics	3	Information	8	Asset prices	5
Predictability	3	Talk	7	Dow jones	5
Vix	3	Information-content	6	Media coverage	5
Analysis	2	Attention	5	Predict stock	5
Behavioral finance	2	Behavior	5	Price movements	5
Efficiency	2	Earnings	5	Returns volatility	5
Financial news	2	Heteroskedasticity	5	Stock price	5

Note: G14 is the Journal of Economic Literature (JEL) topic code for "Information and Market Efficiency", "Event Studies", and "Insider Trading".

2.6 Empirical Findings

2.6.1 Grouping of Empirical Findings

The purpose of this literature review is to determine the current state of research on the use of news analytics for low-frequency investment decisions and the economic relevance of the news data on the outcomes of these decisions. More specifically, we are interested in the cross-sectional impacts of different firm-specific news environments and the degree to which these can be exploited in the context of systematic portfolio management. As portfolio management is fundamentally about balancing risk against performance, the relationship between news information and expected return and risk is fundamental to the current work.

A first-level grouping was applied on the basis of the investment or estimation horizon analysed within each publication, as this feature largely governs the relevance of a given finding to the current work. Generally, any study in which the aggregation period of news information was beyond one week was considered low-frequency while those observing the market response to single items or single days worth of news were considered high-frequency. Most publications within the review corpus have tended to focus on either expected return or risk, rather than both, and this feature was used as a second-level grouping in which to review the literature and discuss findings. The findings of all review items are summarised in Table 2.16.

2.6.2 High-frequency Findings

News and Expected Return

One of the earliest publications documenting automated analysis of news content for the purpose of stock market prediction was that of Wüthrich et al. (1998). The authors showed that statistical classifiers based on the occurrence of predefined phrases in online news articles could be used to predict the direction of next-day market returns for a number of international indices. Nuij et al. (2014) found that trading signals derived from the automated identification of news events at the firm level provide more predictive power than common technical indicators.

Unlike earlier empirical analyses such as Niederhoffer (1971) and Wüthrich et al. (1998), which sought to identify specific information events in news items and link them to returns, Tetlock (2007) examined the non-informational influence of news media. Tetlock (2007) hypothesized that news content could be used to measure investor sentiment. Tetlock used the frequency of negatively associated words in occurring in the popular *Wall Street Journal* (WSJ) column *Abreast of the market* (AM) to construct a daily measure of linguistic media tone, and tested whether this measure was related to the behaviour of the DJIA through a simple vector autoregression (VAR) framework. Consistent with the interpretation of media tone as a reflection of investor sentiment, Tetlock found that high values of media pessimism were associated with lower same-day and next-day stock returns, followed by a reversion to original prices within one week.

Dougal et al. (2012) built upon the conclusions of Tetlock (2007) by showing that the AM column influenced short-run DJIA returns through market-exogenous content. Dougal et al. (2012) demonstrated that accounting for the exogenous rotation of AM columnists with different writing styles explained more than an additional 35% of next-day DJIA returns, compared to that explained by common control measures. Further evidence in support of Tetlock's (2007) hypothesis are provided by Garcia (2013), who studied the positive and negative word content of two NYT columns from 1905 to 2005. Garcia similarly found that media tone leads next-day DJIA returns, and the effect is partially, but significantly, reversed over the following week. He also showed that the strength of the relationship between media tone and market returns is time varying and is most pronounced during economic hardship. Narayan and Bannigidadmath (2017) provided some evidence suggesting that Garcia's (2013) NYT news tone data predicts stock returns across a range of international market indices.

A related analysis was conducted by Hanna, Turner, and Walker (2020), who studied influence of FT news commentary on the UK blue-chip market from 1899 to 2010. They found that the tone of the news commentary led next-day market returns throughout the entire sample, but influenced trading volume only during bull markets.

Johnman, Vanstone, and Gepp (2018) examined the influence of news tone from the business section of The Guardian on the FTSE 100 index from 2000 to 2016, and found that news tone had no significant impact on market return. Kelly and Ahmad (2018) closely followed Tetlock's 2007 analysis and used sentiment derived from the WSJ AM column and the Financial Times Lex column to predict next-day DJIA return using a VAR model. Pyo and Kim (2019) found evidence of return predictability of the Korean stock market (KOSPI) using sentiment derived from market-level news items.

In contrast to the market commentary studied by Tetlock (2007), Dougal et al. (2012), and Garcia (2013), firm-specific news contains qualitative descriptions of value-relevant events, corporate announcements, and economic releases. In one of the first studies to investigate news content at the firm level, Tetlock, Saar-Tsechansky, and Macskassy's (2008) analysis of S&P 500 firms showed that the frequency of negative words embedded within firm-level news provides significant explanatory power for future earnings above and beyond more recent analysts' forecasts and market price information. Tetlock, Saar-Tsechansky, and Macskassy (2008) demonstrated that the negative word content of newswires is negatively related to next-day abnormal returns. Leinweber and Sisk (2011) confirmed a price response to firm-specific news tone qualitatively similar to that documented by Tetlock, Saar-Tsechansky, and Macskassy (2008) among S&P 1500 constituents, and additionally found that post-news drift is more pronounced for novel news items with extreme tone. Zhang, Härdle, et al. (2016) studied the sentiment of online message boards, blogs and news items related to 100 S&P 500 constituents and found that positive online sentiment is positively related to next-day returns, while negative sentiment is related to returns with a lag of two days.

A number of studies have also confirmed such firm-level news effects in international markets. Following the results of Tetlock, Saar-Tsechansky, and Macskassy (2008), Ferguson et al. (2015) examined the separate and combined effects of the tone and volume of firm-specific news on stock returns among FTSE100 constituents. They found that positive (negative) news tone positively (negatively) predicts next-day return, and that the volume of news amplified these effects. Zhang, Song, et al. (2016) found that the returns of Chinese stocks responded differently to news tone when mentioned in two different online financial news columns. Khuu, Durand, and Smales (2016) showed that firm-level news tone is positively contemporaneously related to stock returns on the Tokyo Stock Exchange and that the strength of the relationship tends to decrease with firm size. Wang, Chen, and Wei (2015) found that daily new coverage, news tone and stock returns in the Taiwanese market, were dynamically related at the company level for up to five days. Using a sample period from 1989-2010, Hillert, Jacobs, and Mueller (2018) found that increased journalist disagreement of firm-specific news articles is negatively related to the cross-section of US stock returns. Minh et al. (2018) and Ahmed, Sriram, and Singh (2020) used used firm-level news sentiment with advanced machine learning methods to predict price movements of four and two US stocks, respectively. Both studies found predictive power in the sentiment-based inputs.

Chan and Chong (2017) found that sentiment of online news articles and blogs were associated with daily closing prices of the Hang Seng Index up to five days in advance. Li, Xie, et al. (2014) used the sentiment of firm-specific news items to predict the daily return direction of Hang Seng Index constituents using SVM, and found that word sentiment increased prediction accuracy compared to using word frequency alone. Zhang, Zhang, et al. (2018) and Zhou et al. (2020) demonstrate short-term price predictability of Chinese stocks using multiple data types, including news content. Chen, Liao, and Hsieh (2019) found next-day return predictability among four Taiwan 50 Index constituents using news sentiment. In a 2015-2018 study of ASX 20 Index constituents, Vanstone, Gepp, and Harris (2019) provide evidence that news sentiment enhances forecasts of stock price, but not direction, relative to price information alone.

In an analysis of 20 large non-financial US stocks, Ahmad et al. (2016) showed that the relationship between news tone and firm-level stock returns passes through periods of significance and non-significance, and that each episode of significance can be characterised by either a transitory (short-term reversal) or lasting response of returns to news tone. Using aggregate measures of industry-specific news tone, Smales (2015[b]) showed that the news-stock return relationship varies through time, across industries, is related to industry beta, and, consistent with the findings of Garcia (2013), is responsive to wider investor fear. Hillert, Jacobs, and Mueller (2018) found that the influence of aggregate firm-level journalist disagreement on US market returns is strengthened during recessions, supporting Garcia's 2013 findings that used market-level commentary. Gan et al. (2020) showed that the influence of aggregate news sentiment on S&P 500 returns declined throughout their 2011-2017 sample period, while the influence of UGC increased.

Together, the findings of Tetlock (2007) and Tetlock, Saar-Tsechansky, and Macskassy (2008) demonstrated that simple and generalizable measures of news content can capture both market-wide sentiment effects and information about firm fundamentals. Following these studies a number of researchers have aggregated firm-level news data to produce market-level measures of news tone which appear to be largely uncorrelated with other proxies of investor sentiment (Chan, Durand, et al., 2017; Cahill, Wee, and Yang, 2017). Allen, McAleer, and Singh (2017) studied the daily interaction between DJIA return and aggregate news sentiment derived from the TRNA database. They found that aggregate news sentiment is positively associated with daily market returns, and that this relationship was strongest during the GFC. Yang, Song, et al. (2015) found that shocks and trends in the aggregate tone of US financial news items are positively associated with S&P 500 returns for the following 4 and 10 days respectively. In a trading-focused analysis similar to Nuij et al. (2014), Yang, Mo, Liu, and Kirilenko (2017) found that indicators based on the feedback between news and social media sentiment are able to predict S&P 500 returns more effectively than common technical indicators. Gan et al. (2020) and Griffith, Najand, and Shen (2020) used TRMI aggregate sentiment indices in a VAR framework to predict S$P 500 index returns, with both studies verifying some degree of predictive efficacy of sentiment.

Counter-evidence for the news tone-stock price relationship is provided by Strauß, Vliegenthart, and Verhoeven (2016). They found no consistent relationship between next-day opening prices and either the tone or volume of news across their sample of 21 high media coverage firms listed on the Amsterdam exchange. Hendershott, Livdan, and Schuerhoff (2015) analysed the market response surrounding the arrival of news items for 1667 NYSE firms and found that institutional trading predicts the occurrence, tone and news-day return of firm-level news stories. This is consistent with the event studies of Tetlock, Saar-Tsechansky, and Macskassy (2008) and Leinweber and Sisk (2011) which show that the majority of the abnormal returns surrounding news occur in the days prior to the news item being released. Hendershott, Livdan, and Schuerhoff's (2015) results are also consistent with the findings of Tetlock (2011), who showed that overreactions in response to stale news are driven by individual investors rather than institutions. Evidence of the endogeneity of news is also provided by Gan et al. (2020), who found that market variables tend to exert stronger impact on sentiment indices than the other way around. In their analysis of over 1000 stocks on US exchanges, Coqueret (2020) conclude that news sentiment does not carry meaningful economic value for investors at the stock level and further observe that stock returns lead sentiment. These findings were stated to also be reflected in unreported results of sentiment-driven factor portfolios.

News and Risk

Studies examining the empirical relationship between news content and measures of risk typically focus on either return variance, idiosyncratic volatility or implied volatility indices. Bianconi, Hua, and Tan (2015) showed that Garcia's (2013) media pessimism is dynamically related to the VIX over one to two day horizons and that the VIX appears to better predict the systematic risk (tail covariance) of large US financial stocks than media pessimism. In their analysis of 100 S&P 500 firms, Zhang, Härdle, et al. (2016) found that negative news and UGC sentiment was positively related to next-day trading volume and volatility and positive sentiment was negatively related to next-day volatility.

Unlike other analyses in the field which have investigated the influence of news on investor fear using aggregate implied volatility (i.e. VIX), Uhl (2018) broke down investors' response to news sentiment using individual S&P 500 index options across varying levels of moneyness and time-to-maturity. He found that as positive (negative) sentiment increases, implied volatility decreases (increases). This effect was shown to decrease with maturity of the option, and was more pronounced for OTM put options than call options.

In their study of the Korean stock market, Pyo and Kim (2019) found that the sentiment of news headlines is negatively associated with the one-day-ahead level of the volatility index.

Borovkova and Mahakena (2015) and Smales (2015[a]) studied the tone of macroeconomic news and its effect on volatility dynamics in the natural gas and gold futures markets, respectively. Borovkova and Mahakena (2015) found that incorporating news sentiment information into one-day-ahead volatility estimates significantly improves the accuracy of the forecasts. Smales (2015[a]) revealed a strong asymmetry in the impact of news, with negative news stories demonstrating a larger impact on both returns and volatility compared to positive messages. He also found that the news-volatility relationship strengthened during periods of economic stress.

Yu, Duan, and Cao (2013) provided evidence that news sentiment and the interaction of UGC and news sentiment to be statistically significant predictors of next-day idiosyncratic volatility across a sample of 824 US firms. Mitra, Mitra, and Dibartolomeo (2009) demonstrated how firms' recent news sentiment history can be incorporated into volatility factor models to provide next-day volatility estimates that may respond more rapidly to market developments than similar model specifications without news. Nooijen and Broda (2016) found that the inclusion of news sentiment into volatility estimates resulted in only modest improvements in the accuracy of VaR forecasts for MSCI US Equity Sector indices. In-sample fits indicated that news sentiment was still a stronger predictor of volatility relative to social media sentiment.

Johnman, Vanstone, and Gepp (2018) showed that sentiment of The Guardian business is predictive of the volatility, but not direction, of next day return of the FTSE 100 index. An in-sample trading strategy exploiting this relationship was shown to outperform buy-and-hold. In their analysis of UK and European stock indices, Sadik, Date, and Mitra (2019[b]) showed that incorporation of news sentiment into the GARCH model leads to superior prediction of volatility than simple GARCH and EGARCH models. This result was also demonstrated at the stock-level.

Myskova, Hajek, and Olej (2018) used meta-learning models with sentiment variables to forecast return volatility of 14 large US stocks. They found that article length was a more useful feature in the predictive models than the story sentiment. Griffith, Najand, and Shen's 2020 VAR analysis showed that negative news sentiment strongly predicts conditional market volatility of the S&P 500 index, even when controlling for VIX. Al-Maadid et al. (2020) found that negative business news from larger GCC economies have domestic and cross-boarder impact on stock return and volatility. Whereas only weak evidence of causality running from political news to market return could be demonstrated.

2.6.3 Low-Frequency Findings

Expected Return

Whereas the intraweek price effects discussed in the studies above were typically assessed in response to news metrics measured on a per-story or daily basis, a number of researchers have found low-frequency pricing effects at the firm and market level by aggregating news information over formation periods spanning weeks and months. These studies have found that when measured at lower-frequencies, news information can explain future returns at horizons of months and even years. Of these, some have demonstrated that there exists inefficiencies to news at discrete lags while others have observed what appear to be slowly building under- and over-reactions to news flow.

Uhl (2014) was among the first to test the impact of news sentiment on stock returns at timeframes comparable to those used in the investor sentiment literature. He showed that aggregate news sentiment is incorporated into DJIA returns with a lag of up to three months, and that conditioning on lagged sentiment improves the accuracy of next-month market return forecasts. Wei et al. (2017) followed an analysis broadly similar to Uhl

(2014) in the Taiwan market and found that weekly and monthly news sentiment only influenced future market returns between particular levels of sentiment. At the firm level, Wu and Lin (2017) found that monthly abnormal returns in the Taiwanese stock market are positively related to the volume and tone of news articles released that month. Xiong and Bharadwaj (2013) found that the impact of positive monthly news flow on the abnormal stock return of large US companies are amplified by firms' advertising spending.

Uhl, Pedersen, and Malitius (2015) showed that trend changes in the weekly level of aggregate macro-economic and firm-level news tone forecasts changes in global equities, and that a signal based on global news tone momentum is an effective means of timing global equity asset allocations. Using a news-driven, analytic, agent-based market model, Kroujiline et al. (2016) showed that investors operating on time scales of less than one day cause prices to rapidly adjust to news releases, whereas nonlinear feedback dynamics resulting from traders with longer investment horizons results in market inefficiency to news at longer time scales.

Kraussl and Mirgorodskaya (2017) examined the individual and joint significance of lagged negative news tone on future stock returns of the S&P 500 and MSCI World Indices. Their results suggest that negative news is incorporated into market prices gradually over one-and-a-half years followed by an overreaction and subsequent reversal at two-and-a-half years. Ammann, Frey, and Verhofen (2014) demonstrated the capacity of words occurring in the Handelsblatt, a leading German financial newspaper, to predict Deutscher Aktienindex (DAX) returns at one-month and 3-month horizons. Sub-period analysis revealed that the significance of certain words varied through time, as did the ratio of those which were positively or negatively related to return.

Calomiris and Mamaysky (2019) analysed 51 developed and emerging equity markets over a sample from 1996-2015. Using topic-specific measures of news sentiment and entropy (novelty), they found news content forecasts market-level one-year ahead returns in sample and out-of-sample, and that the predictive content of sentiment, frequency, and entropy is time-varying and topic specific. They also found the predictive capacity of news content to be stronger in emerging than developed markets.

Narayan (2019) studied the impact of oil price news on 45 global equity indices. They found mixed evidence that oil price news provides marginal predictive power on monthly market returns, relative to oil price growth rate.

Coqueret (2020) analysed the impact of news sentiment on over 1000 US stocks, for prediction horizons from 4 to 20 days. They found no evidence that news sentiment has predictive utility for returns at the stock level.

A repeated result in the low-frequency literature is an observed relationship between news coverage, past return and stock price continuation or 'momentum'. Although neither analysis makes use of measures derived from news *content*, two important studies in this area are those of Chan (2003) and Fang and Peress (2009). These studies provide robust evidence in support of incorporating news data into low-frequency trading strategies and together they highlight the importance of considering both news flow and market data in forecasts.

Chan (2003) showed that firms with extreme past return in the previous month tend to exhibit stock price momentum for up to 12 months if accompanied by news during the formation period, whereas those without news tend to reverse in the following month. Chan (2003) suggested that removing no-news stocks from a standard momentum strategy (as in Jegadeesh and Titman, 1993) results in 40%-50% higher returns. Fang and Peress (2009) revealed that unconditionally, no-coverage stocks outperform stocks with above-average coverage (by 0.39% per month in their sample). This effect was shown to be approximately three to four times larger among firms with low market capitalisation and analyst coverage, high individual ownership and high idiosyncratic volatility. The effect was also shown to be more stable for longer portfolio formation (6 months) and holding (12 months) periods. Consistent with Chan (2003), the no-coverage premium was not found for stocks with high current return.

Building on the findings of Chan (2003) and Fang and Peress (2009), Hillert, Jacobs, and Mueller (2014) examined the influence of news coverage and content on stock price momentum in portfolios more closely aligned with the traditional 6/1/1 specification. They showed that momentum profits are most pronounced in high coverage stocks with concordant news tone, and that this media-driven momentum effect is punctuated by larger long-term reversals.

Motivated by Gutierrez and Kelley's (2008) weekly momentum, Sinha (2016), Heston and Sinha (2017), and Huynh and Smith (2017), studied price continuations arising from weekly news and stock return events. Sinha (2016) showed that a portfolio that buys (sells) the top (bottom) decile of stocks ranked on weekly average news tone and holds for 13 weeks generates excess return. Consistent with the findings of Hillert, Jacobs, and Mueller (2014), Sinha also found that stocks with extreme past return and concordant weekly news tone exhibited significantly stronger price continuation than those with discordant news tone. Heston and Sinha (2017) analysed the same news tone portfolio as Sinha (2016). They found that when controlling for a news effect (neutral news), the portfolio is driven primarily by negative news. They also found that most of the delayed reaction to news occurs around earnings announcements. The portfolio was shown to be highly positively correlated with a 6-month momentum strategy (80%) and earnings surprise strategy.

Huynh and Smith (2017) provide evidence that the 52-week momentum effect of weekly return is driven by high attention (or low novelty) positive news. They found that the relationship between news tone on momentum is highly conditional on the attention of the news. A trading strategy that buys one-week winners with high-attention positive news, sells one-week losers with low-attention negative news, and holds for 52 weeks, is shown to be highly profitable over the sample period.

News and Risk

A relatively small number of studies have tested for low-frequency relationships between news sentiment and volatility. This may be a result of the motivating theory, the MDH, having been originally proposed as a means of explaining the non-normality of daily returns, and thus written in terms of intraday price discovery. Another contributing factor may be that the prominent behavioural models of investor over- and under-reaction, which provided the theoretical motivation for much of the low-frequency return literature, do not make strong claims regarding variance. Whatever the case, it is not implausible when considering the longer term impacts of news on expected returns that analogous relationships hold for dispersion too. Empirically, the existing findings are mixed, but point toward news tone being a useful predictor of future volatility at longer horizons, even within simple forecasting models. Studies examining the relationship between news and implied volatility indices have also revealed a relationship between news and market estimates of forward volatility.

Cahan et al. (2017) found that the monthly return volatility and liquidity volatility of NYSE- and AMEX-listed stocks is negatively related to the tone intensity, and positively related to the volume, of news stories released about the firm in the previous month. They provided evidence that investor response to news is mitigated by firm accounting quality. Kothari, Li, and Short (2009) showed that positive (negative) news coverage is negatively (positively) associated with return volatility and analyst forecast dispersion over the contemporaneous quarter. In examining the effect of monthly media pessimism on return volatility, Kraussl and Mirgorodskaya (2017) demonstrated a statistically significant Granger causal effect of media pessimism on MSCI world volatility for lag subsets 1-12 and 12-24 months, and on S&P 500 volatility for the lag subset of 12-24 months. VAR analysis revealed that the direction of this effect was positive (increased media pessimism is followed by an increase in market volatility).

Shi, Ho, and Liu (2016) showed that including the number of positive and negative news items received in a given month results in significantly different forecasts of next-month idiosyncratic volatility, such that it confounds the relationship between estimated idiosyncratic volatility and abnormal return. Monthly volatility was also found to increase (decrease) with the amount of contemporaneous negative (positive) news. Seng and Yang (2017) showed that the number of positive and negative news items attributed to firms listed on the Taiwan Stock Exchange is correlated with return volatility at various horizons. They also provided some evidence that positive and negative news coverage improves volatility forecasts at horizons from one month to one year.

Smales (2016) showed that aggregate news sentiment was significantly contemporaneously related to changes in the VIX at five-minute, hourly and daily frequencies, with positive (negative) news associated with a decrease (increase) in VIX. He also found daily news sentiment to be negatively related to changes in VIX at one- and two-day lags. Wei et al. (2017) found that weekly and monthly aggregate news sentiment only influenced the Taiwanese implied volatility index (TVIX) between particular levels of sentiment. Yang, Song, et al. (2015) showed that the occurrence of an aggregate news sentiment shock is negatively correlated to VIX for the following 10 trading days (with the exception of day 5), while the occurrence of a sentiment trend is negatively correlated with VIX for the following 11 trading days.

In their analysis of 51 developed and emerging equity markets,Calomiris and Mamaysky (2019) showed that topic-specific measures of news sentiment and entropy are significant predictors of volatility and drawdowns over a one-year forecast horizon. These results were demonstrated in both in-sample and out-of-sample tests. Glasserman and Mamaysky (2019) analysed the top 50 global banks, insurance, and real estate firms by USD market capitalization as well as the S&P 500 index and VIX. They found measures of news sentiment and novelty forecast one-month volatility at the company and market level, with the predictive power of news being stronger at the market level. They further demonstrated that news shocks are impounded into volatility over the course of several months.

Using VAR impulse response functions over 20 days, Gan et al. (2020) find that although market volatility and media sentiment mutually cause each other, the evidence suggests that the feedback effects from market volatility on news on sentiment is the dominant direction. Audrino, Sigrist, and Ballinari (2020) examined the addition of news sentiment and attention variables in HAR volatility models in the US market at the stock and index level for prediction horizons spanning one day to one month. They found that the addition of the sentiment and attention variables improved the fit of the volatility models, but that sentiment variables only have short-term effects on volatility, with little evidence for predictive power beyond 2 days.

2.6.4 Summary of Review Items

Table 2.16: Summary of survey items

Author	Assets	Timeframe	Findings
Wüthrich et al. (1998)	DJIA, Nlkkei 225, FT100, Hang Seng, Singapore Straits	Dec 1997-Mar 1998	The occurrence of predefined words and phrases in news headlines can help predict direction of next-day market return through statistical learning techniques. A probabilistic rule-based classifier was found to outperform rival techniques.
Chan (2003)	Random 25/% of CRSP stocks	1980-2000	Stocks with extreme monthly return accompanied by news exhibit stock price momentum for the following 12 months, while those without news tend to reverse the subsequent month; the effect is focused among small stocks with negative return; excluding no-news stocks from a standard momentum portfolio can significantly improve performance.
Tetlock (2007)	DJIA	1980-2008	High values of media pessimism are associated with lower same-day and next-day stock returns, followed by a reversion to fundamentals within one week; unusually high or low values of media pessimism lead to temporarily high market trading volume; media pessimism weakly predicts market volatility; market responses to media pessimism are unlikely to be due to information content.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Tetlock, Saar-Tsechansky, and Macskassy (2008)	S&P 500 constituents	1980-2004	The fraction of negative words in firm-specific news stories explains future earnings above and beyond stock analysts' forecasts and accounting data; stock prices respond to negative words in news items with a one day delay; negative words in news stories about earnings are more predictive of earnings and return than negative words in other news stories.
Mitra, Mitra, and Dibartolomeo (2009)	Selected constituents of EUROSTOXX 50 and Dow Jones 30	17-23 Jan 2008; 18-24 Sep 2008	Incorporation of news sentiment into volatility models can increase the responsiveness of volatility estimates to market developments.
Kothari, Li, and Short (2009)	889 US listed firms	1996-2001	Positive and negative news media impacts cost of capital, return volatility and analyst forecast dispersion; negative disclosures from management lead to higher return volatility and analyst forecast dispersion; analysts' reports and positive disclosures from management appear to be heavily discounted by the market.
Fang and Peress (2009)	All firms listed on NYSE + 500 randomly selected firms listed on NASDAQ	1993-2002	Stocks with no media coverage earn higher returns than stocks with high media coverage after controlling for well-known risk factors including the probability of informed trading; a profitable zero-cost portfolio can be constructed by going long no-coverage stocks and short high-coverage stocks; the media coverage effect is largest among small, low-priced stocks

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Tetlock (2010)	NYSE, AMEX and NASDAQ common stocks	1979-2007	Ten-day return reversals are significantly lower on news days; ten-day volume-induced momentum exists only on news days for small and illiquid stocks; news stories with the most informative characteristics are the most powerful predictors of lower return reversal; results consistent with model in which news resolves asymmetrically held information.
Leinweber and Sisk (2011)	S&P 1500 consituents	2004-2008; Jan-Sep 2010	Stock prices respond to the tone of firm-specific news with some delay; post-news drift is more pronounced for novel news items with extreme tone; post announce drift varies by firm market capitalisation and industry; going long (short) stocks with extremely positive (negative) news can form the basis of a profitable trading strategy.
Tetlock (2011)	10,187 US firms	1996-2008	Stock prices respond to stale news less than novel news; one week return reversal increases with the staleness of the news; the impact of news staleness on return reversal is much larger in stocks with above-median individual ownership; individual investors trade more aggressively on stale news.
Dougal et al. (2012)	DJIA	1970-2007	WSJ columnists exhibit measurable differences in writing style; unconditional and conditional columnist fixed effects have significant explanatory power for next-day market returns.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Garcia (2013)	DJIA	1905-2005	Media tone helps predict market returns at the daily frequency, and this predictability is greater during recessions; the effect of media tone on next-day return partially reverses over the following four days; news tone leads to an increase in trading volume.
Xiong and Bharadwaj (2013)	141 large US firms	2004-2010	Positive (negative) news sentiment is associated with higher (lower) abnormal returns at a monthly frequency. Firm advertising amplifies the impact of positive news on stock price but does not affect the impact of negative news on stock price.
Yu, Duan, and Cao (2013)	824 randomly selected US firms	Jul-Sep 2011	Social media appears to have a stronger effect on stock price than news media; social media and news media have strong interaction effects on stock price; both social media and news media appear to be more strongly related to stock price volatility than abnormal return.
Hillert, Jacobs, and Mueller (2014)	All common stocks traded on NYSE, AMEX or NASDAQ subject to price and size constraints	1989-2010	Stock price momentum increases with firm media coverage; the stock price momentum of high-coverage firms is stronger for firms with concordant news tone; the momentum of high-coverage firms reverses in the long run; the momentum of high-coverage firms appears to be the result of investor overreaction.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Uhl (2014)	DJIA	2003-2010	Changes in aggregate news tone can predict and explain changes in monthly stock market returns; negative news tone has greater predictive capacity than positive news tone; aggregate news tone explains and predicts changes in stock market returns more effectively than market trading volume and certain macroeconomic factors; a trading strategy based on market forecasts which incorporate aggregate news tone can significantly outperform the market.
Ammann, Frey, and Verhofen (2014)	DAX	1989-2011	Word count indices of financial news can help predict monthly stock market returns; the predictive capacity of specific words contained in news varies through time; the overall prediction accuracy of word count indices varies through time and has increased since 2000; clustering of word count indices increases the forecast accuracy of estimates relative to those made with raw word counts.
Li, Xie, et al. (2014)	22 Hang Seng Index constituents	2003-2008	In predicting the direction of daily stock movements using SVM: the use of sentiment dictionaries outperformed using term frequencies alone, using multiple dimensions (word categories) of dictionaries outperformed using only positive and negative categories, LM slightly outperformed H-IV4.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Nuij et al. (2014)	FTSE 350 constituents	Jan-Apr 2007	Event-based variables extracted from news items are often included in optimal trading rules selected among technical indicators by genetic algorithms based on strategy performance; authors present framework for automatically incorporating news into stock trading strategies.
Smales (2015[b])	12 US industry indices	2004-2010	The strength of the relationship between industry-specific news tone and stock returns varies through time and by industry and is related to the market risk premium; during periods of high investor fear, the relationship between industry-specific news and stock returns is weak.
Hendershott, Livdan, and Schuerhoff (2015)	1667 firms listed on NYSE stock exchange	2003-2005	Institutional order volume predicts the occurrence of news stories; institutional order flow predicts the tone and stock market reaction of next-day news items; Institutional order flow predicts event-day returns for unexpected, value-destroying news events; institutions do not appear to be responsible for overreactions following hype news.
Ferguson et al. (2015)	FTSE 100 constituents	1981-2010	Positive and negative news tone predict next-day stock returns; news volume is a stronger predictor of next-day stock return than news tone; the predictive power of news tone is stronger among smaller firms and those with high book-to-market ratios; a trading strategy based on firms' average daily news tone generates positive excess return over the sample period.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Uhl, Pedersen, and Malitius (2015)	MSCI World	2004-2013	A tactical asset allocation strategy driven by news sentiment momentum out-performed common benchmarks in terms of total returns, volatility and maximum drawdown. Macroeconomic and company specific news hold approximately the same predictive power at global equities level, both combined performs best.
Wang, Chen, and Wei (2015)	TSEC-listed (Taiwanese) companies	2001-2012	Daily news coverage and tone are related to future daily stock returns for up to five days; news coverage and amount are autocorrelated; daily news coverage is related to past stock return for up to five days.
Yang, Song, et al. (2015)	SPY	2012-2014	Shocks and trends in aggregate news tone predict stock market return and implied volatility over a number of days; news tone trends are stronger predictors of stock market movements than news tone shocks; a trading strategy based on the occurrence of news tone shocks and trends outperforms the market over the sample period.
Smales (2015[a])	Gold futures	2003-2012	The tone of macroeconomic news is significantly related to the volatility of gold futures; the volatility response of gold futures to news is much stronger for negative news than positive news; the volatility response to news increased dramatically during the GFC recessionary period.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Bianconi, Hua, and Tan (2015)	14 large US financial firms	1992-2006	Media pessimism impacts financial firms' tail risk through tail covariance of other firms' risk (externality); VIX has a direct effect on systematic risk and via other firms' financial distress; media pessimism and VIX exhibit dynamic feedback at one to two day horizons; VIX appears to have greater predictive capacity than media pessimism when forecasting systematic risk.
Borovkova and Mahakena (2015)	Natural gas futures	2006-2010	News tone Granger-causes price jumps; news tone is Granger-caused by volatility (especially downside semi-volatility); arrival of news in non-trading periods impacts overnight returns; including news tone in volatility models significantly improves the accuracy of volatility forecasts.
Ahmad et al. (2016)	20 large non-financial US firms	2001-2010	The prediction strength of negative news tone on stock returns at the firm level passes through periods of significance and non-significance; each epsiode of significance may be characterised by transitory (short-term reversal) or lasting responses to news tone; market response to newswires appears to be more permanent than that of other types of news; price response to news is most pronounced at the one-day horizon.
Zhang, Härdle, et al. (2016)	100 S&P constituents	2009-2014	Positive and negative news tone impacts return, volume, and volatility at one to two day horizons; the strength of the relationship between news tone and market response varies by sector and firm media coverage; the market response to news tone is asymmetric.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Strauß, Vliegenthart, and Verhoeven (2016)	21 firms listed on Amsterdam exchange	2002-2013	Neither the tone or volume of news stories stories demonstrate consistent effects on next-day opening prices; news tone appears to be influences by changes in opening prices rather than the other way around.
Zhang, Song, et al. (2016)	242 Chinese stocks	Nov 2013-Jan 2014	Market responses to internet news differ depending on the type of the news (market commentary versus firm-recommendations); firms mentioned in internet news experience abnormal return and excess trading volume the day of the news.
Khuu, Durand, and Smales (2016)	Firms traded on Tokyo stock exchange	2003-2012	Negative news sentiment is associated with negative returns at the firm and market level; the effect of news sentiment is largest among smaller firms; news sentiment for Japan-listed firms was predominantly negative over the sample period.
Sinha (2016)	US common stocks	2003-2010	Weekly news sentiment predicts stock returns for 13 weeks; drift following weekly news sentiment is more pronounced for negative news; a portfolio long (short) stocks with highly positive (negative) news tone generates positive abnormal returns.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Kroujiline et al. (2016)	SPY	2003-2016	A news-driven, analytic, agent-based market model with heterogenous investors is able to capture stock market price dynamics at various horizons; the stock market exhibits efficiency to news at intraday timescales but becomes inefficient at longer horizons; a trading strategy based on the model's forecasts, using measured news tone as an input, generates significant abnormal returns.
Nooijen and Broda (2016)	MSCI US Equity Sector Indices	1998-2013	News and social media tone more strongly predict volatility than returns; news tone is a better predictor of volatility than social media tone; stock price sensitivity to news tone varies by sector; news and social media tone appear to have a larger impact on prices during stressed (high volatility) markets.
Smales (2016)	S&P 500 Index	2000-2010	Changes in VIX are significantly negatively related to changes in news sentiment; the VIX-tone relationship is stronger during periods of high volatility; current and lagged news tone is significantly positively related to stock returns and significantly negatively related to VIX; VIX appears better at forecasting future returns thatn news tone.
Shi, Ho, and Liu (2016)	NYSE, AMEX and NASDAQ common stocks	2000-2011	The idiosyncratic volatility puzzle (negative risk premium) appears to be due to the confounding relationship between news (volume and tone), abnormal return and expected volatility; stock return is negatively (positively) related to the volume of negative (positive) news; idiosyncratic volatility is positvely (negatively) related to the volume of negative (positive) news.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Allen, McAleer, and Singh (2017)	DJIA	2007-2012	Aggregate news tone is positively related to market returns; the news tone-return relationship was significantly stronger during the GFC; the entropy of news tone is greater than the entropy of market returns.
Yang, Mo, Liu, and Kirilenko (2017)	SPY	Aug 2012-Jan 2015	Sentiment feedback indicators extracted from news and social media are recognised as optimal trading rules (among common technical indicators) by genetic algorithm selection based on strategy performance; sentiment feedback indicators may help reduce the maximum drawdown of algortihmic trading strategies.
Kraussl and Mirgorod-skaya (2017)	MSCI World and S&P 500 Indices	1990-2012	Media pessism Granger-causes global market returns for 12 to 24 in advance and volatility for 1-24 months in advance; media pessimism is associated with positive (negative) market returns 24 to 25 (14 to 17) months in advance and positive market volatility 1 to 20 months in advance.
Wei et al. (2017)	All firms on Taiwan Stock Exchange	2003-2012	Aggregate news tone appears to be related to market behaviour only within a particular range of news tone. News tone is related to stock return at the firm level only for a subset of stocks with particular characteristics.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Heston and Sinha (2017)	US common stocks	2003-2010	Weekly news tone predicts stock returns for 13 weeks, whereas daily news tone predicts stock returns for only one to two days; drift following weekly news tone is concentrated on negative news; drift following weekly news tone is slightly more pronounced for firms with multiple news items; a portfolio long (short) stocks with highly positive (negative) news tone generates positive abnormal returns and is highly correlated with 26-week momentum and earnings surpise decile spreads.
Song, Liu, and Yang (2017)	512 heavily traded US stocks	2003-2014	News tone, including news shock and trend indicators, can be incorporated into relative performance forecasts to produce reliable rankings of the best and worst performing stocks at one week horizons; these news-informed rankings can be incorporated into trading strategies that generate positive excess returns over the sample period.
Huynh and Smith (2017)	Intersection of TRNA and TRTH database	2003-2011	Market underreaction to positive news is a driver of weekly momentum returns (as in Gutierrez and Kelly, 2008); underreaction is greater for stories with high journalist attention; a trading strategy formed on the basis of news tone and past return is no longer profitable after high-attention news is controlled for; a strategy buying (selling) winners (losers) with high attention (low attention) positive (negative) news is highly profitable in the US and 21 developed markets.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Cahan et al. (2017)	2683 NYSE- and AMEX-listed stocks	2003-2011	Contemporaneous and lagged news tone is negatively associated with return volatility over monthly horizons; positive and negative news tone reduces liquidity volatility; the liquidity volatility and news tone relationship is conditional on the firm's accounting quality.
Seng and Yang (2017)	Companies listed on Taiwanese stock exchange excluding F-stocks an TDRs	2013-2014	News tone is associated with future stock return volatility at monthly, quarterly, semmiannual and annual horizons.
Chan and Chong (2017)	Hang Seng Index	Sep 2014-Mar 2015	Tone extracted from online news articles and blogs is associated with market returns at three- and five-day horizons. SentiWordNet outperforms H-IV4 dictionary in sentiment classification for a range of models. Grammar-based models outperform word-based models in sentiment classification.
Manela and Moreira (2017)	S&P 500 Index	1890-2009	Front page macroeconomic and political news captures investor uncertainty. Time variation in this uncertainty drives expected returns at horizons of 6-24 months. News implied volatility is associated with disaster regime transition probability.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Narayan and Bannigidad-math (2017)	18 Dow Jones stock indices	1996-2012	News tone predicts stock returns across Asia-Pacific, Emerging Markets, Japan, Canada and UK, but not USA. News tone more accurately predicts islamic stocks than conventional stocks.
Wu and Lin (2017)	Taiwanese stocks (TSE)	2001-2014	Foreign institutional investment is associated with the quantity and content of news. News tone is contemporaneously associated with abnormal return.
Hillert, Jacobs, and Mueller (2018)	All common stocks traded on NYSE or NASDAQ subject to price and size constraints	1989-2010	Journalist disagreement predicts next-day market return. This effect is strengthened during recessions. Journalist disagreement is predictive of the cross-section of next-day stock returns; effect is stronger for stocks with greater sensitivity to dispersion of beliefs and limits to arbitrage. Increased journalist disagreement is negatively related to return at cross-sectional and market level.
Johnman, Vanstone, and Gepp (2018)	FTSE 100 Index	2000-2016	Business news sentiment predicts magnitude (volatility) but not direction of next day return. Tone of sentiment is negatively related to return magnitude. An in-sample return strategy exploiting this relationship was shown to outperform buy-and-hold.
Kelly and Ahmad (2018)	DJIA, WTI futures	1989-2015	Sentiment has predictive power for next-day market returns. A trading strategy exploiting the relationship between negative sentiment and negative next-day return outperforms the index over the test period.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Minh et al. (2018)	S&P 500 Index, VN Index, 4 S&P 500 constituents	2006-2013	Sentiment has predictive power for equity returns at index and firm level. Predictive power is strongest within 24 hours and is negligible beyond one week of news article. Technical indicators increase performance of sentiment-based model. TGRU architecture outperforms alternative architectures and models for prediction task.
Myskova, Hajek, and Olej (2018)	14 US mega-cap stocks	Feb 2016	Meta-Learning approach better predicted volatility than comparison models. Sentiment variables contributed negligibly to predictive accuracy, while article length provided some predictive merit.
Uhl (2018)	S&P 500 Index	2007-2016	As positive (negative) sentiment increases, implied volatility decreases (increases). Effect of sentiment on implied volatility decreases with maturity of the option. Effect is more pronounced for OTM put options than call options.
Zhang, Zhang, et al. (2018)	78 Stocks from CSI 100, 13 Stocks in HK Market	Jan-Dec 2015	A novel coupled matrix and tensor factorization (CMT) method is an effective way to combine news, social media, and price information for stock prediction. CMT out-performs comparison models using the same data sources.
Calomiris and Mamaysky (2019)	51 developed and emerging equity markets	1996-2015	News forecasts market-level one-year ahead returns and drawdowns in sample and out-of-sample. The predictive content of sentiment, frequency, and entropy is time-varying and topic specific. Predictive capacity of news content is stronger in emerging than developed markets.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Chen, Liao, and Hsieh (2019)	Four Taiwan 50 Index constituents	2016-2017	News content predicts daily return at stock level. Custom lexicon trained on financial data outperformed general lexicon (NTU Sentiment Dictionary) when predicting daily return direction.
Glasserman and Mamaysky (2019)	Top 50 global banks, insurance, and real estate firms by USD market capitalization	1996-2014	News forecasts volatility at the company and market level. News shocks are impounded into volatility over the course of several months. Interacted measures of unusualness and sentiment are the strongest predictors of volatility among the tested measures. Predictive power of news was stonger at market level than firm level.
Griffith, Najand, and Shen (2020)	S&P 500 Index	1998-2014	Negative sentiment (fear) weakly predicts market return up to five (daily) lags. Market return also predicts sentiment measures through feedback effects. Negative sentiment (fear) strongly predicts conditional market volatility.
Narayan (2019)	45 global equity indices	1995-2013	Oil price news provides marginal predictive power on market returns relative to oil price growth rate. Predictive power of news is limited to a subset (17 of 45) of the tested countries. Results hold through a variety of robustness checks. Evidence suggests oil price news impacts returns through the expected discount rate and cashflow channels.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Sadik, Date, and Mitra (2019[a])	WTI Crude Oil	2014-2016	Novel incorporation of macroeconomic news into state-space model out-performs comparison approaches in predicting next-day spot price. Addition of macroeconomic news enhances model forecasting accuracy beyond novel model structure. Results are consistent between in-sample and out-of-sample tests.
Vanstone, Gepp, and Harris (2019)	S&P ASX 20 constituents	Jan 2015-Jul 2018	Addition of news features into neural network autoregression (NNAR) models enhances forecasting accuracy of stock price level but not direction, relative to NNARs using price information alone.
Ahmed, Sriram, and Singh (2020)	Amazon (AMZN) and Apple (AAPL) stock	2004-2014	News sentiment improves predictive performance across a range of different learning algorithms, and almost every one tested. Learning vector quantization (LVQ) was the best performing model among those tested.
Audrino, Sigrist, and Ballinari (2020)	18 US stocks, DJIA	2012-2016	The addition of sentiment and attention variables improves the fit of HAR volatility models when controlling for macroeconomic and financial data. Sentiment and attention variables improve prediction accuracy of HAR volatility models in out-of-sample forecasts. Meaures of investor attention (rather than sentiment) have most significant impact on future volatility. Sentiment variables only have short-term effects on volatility, with little evidence for predictive power beyond 2 days.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Gan et al. (2020)	S&P 500	2011-2017	Influence of news media on market dynamics decreased throughout the sample period, with social media becoming a more influential source of information by 2016. Market variables exert stronger impact on sentiment variables than the other way around. The link between volatility and sentiment is much more persistent than that between returns and sentiment.
Sadik, Date, and Mitra (2019[b])	12 stocks from FTSE100 and EUROSTOXX50 indices	2005-2015	Incorporation of news sentiment into the GARCH model leads to superior prediction of volatility than simple GARCH and EGARCH models. Positive (negative) sentiment is negatively (positively) related to next-day volatility.
Zhou et al. (2020)	Seven China A-share stocks	Sep 2017-Jun 2018	Combining multiple data sources, including news sentiment, can improve directional forecasting performance.
Coqueret (2020)	1009 stocks on US exchanges	2007-2017	Sentiment does not carry preeminent economic value for investors at the stock level. News sentiment is not a powerful predictor of future returns at the stock level and returns appear to drive sentiment rather than the other way around. Sentiment-driven predictability has increased from 2012 to 2017. In unreported results, returns of monthly cross-sectional portfolios built on smoothed news sentiment further suggest that sentiment is not a priced factor.

Continued on next page

Table 2.16: Summary of survey items (continued)

Author	Assets	Timeframe	Findings
Hanna, Turner, and Walker (2020)	UK market blue-chip index	1899-2010	News sentiment influences trading volume in bull markets, but not in bear markets. News sentiment influences return in all periods, with mixed evidence as to whether the effect is different between market states.
Pyo and Kim (2019)	KOSPI index, KOSPI VIX, KRW/USD exchange rate	2011-2017	Positive change in news sentiment predicts positive market return and lower volatility. Higher sentiment also leads to an appreciation of the exchange rate and lower exchange rate volatility.
Al-Maadid et al. (2020)	7 Gulf Cooperation Council (GCC) stock market indices	2010-2018	Negative business news from larger GCC economies have domestic and cross-boarder impact on stock return and volatility. A regime switching model specification fit the sample data more closely than the benchmark linear model. Only weak evidence of causality running from political news to stock returns.

Chapter 3

Data and Processing

3.1 Data and Processing

3.1.1 Price Data

Daily price, trading volume, market capitalisation (price times number of shares outstanding), analyst following, book value and index membership data for the historical constituents of the S&P 500 were sourced from Bloomberg. All market data have been corrected for capital changes and corporate actions such as stock splits and dividends. Bloomberg's historical index membership and pricing data are adjusted for symbology changes internally and remain consistent between the two. This negates the need to reconstruct price and membership time series for stocks which have changed identifiers and any potential errors that may be introduced during the process.

The Bloomberg historical constituent data contained 867 unique tickers for the sample period, one of which (ticker 1437355D) was removed as no price data was available. Stock price data missing on the last business day of each month were filled with the last available price, as long as the last available price was no more than five business days prior to the date being filled. This resulted in changes to 20 firm-month observations, none of which coincided with the firms' index membership. Returns were calculated arithmetically from price data, unless stated otherwise. Firm characteristics such as market capitalisation, turnover and price-to-book ratios for multi-period formation periods were taken as their average value over that period, excluding missing values. In the case of analyst following, missing observations were zero-filled prior to averaging, as analyst following in Bloomberg is only recorded when the value is at least one.

3.1.2 News Data

Intraday news data for the period 1 January 2003 to 31 December 2017 is sourced from Thomson Reuters News Analytics (TRNA). As stated earlier, the use of an institutional-grade news analytics platform such as TRNA represents a realistic option for asset managers (compared to say, those based on the archive subscriptions of educational institutions) and reduces the idiosyncrasies of custom news-processing techniques often applied in the literature. However, it should be noted that other institutional options are available, and of these TRNA was chosen largely because Thomson Reuters generously provided access to the data to support academic research in the area.

TRNA transforms unstructured, real-time news from various major news sources into a machine-readable feed of quantitative scores and meta data. Due to entity level scoring (discussed below), the TRNA feed consists of two top-level datasets, a dataset containing information about each news item, and another containing the analytics scores for the entities (e.g. companies) covered in each of the news items. The two datasets are linked through unique identification numbers, resulting in over 40 pieces of data for every firm-story observation.

A story on the TRNA feed often consists of multiple news items. The first part of a story may be an alert–a single line containing only essential information about an emerging story. Several alerts for a single story may be filed in quick succession. Alerts are typically followed by a story take, which includes a headline and a body. Multiple takes may be used to describe a particular story if additional text is needed. All parts of a story contain a common identifier which can be used in conjunction with date/time fields to identify the story.

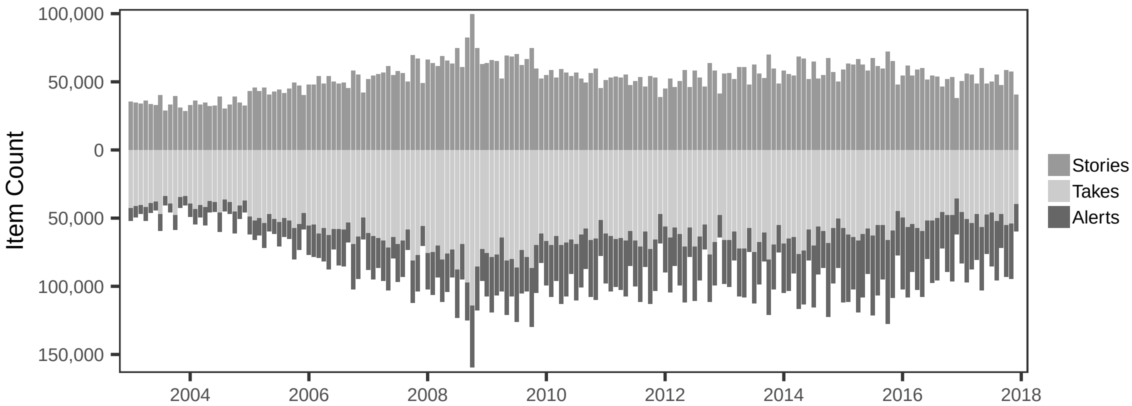

Figure 3.1: Number of stories, takes, and alerts by month for all firms listed on major US exchanges within the TRNA dataset from 01 Jan 2003 to 31 Dec 2017. Stories (above) are typically comprised of multiple takes and alerts (stacked, below).

Typically, it is only news agencies such as Reuters News that will publish stories over multiple news items, while stories from press wires and exchange wires will publish in a single item and do not issue alerts. Although some authors elect to remove alerts (e.g. Clements and Todorova, 2016; Sinha, 2016) from their sample, due their short length, we do not find any compelling reason to do so. And, as discussed below, our aggregation of news items to the story-level attenuates the influence of alerts in sentiment and coverage scores, relative to treating them as individual items, and so can be seen as a compromise between excluding them completely and treating them as stand-alone items. Figure 3.1 visualises the number of stories, takes, and alerts each month, for all firms trading on major US exchanges.

In addition to timestamps and asset identifiers, the most important fields for the current work are the sentiment and relevance scores. The primary variable used to characterise news content both in this paper and the wider literature is a measure of the sentiment or tonal valence of the news. The process used within TRNA to construct sentiment scores involves three main stages: linguistic preprocessing, feature extraction and classification.

The preprocessing stage includes a number of subroutines that collectively perform a shallow parse of the news item, transforming the raw text into a representation that can be mined for sentiment-relevant features. The feature extraction module then traverses the output of the shallow parse and identifies lexical and tonal patterns occurring within the text. This involves assigning a set of atomic features obtained to each token (word) and evaluating these features within the context of the sentence in which they occurred.

Evaluating text at the sentence level in this way means accounting for features of natural language such as word ambiguity, negation, intensification and subject-object relationships that are ignored by bag-of-word approaches. The output of the feature extraction phase is a vector representing the sentiment features for the entity being scored. This feature vector is used as input to the classifier, a three layer neural network trained on 5000 news articles triple annotated by financial analysts. The classifier produces the final sentiment score for the text, which comprises three real-valued outputs between 0.0 and 1.0 reflecting the probabilities that the story is positive, negative or neutral. As they are probabilities, these outputs sum to 1.0.

An example of a news item scored positive by the TRNA is the article from Canada NewsWire concerning Weight Watchers International Inc:

> *Weight Watchers Ranked #1 in Four Best Diet Categories by U.S. News & World Report Including "Best Diet for Fast Weight Loss"*

which was classified as being positive with a probability of 0.80, negative with a probability of 0.06 and neutral with a probability of 0.14. An example of a negatively scored news item is the article from Reuters News concerning Coca-Cola Co:

> *Lawsuit in U.S. says Coca-Cola downplays risks of sugary drinks*

which was classified as being positive with a probability of 0.06, negative with a probability of 0.81 and neutral with a probability of 0.13.

In contrast to most of the manual news processing techniques used in the literature, TRNA produces scores at the entity level. This means that only the parts of a news story relevant to a particular company will be used to calculate the scores for that company. This is an important feature considering that many news items will express several opinions about multiple subjects. For example, an article discussing two competing companies may be positive about one and negative about the other and may also focus on one company much more heavily than the other. The entity-level analysis employed by TRNA is designed to identify these differences and attribute story sentiment and relevance to each company accordingly. This is opposed to the document-level analysis found in Tetlock (2007) and Hillert, Jacobs, and Mueller (2014) among others, which would produce a single score reflecting the overall tone of the story, and be assigned to one or both companies.

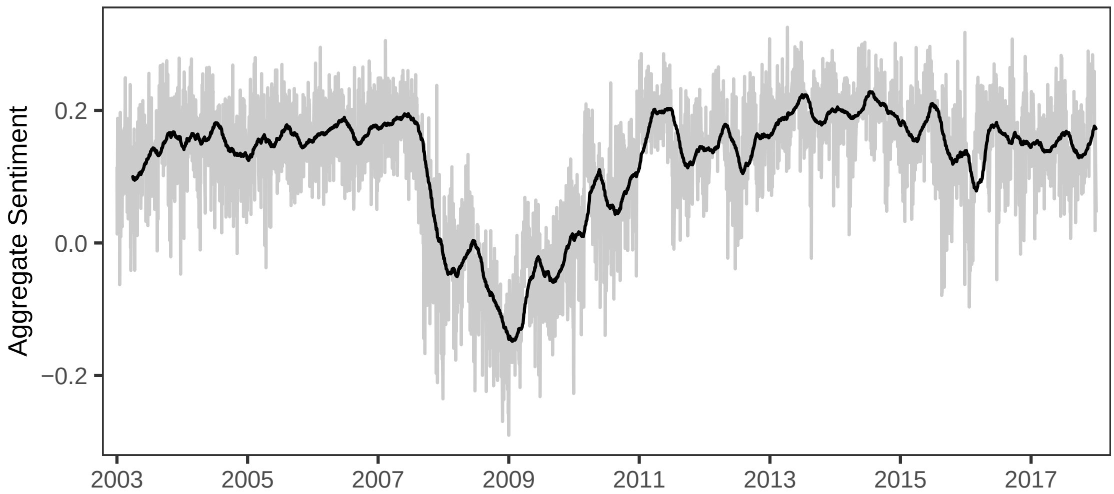

Figure 3.2: Daily (grey) and 60-day simple moving average (black) aggregate news sentiment of S&P 500 Index constituents.

Figure 3.2 shows the mean daily news story sentiment of S&P 500 index constituents for each day in the sample period, and the 60-day simple moving average of these scores. It can be seen that market-wide news sentiment is responsive to salient economic and political events such as the Global Financial Crisis (GFC) and the 2016 US presidential elections. It is also apparent that aggregate news coverage is positive (at least numerically) in general throughout the sample period, with negative aggregate story sentiment appearing only for an eight month period during the GFC.

The relevance of a news item to a given company is a 0.0-1.0 score determined by the number of times the company is mentioned in the item compared to the number of mentions of the most common company and the total number of other company mentions within the same item. Additionally, if the company is mentioned in the article headline it will receive a relevance score of 1.0. The relevance score provides a means of attenuating the noise created by passing mentions and giving more weight to predominant articles. It is typical in the news analytics literature to remove items below a certain relevance threshold. Table 3.1 summarises the effect of different relevance thresholds on the number of items in our data set. Following Leinweber and Sisk (2011), we exclude all news items with a relevance below 0.6.

Table 3.1: News item relevance thresholds

			Percent Remaining			
Type	No.	Mean *Rel*	*Rel*=0.6	*Rel*=0.7	*Rel*=0.8	*Rel*=0.9
Alert	1,481,005	0.98	0.99	0.99	0.93	0.93
Take	3,821,806	0.59	0.46	0.45	0.44	0.43
Undefined	194	0.66	0.36	0.36	0.35	0.35

This table summarises the impact of applying different relevance (*Rel*) thresholds on the number of news items in the data set for constituents of the S&P 500 Index between January 2003 and December 2017.

TRNA also provides a measure of item novelty, indicating how many news items over selected look-back windows contain textually-similar asset-specific text to the current item. For example, a seven-day novelty count of two means that there were two news items within the previous seven days that contained similar asset-specific text to the current item. Some authors, such as Smales (2015[b]) and Leinweber and Sisk (2011), remove items with a novelty count above a certain threshold, in order to focus on the arrival of new information only. Although novelty should be relevant to high-frequency market reactions, we are considering the low-frequency impact of news aggregated over multiple-month formation periods and portfolios employing monthly rebalancing, we therefore include all news items regardless of their novelty count. Further, it is plausible that content repeated in the news cycle across multiple stories and news agencies represents particularly important or hard-to-quantify events, while the heightened media exposure of such events may have an increased effect on investor reaction in itself. This line of reasoning is consistent with the findings of Huynh and Smith (2017), who explicitly use novelty count as a measure of journalist attention and find it to be the most relevant characteristic of news for 12-month forward return.

Construction of news scores

As discussed in Section 2.4 of the literature review, one of the first decisions to be made when dealing with news data is how news scores should be constructed over multi-period horizons. One option is to treat the entire formation period as one unit, and calculate scores by aggregating over the entire period at once. More commonly, daily scores are computed by first aggregating at the daily level (for interday analyses), and then combining daily scores over the formation period.

This is the approach we take, as daily data is straight-forward to work with and provides a flexible baseline to aggregate over any longer horizon of interest. To summarise our process: We first calculate sentiment scores for each individual intraday news item and then hierarchically aggregate scores to the story level, daily level, and finally to our multi-month formation periods. News items published after market close or on non-trading days are allocated to the next trading day.

The sentiment score of an individual item (such as an alert or take) is given by its positive sentiment score minus its negative sentiment score,

$$item\_sentiment = P_{positive} - P_{negative}. \tag{3.1}$$

As the sentiment scores are probabilities, each item's sentiment score will range between -1 and 1. After computing item-level sentiment, the sentiment of story j is calculated as a relevance-weighted average of item sentiment:

$$story\_sentiment_j = \frac{\sum_{i=1}^{N_j} relevance_i \times item\_sentiment_i}{\sum_{i=1}^{N_j} relevance_i}, \tag{3.2}$$

and relevance of story j is taken to be the average relevance of its constituents:

$$story\_relevance_j = \frac{1}{N_j} \sum_{i=1}^{N_j} relevance_i. \tag{3.3}$$

Daily sentiment for a given day d is calculated as the relevance-weighted average of story sentiment:

$$daily\_sentiment_d = \frac{\sum_{i=1}^{N_d} story\_relevance_i \times story\_sentiment_i}{\sum_{i=1}^{N_d} story\_relevance_i}, \tag{3.4}$$

Finally, to calculate sentiment over a given formation period, we average over all daily sentiment scores within the formation period, ignoring no-news days:

$$formation\_sentiment = \frac{1}{N} \sum_{i=1}^{N} daily\_sentiment_i.$$

Summary Statistics

Median item-level characteristics are detailed in Table 3.2. The number of sentiment words ("Sent. Words") is the number of words in a news item that are relevant to the sentiment score of a particular company. If the entire article is relevant to a given company, the number of words will equal the number of sentiment words. It can be seen that alerts and takes tend to have approximately the same sentiment bias on average.

Undefined items (neither takes or alerts) are discarded from the data set, as they are not defined in the TRNA documentation and represent a negligible fraction of the data.

Table 3.2: News item summary statistics

		Median Value						
Item Type	No.	Company Count	Pos. Sent.	Neg. Sent.	Sent.	No. Words	No. Sentences	No. Sent. Words
Alert	1465471	1	0.21	0.09	0.10	16	1	16
Take	1765016	1	0.31	0.16	0.11	340	14	199
Undefined	69	1	0.34	0.15	0.16	110	4	95

This table includes item-level summary statistics of news data for constituents of the S&P 500 Index between January 2003 and December 2017.

Summary statistics of story-level features are included in Table 3.3 and histograms of news attributes by company are provided in Figure 3.1.2. Company news coverage, in terms of the number of stories relating to the firm over the sample period, is approximately normally distributed across the index constituents. Additionally, the histograms show that most firms' news sentiment is positive on average, with a standard deviation between 0.4 and 0.55, and with negative skew.

Table 3.3: Summary statistics of news story features

	Pos. Sent.	Neg. Sent.	Sent.	Story Relevance	No. Alerts	No. Takes	No. Words	No. Sentences	No. Sent. Words
Min.	0.02	0.01	-0.78	0.60	0.00	0.00	1.0	1.00	1.00
1st Qu.	0.12	0.05	-0.33	1.00	0.00	1.00	85.0	5.00	70.00
Median	0.30	0.16	0.11	1.00	0.00	1.00	253.0	11.00	169.00
Mean	0.37	0.27	0.11	0.98	0.78	0.94	537.9	21.04	320.51
3rd Qu.	0.56	0.45	0.52	1.00	1.00	1.00	686.0	26.00	417.00
Max.	0.86	0.82	0.83	1.00	66.00	65.00	52122.0	2988.00	41924.00

This table includes summary statistics for features of story-level news items for constituents of the S&P 500 Index between January 2003 and December 2017.

While market news (and our data) reflects a 24-7 stream of information, the publication of news largely conforms to a weekly cycle, with the majority of news published during the work week. Figure 3.4 displays the periodogram (using fast Fourier transform, or *FFT*) of mean daily sentiment, boxplot of number of news stories by day of week, and density estimates of mean story sentiment by day of week. The most significant seasonality in

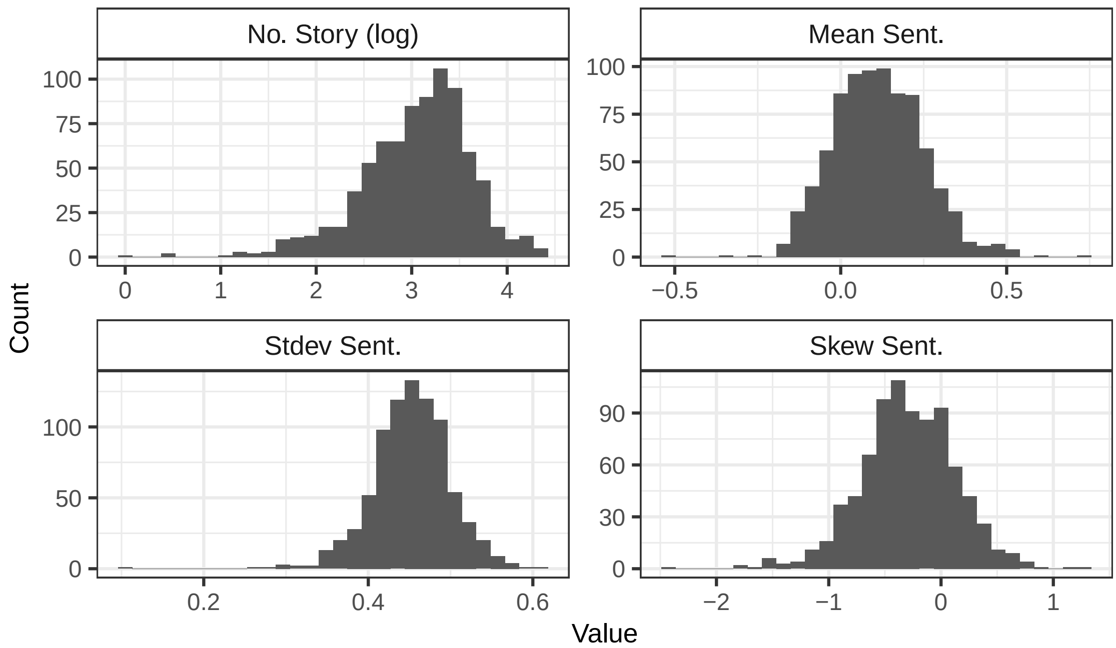

Figure 3.3: Histogram of news coverage and first three moments of news story sentiment across S&P 500 constituents.

daily sentiment as per the FFT occurs at a frequency of seven days. This is also evidenced by the density plots, which show the relative positivity of news on Mondays, followed by a decline in sentiment throughout the week until reaching a low on Saturday. From the boxplots we see that mid-week (Wednesday, Thursday) corresponds to the peak in terms of news coverage, and relatively few articles are published on weekends. Summary statistics by day-of-week are provided in Table 3.4.

Table 3.4: Story features by day-of-week

	Mean		Median	
Weekday	Sent	No.	Sent	No.
Mon	0.16	431.98	0.19	428.5
Tue	0.14	510.22	0.17	487.0
Wed	0.13	503.27	0.15	483.0
Thu	0.13	516.71	0.15	493.0
Fri	0.07	380.67	0.10	351.0
Sat	0.00	14.41	0.00	11.0
Sun	0.10	15.75	0.10	13.0

This table includes summary statistics for story-level news items by weekday for constituents of the S&P 500 Index between January 2003 and December 2017.

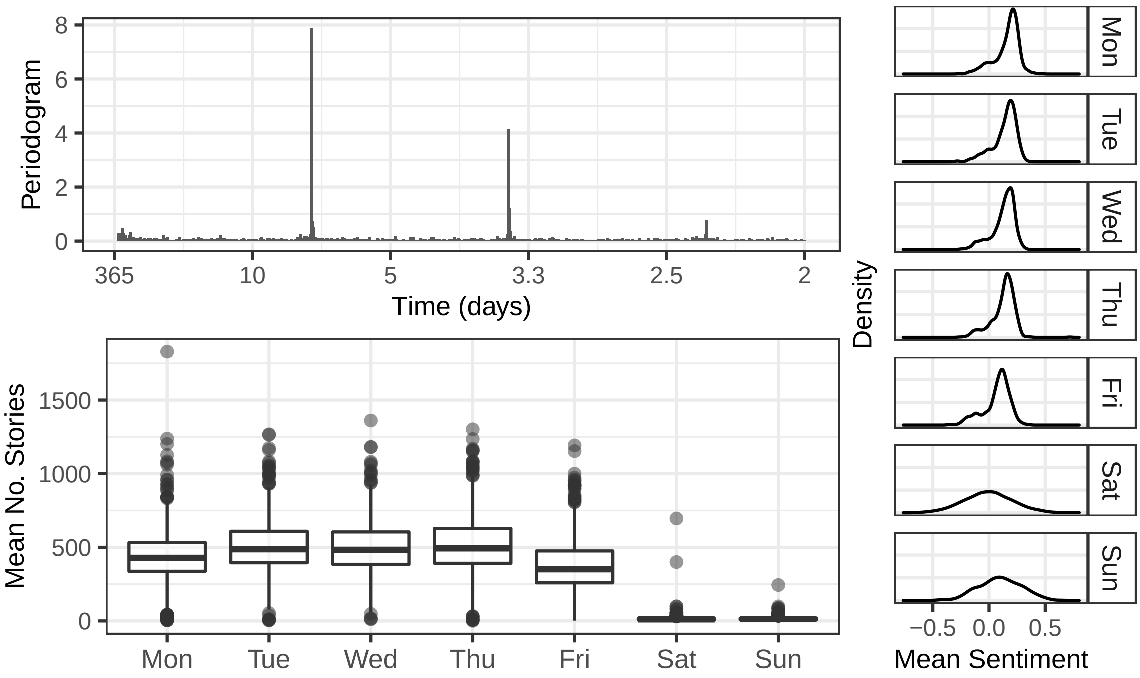

Figure 3.4: Day-of-week seasonality of aggregate news characteristics. Top) periodogram of daily average news sentiment. Bottom) boxplot of number of news stories by day-of-week. Right) density estimates of average news sentiment by day-of-week.

To align daily sentiment scores with price data and trading activity, and to protect against data leakage (look-ahead bias), we allocate news items published after market close (1600 EST) or on non-trading days to the next trading day, as determined by the NYSE. Due to the removal of weekends, the variation in coverage across weekdays is significantly reduced. The seven-day seasonality in average sentiment is also attenuated and shifted to a five-day (4.8) period as shown in Figure 3.5. The corresponding business-day summary statistics are given in Table 3.5.

Pooled summary statistics for features of daily news sentiment scores are given in Table 3.6. Typically, a firms daily sentiment score reflects the content of two stories, and at the third-quartile this becomes three. As we would expect from the item and story-level summaries, daily news scores are most often positive. In cases where more than one story about a firm is published in a single day, the standard deviation in sentiment between these items is relatively high at a median value of 0.26—considering the median daily sentiment score is 0.15.

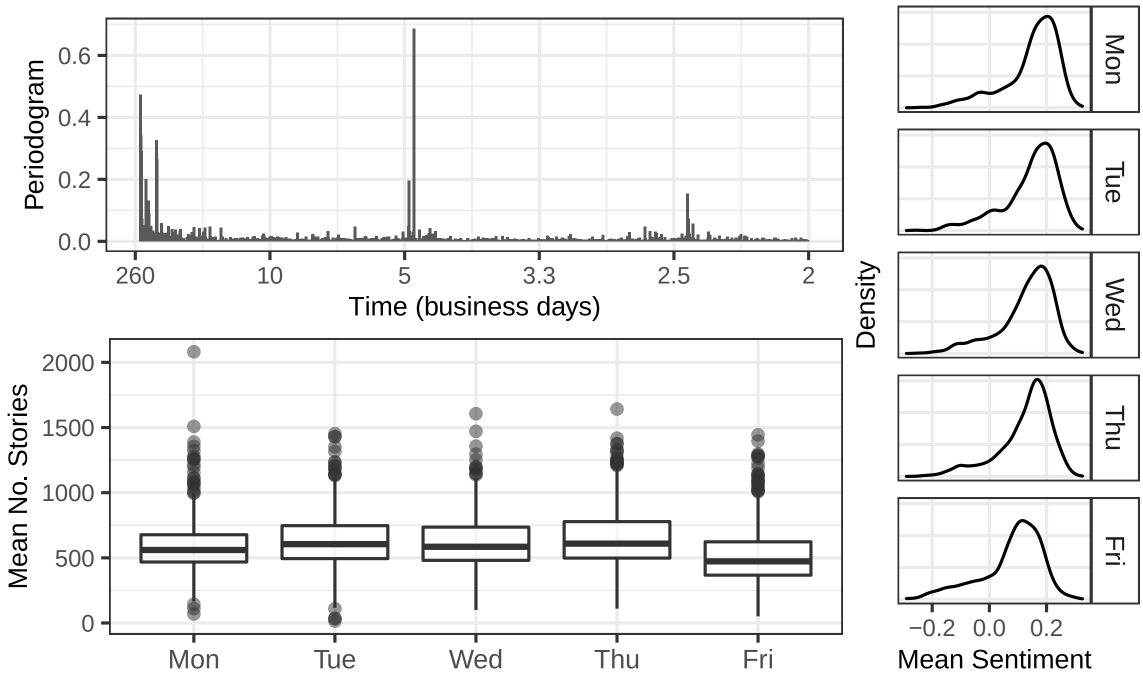

Figure 3.5: Day-of-week seasonality of aggregate news characteristics after news-item allocation to NYSE trading days. Top) periodogram of daily average news sentiment. Bottom) boxplot of number of news stories by day-of-week. Right) density estimates of average news sentiment by day-of-week.

Table 3.5: Story features by business day

	Mean		Median	
Weekday	Sent	No.	Sent	No.
Mon	0.15	483.97	0.17	459.0
Tue	0.14	520.45	0.17	494.5
Wed	0.13	508.81	0.15	482.5
Thu	0.12	533.12	0.15	499.0
Fri	0.07	425.63	0.10	388.0

This table includes summary statistics for story-level news items by business day for constituents of the S&P 500 Index between January 2003 and December 2017.

Table 3.6: Summary statistics of daily sentiment scores

	Sent. Stdev	Pos. Sent.	Neg. Sent.	Sent.	No. Alerts	No. Takes	No. Story
Min.	0.00	0.02	0.01	-0.78	0.00	0.00	1.00
1st Qu.	0.06	0.18	0.07	-0.16	0.00	1.00	1.00
Median	0.26	0.36	0.18	0.15	0.00	2.00	2.00
Mean	0.28	0.38	0.25	0.13	2.07	2.49	2.63
3rd Qu.	0.43	0.55	0.41	0.45	2.00	3.00	3.00
Max.	1.13	0.86	0.82	0.83	449.00	481.00	362.00

This table includes summary statistics for features of daily news sentiment scores for constituents of the S&P 500 Index between January 2003 and December 2017.

Table 3.7 presents averages of firm-level features of daily news sentiment scores for each year of the sample period. For example, the table shows that in 2005 the average skewness of firms' daily sentiment scores across the year was -0.25.

Table 3.7: Summary statistics of daily sentiment scores by year

		Mean Company Value					
Year	No. Company	No. News Days.	No. Story.	Mean Sent.	Stdev Sent.	Skew Sent.	Kurt. Sent.
2003	477	72.86	163.87	0.18	0.37	-0.47	-0.09
2004	487	71.78	166.34	0.20	0.34	-0.47	0.06
2005	489	94.08	250.93	0.19	0.31	-0.25	0.33
2006	502	94.73	257.51	0.19	0.32	-0.28	0.33
2007	512	101.17	277.75	0.13	0.35	-0.14	-0.42
2008	511	122.91	368.15	-0.06	0.37	0.53	0.33
2009	502	119.97	356.98	-0.03	0.36	0.48	0.33
2010	492	112.84	278.75	0.06	0.38	0.07	-0.33
2011	493	87.06	220.83	0.20	0.38	-0.60	0.00
2012	496	84.29	212.12	0.20	0.39	-0.60	-0.01
2013	504	91.25	217.46	0.23	0.37	-0.63	0.02
2014	500	95.34	230.30	0.23	0.37	-0.70	0.23
2015	518	93.24	255.45	0.15	0.41	-0.54	-0.20
2016	522	87.42	235.91	0.17	0.44	-0.62	-0.34
2017	523	79.98	208.58	0.21	0.41	-0.74	0.06

This table includes summary statistics for of daily news sentiment scores for constituents of the S&P 500 Index between January 2003 and December 2017. Columns three to eight reflect (cross-sectional) average values of the listed company-level summary statistics (longitudinal) for that year.

Sensitivity of Sentiment Scores

In addition to reflecting the inherent structure of TRNA data, aggregating items to the story-level prior to aggregating to the daily-level ensures that stories with a larger number of takes and alerts are not weighted more heavily than stories published in a single take and similarly avoids the problem of biasing sentiment toward news agencies (i.e. Reuters News) that release alerts. Yet, beyond this data processing decision, there are still methodological choices to be made with respect to story sentiment scores. For instance, it was previously stated that some authors elect to remove alerts from TRNA data before calculating news sentiment or volume measures, due to them being single-line items. Applying similar logic, we could have calculated story scores by weighting each item by its length, in addition to its relevance.

Table 3.8: Summary of absolute differences between sentiment specifications

	Takes-Only Vs. Story	Story Vs. Word-Weighted	Takes-Only Vs. Word-Weighted
Min.	0.00	0.00	0.00
1st Qu.	0.00	0.00	0.00
Median	0.00	0.00	0.00
Mean	0.03	0.02	0.02
3rd Qu.	0.00	0.00	0.00
Max.	0.86	0.69	0.61

This table includes summary statistics of absolute differences between three approaches to calculating news story sentiment. *Takes-Only* is calculation after removing alerts, *Story* includes both takes and alerts (default), and *Word-Weighted* includes takes and alerts with additional weighting of each item given by the number of sentiment words it contains. Statistics were calculated using a random sample of 10,000 news stories.

Fortunately, we found very little difference between alternative aggregation techniques. As evidence for this, Table 3.8 details summary statistics for the absolute differences in news sentiment scores across the three aforementioned approaches: 1) the "baseline" approach, which includes takes and alerts in the aggregate relevance-weighted sentiment score, 2) the baseline calculation following removal of alerts, and 3) weighting each item by the number of sentiment words it contains, in addition to weighting by item relevance. Note that the third quartile absolute difference between the three approaches is zero (to at least two decimal places), with mean absolute difference less than or equal to 0.03. Table 3.9 presents the correlation matrix between the three sentiment measures, the near-unity elements of which further convey the similarity between the approaches. Sensitivity to variable construction is covered again in Chapters 4 and 5.

3.1.3 Bloomberg and Thomson Reuters Mapping

Bloomberg and Thomson Reuters identifiers were matched through Thomson Reuters' record matching service at permid.org/match. The service allows users to upload entity information in terms of local identifiers in csv format which is then matched server-side,

Table 3.9: Correlation between sentiment specifications

	Story	Takes-Only	Word-Weighted
Story	1.00	0.98	0.99
Takes-Only	0.98	1.00	0.99
Word-Weighted	0.99	0.99	1.00

This table represents the correlation matrix between values obtained using three different approaches to calculating news story sentiment. *Takes-Only* is calculation after removing alerts, *Story* includes both takes and alerts (default), and *Word-Weighted* includes takes and alerts with additional weighting of each item given by the number of sentiment words it contains. Statistics were calculated using a random sample of 10,000 news stories.

returning a csv containing the output of the matching algorithm. The matching report contains the Reuter's PermID, match level (one of "Possible", "Good", "Excellent", or "No Match"), and a match score (0 - 100).

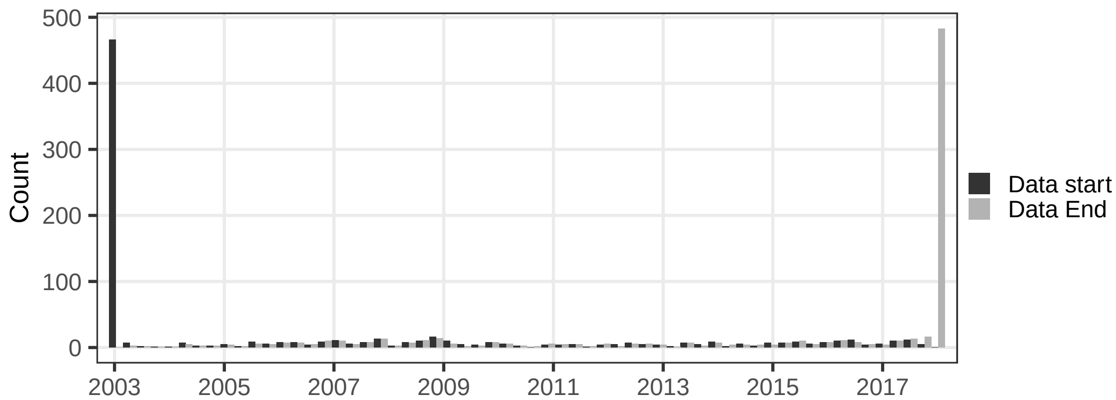

Figure 3.6: Chronological range of TRNA data coverage for sample firms. The histogram shows the distribution of beginning and end dates of news coverage across firms.

We used company name and ticker, provided by Bloomberg, as inputs to the matching service. For the 867 input companies, the output contained 772 Excellent matches, 93 Possible matches, and two unmatched companies. Each company match was then verified manually. This resulted in 50 manual PermId changes. Figure 3.6 illustrates the distribution of data availability across the index constituents via histogram of start-dates and end-dates of TRNA coverage for each firm. The high density peaks at either end of the sample period suggests that the news data for most firms spans the entire sample period.

Chapter 4

Longitudinal Analysis

Firm-level predictability is one pathway to the construction of news-informed portfolios that has largely been overlooked by the low-frequency news analytics literature. Depending on the degree of predictability offered by news, firm-specific models may inform longitudinal trading strategies or simply provide a better basis for portfolio sorts than the non-parametric conditioning studied so far. However, a basic prerequisite to such use-cases is some demonstration that news has predictive capacity at the firm-level. This is the focus of the current analysis, which is primarily regression-based. Implementations of "momentum-style" portfolios, as previously referenced, are not examined until Chapters 5 and 6.

Herein, we test the longitudinal predictability of forward returns through a series of firm-level regressions, including single period, multi-period, and VAR specifications. We find no supporting evidence that either news content or news volume are useful predictors of future return at the firm-level. Rates of statistical significance for predictive model coefficients were comparable to those of random covariates, and the majority of effect sizes are economically irrelevant.

4.1 Introduction

The focus of the current work is cross-sectional analysis, due to it being the basis of the momentum literature and its amenability to practical portfolio construction techniques. However, the application of news sentiment to firm-level predictability remains an important consideration in any determination regarding its potential utility to portfolio managers. Furthermore, firm-level predictive regressions can serve as a useful precursor and point of comparison to deeper cross-sectional analyses by revealing methodological sensitivities, conditional relationships, and calibrating expectations.

We performed a series of firm-level predictive regressions for historical S&P 500 index constituents, using monthly news sentiment and news coverage as regressors and one-month excess return as the dependent variable. We then subjected these regressions to a variety of subset and specification robustness tests. Our results from these monthly regressions do not support the hypothesis that news leads returns at the monthly horizon.

We then extended the formation and prediction horizons of our tests, using a momentum-style regression specification. Rates of statistical significance of news regressors and the explanatory power of the extended horizon models were similar to the baseline regression, presenting no compelling evidence of predictability.

Finally, we dropped the assumption of exogeneity of news and turned to a richer dynamic model described by firm-level six-lag vector autoregressions (VARs). The VAR models were analysed on the basis of individual and joint coefficient hypothesis tests, causality tests, and forecast error variance accounting. The results demonstrated autodependence and cross-dependence between news sentiment and news coverage, though very little evidence of predictive cross-dependency from news to excess return was found. If anything, return was found to be a stronger predictor of news than vice versa. Overall, our analysis fails to find evidence that news sentiment or news coverage are useful for longitudinal prediction of excess returns at the firm-level, for one- to six-month horizons.

The existing research on firm-specific predictability using news variables has been focused on short (typically daily) prediction horizons, with mixed results. Ferguson et al. (2015), Zhang, Härdle, et al. (2016), and Wang, Chen, and Wei (2015) found that news tone and coverage are related to next-day returns among FTSE 100 constituents, 100 S&P 500 constituents, and TSEC-listed Taiwanese companies, respectively. In small samples of two and four stocks respectively, Ahmed, Sriram, and Singh (2020) and Minh et al. (2018)

found that news variables improved the performance of machine learning models in predicting daily return. Ahmad et al. (2016) showed evidence of time-varying predictability using news sentiment in daily returns of 20 large non-financial US firms.

Counter-evidence concerning daily firm-level return predictability is offered by Vanstone, Gepp, and Harris (2019), Strauß, Vliegenthart, and Verhoeven (2016) and Coqueret (2020). Using neural network autoregression, Vanstone, Gepp, and Harris (2019) found news sentiment improved the forecasting accuracy of stock price level, but not direction, among S&P ASX 20 constituents. Strauß, Vliegenthart, and Verhoeven (2016) found that neither the tone or volume of news stories had any effect on the next-day opening prices of 21 firms listed on the Amsterdam exchange. Coqueret (2020) performed daily predictive regressions on 1009 stocks on US exchanges and concluded that firm-specific news sentiment is not an economically relevant predictor of forward returns.

To our knowledge, no studies have examined firm-level predictability at horizons of one-month or more. Xiong and Bharadwaj (2013) and Wu and Lin (2017) study the impact of news on return at monthly frequency, but they study contemporaneous effects only. Coqueret (2020) tested prediction horizons up to one month, but the dependent variables still consisted of daily news scores. Similarly, Ahmad et al. (2016) examined the impulse response of news up to a 250-day horizon, using a daily VAR specification. In our view, given the literature on price momentum and existing results on 'news momentum', these specifications are less plausible than ones in which dependent and independent variables are aggregated over comparable horizons.

Our baseline regression and subsequent robustness tests are most similar to Coqueret (2020), we then extend the analysis by investigating predictability in multi-period 'momentum-style' regressions. Our use of VAR is most similar to Ahmad et al. (2016) in being at the firm level, and most resembles Uhl (2014) and Kraussl and Mirgorodskaya (2017) in the subsequent tests performed.

4.2 Data

Price data for historical constituents of the S&P 500 index was sourced from Bloomberg, and news data was sourced from Thomson Reuters News Analytics. Market return and the risk-free return were sourced from Ken French's data library. Data was processed using the methodology described in Chapter 3. Briefly, this involved aggregating intraday news items (alerts and takes) to the story level, and then aggregating the intraday stories to the daily level. Items released after market close were attributed to the next trading day. Story-level and daily sentiment scores were calculated as relevance-weighted averages. The number of stories for each firm were also recorded, resulting in a daily series of sentiment scores and news coverage for each firm. Unless stated otherwise, sentiment scores over horizons longer than one day are given by a simple average of the daily scores over the period, while news coverage is given by the total number of stories released over the period.

When no news data is present in a given month, sentiment is carried forward from the previous month, up to a maximum of five months. The assumption being that sentiment about a firm remains constant until new news media is published. Alternatives to this forward-filling procedure are discussed in later sections. We remove firms that have news data in less than 20% of the months for which they were index constituents, prior to the forward-filling of sentiment.

The coverage variable (number of stories) was transformed as $f(x) = \log(1 + x)$. We made this transformation choice by applying the Box and Cox (1964) procedure to the coverage series of all (filtered) firms and recording the scaling parameter (λ) that gave the highest likelihood estimate. The mean and median lambda estimates were close enough to zero (0.2 and 0.042, respectively) to warrant the log transform.

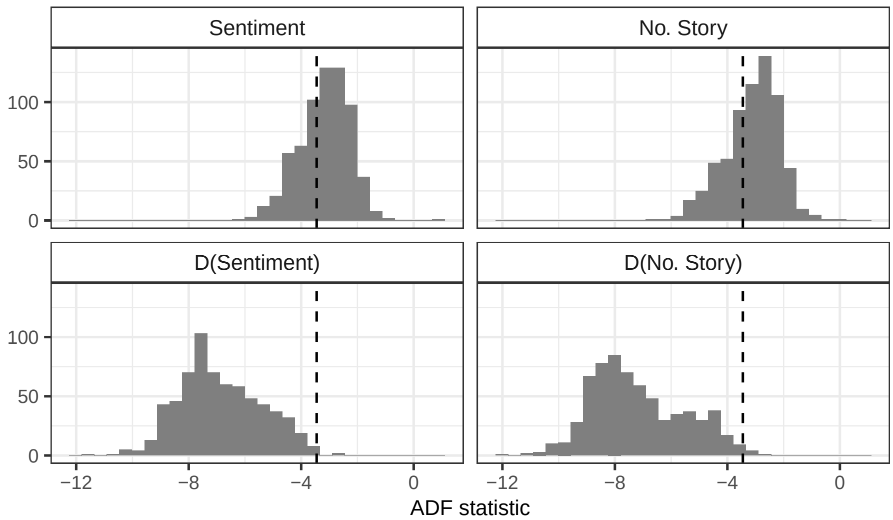

Figure 4.1: Distribution of ADF statistics for original (top) and first-differenced (bottom) news variables. The test is performed with constant and linear trend with maximum lag of floor$((T-1)^{1/3})$. The dotted line in each plot corresponds to the 0.05 significance level (Hamilton, 2014, Table B.6 Case 4).

All series were standardised (centered and scaled) at the firm level after any transformations were applied, prior to use in any regressions. The exception is the momentum-style regressions discussed later, for which the momentum variables were summed post standardization (however, these variables, being sums of standardised variables, have an expectation close to zero anyway).

We tested the stationarity of our variables to avoid performing our regressions with variables of different integration order. Specifically, we performed the Augmented Dickey-Fuller (ADF) unit root test on the sentiment, return, and coverage series for each stock. Figure 4.1 shows the distribution of ADF statistics for news sentiment and coverage, pre- and post-differencing.

The null hypothesis of a unit root was rejected at the 5% level in 88%, 35% and 33% of cases for return, sentiment, and coverage, respectively. Notably, these results suggest that more often than not, sentiment and coverage are integrated processes. Conducting the ADF test on the first differences of these variables resulted in a rejection rate of 99% for both variables at the 5% level, indicating that they are at most I(1) processes. This also confirms that by taking first differences across all stocks, we are not introducing non-stationarity into those for which the existence of a unit root was initially rejected.

We consider using first-differences to be a conservative choice, and expect that most of the raw sentiment and coverage series are in fact stationary, despite the failure to reject the existence of a unit root in a majority of cases. This is both from visual inspection of many of the series and taking into account the low power of unit root tests in finite samples, particularly for autocorrelated series (Cochrane, 1991; DeJong et al., 1992). For instance, in simulations of 10,000 stationary ARIMA(1, 0, 1) series, with coefficients based on the average model fit across all stocks, unit roots were rejected in only 75% of cases at the 5% level, and 54% of cases at the 1% level. The number of observations in these simulations were sampled from the number of observations in our data set, which ranges from 36 to 180, after filtering. In any case, we find no meaningful differences in baseline regression results whether we use the raw or differenced news variables, and continue with the differenced versions.

4.3 Baseline regression

Our baseline procedure tests the in-sample predictive content of news variables through longitudinal firm-specific regressions. For each firm we test the following predictive specification:

$$\tilde{r}_{t+h,i} = \alpha_i + \beta_{1_i} Sent_{t-h,i} + \beta_{2_i} Coverage_{t-h,i}, \tag{4.1}$$

where $\tilde{r}$ is the excess return of the firm relative to the market, *Sentiment* is firm-specific news sentiment, and *Coverage* is the number of news stories related to the firm. The subscript i denotes firm i.

Our baseline regression uses a one-month return horizon and variable formation period. As the sampling frequency of the data for the regressions is monthly, this keeps observations non-overlapping, and the relatively short horizon is a generous starting point from which to demonstrate any predictive content of news. We use the standard OLS regression procedure, with HAC standard errors following Andrews (1991).

In Figure 4.2 we plot the histogram of beta estimates and test-statistics for the beta coefficients in the baseline regressions, along with the boxplot of adjusted r-squared values which shows the 25th, 50th (median), and 75th percentiles, with outliers plotted individually. The Durbin-Watson test for autocorrelation in residuals (Durbin and Watson, 1950a; Durbin and Watson, 1950b; Durbin and Watson, 1971) resulted in the null of no autocorrelation being rejected in 6.4% of cases at the 5% level and 2.4% of cases at the 1% level, which is reasonable given our use of HAC robust standard errors.

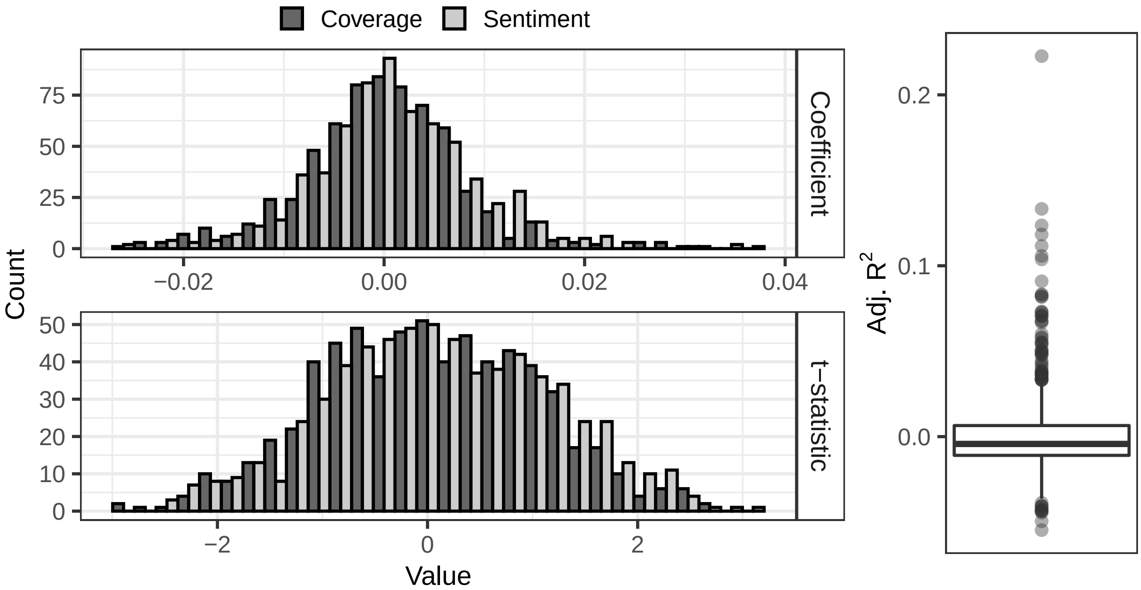

Figure 4.2: Left) Histogram of coefficients (top) and test statistics (bottom) and, Right) boxplot of adjusted R^2 values, for the Equation 4.1 regression model.

As far as predictability is concerned, important metrics are the proportion of t-statistics that reach a significant two-sided threshold, and the distribution of adjusted r-squared. These are also summarised in Table 4.1.

From the histogram of t-statistics it is clear that in the vast majority of cases, neither news sentiment or coverage are statistically significant predictors of forward return. Table 4.1 confirms that news sentiment is significant at the 5% level in approximately 8% of cases and at the 1% level in approximately 1% of cases. News coverage is significant at 5% level in approximately 6% of case and at the 1% level in approximately 0.8% of cases.

The low number of significant coefficients is accompanied by low explanatory power of the model; the median adjusted r-squared is negative, with less than one quarter of the regressions yielding a value near 1%. Another feature of the regression data is that the coefficients are approximately centered around zero. This matters for predictability; ex ante we would not know the sign of these relationships for any given firm and makes it more difficult to describe a plausible conceptual model of how news influences returns on a month-to-month basis.

Due to the large number of tests being performed, we provide a synthetic point of comparison. While we know what rates of significance to expect under theoretically-pure conditions given the null hypothesis of zero predictability, there are well known reasons why our results could depart from these expectations—even when there is no true predictability. We therefore benchmark regressions against simulated results based on an identical empirical setup, but using random regressors. For news sentiment and news coverage we use a randomly resampled set of the original values. This maintains the empirical distribution of the regressors but disrupts any potential predictive relationship.

From Table 4.1 it is apparent the randomised features compare very closely to the realised features in terms of explanatory power and rate of statistical significance. This is further evidence against the predictive capacity of the firm-specific regression model.

Table 4.1: 1-Month Predictive Regression Summary

	t-statistic			
	Sentiment	Coverage	R^2	Adj. R^2
Panel A: Realised covariates				
Min.	-3.4072	-3.4352	0.0000	-0.0547
1st Qu.	-0.6272	-0.7419	0.0050	-0.0110
Median	0.0909	0.0259	0.0124	-0.0043
Mean	0.1327	0.0259	0.0215	0.0009
3rd Qu.	0.9027	0.7713	0.0264	0.0065
Max.	3.5272	3.7518	0.2580	0.2227
Prop.$\|t\| > 2$	0.0773	0.0621		
Prop.$\|t\| > 3$	0.0091	0.0076		
Panel B: Random covariates				
Min.	-3.6975	-2.9853	0.0000	-0.0499
1st Qu.	-0.6737	-0.6974	0.0046	-0.0110
Median	0.0139	0.0224	0.0111	-0.0051
Mean	-0.0033	-0.0042	0.0200	-0.0005
3rd Qu.	0.6875	0.6674	0.0244	0.0049
Max.	3.3191	3.2176	0.2222	0.1884
Prop.$\|t\| > 2$	0.0561	0.0621		
Prop.$\|t\| > 3$	0.0045	0.0045		

This table presents distributional summaries of test-statistics and adjusted R^2 for the predictive regression represented by Equation 4.1. Panel A details results using observed news variables. Panel B details results using simulated noise as in place of the news variables. Test-statistics are based on Newey-West corrected standard errors.

4.4 Robustness Checks

In the following analyses we apply alternate specifications and cut the data in different ways to apply pressure to the null result regarding firm-specific predictability with news variables.

4.4.1 Specification Robustness

Here we investigate whether our results revealing the low predictive capacity of news variables are merely a product of the way the variables were constructed. We do this by testing four additional model specifications (specification 1 being the baseline model):

- Specification 2: Only the negative component of sentiment scores are used. This is motivated by the finding in the literature that negative tone may be a less noisy signal of news content than positive tone (Tetlock, 2007; Loughran and McDonald, 2011; Ferguson et al., 2015). For the sake of comparison, we multiply the negative sentiment score by negative one so that its direction is consistent with the original sentiment score.

- Specification 3: Rather than forward-filling, missing sentiment observations are filled using the median firm sentiment that month. Observations are then median-centered (using original median). This is equivalent to zero-filling median-centered data. As all observations are centered using the cross-sectional median, this specification controls for market-wide sentiment.

- Specification 4: When aggregating daily sentiment scores to the monthly level, instead of taking the simple average, daily sentiment is weighted by the number of stories driving the sentiment score for that day.

- Specification 5: News coverage or volume is represented by the number of days with news, rather than the number of stories released.

Figure 4.3 presents the histogram of test statistics for the sentiment variable coefficient across each of the regression specifications. While some variation exists, the distribution of test-statistics between specifications appears to be overwhelmingly similar. This is supported by corresponding summary statistics for each specification provided in Table 4.2.

Although the pooled distributional characteristics are similar between specifications, it could still be the case that the differences at the firm-level are meaningful. While this

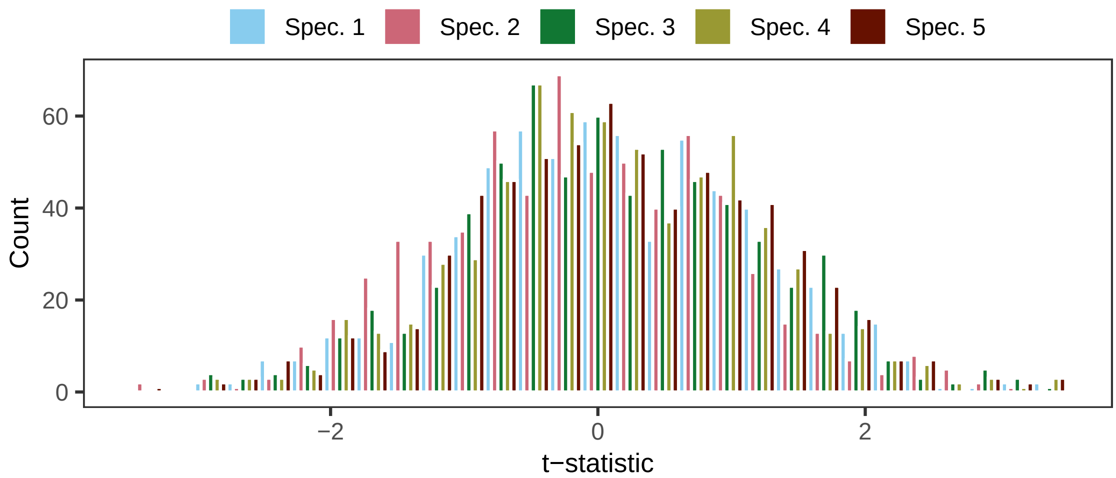

Figure 4.3: Histogram of test statistics for sentiment variable for alternate variable specifications of the Equation 4.1 regression model.

would not change our overall conclusions about the predictive capacity of news sentiment, it could point to a sensitivity to variable construction that would become relevant later.

Table 4.3 shows the correlation matrix of firms' sentiment test statistics between regression specifications. The high correlation across the different regressions indicate that the significance and direction of influence of sentiment for each firm is robust to variations in variable definition.

Figure 4.4 illustrates the distribution of adjusted r-squared values for each specification through violin plots. Consistent with the results concerning test statistics, there are no meaningful differences in explanatory power across the regressions.

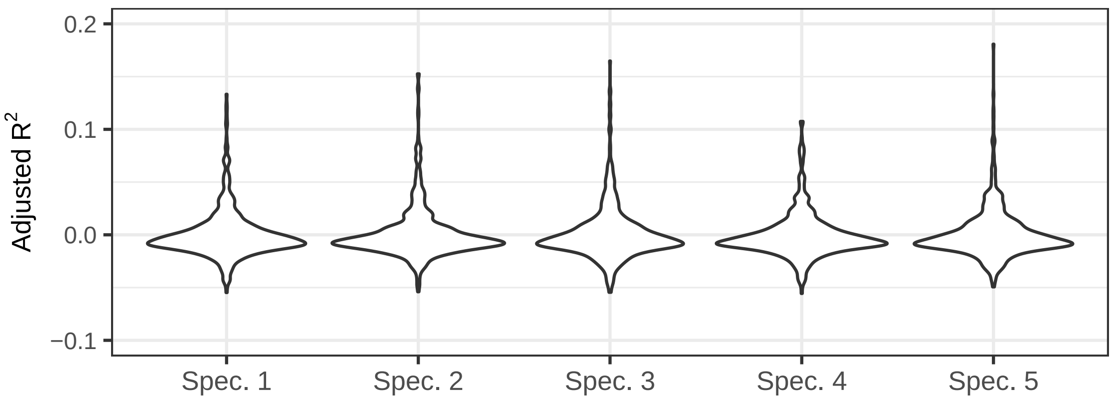

Figure 4.4: Violin plots of adjusted R^2 statistics for sentiment variable for alternate variable specifications of the Equation 4.1 regression model.

Table 4.2: Summary of sentiment coefficient t-statistics

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5
Min.	-3.4072	-4.3976	-2.9015	-2.8863	-3.2795
1st Qu.	-0.6272	-0.8628	-0.6361	-0.5903	-0.6811
Median	0.0879	-0.1435	0.0667	0.0170	0.0776
Mean	0.1272	-0.1437	0.0992	0.0922	0.1208
3rd Qu.	0.8974	0.6759	0.8703	0.8751	0.9289
Max.	3.5272	3.1690	3.3876	3.5739	3.3700
Prop. $\|t\| > 2$	0.0765	0.0841	0.0642	0.0703	0.0688
Prop. $\|t\| > 3$	0.0092	0.0138	0.0061	0.0076	0.0092

This table presents summary statistics for test-statistics of news sentiment coefficients for regression specifications one through five. Test statistics are based on Newey-West corrected standard errors.

Table 4.3: Correlation of sentiment coefficient t-statistics

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5
Spec. 1	1.00	0.91	0.96	0.91	0.99
Spec. 2	0.91	1.00	0.88	0.84	0.90
Spec. 3	0.96	0.88	1.00	0.88	0.95
Spec. 4	0.91	0.84	0.88	1.00	0.90
Spec. 5	0.99	0.90	0.95	0.90	1.00

This table represents the correlation matrix of test-statistics of news sentiment coefficients for regression specifications one through five. Test statistics are based on Newey-West corrected standard errors.

4.4.2 High-coverage Subset

It is plausible that impact of news on return is more prominent among firms with the most consistent news flow. At the same time, sparsity may empirically dampen results due to the filling of missing observations.

To investigate this further, we partition results based on news coverage, and compare firms in the high-coverage subset with those in the low-coverage subset. Specifically, we calculate the average number of news days for each firm over the sample period, and sort into three groups; those in the tercile with the highest (lowest) average news coverage are taken to be the high coverage (low coverage) subset. Breakpoints were determined after the minimum coverage threshold was applied, so that each tercile had the same number of firms.

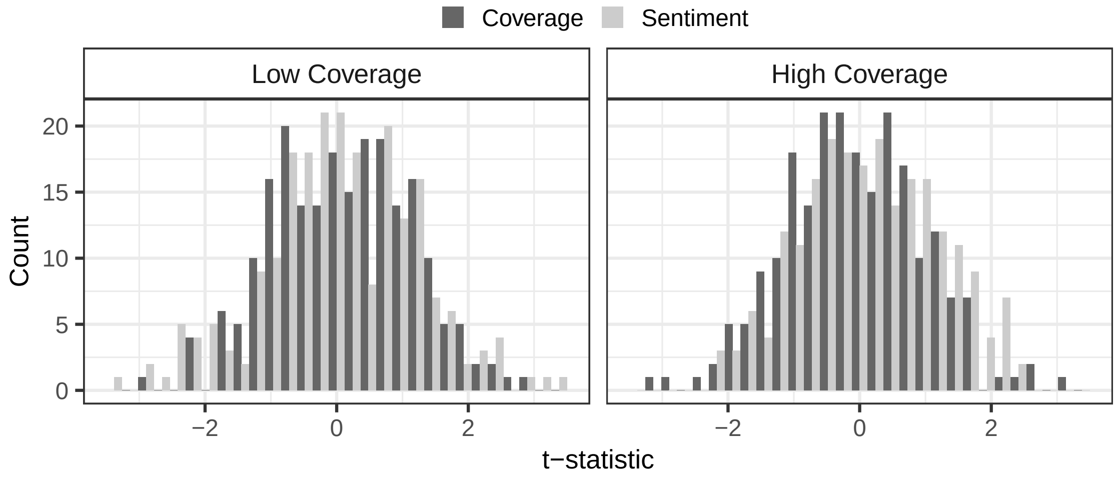

Figure 4.5: Histogram of coefficient test statistics for high coverage (top tercile) and low coverage (botom tercile) stocks, for the Equation 4.1 regression model. News coverage breakpoints are based on the average number of news days each month for that stock.

Figure 4.5 compares the histogram of test statistics for coverage and sentiment for the high- and low-coverage groups. The test statistics for both groups are centered around zero, and predominantly bound within ± 2. In short, there does not appear to be meaningful differences across the subsets or between the subsets and the aggregate results, in terms of significance. However, one difference worth noting is the higher rate of significance for the sentiment variable among low coverage firms at both levels of significance. The low coverage group also has a larger average R^2 than the high coverage group. This difference is not explained by a temporal bias between the groups, nor differences in the number of observations. It may be the case that the less frequently a firm has news, the larger the impact of that news on stock prices. In any case, the rate of significance is still low, comparable to what we would expect by chance, with only a marginally positive adjusted R^2.

4.4.3 Residual News

News coverage, and to a lesser degree, news sentiment, are partially explained in the cross-section by firm characteristics such as market capitalisation. This is discussed further in Chapter 5, but for now we want to correct for this explainable component of news in our regressions. The rationale being that the market may only respond to "unexpected" news—the difference between realised news and that expected on the basis of

Table 4.4: Regression statistics by news coverage group

			Mean		
Variable	$\lvert t \rvert > 2$	$\lvert t \rvert > 3$	Coef.	R^2	Adj. R^2
High Coverage					
Coverage	0.0545	0.0091	-0.0001	0.0158	-0.0008
Sentiment	0.0636	0.0045	0.0015		
Medium Coverage					
Coverage	0.0636	0.0000	-0.0004	0.0182	-0.0009
Sentiment	0.0636	0.0091	0.0012		
Low Coverage					
Coverage	0.0682	0.0136	0.0008	0.0304	0.0045
Sentiment	0.1045	0.0136	0.0005		

This table presents baseline regression summary statistics for firms grouped by news coverage tercile. Test statistics are based on Newey and West (1987) standard errors.

firm characteristics. To do this, we perform the following cross-sectional section regression each month:

$$\ln(1 + no.\ stories) = \beta_1 \ln(Size) + \beta_2 \ln(1 + Analyst) + \beta_3 BM + \epsilon_{Coverage}, \tag{4.2}$$

$$Sentiment = \alpha_1 \ln(1 + Analyst) + \alpha_2 BM + \epsilon_{Sentiment} \tag{4.3}$$

We then use the residual terms, $\epsilon_{Coverage}$ and $\epsilon_{Sentiment}$, in place of the original variables. Missing values are replaced with the median residual. Table 4.5 documents the summary statistics for the residual based regression. Again, rates of statistical significance are on the order of what we expect from noise, and coefficients of determination are very small.

4.4.4 Firm Characteristics

While there were no meaningful differences in predictability between high- and low-coverage firms, other attributes, such as market capitalisation, may interact with the way investors respond to media coverage and lead to greater predictability in certain subsets of the market.

Table 4.5: 1-Month Predictive Regression with Residual News

	t-statistic			
	Sentiment	Coverage	R^2	Adj. R^2
Min.	-3.3798	-3.5100	0.0000	-0.0554
1st Qu.	-0.6702	-0.6598	0.0054	-0.0109
Median	0.0777	0.0346	0.0116	-0.0046
Mean	0.0960	0.0424	0.0216	0.0011
3rd Qu.	0.8623	0.7983	0.0259	0.0066
Max.	4.2101	3.8362	0.2577	0.2247
Prop.$\|t\| > 2$	0.0750	0.0510		
Prop.$\|t\| > 3$	0.0135	0.0105		

This table presents distributional summaries of test-statistics and adjusted R^2 for the baseline predictive regression using residual news variables. Test-statistics are based on HAC standard errors Andrews (1991).

Table 4.6 presents median regression statistics for market partitions defined by market capitalisation, price-to-book ratio, analyst following, and turnover. Terciles were formed by first standardising each firm characteristic cross-sectionally in each month to prevent temporal biases resulting from mean non-stationarity, and then taking the firm's total average of the standardised variable. Terciles with approximately the same number of firms were chosen on the basis of these averages.

Considering the absolute value of test-test statistics, it is apparent that no group is associated with a significantly higher rate of statistical significance. In terms of the coefficient estimates, the greatest variation occurs across breakpoints defined by price-to-book ratio and analyst following. This is also reflected in the adjusted r-squared values. However, the magnitude of the values remain so small that it is hard to derive any meaning from the differences.

4.4.5 Subperiod Analysis

Previous studies have found the news-return relationship to be time-varying. This is not too surprising—the characteristics of news coverage itself develop, and investor consumption and response to news is entangled with the continual evolution of the digital landscape and changes in the information environment more broadly. In any case, the length of the time series is a logical and commonly used dimension for subset analysis.

Table 4.6: Predictive Regression Summary by Firm Characteristic

	Abs. t-statistic		Coefficient		
Tercile	Coverage	Sentiment	Coverage	Sentiment	Adj.r^2
Panel A: Mkt Cap					
1	0.7195	0.8249	0.0003	0.0003	-0.0047
2	0.7750	0.7301	-0.0000	0.0007	-0.0047
3	0.7889	0.7533	0.0002	0.0005	-0.0037
Panel B: Price/Book					
1	0.8147	0.8208	-0.0012	0.0006	-0.0045
2	0.7669	0.8948	0.0007	0.0009	-0.0034
3	0.6684	0.6842	0.0003	-0.0001	-0.0045
Panel C: Analyst					
1	0.8116	0.8550	0.0009	0.0005	-0.0037
2	0.7210	0.8063	-0.0008	0.0002	-0.0056
3	0.7502	0.6911	0.0003	0.0008	-0.0039
Panel D: Turnover					
1	0.7496	0.8072	0.0004	0.0001	-0.0050
2	0.8019	0.8073	0.0000	0.0011	-0.0042
3	0.7361	0.7069	0.0002	0.0002	-0.0041

This table presents median test-statistics, coefficients, and adjusted R^2 for the 4.1 regression model, for terciles sorted by firm characteristic. Terciles are numbered such that 1 corresponds to the largest values for the firm characteristic and 3 the smallest. Test-statistics are based on Newey-West corrected standard errors.

A natural breakpoint in the current sample is the GFC—by testing predictability in pre-GFC and post-GFC environments, we are able to compare two natural subperiods while also removing any distorting effect the GFC may have had on results.

The first subperiod is from the beginning of the sample (Jan 2003) to until the fall of Lehman Brothers (Sep 2008). The second subperiod is from the end of the recessionary period as defined by NBER (Jun 2009) until the end of the sample. We also remove an extra month from each subperiod such that neither the dependent nor independent variables are formed from the GFC data.

Figure 4.6 illustrates the histogram of test statistics for coverage and sentiment coefficients for the two subperiods. The proportion of sentiment test statistics with absolute

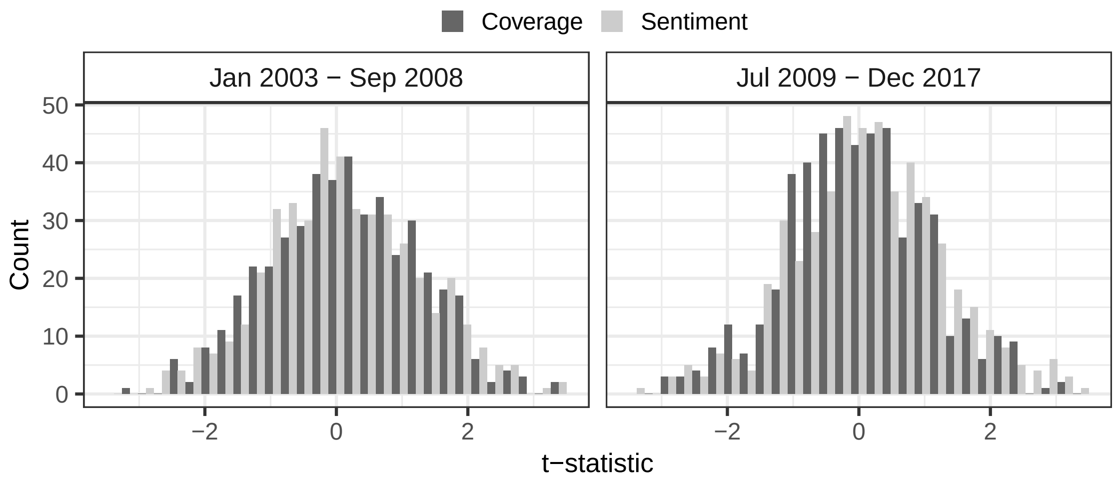

Figure 4.6: Histogram of coefficient test statistics for pre- and post-GFC subperiods, for the Equation 4.1 regression model.

value greater than two is 0.070 and 0.083 for the pre- and post-GFC periods, respectively. As shown in Table 4.7, the proportion of coverage test statistics with absolute value greater than two is 0.092 and 0.097 for the pre- and post-GFC periods, respectively. Again, there does not appear to be an important difference in the regression results between the two subsets.

Table 4.7: Regression statistics by subperiod

			Median		
Variable	$\|t\| > 2$	$\|t\| > 3$	Coef.	R^2	Adj. R^2
			Jan 2003 - Sep 2008		
Coverage	0.0699	0.0175	0.0014	0.0272	-0.0044
Sentiment	0.0917	0.0131	-0.0001		
			Jul 2009 - Dec 2017		
Coverage	0.0838	0.0058	-0.0002	0.0167	-0.0066
Sentiment	0.0975	0.0156	0.0006		

This table presents baseline regression summary statistics for pre- and post-GFC subperiods. Test statistics are based on Newey and West (1987) standard errors.

4.5 Extended Forecast Horizon

One of the main questions underlying this thesis is whether news flow is a relevant conditioning variable for forward returns over horizons comparable to other common factors, i.e. 6 - 12 months. In this section we turn our attention to a regression specification that more closely addresses that question. We use the structure of Equation 4.1 (repeated below),

$$\tilde{r}_{t:t+k} \sim \gamma_1 X_{1,t:t-k} + \gamma_2 X_{2,t:t-k} + \epsilon_{t:t+k},$$

but now with a formation period of $k = 6$. The complication of evaluating such a specification lies in the fact that our sampling frequency is higher than our aggregation frequency, resulting in deterministically overlapping observations.

The problem of inference in the presence of overlapping observations and estimation of long horizon predictability is widely recognised in the financial literature. For example, Mankiw and Shapiro (1986), Kendall (1954), Stambaugh (1993), and Stambaugh (1999) all examine the pathologies of OLS estimators in serial-dependent processes.

In addition to the generally applicable heteroskedasticity and auto-correlation consistent covariance estimators of Newey and West (1987) and Andrews (1991) among others, a number of papers develop procedures aimed more specifically at (overlapping) regressions with multi-period horizons. Richardson and Smith (1991) develop an estimation procedure that explicitly models the moment conditions arising from overlapping observations, with excellent small-sample properties. However, the assumptions under which their approach is developed do not reasonably extend to the current case.

Hodrick (1992) generalises the approach of Richardson and Smith (1991) and suggests a procedure in which implied long-horizon coefficients are derived from short-horizon regressions. The derivation is complicated and the procedure is not straight-forward to implement and scale, which appears to be why it has received relatively little uptake in the empirical finance literature. Valkanov (2003) and Hjalmarsson (2006) propose rescaling the OLS t-statistic by a factor of $1/\sqrt{j}$, where j is the number of overlapping periods. This addresses the problem of downward biased standard errors but is demonstrably conservative (Boudoukh, Israel, and Richardson, 2019).

A recent approach in a similar vain to Hodrick's (1992) estimator is presented by Britten-Jones, Neuberger, and Nolte (2011). They make use of the fact that the multi-period horizon regression equation can be represented in terms of single-period observations, and suggest inference be performed on the latter. The benefit of using the transformed regression is that it explicitly accounts for the autocorrelation induced by the deterministic aggregation scheme, leaving a less complex residual structure to be estimated and corrected for, if required. Their results suggest that OLS estimates on the transformed regression provide better small sample properties than common corrective procedures, such as White (1980), Newey and West (1987), and Hansen and Hodrick (1980), applied directly to the original regression.

Boudoukh, Israel, and Richardson (2019) argue that the efficiency gain of using overlapping observations is often negligible and is counteracted by the difficulty in getting robust estimates. Predictive regressions with overlapping regressors are the target of their paper, but their analysis is based on an AR(1) process, with AR parameter chosen based on the fraction of overlap between adjacent observations of the regressor. This process gives ith order autocorrelation of $\rho_i = \rho_1^i$, however, the structural autocorrelation induced by overlapping observations is $\rho_i = (j\tilde{\ }i)/j$ for $1 \leq i \leq j-1$, and zero otherwise, which is a much less persistent series.

As the efficiency of different estimators depends significantly on the dynamic structure of the regressor (Boudoukh and Richardson, 1994), and no results exist for our particular case, we perform a series of simulations with the aim of choosing a reasonable (existing) approach to estimate our regression.

4.5.1 Monte Carlo Simulations

Here we perform Monte Carlo simulations in which the underlying data generating process of forward returns takes the following form:

$$\tilde{r}_{t:t+k} \sim \gamma_1 X_{1,t:t-k} + \gamma_2 X_{2,t:t-k} + \epsilon_{t:t+k},$$

where $\tilde{r}_{t:t+k}$, is the k period forward log return, and the regressors reflect the trailing k period aggregate given by $X_{t:t-k} = \sum_{i=0}^{k-1} X_{t-i}$. This 'momentum' DGP is modeled through the following single-period return process:

$$\tilde{r}_t \sim \gamma_1 X_{1,t-k} + \gamma_2 X_{2,t-k} + \epsilon_{t+1}$$

We perform this simulation with $k = 6$ under five scenarios:

1. No predictability at any horizon ($\gamma_1 = \gamma_2 = 0$). X_1 and X_2 are independent $N(0,1)$ variables, and ϵ_{t+1} is distributed $N(0,1)$
2. No predictability at any horizon ($\gamma_1 = \gamma_2 = 0$). X_1 is an ARIMA (1, 0, 1) process with unit variance, AR parameter 0.1 and MA parameter -0.9. This is representative of our differenced sentiment score. X_2 is an ARIMA (2, 0, 1) process with unit variance, AR parameters (-0.3, -0.3), and MA parameter -0.7. This is representative of our differenced news coverage score. ϵ_{t+1} is distributed $N(0,1)$.

3. Predictability given by ($\gamma_1 = \gamma_2 = 0.1$). X_1 and X_2 are independent $N(0,1)$ variables as in (1), and $\epsilon_{t+1} \sim N(0, \sigma_\epsilon)$, with σ_ϵ chosen so that R^2 is 20%.

4. Predictability given by ($\gamma_1 = \gamma_2 = 0.1$). X_1 and X_2 follow ARIMA processes as in (2), and $\epsilon_{t+1} \sim N(0, \sigma_\epsilon)$, with σ_ϵ chosen so that R^2 is 20%.

5. No predictability at any horizon ($\gamma_1 = \gamma_2 = 0$). X_1 and X_2 are independent $N(0,1)$ variables, and disturbances are heteroskedastic through $\epsilon_{t+1} \sim X_{1,t} N(0,1)$.

The specifications for the simulated ARMA dependent variables were determined by first fitting an ARMA model to the news series of each stock. The model order for each individual series was chosen on the basis of AIC using a stepwise selection procedure (Hyndman and Khandakar, 2008). The model order to use for the simulation was selected as the (rounded) average order across all models chosen by the stepwise selection procedure. A model of this order was then fitted to the news series of each stock. The most common of these models up to the sign of the coefficients was then selected, and the median coefficient values were taken to be the simulation coefficients. This process was applied separately for news sentiment and news coverage.

In each simulation case, we present results of covariance estimators applied to the overlapping OLS regression: the standard 'OLS' estimator, the 'HC3' refinement of White's (1980) heteroskedasticity-consistent estimator (Long and Ervin, 2000), the 'NW' estimator (Newey and West, 1987) with automatic bandwidth selection described in Newey and West (1994), and the HAC estimator with quadratic spectral kernel as per Andrews (1991) ('HAC'). We then present results for the Britten-Jones, Neuberger, and Nolte (2011) transformed regression using the same four estimators: 'OLS', 'HC3', 'NW', and 'HAC'. We use implementations of the covariance estimators described in Zeileis (2004). Each scenario was performed separately with $T = 100$ and $T = 180$. 10,000 simulations were performed for each scenario.

For each covariance estimator and each scenario we report the bias, standard deviation, root mean squared error (RMSE), and the true confidence levels of the nominal 99%, 95%, and 90% regression coefficient confidence intervals. For simplicity of presentation, results are detailed for γ_1 only, as results for γ_2 were essentially the same as for γ_1 across all values. Results for γ_2 are available upon request.

Table 4.8 shows results for simulation case 1. As expected, the overlapping OLS estimator is strongly downward biased. This is reduced by the NW estimator, but remains significantly biased. The smallest bias is achieved by the transformed regression with OLS estimator, followed by the HAC estimator applied to the original regression. These results are reflected for both $T = 100$ and $T = 180$.

Table 4.8: Monte Carlo Simulations: No return predictability, IID Regressors.

Obs.	Variance Est.	Bias	Std.	RMSE	99%	95%	90%
		Overlapping Regression					
180	OLS	-0.082	0.012	0.083	0.793	0.664	0.581
	HC3	-0.083	0.015	0.084	0.785	0.658	0.575
	NW	-0.033	0.030	0.045	0.946	0.865	0.798
	HAC	0.014	0.055	0.057	0.982	0.942	0.895
		Transformed Regression					
	OLS	-0.002	0.017	0.017	0.988	0.948	0.898
	HC3	0.000	0.021	0.021	0.988	0.948	0.901
	NW	-0.017	0.030	0.034	0.970	0.906	0.850
	HAC	-0.005	0.027	0.028	0.984	0.935	0.880
		Overlapping Regression					
100	OLS	-0.115	0.024	0.117	0.792	0.662	0.576
	HC3	-0.116	0.027	0.119	0.780	0.651	0.566
	NW	-0.065	0.047	0.081	0.910	0.816	0.745
	HAC	0.004	0.162	0.162	0.964	0.910	0.861
		Transformed Regression					
	OLS	-0.003	0.032	0.032	0.991	0.946	0.903
	HC3	0.002	0.040	0.040	0.991	0.948	0.904
	NW	-0.036	0.051	0.062	0.951	0.880	0.818
	HAC	-0.010	0.050	0.051	0.982	0.931	0.875

The single period data generating process for this specification is $\tilde{r}_t \sim \gamma_1 X_{1,t-k} + \gamma_2 X_{2,t-k} + \epsilon_{t+1}$, where $X_{1,t}$ and $X_{2,t}$ are $\sim N(0,1)$, $\gamma_1 = 0$, and $\gamma_2 = 0$. The overlapping regression specification is $\tilde{r}_{t:t+k} \sim \gamma_1 X_{1,t:t-k} + \gamma_2 X_{2,t:t-k} + \epsilon_{t:t+k}$, with k=6. Results listed refer to estimates of γ_1. For each estimator, the bias, standard error, root mean square error, and true confidence levels of 99%, 95% and 90% are shown.

Table 4.9 shows results for simulation case 2, where there is auto-dependency in the pre-aggregation regressors. Again, the transformed regression with OLS estimates exhibits the smallest bias and RMSE. Good results are also achieved in the overlapping regression with HAC estimator. It can be seen that use of NW or HAC estimators on the transformed regression have a detrimental effect on standard errors.

Table 4.10 includes results for simulation case 3, where the regressors are predictive and normally distributed iid. In this case, best results are achieved for $T = 100$ and $T = 180$ with the transformed regression and HAC estimator. For $T = 180$ this is followed by the original regression with HAC estimator and then the transformed regression with OLS estimator. For $T = 100$, the transformed regression with HAC estimator is followed

Table 4.9: Monte Carlo Simulations: No return predictability, ARIMA Regressors.

Obs.	Variance Est.	Bias	Std.	RMSE	99%	95%	90%
			Overlapping Regression				
180	OLS	-0.024	0.022	0.033	0.976	0.915	0.852
	HC3	-0.023	0.026	0.034	0.975	0.916	0.851
	NW	-0.017	0.034	0.038	0.978	0.922	0.860
	HAC	-0.002	0.040	0.040	0.986	0.942	0.888
			Transformed Regression				
	OLS	0.000	0.018	0.018	0.990	0.954	0.901
	HC3	0.003	0.024	0.024	0.989	0.953	0.903
	NW	-0.012	0.032	0.034	0.978	0.925	0.870
	HAC	-0.003	0.031	0.031	0.984	0.939	0.887
			Overlapping Regression				
100	OLS	-0.035	0.040	0.054	0.975	0.910	0.849
	HC3	-0.030	0.048	0.057	0.974	0.913	0.850
	NW	-0.037	0.057	0.068	0.967	0.901	0.838
	HAC	-0.003	0.074	0.074	0.982	0.935	0.881
			Transformed Regression				
	OLS	0.001	0.034	0.034	0.992	0.952	0.904
	HC3	0.007	0.045	0.046	0.990	0.954	0.906
	NW	-0.028	0.055	0.061	0.967	0.909	0.849
	HAC	-0.007	0.056	0.057	0.982	0.935	0.882

The single period data generating process for this specification is $\tilde{r}_t \sim \gamma_1 X_{1,t-k} + \gamma_2 X_{2,t-k} + \epsilon_{t+1}$. $X_{1,t}$ is an ARIMA (1, 0, 1) process with AR=0.1, MA=-0.9. $X_{2,t}$ is an ARIMA (2, 0, 1), process with AR=(-0.3, -0.3), and MA=-0.7. $\gamma_1 = 0$, and $\gamma_2 = 0$. The overlapping regression specification is $\tilde{r}_{t:t+k} \sim \gamma_1 X_{1,t:t-k} + \gamma_2 X_{2,t:t-k} + \epsilon_{t:t+k}$, with k=6. Results listed refer to estimates of γ_1. For each estimator, the bias, standard error, root mean square error, and true confidence levels of 99%, 95% and 90% are shown.

by the transformed regression with OLS estimator and then overlapping regression with HAC estimator.

Table 4.11 includes results for simulation case 4, where the regressors are predictive and auto-dependent. In this case, best results are achieved for $T = 100$ and $T = 180$ with the transformed regression and HAC estimator, followed by the original regression with HAC estimator and then the transformed regression with OLS estimator.

Table 4.12 shows results for simulation case 5, with no predictability, iid regressors, and heteroskedastic errors. As expected, the heteroskedasticity increases the bias of most estimators. In the original regression, the HC correction does not go far in correcting for this

Table 4.10: Monte Carlo Simulations: Return predictability, IID Regressors.

Obs.	Variance Est.	Bias	Std.	RMSE	99%	95%	90%
			Overlapping Regression				
180	OLS	-0.023	0.003	0.023	0.788	0.659	0.574
	HC3	-0.023	0.004	0.024	0.781	0.655	0.572
	NW	-0.009	0.009	0.013	0.946	0.865	0.792
	HAC	0.004	0.015	0.016	0.983	0.941	0.899
			Transformed Regression				
	OLS	0.004	0.005	0.006	0.995	0.967	0.927
	HC3	0.004	0.006	0.008	0.995	0.967	0.927
	NW	-0.003	0.009	0.009	0.974	0.919	0.860
	HAC	0.001	0.008	0.008	0.989	0.950	0.901
			Overlapping Regression				
100	OLS	-0.033	0.007	0.033	0.792	0.658	0.572
	HC3	-0.033	0.008	0.034	0.781	0.651	0.566
	NW	-0.018	0.014	0.023	0.910	0.818	0.740
	HAC	0.001	0.036	0.036	0.964	0.910	0.860
			Transformed Regression				
	OLS	0.005	0.010	0.011	0.995	0.968	0.927
	HC3	0.006	0.012	0.014	0.995	0.968	0.931
	NW	-0.008	0.016	0.018	0.949	0.890	0.830
	HAC	0.001	0.015	0.015	0.985	0.944	0.895

The single period data generating process for this specification is $\tilde{r}_t \sim \gamma_1 X_{1,t-k} + \gamma_2 X_{2,t-k} + \epsilon_{t+1}$, where $X_{1,t}$ and $X_{2,t}$ are $\sim N(0,1)$, $\gamma_1 = 0.1$, and $\gamma_2 = 0.1$. The overlapping regression specification is $\tilde{r}_{t:t+k} \sim \gamma_1 X_{1,t:t-k} + \gamma_2 X_{2,t:t-k} + \epsilon_{t:t+k}$, with k=6. Results listed refer to estimates of γ_1. For each estimator, the bias, standard error, root mean square error, and true confidence levels of 99%, 95% and 90% are shown.

bias unless the deterministic autocorrelation is corrected for, either through HAC standard errors or by transforming the regression. For this DGP, the HC3 estimator applied to the transformed regression is essentially unbiased and has the closest true confidence intervals, followed by the OLS regression with HAC estimator, and thirdly the transformed regression with HAC estimator.

In summary, the transformed regression with OLS estimator, and the overlapping regression with HAC standard errors gave good results across each of the cases. Additionally, the HC3 estimator improves standard errors in the presence of heteroskedasticity with very little degradation in its absence. The estimator which exhibits the lowest bias and tightest confidence intervals under the case of no predictability is the transformed regres-

Table 4.11: Monte Carlo Simulations: Return predictability, ARIMA Regressors.

Obs.	Variance Est.	Bias	Std.	RMSE	99%	95%	90%
			Overlapping Regression				
180	OLS	-0.008	0.006	0.010	0.976	0.909	0.843
	HC3	-0.007	0.007	0.010	0.975	0.908	0.843
	NW	-0.006	0.009	0.011	0.977	0.917	0.856
	HAC	-0.001	0.011	0.011	0.985	0.939	0.883
			Transformed Regression				
	OLS	0.005	0.006	0.008	0.996	0.969	0.928
	HC3	0.006	0.007	0.009	0.996	0.968	0.927
	NW	-0.001	0.009	0.009	0.986	0.940	0.888
	HAC	0.001	0.009	0.009	0.989	0.950	0.900
			Overlapping Regression				
100	OLS	-0.010	0.012	0.015	0.973	0.910	0.848
	HC3	-0.009	0.014	0.016	0.973	0.913	0.849
	NW	-0.011	0.016	0.020	0.966	0.900	0.836
	HAC	-0.001	0.021	0.021	0.982	0.934	0.879
			Transformed Regression				
	OLS	0.008	0.011	0.013	0.996	0.970	0.934
	HC3	0.010	0.014	0.017	0.996	0.967	0.933
	NW	-0.003	0.016	0.017	0.976	0.926	0.874
	HAC	0.001	0.016	0.016	0.985	0.943	0.895

The single period data generating process for this specification is $\tilde{r}_t \sim \gamma_1 X_{1,t-k} + \gamma_2 X_{2,t-k} + \epsilon_{t+1}$. $X_{1,t}$ is an ARIMA (1, 0, 1) process with AR=0.1, MA=-0.9. $X_{2,t}$ is an ARIMA (2, 0, 1), process with AR=(-0.3, -0.3), and MA=-0.7. $\gamma_1 = 0.1$, and $\gamma_2 = 0.1$. The overlapping regression specification is $\tilde{r}_{t:t+k} \sim \gamma_1 X_{1,t:t-k} + \gamma_2 X_{2,t:t-k} + \epsilon_{t:t+k}$, with k=6. Results listed refer to estimates of γ_1. For each estimator, the bias, standard error, root mean square error, and true confidence levels of 99%, 95% and 90% are shown.

sion with OLS standard errors. This estimator also exhibits the lowest RMSE across all regressions, even when its bias is not. We continue with the transformed regression and HC3 standard errors, to account for cases of heteroskedasticity[1].

[1] In Breusch and Pagan (1979) tests against heteroskedasticity in the residuals of the transformed regression, with the null of no heteroskedasticity was rejected in 14% of cases at the 5% level of significance.

4.5.2 Extended Horizon Results

Table 4.13 includes the distributional summary of test-statistics and r-squared values for the six-month regression. As was done for the one-month regression, we include results obtained from using randomly resampled covariates in place of the realised observations for a synthetic point of comparison.

The proportion of significant test-statistics is similar between the two sets, and both are close to what we would expect rejection rates to be at the associated levels of significance. The R^2 statistics between the two sets are also very similar, and both are similar to the baseline regression figures. Based on the upper bounds of R^2 statistics we should expect to see meaningfully higher values for the extended horizon results (Ross, 2009; Zhou, 2010; Huang and Zhou, 2017). The reason we do not here is because we used the transformed regression, which gives numerically identical coefficient estimates to the overlapping specification but not the variance explained.

Based on these results, news momentum does not appear to be longitudinally predictive of forward returns at the six month horizon.

Table 4.12: Monte Carlo Simulations: No return predictability, IID Regressors, Heteroskedasticity.

Obs.	Variance Est.	Bias	Std.	RMSE	99%	95%	90%
			Overlapping Regression				
180	OLS	-0.095	0.012	0.096	0.760	0.621	0.546
	HC3	-0.092	0.015	0.093	0.776	0.640	0.562
	NW	-0.036	0.033	0.049	0.939	0.863	0.798
	HAC	0.019	0.354	0.355	0.979	0.937	0.896
			Transformed Regression				
	OLS	-0.015	0.014	0.021	0.980	0.924	0.868
	HC3	0.002	0.023	0.023	0.990	0.948	0.902
	NW	-0.018	0.032	0.036	0.968	0.908	0.854
	HAC	-0.004	0.030	0.030	0.982	0.933	0.887
			Overlapping Regression				
100	OLS	-0.136	0.022	0.138	0.741	0.611	0.529
	HC3	-0.133	0.027	0.136	0.751	0.619	0.537
	NW	-0.077	0.049	0.091	0.897	0.798	0.727
	HAC	0.006	0.707	0.707	0.954	0.895	0.841
			Transformed Regression				
	OLS	-0.024	0.028	0.036	0.980	0.925	0.863
	HC3	0.000	0.042	0.042	0.990	0.947	0.897
	NW	-0.043	0.054	0.069	0.941	0.869	0.806
	HAC	-0.015	0.052	0.054	0.977	0.919	0.865

The single period data generating process for this specification is $\tilde{r}_t \sim \gamma_1 X_{1,t-k} + \gamma_2 X_{2,t-k} + \epsilon_{t+1}$, where $X_{1,t}$ and $X_{2,t}$ are $\sim N(0,1)$, $\gamma_1 = 0$, and $\gamma_2 = 0$. Heteroskedasticity is introduced via $\epsilon_{t+1} \sim X_{1,t} \cdot N(0,1)$. The overlapping regression specification is $\tilde{r}_{t:t+k} \sim \gamma_1 X_{1,t:t-k} + \gamma_2 X_{2,t:t-k} + \epsilon_{t:t+k}$, with k=6. Results listed refer to estimates of γ_1. For each estimator, the bias, standard error, root mean square error, and true confidence levels of 99%, 95% and 90% are shown.

Table 4.13: 6-Month Predictive Regression Summary

	t-statistic			
	Sentiment	Coverage	R^2	Adj. R^2
Panel A: Realised covariates				
Min.	-2.7600	-3.2627	0.0000	-0.0846
1st Qu.	-0.6988	-0.7772	0.0053	-0.0197
Median	0.0240	-0.0907	0.0130	-0.0121
Mean	-0.0112	-0.0971	0.0226	-0.0088
3rd Qu.	0.6963	0.5719	0.0281	-0.0011
Max.	5.0849	2.8001	0.3640	0.3305
Prop.$\|t\| > 2$	0.0549	0.0565		
Prop.$\|t\| > 3$	0.0031	0.0063		
Panel B: Random covariates				
Min.	-3.0561	-4.2830	0.0001	-0.0831
1st Qu.	-0.6288	-0.7302	0.0051	-0.0215
Median	0.0331	-0.0769	0.0119	-0.0124
Mean	0.0296	-0.0819	0.0202	-0.0116
3rd Qu.	0.7045	0.5275	0.0242	-0.0035
Max.	2.6789	2.9219	0.2157	0.1553
Prop.$\|t\| > 2$	0.0435	0.0373		
Prop.$\|t\| > 3$	0.0016	0.0031		

This table presents distributional summaries of test-statistics and adjusted R^2 for the predictive regression represented by Equation 4.1 with six-month prediction and formation periods. Panel A details results using observed news variables. Panel B details results using simulated noise as in place of the news variables. Test-statistics are based on OLS estimates of the Britten-Jones, Neuberger, and Nolte (2011) transformed regression.

4.6 VAR Analysis

So far, we have assumed exogeneity of the news variables in evaluating their predictive capacity for future returns. By taking advantage of vector autoregression models (Sims, 1980) we can relax this assumption while obtaining a richer understanding of the joint dynamics governing news coverage, sentiment, and stock return. Unlike autoregressive distributed lag models, VARs are natural tools for forecasting and are widely used in the economic literature for this purpose (Christiano, 2012).

VAR have also been used heavily in the news sentiment literature, but rarely at the firm level. Ahmad et al. (2016) applies individual VAR models to 20 US non-financial firms and Strauß, Vliegenthart, and Verhoeven (2016) applied individual VAR models to 21 firms on the Amsterdam stock exchange. Both studies were performed at the daily level. To our knowledge, no other studies have examined the relationship between news and stock prices using firm-level VARs at the scale of the current analysis.

It was stated earlier that based on unit root tests, monthly news sentiment and coverage are often (in 35% and 33% of firms, respectively) I(1) processes. If sentiment and coverage form a cointegrating relationship with price, then vector-error-correction models (VECM), rather than VAR, would be an appropriate fit. We therefore performed Engle-Granger cointegration tests for price, news sentiment, and news coverage, across all stocks meeting the filtering requirements. The null hypothesis of no cointegration was rejected at the 5% level in only 1.3% of firms, and at the 1% level in zero firms, indicating that VECM is inappropriate.

The VAR to be fit to each firm can be expressed as follows:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_1\mathbf{y}_{t-p} + \mathbf{u}_t$$

where $\mathbf{y}_t$ is a 3×1 vector of endogenous variables (return, sentiment, and coverage), $\mathbf{u}_t$ is a spherical disturbance term of the same dimension, and p is the lag order. The coefficient matrices $A_1, \ldots, A_p$ are of dimension 3×3. We continue to work with first differences of the news variables to ensure that all variables are of the same integration order (return being I(0)).

The next issue is selection of the lag order. In order to compare results across all models, we require the same lag order for each firm. With a maximum lag of six, the average optimum lag (in terms of AIC) was three, and with a maximum lag of 12, the average optimum lag was 5, with approximately 75% of firms having an optimum lag order of six or less. Experimenting further, we found that the proportion of firms with higher optimum lag orders increased as the maximum lag of the search increased. As the number of parameters increases exponentially with the lag length, and we are dealing with at most 180 observations per firm, a lag order of six was chosen as a reasonable compromise between bias and variance. Ahmad et al. (2016) followed a similar process to select their maximum lag order. As a VAR with three endogenous variables and six lags contains 54 parameters, we removed firms with less than 65 observations. VAR equations were fit via OLS using the implementation described by Pfaff (2008).

Having selected and fit the models, the next issue is how to calculate standard errors for the various significance tests. HAC covariance estimators such as Newey and West (1987) require estimation (and subsequent summation) of many autocovariances from the estimated residuals. In absence of strong positive autocorrelation, biases in these estimates can result in overall covariance estimates that are downwardly biased relative to the "true" covariance and even to estimators that simply ignore the autocorrelation (as we saw in the earlier Monte Carlo analysis). This is to say that use of HAC covariance estimates is not necessarily the conservative choice, particularly when the number of parameters is large relative to the number of observations, as is the current case with our VAR models. Heteroskedasticity consistent (HC) estimators such as (White, 1980) are not as noisy, as they do not estimate the cross-product terms required to account for autocorrelation. Depending on the sample size and presence of autocorrelation and heteroskedasticity, HC estimators may offer the most robust compromise between OLS and HAC estimators.

In multivariate ARCH-LM (Lütkepohl, 2006; Engle, 1982) tests with five lags, the null hypothesis of no heteroskedasticity was rejected in 19% of firms at the 5% level. In adjusted Breusch-Godfrey tests (Edgerton and Shukur, 1999) with five lags, the null of no autocorrelation was rejected at the 5% level in 31% of firms. It is not intuitive to us what these intermediate nominal rates of residual autocorrelation and heteroskedasticity imply about the relative bias or noisiness of different estimators for our system, so we compared the standard errors for OLS, HC (White, 1980; Cribari-Neto and Silva, 2011), and HAC (Andrews, 1991) estimators applied to the VAR models.

The first row of Table 4.14 displays the mean standard error across all parameters in each of the return, sentiment, and coverage equations. The second and third rows display the average ratio of the HC and HAC standard errors relative to the OLS standard errors, respectively. Consistent with our concern of applying HAC estimates to our relatively large-parameter/low-sample-size models, HAC standard errors are in fact smaller than those of the OLS estimator. On the other hand, accounting for heteroskedasticity alone significantly increases the standard error estimates on average. We continue with HC standard errors as they are the most conservative in this case, though results do not qualitatively change when using OLS estimates.

Table 4.14: Mean standard errors for VAR submodels

	Return	Sentiment	Coverage
OLS	0.040	0.468	0.411
HC/OLS	1.213	1.157	1.145
HAC/OLS	0.998	0.983	0.978

This table lists the mean standard errors for each of the VAR equations for the OLS estimator (first row), and the mean ratio of HC4m (Cribari-Neto and Silva, 2011) and HAC (Andrews, 1991) standard errors relative to the OLS estimates (second and third rows, respectively).

Parameter significance

Table 4.15 presents the mean estimated coefficients for the full VAR model. Panels are separated by the dependent variable of the sub model, and coefficients are grouped by regressor variable (first column) and lag (columns two - seven). In light of our results from the baseline regression, average estimates provide information on the consistency of the relationship across firms. An average close to zero suggests that the direction of the effect is unpredictable prior to conditioning on the firm. The largest aggregate effects appear to be the negative autocorrelation in sentiment and news coverage, and the largest effect in the return equation are the second and sixth lags of return. The aggregate effect of return on news sentiment appears to be larger than the effect of news sentiment on return. Coverage appears to be driven by coverage and sentiment.

Table 4.16 presents rates of statistical significance of the estimated coefficients for the full VAR model, at the 5% level of significance. For the return model, rates of significance are low for all variables. The persistence of sentiment and coverage is reflected in the significance rates, with the majority of firms exhibiting statistically significant autocorrelation in these variables up to the fourth lag. The highest rate of cross-variable significance is from the first lag of returns in the sentiment equation, at 11.8%.

Joint tests

As we are interested in relatively low-frequency news-return relationships and potential news-driven analogues to stock price momentum, we are less concerned with the impact of individual lags of news variables than that of the entire formation period as a whole. To represent this interest in terms of statistical inference, we perform a joint hypothesis test of the coefficients for all lags of news sentiment in the return equation of the

Table 4.15: Mean Parameter Estimates: Individual VAR Coefficients

	Lag					
Regressor	1	2	3	4	5	6
Panel A: Return Equation						
Return	-0.015	-0.040	-0.010	-0.006	-0.007	-0.039
Sentiment	0.024	0.021	0.026	0.017	0.008	0.005
Coverage	0.000	0.004	-0.004	-0.013	-0.019	-0.004
Panel B: Sentiment Equation						
Return	0.065	0.021	0.003	-0.012	-0.019	-0.019
Sentiment	-0.711	-0.542	-0.380	-0.277	-0.182	-0.090
Coverage	0.005	0.006	-0.022	-0.003	-0.009	-0.032
Panel C: Coverage Equation						
Return	-0.002	-0.002	-0.001	-0.007	-0.009	0.000
Sentiment	-0.023	-0.024	-0.042	-0.033	-0.013	-0.015
Coverage	-0.841	-0.728	-0.408	-0.354	-0.275	-0.047

This table lists mean paraneter estimate for each term in the full VAR equation. Each panel displays the model for that regressor, in terms of its own lags and the lags of the other regressors. Standard errors are based on heteroscedasticity-consistent covariance estimates (Cribari-Neto and Silva, 2011).

VAR. We repeat this test for each explanatory variable in each VAR equation to provide a full picture of the joint cross-relationships. Table 4.17 lists the rates of significance for each of the joint hypothesis tests, at the 5% and 1% levels. Firstly, we can see that each explanatory variable is more reliably explained by its own lags than those of the other variables. We can also see that return and news coverage tend to be more relevant for explaining sentiment than the other way around. For instance, at the 5% level, news sentiment coefficients are jointly significant in the return equation 6.7% of the time, while return coefficients are jointly significant in the sentiment equation 10.2% percent of the time. Consistent with our results from the momentum-style regression, we do not have compelling evidence that news drives return at the firm-level.

Causality Tests

To further disentangle the dynamic interactions between returns and news flow, and to more precisely investigate whether news sentiment and coverage contribute to return

Table 4.16: Significance Rates: Individual VAR Coefficients

	Lag					
Regressor	1	2	3	4	5	6
Panel A: Return Equation						
Return	0.022	0.031	0.031	0.049	0.016	0.031
Sentiment	0.033	0.024	0.035	0.031	0.026	0.037
Coverage	0.022	0.028	0.033	0.035	0.033	0.041
Panel B: Sentiment Equation						
Return	0.118	0.037	0.033	0.035	0.031	0.028
Sentiment	0.994	0.945	0.740	0.488	0.289	0.093
Coverage	0.035	0.053	0.067	0.055	0.047	0.041
Panel C: Coverage Equation						
Return	0.057	0.026	0.033	0.033	0.028	0.028
Sentiment	0.071	0.045	0.079	0.053	0.045	0.037
Coverage	0.992	0.978	0.720	0.636	0.526	0.069

This table lists the fraction of significant p-values for each coefficient estimate in the full VAR equation. Each panel displays the model for that regressor, in terms of its own lags and the lags of the other regressors. Standard errors are based on heteroscedasticity-consistent covariance estimates (Cribari-Neto and Silva, 2011).

forecasts above and beyond returns alone, we perform Granger causality tests between each of the VAR variables. Table 4.18 details the fraction of firms for which each regressor (table columns) statistically significantly Granger-causes the remaining variables. News sentiment and coverage Granger-cause returns in 10.4% of cases at a 5% level of significance and in 3.3% of firms at a 1% level of significance. At a similar rate, returns are shown to Granger-cause news sentiment and news coverage in 10.8% and 3.2% of firms at 5% and 1% levels of significance, respectively. Consistent with the joint coefficient test results, news coverage appears to drive news sentiment more than the other way around, with the highest rate of Granger causality originating from coverage at both levels of significance.

Also provided in Table 4.18 are analogous results for instantaneous causality, which captures logical, rather than causal, dependency (Jaynes, 2003, pp. 92). At both levels of significance we see high rates of instantaneous causality relative to Granger causality. One way to interpret this is that contemporaneous, rather than predictive relationships

Table 4.17: Significance Rates: VAR Joint Hypothesis Tests

	Dependent Variable		
Regressor	Return	Sentiment	Coverage
$\alpha = 0.05$			
Return	0.183	0.102	0.057
Sentiment	0.067	0.998	0.112
Coverage	0.061	0.134	0.998
$\alpha = 0.01$			
Return	0.087	0.030	0.012
Sentiment	0.014	0.988	0.031
Coverage	0.014	0.049	0.998

This table lists the fraction of significant p-values under the null hypothesis that all coefficients of the regressor variable in explaining the dependent variable are zero. Rejection rates are provided for 95% and 99% confidence intervals. Test are based on heteroscedasticity-consistent covariance estimates (Cribari-Neto and Silva, 2011).

dominate the system dynamics. The highest rate of instantaneous causality is from news sentiment, in providing information about return and coverage. Which, given the rates for sentiment and coverage causality in the direction of returns, appears to be largely driven by the sentiment-coverage dynamic.

Forecast error decomposition

Another angle through which to investigate predictive content given our VAR results is through forecast error variance decomposition (FEVD), which separates VAR innovations into a set of uncorrelated disturbances of the component variables. FEVD allows users to analyse the contribution of one variable to the h-step forecast error variance (MSE) of another, thereby isolating its relevance as a forecasting variable. Table 4.19 documents the mean forecast error decompositions for returns, news sentiment, and news coverage over a 12-month forecast horizon. These numbers are also visualised as bar charts in Figure 4.7.

We can see that after nine months the relative contributions have reached their terminal values. On average, by the end of the forecast horizon, 91% of the forecast error variance of returns is accounted for by its own innovations, and 9% by combined sentiment and

Table 4.18: Significance Rates: VAR Causality Tests

	Cause			
Test	Return	Sentiment	Coverage	Sentiment & Coverage
$\alpha = 0.05$				
Granger	0.108	0.126	0.140	0.104
Instantaneous	0.394	0.531	0.287	0.287
$\alpha = 0.01$				
Granger	0.031	0.033	0.049	0.033
Instantaneous	0.203	0.315	0.138	0.138

This table lists the fraction of significant p-values under the null hypothesis of no causality from the regressor (column variable) to the remaining dependent variable(s), for Granger and instaneous causality tests (Granger, 1969; Lütkepohl, 2006). Rejection rates are provided for 95% and 99% confidence intervals. Test are based on heteroscedasticity-consistent covariance estimates (Cribari-Neto and Silva, 2011).

coverage innovations. At the six month horizon, these values 93% and 7% respectively. Similar results are found for news sentiment and coverage—at a six month horizon, 89% of news coverage forecast error variance and 90% of sentiment forecast error variance are accounted for by their own values, with only 13% being driven by innovations of the other variables at the nine month horizon. In line with our previous results, return appears to be a marginally stronger predictor of news than the other way around, with greater forecast error variance attributable to return than either sentiment or coverage at all lags.

Table 4.19: VAR Forecast Error Decomposition

	Forecast Horizon (Months)											
Regressor	1	2	3	4	5	6	7	8	9	10	11	12
Panel A: Return Forecast												
Return	1.00	0.98	0.97	0.95	0.94	0.93	0.92	0.91	0.91	0.91	0.91	0.91
Sentiment	0.00	0.01	0.02	0.02	0.03	0.04	0.04	0.04	0.05	0.05	0.05	0.05
Coverage	0.00	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.05
Panel B: Sentiment Forecast												
Return	0.04	0.04	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.07
Sentiment	0.96	0.96	0.94	0.92	0.91	0.90	0.89	0.88	0.87	0.87	0.87	0.87
Coverage	0.00	0.01	0.02	0.02	0.03	0.04	0.05	0.05	0.06	0.06	0.06	0.06
Panel C: Coverage Forecast												
Return	0.01	0.02	0.03	0.03	0.04	0.05	0.05	0.05	0.05	0.05	0.06	0.06
Sentiment	0.03	0.03	0.04	0.05	0.05	0.06	0.07	0.07	0.08	0.08	0.08	0.08
Coverage	0.96	0.95	0.94	0.92	0.91	0.89	0.88	0.87	0.87	0.87	0.87	0.86

This table lists the mean proportion of forecast error variance explained by each of the VAR regressors, up to a horizon of 12 months. Panels indicate the variable being forecast.

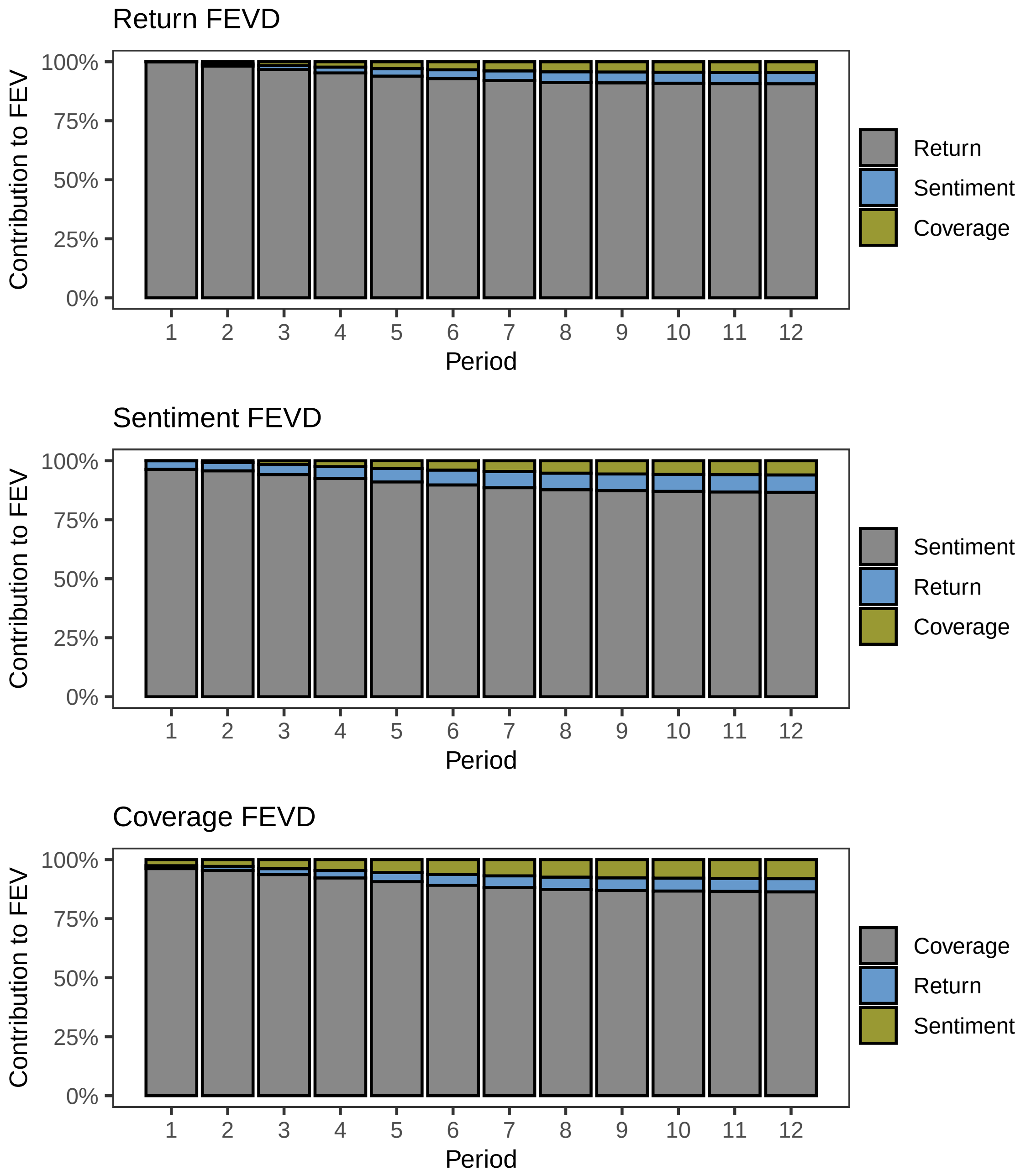

Figure 4.7: Aggregate forecast error variance decomposition for return, news sentiment, and news coverage forecasts, over a 12 month horizon.

4.7 Discussion

In this analysis we investigated the extent to which news sentiment and news coverage are useful conditioning variables for future excess returns at the firm level, for historical constituents of the S&P 500 index. We began with a one-month predictive regression model in which forward return was regressed against single-period lags of news sentiment and coverage. Rates of statistical significance for both news variables were similar to those with random regressors and the average explanatory power of the model was close to zero. Additionally, coefficient estimates for both variables were centered around zero, a feature that undermines the usefulness of either as a predictor in the given context.

This result was subject to a series of robustness tests involving variations of specification, and cross-sectional and temporal subset analysis. None of these robustness tests led to qualitatively different result. A possible exception was a two-three fold increase in the rate of statistical significance for news sentiment and higher R^2 for low-coverage firms relative to high-coverage firms. As the rates of statistical significance were still low in absolute terms (10.45% and 1.36% at the 5% and 1% levels of significance, respectively), with an average adjusted R^2 of 0.45%, we leave further exploration of this result to future analysis.

We then tested the predictive capacity of news in a momentum-style regression in which cumulative six-month excess return was regressed against six-month aggregates of news sentiment and news coverage. Informed by a series of Monte Carlo simulations, this overlapping regression was performed using a Britten-Jones, Neuberger, and Nolte (2011) transformed specification. Rates of statistical significance and R^2 of this extended horizon regression were similar to those of the baseline regression and were not meaningfully different from an analogous specification with random regressors.

Dropping the assumption of exogeneity of news, we fit firm-level VAR models to each firm, linking return, news sentiment, and news coverage up to six lags. Analysing rates of statistical significance in individual coefficient estimates suggested autodependencies to be a far stronger driver of system behaviour than predictive cross-relationships. This was confirmed in joint hypothesis tests for each variable pairing, Granger and instantaneous causality tests, and forecast error variance decomposition analysis.

All-in-all, we fail to find evidence for the hypothesis that commonly used news variables—sentiment and coverage—are useful predictors of future returns at the firm-level, at one- to six-month horizons. However, the apparent lack of longitudinal predictability does not preclude useful cross-sectional relationships. We investigate the utility of news as conditioning variable in a cross-sectional context in the following analysis.

Chapter 5

Portfolio Sorts

Several studies have provided evidence that firms' news exposure is an economically useful data source in the formation of low-frequency equity portfolios. Together, they suggest news sentiment and attention are important features in predicting the continuation or 'momentum' of stock prices in the cross-section. However, a major difficulty in drawing practical conclusions from this literature is that it comprises strategies based on different features of news, and different combinations of news and momentum, documented over significantly varied empirical settings. Moreover, none of the relevant studies were conducted on a highly liquid, well-understood index—leaving the comparison and generalisability of these results on shaky ground.

In this analysis, we test the performance of news-informed portfolios motivated by the aforementioned literature, in a consistent and institutionally-relevant experimental setting. More specifically, we examine the performance of investment strategies employing nonparametric conditioning on news sentiment, news attention, and stock price momentum using single, double, and triple-sorted decile portfolios.

Our results offer little evidence that news provides economically valuable conditioning information in momentum-style or momentum-enhanced portfolios at the horizons studied. Plausibly ex-ante-identifiable strategies, such as those motivated by the literature, failed to generate risk-adjusted excess returns in our sample, even after controlling for the GFC.

5.1 Introduction

Research on market responses to news and price information suggests that the strength of firms' stock price continuation is highly dependent on the news exposure of the firm during the formation period. Chan (2003) found that among stocks with extreme monthly price movements, those which were accompanied by news exhibited strong price continuation for over 12 months following the event, while those without news tended to reverse in the following month. Using a formation and holding period of six months, Hillert, Jacobs, and Mueller (2014) found that firms with high media coverage exhibited significantly stronger momentum than low media coverage firms, and that momentum was stronger still when the tone of the news was consistent with the direction of past return. Similar findings have been made in the context of weekly momentum, where it has been shown that sorting firms on the basis of news content characteristics such as tone and journalist attention generates significant abnormal returns and outperforms analogous strategies sorting on past prices alone (Huynh and Smith, 2017; Sinha, 2016). Fang and Peress (2009) found a cross-sectional relationship between news coverage and expected stock return that was most stable over formation and holding periods of six and 12-months, respectively. For ease of reference, the major findings and portfolio specifications of these studies are summarised in Tables 5.1 and 5.2.

Table 5.1: Summary of selected portfolio analyses

Author	Major Finding
Heston and Sinha (2017)	Buying (selling) stocks with highly positive (negative) news tone generates positive abnormal returns.
Huynh and Smith (2017)	Buying (selling) winners (losers) with high attention (low attention) positive (negative) news generates positive abnormal returns.
Sinha (2016)	Buying (selling) stocks with highly positive (negative) news tone generates positive abnormal returns.
Hillert, Jacobs, and Mueller (2014)	Stock price momentum is strongest for firms with high media coverage and is stronger still for those also with concordant news tone.
Fang and Peress (2009)	Stocks with no media coverage generate higher returns than stocks with high media coverage.
Chan (2003)	Momentum returns for stocks accompanied by news are significantly larger than those without news.

These findings point toward the potential use of news analytics to enhance the performance of portfolio managers operating momentum-style investment strategies. How-

Table 5.2: Specifications of selected cross-sectional portfolio analyses

Author	J/S/K	Sorting variables
Heston and Sinha (2017)	1W/[0,1W]/13W	tone
Huynh and Smith (2017)	1W/[0,4W]/52W	attention, tone, return
Sinha (2016)	1W/0W/13W	tone
Hillert, Jacobs, and Mueller (2014)	6M/1M/[6-36M]	attention, return, tone
Fang and Peress (2009)	[1-6M]/0M/[1-12M]	incidence, attention
Chan (2003)	1M/1W/[1-36M]	incidence, return

Note: *J*, *S*, and *K* represent formation, skip, and holding period, respectively. *M* and *W* reflect months and weeks, repsectively.

ever, from the perspective of a practitioner interested in making use of such research, a number of important practical considerations remain unresolved. For instance, from the current literature it is unclear what role news tone, news volume, stock price momentum, and the interaction between these variables, play in the performance of momentum-style portfolios or the horizons over which these variables matter. Furthermore, the economic relevance of news-enhanced strategies, in terms of dollar capacity and performance when constrained to only large, highly liquid firms and realistic trading constraints, is not understood. These uncertainties are largely due to differences in experimental design between the relevant studies—sources of news, variable definitions, data filtering, formation and holding periods, universes of stocks, and sample periods—that preclude inter-study comparisons, generalisations and inference.

In this analysis we attempt to alleviate these uncertainties by benchmarking the performance of news-enhanced portfolios in a consistent experimental setting and under realistic trading constraints. This allows direct assessment of the relative economic significance of the different components of news data shown to be of influence in the previous literature. We choose as our universe the uniquely investable and ex ante-identifiable strata that is the S&P 500. By considering only the constituents of major S&P indices, we ensure stock selection criteria are defined ex ante, realistic, obtainable and are relevant to both academics and practitioners.

Factor-style or quantile-conditioned portfolios are used as the primary subject of investigation in this chapter. We believe this is appropriate for several reasons. First, they are the basis for the overlapping portfolio approach used heavily within the academic market anomaly literature, and thus provide a widely recognised and econometrically sound approach to testing predictability in the cross-section of returns. Second, and equally import for this thesis, is the practical relevance of the approach to portfolio formation. This real-world viability is evident in the proliferation of factor-based investment products—a $1.9 trillion industry projected to grow to $3.4 trillion by 2022 (BlackRock, 2021). The quantile conditioning approach is also the basis for major practitioner-focused texts on active portfolio design and management (Grinold and Kahn, 2000; Fabozzi, Focardi, and Kolm, 2010; Qian, Hua, and Sorensen, 2007; Chincarini and Kim, 2010).

The core of this analysis involves tests of univariate, bivariate, and trivariate portfolios formed on the basis of price momentum, news sentiment, and news attention. Some of the formation procedures we test have been examined previously in the literature while others have not. We investigate changes to sorting order, the use of residual news variables, and provide risk-adjusted returns under several factor models. We investigate the impact of the GFC on strategy performance and provide results for pre-, post-, and ex-GFC periods. We document results for additional formation procedures based on numerical rather than cross-sectional breakpoints, and for sentiment portfolios utilising a temporal weighting scheme whereby more recent news is given higher weight.

To the question of whether conditioning on news variables can enhance or outperform momentum returns, our results are mixed. In univariate sorts, momentum wins. In double and triple sorts, there were portfolios that outperformed over the sample, but we don't consider these to be plausible investment choices ex-ante. In particular, we document large variation within the tercile of stocks with the lowest news sentiment; those with high news attention and/or momentum outperform, while those with low news attention and/or momentum underperform, and a strategy that buys (sells) the former (latter) generates significant risk-adjusted returns over the sample.

In additional tests, we find evidence that there is value in using news sentiment as a screening mechanism within a traditional momentum framework. Beginning with a standard momentum portfolio, the adjusted procedure seeks to remove stocks with discordant news sentiment, i.e. momentum winners (losers) with numerically negative (positive) news sentiment. Such a portfolio (and several variants) generate statistically significant risk-adjusted returns over the ex-GFC sample. The effect of sentiment-based screening is found to be independent from momentum screening (timeseries momentum), which has almost no effect in our sample. However, due to the positive bias of our news data, removing positive sentiment stocks from the short side requires a portfolio with variably higher concentration.

5.2 Data and Methodology

5.2.1 Data

The data sourcing and processing used for this analysis is as described in Chapter 3, and we again rely on price return, news sentiment, and news coverage (or attention) as our variables of interest. We differ slightly from the previous analysis in our calculation of the formation period sentiment score—rather than aggregating to monthly observations and then averaging the monthly scores over the formation period (e.g. six months), we simply take the average of all daily scores within the formation period at once. In the previous analysis we began with monthly regressions and needed to be deliberate about the timeseries properties of regressors. This is not the case for the current analysis, and aggregating from the daily level allows us to be more granular when applying temporal weighting schemes later on. In any case, we find only very slight numerical differences, and no qualitative differences, between the two approaches.

5.2.2 Portfolio Construction

Our central analysis compares the performance of decile portfolios formed through univariate and bivariate sorts on past news sentiment, stock return, and news attention (coverage). In most cases we document the long-only performance of each decile individually, as well as a zero-cost portfolio constructed from the extremes. The performance of individual deciles is useful in examining the cross-sectional distribution of returns with respect to the conditioning variable(s), and is of interest to practitioners bound to long-only investment strategies. Zero-cost portfolios are standard practice in the literature analyzing the determinants of the cross-section of stock returns and are the main focus of related studies examining the performance of news-driven portfolios.

Here we briefly describe our portfolio formation procedure using two examples–a single sort on news sentiment and a double sort on sentiment and momentum. On the last trading day of the month, we sort all current index members by their formation period news sentiment and divide them into ten equal groups. The high news sentiment portfolio buys all firms in the top (most positive) sentiment group, and the low news sentiment portfolio buys those in the bottom (most negative) news sentiment group. The high-low sentiment portfolio buys those in the top sentiment group and sells those in the bottom sentiment group, with equity distributed equally within and between the long and short legs.

To construct portfolios with exposure to the joint effects of news sentiment and momentum, we use a sequential double-sorting procedure. On the last trading day of the month, we sort all current index members by their formation period news sentiment and divide them into three equal groups. Within each of these news sentiment terciles, firms are then sorted by their formation period return (momentum) and divided into a further three groups. This results in nine portfolios, each representing different portions of the sentiment-momentum feature space. The high-sentiment, high-momentum portfolio buys firms in the top momentum tercile of the top sentiment tercile. The high-sentiment, low-momentum portfolio buys firms in the bottom tercile of the top sentiment tercile. The difference in performance of these two portfolios is captured by a high-sentiment momentum portfolio, which buys (sells) the top (bottom) momentum tercile within the top sentiment tercile. In this example, as we sorted firms on news sentiment first, and momentum second, we would refer to news sentiment as our primary sorting variable and momentum as our secondary sorting variable.

Note that since the feature space is being defined cross-sectionally (i.e ."high" means *relatively* high) and because sorting occurs sequentially, the resulting relationships between variables are conditional, with sorting order dictating the direction of conditioning. For example, sorting on sentiment and then momentum helps answer the question "what is the effect of past return, *conditioning* on high, mid, or low news sentiment history?", rather than "what is the effect of news sentiment history, conditioning on high, mid or low past return?". For each portfolio formed through two-way sorting, we also present results to the analogous portfolio formed using the reversed sorting order.

Our reasons for using a sequential sorting procedure are threefold: it results in portfolios with clear, intuitive interpretations, it is the standard approach in the relevant literature, and it allows for granular investigation of conditional relationships unavailable to other portfolio formation procedures. However, the double-sorting procedure does not necessarily let us examine the most extreme unconditional combinations, such as stocks with the highest joint combination of sentiment and momentum, and can only practically be applied to a small number of variables for the given universe. This latter point is due to the fact that group size decreases geometrically with each sorting layer (which also places a limit on the "sharpness" of our conditional statements). We address some of these concerns by performing additional tests with simultaneous ranking and numeric thresholds in section 5.3.6.

The tercile breakpoints were chosen so that each of the resulting return terciles contain approximately the same number of stocks as a decile portfolio[1], while conditioning on each sorting variable with equal cross-sectional precision.

For the baseline analysis we use a formation and holding period of six months, with no skip period between. Portfolios are equal-weighted at the time of formation and positions are held until the end of the holding period without rebalancing. To be considered for portfolio formation, each stock must have began trading before the beginning of the formation period and must have traded within the last five trading days at the time of formation. If a stock loses index membership during the holding period, we exit the position at the close of the last day of membership and assume a 0% return on the proceeds for the remainder of the holding period.

5.2.3 Testing Procedure

We evaluate each trading strategy using the calendar-time overlapping portfolio approach of Jegadeesh and Titman (1993). In this approach, portfolios are formed every month such that each strategy contains portfolios with overlapping holding periods. That is, at a given point in time, a strategy with a holding period of K months will hold the portfolio formed in the current month as well as those formed in the previous K-1 months. Additionally, at the end of each month, the portfolio formed K months ago will be closed. Thus, the weights on 1/K of the stocks in the strategy are revised each month, while the rest are carried over from the previous month.

The time-series of the average return of the K portfolios in each calendar month captures the effects of the correlation between portfolios and can be appropriately used to test the average monthly return generated by the portfolio formation procedure over the K months following formation(Fama, 1998; Chan, 2003). Test statistics are given by the time-series mean of the overlapping return series divided by the time-series standard error. Risk-adjusted returns are attained by regressing excess portfolio returns on contemporaneous factors according to the relevant factor model being used. These regressions are conducted in a panel format comprising each of the K portfolio return series. Standard errors are corrected using Driscoll and Kraay's (1998) spatial correlation consistent (SCC) estimator as implemented by Millo (2017). We briefly explain this choice.

Consider the panel regression:

$$r_{it} = \mathbf{x}^{\mathsf{T}}{}_{it}\boldsymbol{\beta} + \epsilon_{it}$$

[1]The difference is between group sizes of 1/9th of the index for the double sorting procedure and 1/10th of the index for the decile portfolios, which is approximately five stocks for the S&P 500 and 2 stocks for the ASX 200.

In addition to heteroskedasticity, residuals in a panel setting can be correlated between observations in three main ways:

Group effects (Firm effects): Residuals are correlated through time for a particular firm (time-series dependence), so that $E[\epsilon_{it}\epsilon_{is}|\boldsymbol{x}_{it}, \boldsymbol{x}_{is}] \neq 0$, $\forall t \neq s$.

Time effects: Residuals are correlated across firms in a given time period (cross-sectional dependence), so that $E[\epsilon_{it}\epsilon_{jt}|\boldsymbol{x}_{it}, \boldsymbol{x}_{jt}] \neq 0$, $\forall i \neq j$.

Persistent common shocks: Residuals are correlated between different firms in different time periods. These correlations are assumed to decay over time such that they may be ignored after a maximum lag length L, so that $E[\epsilon_{it}\epsilon_{js}|\boldsymbol{x}_{it}, \boldsymbol{x}_{js}] = 0$, if $i \neq j$ and $|t - k| > L$.

Failing to account for these effects can lead to significantly under-estimated standard errors and consequent over-rejection using standard hypothesis tests. Petersen (2009) contrasts various approaches to panel data commonly applied in the empirical finance and asset pricing literature. He shows that in the presence of both firm and time effects, the application of Newey-West (adapted for panel data), Fama-Macbeth, and Fama-Macbeth corrected for first-order autocorrelation, procedures leads to downward biased estimates of standard errors. Petersen (2009), Thompson (2011), and Cameron, Gelbach, and Miller (2011) show that standard error estimates robust to firm and time effects can be achieved by clustering residuals by time and group simultaneously.

The standard errors obtained from the multi-way clustering procedure are robust to any correlation structure within each group or within each time period, but rely on the asymptotics in each direction: the number of clusters, rather than just the number of observations, is assumed to go to infinity. Thompson (2011) extends the two-way clustering approach to accommodate persistent common shocks as defined above. Thompson's estimator is robust to cross-sectional and time-series dependence within groups and time periods, respectively, and dependence between different groups in different time periods, up to a maximum lag L.

Kernel-based methods account for dependence structures in essentially the same way as clustering approaches, but weight correlations by a distance-decreasing smoothing function (Foote, 2007). The underlying assumption being that correlations between observations decay as the lag distance between them increases. The Newey and West (1987) nonparametric estimator, developed for the time series context, is a well-known example of this. Driscoll and Kraay's (1998) spatial correlation consistent estimator extends the Newey-West estimator to a panel time series context, allowing for very general forms of cross-sectional and temporal dependence. More specifically, the SCC estimator is robust to contemporaneous and lagged cross-sectional dependence as well as serial dependence, provided that these correlations decay quickly enough in the time dimension, which is therefore assumed to be relatively large. However, unlike the multi-way clustering approaches, the SCC estimator does not rely on group asymptotics and so the number of firms relative to time periods does not affect convergence.

For multi-way clustering approaches, the practical minimum group size necessary for suitable standard error estimates appears to range from between 25 to 50 depending on the situation (see Cameron and Miller, 2015; Thompson, 2011). For the SCC estimator, Driscoll and Kraay (1998) show that a time dimension of 50 leads to comparable finite sample performance as standard time series HAC estimators. As most of the strategies we test have a group size (number of holding period months, K) between 6 to 12, with approximately 170 time periods, we use the SCC estimator for these regressions. Newey and West (1987) and Driscoll and Kraay (1998) use the Bartlett kernel as their weighting function. We apply the quadratic spectral kernel following the results of Andrews (1991), with a bandwidth parameter (maximum lag length) of the closest integer equal to or greater than $T^{1/4}$ (Greene, 2005; Driscoll and Kraay, 1998). Figure 5.1 compares common kernels for a range of lag/bandwidth values, normalised by asymptotic variance (Figure 1 in Andrews, 1991).

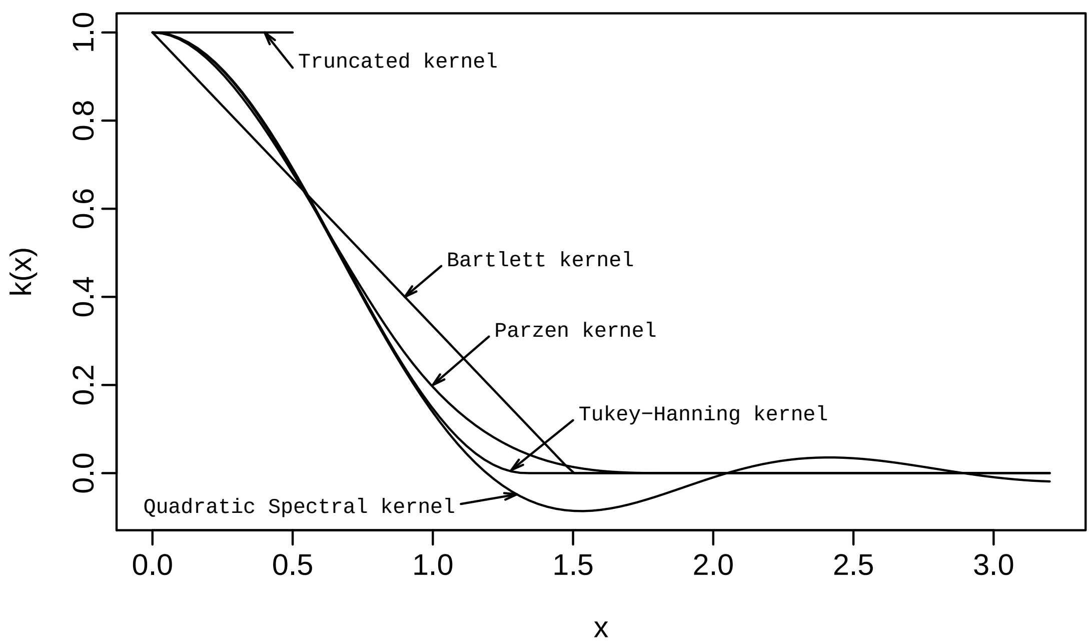

Figure 5.1: Comparison of common heteroskedasticity and autocorrelation consistent estimator smoothing kernels as a function of lag/bandwidth, normalised by asymptotic variance (Andrews, 1991).

5.3 Results

5.3.1 Sentiment and Momentum

Table 5.3 presents mean monthly returns for decile portfolios sorted separately by sentiment and momentum. These figures suggest that the unconditional relationships between expected return and the sorting variables, if they exist, are not monotonic (and are therefore also nonlinear) within the observed sample. Both sentiment and momentum do, however, appear to contain information on the worst performing stocks; the lowest sentiment and momentum deciles are the only long-only portfolios with a mean monthly return statistically indistinguishable from zero (the fact that none of the decile sorts produce a negatively performing portfolio is also telling of the sample period being examined). In terms of capturing high performing stocks with a single variable, sentiment appears to do a better job than momentum, as the highest sentiment decile produces a higher mean return with greater statistical significance than the top momentum decile. Neither of the high-minus-low (HML) portfolios generate a statistically significant return, indicating that the difference in performance between the top and bottom deciles is not large or stable enough to represent a profitable arbitrage opportunity.

Panel A of Table 5.4 presents returns to momentum portfolios formed within high, mid and low sentiment terciles. Surprisingly, it is really only within low sentiment stocks that momentum winners outperform momentum losers in terms of absolute returns. This is the opposite of what we might expect from the weekly momentum results of Huynh and Smith (2017). Further, low sentiment momentum winners appear to outperform high sentiment momentum winners, which is again in contrast to the results of Hillert, Jacobs, and Mueller (2014) and Huynh and Smith (2017), who find positive sentiment momentum winners to outperform low sentiment momentum winners.

Panel B of Table 5.4 presents returns to sentiment portfolios formed within high, mid and low momentum terciles. We see that it is only within low momentum stocks that sentiment is related to performance in the way we expect; a portfolio formed from momentum losers with high sentiment achieves a mean expected return of approximately 1% per month, while the return to a portfolio of momentum losers with low sentiment is statistically indistinguishable from zero. This is in contrast to Sinha's (2016) weekly sentiment portfolios, which performed much stronger among momentum winners than momentum losers.

In both panels, the best and worst performing terciles are those within the lowest primary grouping. Taking into account the results of the unconditional sorts, this suggests

Table 5.3: Univariate Sentiment and Momentum Returns

	Sentiment		Momentum	
Decile	mean	p (t-stat)	mean	p (t-stat)
High	0.0101**	0.0042 (2.9028)	0.0088*	0.0167 (2.4172)
2	0.0088**	0.0097 (2.6165)	0.0093**	0.0038 (2.9373)
3	0.0092**	0.0055 (2.8154)	0.0086**	0.0055 (2.8152)
4	0.0096**	0.0035 (2.9618)	0.0093**	0.0025 (3.0678)
5	0.0088**	0.0097 (2.6168)	0.0096**	0.0019 (3.1594)
6	0.0081*	0.0230 (2.2950)	0.0098**	0.0024 (3.0835)
7	0.0091*	0.0124 (2.5287)	0.0098**	0.0050 (2.8451)
8	0.0104**	0.0044 (2.8888)	0.0093*	0.0145 (2.4692)
9	0.0107**	0.0054 (2.8205)	0.0090*	0.0328 (2.1520)
Low	0.0071	0.0720 (1.8105)	0.0077	0.1640 (1.3979)
High-Low	0.0031	0.1442 (1.4672)	-0.0001	0.9884 (-0.0146)
Index (SPY)	0.0077*	0.0108 (2.5767)		

This table presents raw monthly returns for equal-weighted portfolios sorted separately by news sentiment and stock return (momentum). Portfolios are constructed using formation and holding periods of six months, with no skip period.

that stocks with the highest sentiment consist of both momentum winners and momentum losers. In particular, the worst performing stocks are those with the combination of lowest sentiment and lowest momentum. However, we see again that none of the long-only portfolios are negative, and none of the zero-cost portfolios are profitable over the sample.

5.3.2 News coverage

The determinants of news flow

The amount of news coverage, or news attention, received by firms has been shown to be significantly related to the cross-section of returns (Fang and Peress, 2009), and in particular the performance of momentum portfolios (Hillert, Jacobs, and Mueller, 2014; Chan, 2003; Huynh and Smith, 2017). A consistent finding across these studies is that momentum strategies are more profitable among firms with a high amount of news coverage, which is either evidence of over- or under-reaction, depending on who you ask. In this

Table 5.4: Sentiment and Momentum Returns: Bivariate Comparisons

	Secondary Sort							
	High		Mid		Low		High-Low	
Primary sort	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: Primary sort on sentiment								
High	0.0084**	(2.6134)	0.0096**	(3.0899)	0.0100*	(2.5169)	-0.0025	(-1.1285)
Mid	0.0085**	(2.6390)	0.0090**	(2.8888)	0.0086*	(1.9781)	-0.0020	(-0.6603)
Low	0.0101**	(2.9909)	0.0101**	(2.9712)	0.0080	(1.6818)	-0.0007	(-0.1816)
HH-LL	-0.0019	(-0.4845)						
Panel B: Primary sort on momentum								
High	0.0088**	(2.7052)	0.0083**	(2.6323)	0.0097**	(2.8208)	-0.0010	(-0.7474)
Mid	0.0101**	(3.1950)	0.0091**	(2.8890)	0.0097**	(3.0320)	0.0005	(0.5444)
Low	0.0101*	(2.4422)	0.0086	(1.9522)	0.0077	(1.6975)	0.0023	(1.4303)
HH-LL	-0.0003	(-0.0950)						

This table presents raw monthly returns for double-sorted stock portfolios sorted by news sentiment and momentum. In panel A, news sentiment terciles are formed by sorting firms by their formation period news sentiment. Within each tercile, firms are then sorted by formation period return (momentum). The HH-LL portfolio buys (sells) stocks in the highest (lowest) momentum tercile within the highest (lowest) sentiment tercile. In panel B, the sorting order is reversed–momentum terciles are formed by sorting firms by their formation period returns. Within each tercile, firms are then sorted by formation period news sentiment. The HH-LL portfolio buys (sells) stocks in the highest (lowest) sentiment tercile within the highest (lowest) momentum tercile. Portfolios are constructed using formation and holding periods of six months, with no skip period.

section we examine the relationship between news attention and returns, and the joint effects of news attention with momentum and news tone, within S&P 500 index members. Fang and Peress (2009) and Hillert, Jacobs, and Mueller (2014) found that news coverage was partially explained by various firm characteristics and Chan (2003) noted that stocks with news in a given month are on average much larger than those that do not. If uncorrected, these relationships can lead to inadvertent use of news coverage as a proxy for size or other firm characteristics. We therefore investigate whether this is an issue within our sample.

Panel A of Table 5.5 shows the determinants of news attention (log number of stories)[2]

[2]The log-transform of news coverage follows from Hillert, Jacobs, and Mueller (2014), but is also justified by the Box and Cox (1964) procedure—see section 4.2.

Table 5.5: Sentiment and Momentum Returns: Bivariate Comparisons

Model spec.	Size 1	Analyst 2	BM 3	1+Analyst 4	4+BM 5
Panel A: News attention as dependent variable					
Size	0.5049***			0.4089***	0.4212***
	(24.0999)			(17.8426)	(17.6472)
Analyst following		0.3965***		0.2082***	0.2142***
		(15.6719)		(9.8483)	(9.7223)
Book-market			0.0432		0.1179***
			(1.6607)		(4.6178)
Adj. R-sq	0.2549	0.1572	0.0019	0.2890	0.3026
Panel B: News sentiment as dependent variable					
Size	-0.0440*			-0.0039	-0.0243
	(-2.3060)			(-0.1849)	(-1.1128)
Analyst following		-0.0897***		-0.0879***	-0.0855***
		(-4.9946)		(-4.4125)	(-4.3496)
Book-market			-0.1194***		-0.1276***
			(-6.7691)		(-6.8952)
Adj. R-sq	0.0019	0.0080	0.0144	0.0080	0.0240

This table presents raw monthly returns for double-sorted stock portfolios sorted by news sentiment and momentum. In panel A, news sentiment terciles are formed by sorting firms by their formation period news sentiment. Within each tercile, firms are then sorted by formation period return (momentum). The HH-LL portfolio buys (sells) stocks in the highest (lowest) momentum tercile within the highest (lowest) sentiment tercile. In panel B, the sorting order is reversed–momentum terciles are formed by sorting firms by their formation period returns. Within each tercile, firms are then sorted by formation period news sentiment. The HH-LL portfolio buys (sells) stocks in the highest (lowest) sentiment tercile within the highest (lowest) momentum tercile. Portfolios are constructed using formation and holding periods of six months, with no skip period.

on the basis of cross-sectional regressions, with standard errors clustered by firm and month Petersen (2009)[3]. As the mean value of each of the regression variables varies through time during the sample, and to assist interpretation of the regression coefficients, all regression variables were standardised by month to have mean of zero and standard deviation of one.

[3]We use two-way clustering here as, unlike our portfolio regressions, the panel is of sufficient length and width to satisfy the asymptotic assumptions.

Specification 1 shows that firm size (log market capitalisation) is positively associated with news coverage and explains over 25% of the variation in coverage among S&P 500 members alone. In specification 2, firms' (log) analyst following is also found to be associated with news coverage, explaining over 15% of the variation between firms. Fang and Peress (2009) and Hillert, Jacobs, and Mueller (2014) find that analyst following is negatively related to news attention, whereas our results show the relationship to be positive. This could be due to differences in the sample period, the news sample, or our particular universe of stocks. Huynh and Smith (2017), who use a related but not entirely analogous measure of news attention as their dependent variable, also find analyst following to be positively related to attention. Huynh and Smith (2017) use TRNA and a sample period that is contained within ours, but employ a much wider universe of US stocks. This suggests that effect is not due to our focus on index constituents.

When size and analyst following are applied together, as in specification 4, the coefficient magnitude and t-statistics for both variables decrease, indicating that the two explanatory variables are correlated. Adding book-to-market ratio to the model, as in specification 5, increases R^2 from 29% to just over 30%. To control for these correlations between firm characteristics and news attention, we follow the approach of Hillert, Jacobs, and Mueller (2014) in specifying a measure of residual attention, $\epsilon_{Attention}$, which we define through the following regression (specification 5 in Table 5.5)[4]:

$$\ln(1 + no.\ stories) = \beta_1 \ln(Size) + \beta_2 \ln(1 + Analyst) + \beta_3 BM + \epsilon_{Attention} \tag{5.1}$$

This regression is performed at the end of every month, resulting in a centered time series of residual news attention for each stock. To aggregate residual news attention over the formation period, we simply sum the monthly values together. If a firm's residual attention cannot be calculated for a given month due to missing firm (as opposed to news) data, the value is excluded from the summation. If a firm's residual cannot be calculated for the entire formation period, the stock is excluded from that month's ranking.

Panel B of Table 5.5 presents the determinants of news sentiment, using the same methodological setup as panel A. Compared to the amount of coverage, firm characteristics do a poor job at explaining the tone of the coverage, with a maximum adjusted R^2 of less than 2.5% for the specifications tested. This result is consistent with those of Huynh and Smith (2017), who get the same R^2 for the same specification on weekly data. Notably, size adds no explanatory power to the regression, and drops out of significance when accounting

[4] The transformation $f(x) = \ln(1 + x)$ is used instead of $f(x) = \ln(x)$ for story count and analyst following to handle instances in which these variables equal zero.

for analyst following. To calculate residual news sentiment, $\epsilon_{Sentiment}$, we perform the following rolling monthly regression:

$$Sentiment = \alpha_1 \ln(1 + Analyst) + \alpha_2 BM + \epsilon_{Sentiment} \tag{5.2}$$

Another way to understand the relationship between firm characteristics and news exposure is to examine the distribution of firm variables across groups of stocks sorted by the news variables of interest. Being nonparametric, this method does not impose any structure on the underlying relationships, and can be used to gauge the effectiveness of the linear specification used to define our residual news scores. Table 5.6 presents time series averages of firm characteristics for stock quintiles sorted by raw (panel A) and residual (panel B) media attention.

From panel A it can be seen that mean firm characteristics vary significantly across the raw news attention quintiles, with size and analyst following increasing monotonically with attention and book-to-market ratio increasing monotonically with the exception of the highest attention quintile. In panel B the distribution of firm variables is much more even across quintiles, and both the magnitude and significance of the differences between the highest and lowest attention groups have dropped considerably. Our measure of residual attention appears to have the properties for which it was designed, and we make use of this measure going forward.

Table 5.7 presents time series averages of firm characteristics for stock quintiles sorted by raw (panel A) and residual (panel B) news sentiment. Panel A shows that the variation of firm variables across sentiment quintiles is much less pronounced than for media attention, with the exception of book-to-market ratio, which decreases monotonically as sentiment increases. There is also evidence of weak non-linearity in the distribution of size and analyst following, as the firm values tend largest in quintiles 3 and 4. In panel B, the distribution of firm variables across the residual sentiment quintiles is now more even, although the effect is subtle for firm size and analyst following, which still retain their nonlinearity. The main effect can be seen for book-to-market ratio which shows little variation across the residual quintiles. Due to the relatively weak relationship between firm characteristics and news sentiment captured by the regression, and to keep our portfolio formation procedure as simple as possible, we do not make use of residual sentiment in our main analysis. We find that using residual sentiment in place of raw sentiment results in only minor numerical differences and does not qualitatively change our results.

Table 5.6: Firm characteristics by raw and residual news attention

Quintile	Size (ln)	Analyst (ln)	Book-market
Panel A: Quintile sorts on raw news attention			
1	10.5522	2.5012	0.4503
2	9.6524	2.1868	0.4764
3	9.3459	2.0224	0.4526
4	9.1464	1.8399	0.4355
5	8.9906	1.6144	0.4139
1-5	1.5616***	0.8867***	0.0364***
t-stat 1-5	115.9936	85.4268	4.4447
Panel B: Quintile sorts on residual news attention			
1	9.4252	1.9927	0.4594
2	9.6281	2.1011	0.4515
3	9.6238	2.0814	0.4480
4	9.5526	2.0546	0.4459
5	9.4573	1.9769	0.4248
1-5	-0.0321***	0.0158***	0.0346***
t-stat 1-5	-6.0272	4.0193	4.0373

This table shows time series averages of the monthly mean of firm characteristics for portfolios of S&P 500 index members sorted by news attention (panel A) and residual news attention (panel B). News attention is given by *ln(1+no. stories)* and residual attention is the residual from month-to-month regressions of news attention on *Size* (log market capitalisation), *Analyst* (log analyst following) and *BM* (equity book-to-market ratio). The sample period is from Jan 2003 to Dec 2017.

Table 5.7: Firm characteristics by raw and residual news sentiment

Quintile	Size (ln)	Analyst (ln)	Book-market
Panel A: Quintile sorts on raw news sentiment			
1	9.3332	1.8803	0.3812
2	9.5825	2.0377	0.4204
3	9.6730	2.1084	0.4308
4	9.6958	2.1224	0.4793
5	9.4952	2.0705	0.5182
1-5	-0.1619***	-0.1902***	-0.1370***
t-stat 1-5	-10.7616	-18.4712	-20.8546
Panel B: Quintile sorts on residual news sentiment			
1	9.3604	1.9739	0.4315
2	9.6026	2.0813	0.4401
3	9.6612	2.1102	0.4630
4	9.6776	2.0998	0.4670
5	9.4550	1.9811	0.4285
1-5	-0.0946***	-0.0071	0.0030
t-stat 1-5	-6.2342	-1.6699	0.6846

This table shows time series averages of the monthly mean of firm characteristics for portfolios of S&P 500 index members sorted by news sentiment (panel A) and residual news sentiment (panel B). News sentiment is cumputed as the average of daily news sentiment scores and residual news sentiment is the residual from month-to-month regressions of news sentiment on *Analyst* (log analyst following) and *BM* (equity book-to-market ratio). The sample period is from Jan 2003 to Dec 2017.

News attention: univariate sorts

Table 5.8 reports mean monthly returns to decile portfolios sorted by separately raw and residual news coverage. The two different measures of news coverage are shown to produce markedly different results in unconditional sorts, especially at the extremes where most portfolio formation procedures would focus. The low performance of stocks with high raw media coverage is broadly consistent with the findings of Fang and Peress (2009), who also used a raw measure of news coverage in their analysis. We show the returns to raw coverage portfolios merely as a point of comparison, and do not discuss them any further.

As for the residual attention portfolios, a notable feature of the results is the high performance of the three decile portfolios with the most attention. This is broadly consistent with the findings of Hillert, Jacobs, and Mueller (2014), who find a positive relationship between residual news coverage and performance. In our sample however, performance flattens out beyond decile 4 and returns do not exhibit any further unconditional relationship to news coverage.

Table 5.8: Raw and residual attention returns

	Raw attention		Residual attention	
Decile	mean	p (t-stat)	mean	p (t-stat)
High	0.0078*	0.0319 (2.1639)	0.0108*	0.0229 (2.2970)
2	0.0098**	0.0046 (2.8738)	0.0110**	0.0052 (2.8314)
3	0.0114**	0.0031 (2.9994)	0.0099**	0.0082 (2.6762)
4	0.0083*	0.0307 (2.1798)	0.0090**	0.0098 (2.6123)
5	0.0079*	0.0261 (2.2440)	0.0084*	0.0134 (2.4985)
6	0.0092*	0.0112 (2.5643)	0.0088*	0.0104 (2.5911)
7	0.0094**	0.0070 (2.7310)	0.0081*	0.0147 (2.4641)
8	0.0091**	0.0080 (2.6824)	0.0088**	0.0062 (2.7728)
9	0.0090**	0.0069 (2.7360)	0.0090**	0.0036 (2.9514)
Low	0.0098**	0.0024 (3.0850)	0.0080**	0.0085 (2.6615)
High-Low	-0.0017	0.3046 (-1.0298)	0.0034	0.1811 (1.3431)

This table presents raw monthly returns for equal-weighted portfolios sorted separately by raw and residual news attention. Portfolios are constructed using formation and holding periods of six months, with no skip period.

News attention and momentum

Table 5.9 reports mean monthly returns to double-sorted portfolios sorted by residual news coverage and momentum. Panel A shows the performance of momentum portfolios conditioning on high, mid and low news coverage tercile. As expected from the prior literature, the performance of momentum winners is stronger among high coverage stocks than low coverage stocks. Unlike prior literature, this is also true for the momentum losers in our sample–there is almost no difference between momentum winners and momentum losers after conditioning on high or low news coverage.

Panel B of Table 5.9 documents the performance of news coverage portfolios conditioning on high, mid or low momentum tercile. In both high and low momentum terciles, the high news coverage portfolio outperforms the low coverage portfolio, with the difference between the two being largest among high momentum stocks. In both panels, the biggest performance differential is in the direction of news coverage, suggesting that news attention subsumes momentum as a ranking variable in this sample.

Table 5.9: Attention and Momentum Returns: Bivariate Comparisons

	Secondary Sort							
	High		Mid		Low		High-Low	
Primary sort	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: Primary sort on residual attention								
High	0.0102**	(2.9030)	0.0108**	(2.9237)	0.0103*	(2.0088)	-0.0020	(-0.5517)
Mid	0.0081*	(2.5018)	0.0097**	(3.1242)	0.0076	(1.8652)	-0.0002	(-0.0818)
Low	0.0082**	(2.7266)	0.0092**	(3.1796)	0.0081*	(2.1961)	-0.0005	(-0.2063)
HH-LL	0.0018	(0.8631)						
Panel B: Primary sort on momentum								
High	0.0103**	(2.9612)	0.0081*	(2.5106)	0.0083**	(2.7222)	0.0020	(1.7739)
Mid	0.0097**	(2.8330)	0.0097**	(3.0678)	0.0096**	(3.2577)	0.0002	(0.2357)
Low	0.0099*	(1.9850)	0.0084	(1.9698)	0.0083*	(2.1408)	0.0020	(1.0382)
HH-LL	0.0012	(0.4852)						

This table presents raw monthly returns for double-sorted stock portfolios sorted by residual news attention and momentum. In panel A, news sentiment terciles are formed by sorting firms by their formation period residual news attention. Within each tercile, firms are then sorted by formation period return (momentum). The HH-LL portfolio buys (sells) stocks in the highest (lowest) momentum tercile within the highest (lowest) residual attention tercile. In panel B, the sorting order is reversed–momentum terciles are formed by sorting firms by their formation period returns. Within each tercile, firms are then sorted by formation period residual news attention. The HH-LL portfolio buys (sells) stocks in the highest (lowest) sentiment tercile within the highest (lowest) momentum tercile. Portfolios are constructed using formation and holding periods of six months, with no skip period.

News attention and news sentiment

In this section we investigate the joint effects of news sentiment and residual news attention, without further conditioning on past return. As it stands, this does not appear to have been tested in the literature. Hillert, Jacobs, and Mueller (2014) and Huynh and Smith (2017) examine momentum profits in triple sorted portfolios with exposures to different combinations of attention and sentiment, but do not study sentiment and attention independent of momentum.

Table 5.10 documents returns to portfolios double-sorted by residual news coverage and sentiment. Panel A presents the performance of news coverage portfolios conditioned on

Table 5.10: Attention and Sentiment Returns: Bivariate Comparisons

	Secondary Sort							
	High		Mid		Low		High-Low	
Primary sort	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: Primary sort on residual attention								
High	0.0103*	(2.5968)	0.0096*	(2.4095)	0.0116**	(2.7939)	-0.0014	(-0.9315)
Mid	0.0088*	(2.5486)	0.0090**	(2.6932)	0.0079*	(2.2727)	0.0011	(0.7096)
Low	0.0092**	(3.0183)	0.0086**	(2.9211)	0.0077*	(2.2737)	0.0013	(0.9143)
HH-LL	0.0027	(1.7392)						
Panel B: Primary sort on sentiment								
High	0.0101**	(2.6907)	0.0090*	(2.5782)	0.0090**	(3.0697)	0.0013	(0.8989)
Mid	0.0087*	(2.1483)	0.0091**	(2.7415)	0.0083**	(2.7461)	0.0008	(0.4804)
Low	0.0122**	(2.8908)	0.0081*	(2.2613)	0.0081*	(2.3930)	0.0045**	(2.8076)
HH-LL	0.0020	(1.4299)						

This table presents raw monthly returns for double-sorted stock portfolios sorted by residual news attention and raw news sentiment. In panel A, news sentiment terciles are formed by sorting firms by their formation period residual news attention. Within each tercile, firms are then sorted by formation period news sentiment. The HH-LL portfolio buys (sells) stocks in the highest (lowest) sentiment tercile within the highest (lowest) residual attention tercile. In panel B, the sorting order is reversed–sentiment terciles are formed by sorting firms by their raw formation period news sentiment. Within each tercile, firms are then sorted by formation period residual news attention. The HH-LL portfolio buys (sells) stocks in the highest (lowest) sentiment tercile within the highest (lowest) momentum tercile. Portfolios are constructed using formation and holding periods of six months, with no skip period.

high, mid or low news sentiment. Here, the performance of high coverage stocks is greatest among those with low news sentiment, whereas the performance of low coverage stocks increases with increasing sentiment. Additionally, the difference in performance between high coverage and low coverage stocks is widest among those with low news sentiment, primarily due to the extreme positive returns of low-sentiment, high-coverage stocks within the sample.

Panel B shows the performance of sentiment portfolios conditioned on high, mid or low news coverage. The performance of low sentiment stocks is again shown to decrease significantly with news coverage, with the high coverage, low sentiment portfolio performing the best in the sample, and low coverage, low sentiment stocks performing the worst. Within high coverage stocks, those with low sentiment outperform those with positive sentiment, and this is reversed among low coverage stocks. A behavioural story might be that investors negatively overreact to firms which receive a high volume of negative news and these stocks rebound energetically following the mispricing. Though if this were the case, we would expect high attention stocks with low past returns to exhibit similar performance, or to at least superior performance to those with high past returns; we see neither of these things.

News attention, news sentiment and momentum

So far, we have seen that neither attention, sentiment or momentum are particularly strong unconditional predictors of future return, although they each appear to be positively related to return at the extremes. Among momentum winners, we saw that low sentiment was associated with higher returns, and the opposite was true for momentum losers. Similarly, we saw among high sentiment stocks, momentum was negatively associated with future return. Among low sentiment stocks, momentum was positively associated with future return.

Joint sorts involving attention revealed that the greatest variance in expected return occurred along the attention axis of low sentiment stocks. In particular, firms with low sentiment and high attention exhibited extreme positive returns. We also found that after accounting for attention, momentum had almost no bearing on future return.

From these results alone, it is unclear whether the relationship between sentiment and momentum is still present after accounting for attention, or whether the relationship between attention and momentum is conditionally dependent upon sentiment. In this section we attempt to shed light on these questions by analysing the returns to portfolios sorted by sentiment, attention and momentum. We first rank all stocks by their news sentiment. We take the top tercile as our high sentiment group and the bottom tercile as our low sentiment group, and then within each group we further sort firms into terciles based on their residual attention. Finally, the stocks within each of these attention terciles are sorted by momentum into a further three groups. This can be thought of as conducting a double-sort, as in panel A of Table 5.9, within the top and bottom thirds of stocks ranked by news sentiment. The upper half of Table 5.11 presents returns to the attention-momentum portfolios formed within the high sentiment group, and the bottom half of Table 5.11 presents the returns to attention-momentum portfolios formed within the low sentiment group.

From the double sorts on attention and momentum (Table 5.9) it appeared that momentum had almost no effect on returns after accounting for attention. Table 5.11 shows that after conditioning on sentiment, the relationship between momentum and returns

Table 5.11: Attention, Sentiment and Momentum Returns: Trivariate Comparisons I

	Momentum Tercile							
	High		Mid		Low		High-Low	
Attention Tercile	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: High sentiment								
High	0.0098**	(2.7433)	0.0104**	(2.8455)	0.0101*	(2.0637)	-0.0020	(-0.5839)
Mid	0.0074*	(2.1482)	0.0095**	(2.8913)	0.0097*	(2.3189)	-0.0031	(-1.1746)
Low	0.0082**	(2.7641)	0.0104***	(3.6349)	0.0090*	(2.4609)	-0.0015	(-0.6608)
HH-LL	0.0005	(0.2243)						
Panel B: Low sentiment								
High	0.0126**	(3.2962)	0.0119**	(3.0784)	0.0111*	(2.1025)	-0.0005	(-0.1308)
Mid	0.0091**	(2.6245)	0.0101**	(3.0343)	0.0065	(1.5088)	0.0017	(0.5521)
Low	0.0072*	(2.1083)	0.0086**	(2.8564)	0.0068	(1.7008)	-0.0004	(-0.1499)
HH-LL	0.0054*	(2.0146)						

This table presents raw monthly returns for triple-sorted stock portfolios sorted by residual news attention, news sentiment, and momentum. Stocks are first sorted by news sentiment The top tercile is taken to be the 'high sentiment' group (Panel A), and the bottom tercile is taken to be the 'low sentiment' group (Panel B). Within each sentiment group, firms are then sorted into terciles by residual news attention. Momentum portfolios are then formed within each attention group. Portfolios are constructed using formation and holding periods of six months, with no skip period.

is much more pronounced. Indeed, as each portfolio now consists of fewer stocks compared to the double-sorts (one third), we should expect greater variance between the portfolios, regardless of whether the sorting parameter(s) has any true bearing on returns. Yet the fact that both attention and momentum behave in the same manner as in their respective double-sorts with sentiment suggests that the three variables contain non-overlapping components of information about expected return and can be combined without one being completely subsumed by another[5]. As in Table 5.10, the greatest variance in expected returns occurs within low sentiment stocks, between those with high attention and those with low attention. Here we see that momentum helps further separate the over-performers from the under-performers, with high-attention momentum winners significantly out-performing low-attention momentum losers.

Next, we investigate the relationship between sentiment and attention after conditioning

[5]Is there another simple statistical way to test this point? I.e. conditional independence.

Table 5.12: Attention, Sentiment and Momentum Returns: Trivariate Comparisons II

	Attention Tercile							
	High		Mid		Low		High-Low	
Sentiment Tercile	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: High momentum								
High	0.0105**	(2.9435)	0.0077*	(2.2450)	0.0082**	(2.6972)	0.0025	(1.5741)
Mid	0.0080*	(2.3827)	0.0084**	(2.6221)	0.0083**	(2.6572)	-0.0003	(-0.2120)
Low	0.0122**	(3.3114)	0.0085*	(2.4566)	0.0071*	(2.1472)	0.0051**	(2.8049)
HH-LL	0.0032	(1.6301)						
Panel B: Low momentum								
High	0.0108*	(2.3210)	0.0095*	(2.2954)	0.0097**	(2.6166)	0.0015	(0.8090)
Mid	0.0077	(1.5250)	0.0089*	(2.1838)	0.0086*	(2.1365)	-0.0002	(-0.0715)
Low	0.0105*	(2.0123)	0.0054	(1.2028)	0.0062	(1.4884)	0.0045	(1.8843)
HH-LL	0.0044*	(2.0934)						

This table presents raw monthly returns for triple-sorted stock portfolios sorted by residual news attention, news sentiment, and momentum. Stocks are first sorted by momentum The top tercile is taken to be the 'high momentum' group (Panel A), and the bottom tercile is taken to be the 'low momentum' group (Panel B). Within each momentum group, firms are then sorted into terciles by news sentiment. Residual attention portfolios are then formed within each sentiment group. Portfolios are constructed using formation and holding periods of six months, with no skip period.

on momentum. We first rank all stocks by momentum. We take the top tercile as our high momentum group and the bottom tercile as our low momentum group, and then within each group we further sort firms into terciles based on their news sentiment. Finally, the stocks within each of these sentiment terciles are sorted by residual news attention into a further three groups. This can be thought of as conducting a double-sort, as in panel B of Table 5.10, within the top and bottom thirds of stocks ranked by momentum. The upper half of Table 5.12 presents returns to the sentiment-attention portfolios formed within the high momentum group, and the bottom half of Table 5.12 presents the returns to the sentiment-attention portfolios formed within the low momentum group[6]. Consistent with our earlier findings, the greatest variance in expected return is between different attention terciles of low sentiment stocks, and appears to be relatively unaffected by the stock's momentum.

[6] In untabulated results I find that conditioning just conditioning again on attention captures high performance better than momentum, but momentum is better at forecasting low performance.

Table 5.13: Attention, Sentiment and Momentum Returns: Trivariate Comparisons III

	Sentiment Tercile							
	High		Mid		Low		High-Low	
Momentum Tercile	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: High attention								
High	0.0105**	(2.8752)	0.0083*	(2.3878)	0.0122**	(3.2539)	-0.0020	(-0.9616)
Mid	0.0105**	(2.9774)	0.0090*	(2.3817)	0.0121**	(3.1591)	-0.0016	(-1.1070)
Low	0.0115*	(2.3198)	0.0080	(1.4349)	0.0106*	(1.9834)	0.0011	(0.4403)
HH-LL	-0.0029	(-0.6437)						
Panel B: Low attention								
High	0.0086**	(2.8363)	0.0083**	(2.7163)	0.0068	(1.9696)	0.0014	(0.7416)
Mid	0.0106***	(3.6956)	0.0083**	(2.8538)	0.0086**	(2.8948)	0.0020	(1.8271)
Low	0.0090*	(2.4995)	0.0086*	(2.2889)	0.0053	(1.2635)	0.0032	(1.6510)
HH-LL	0.0023	(0.7741)						

This table presents raw monthly returns for triple-sorted stock portfolios sorted by residual news attention, momentum, and news sentiment. Stocks are first sorted by residual news attention. The top tercile is taken to be the 'high attention' group (Panel A), and the bottom tercile is taken to be the 'low attention' group (Panel B). Within each attention group, firms are then sorted into terciles by momentum. Sentiment portfolios are then formed within each momentum group. Portfolios are constructed using formation and holding periods of six months, with no skip period.

Table 5.13 presents returns to triple-sorted portfolios sorted first by residual news attention, then by momentum, and finally by news sentiment. This can be thought of as conducting a double-sort, as in panel B of Table 5.4, within the top and bottom thirds of stocks ranked by residual news attention. This is analogous to the triple sorting procedure used by Hillert, Jacobs, and Mueller (2014), although they use quintiles, terciles and a median split for attention, momentum and sentiment, respectively.

Table 5.14 presents returns to triple-sorted portfolios sorted first by residual news attention, then by sentiment, and finally by momentum. This can be thought of as conducting a double-sort, as in panel A of Table 5.4, within the top and bottom thirds of stocks ranked by residual news attention. This is analogous to the triple sorting procedure used by Huynh and Smith (2017). In general, the results in Tables 5.13 and 5.14 are qualitatively the same as those already discussed; changing the sorting order does not significantly affect results, although sorting on either sentiment or momentum first appears to result in greater conditional differences between the subsequent sorting variables compared to

Table 5.14: Attention, Sentiment and Momentum Returns: Trivariate Comparisons IV

	Momentum Tercile							
	High		Mid		Low		High-Low	
Sentiment Tercile	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: High attention								
High	0.0089*	(2.5088)	0.0099**	(2.8471)	0.0101*	(2.1308)	-0.0043	(-1.2518)
Mid	0.0086*	(2.3997)	0.0093**	(2.6983)	0.0091	(1.8166)	-0.0024	(-0.6545)
Low	0.0120**	(3.2655)	0.0115**	(2.9933)	0.0094	(1.8104)	0.0003	(0.0828)
HH-LL	-0.0029	(-0.6856)						
Panel B: Low attention								
High	0.0086**	(2.7301)	0.0102***	(3.4620)	0.0091*	(2.3948)	-0.0018	(-0.7636)
Mid	0.0078*	(2.5156)	0.0092**	(3.1217)	0.0101**	(2.7745)	-0.0032	(-1.2424)
Low	0.0076*	(2.2905)	0.0084*	(2.5972)	0.0063	(1.3660)	0.0004	(0.1160)
HH-LL	0.0010	(0.2833)						

This table presents raw monthly returns for triple-sorted stock portfolios sorted by residual news attention, news sentiment, and momentum. Stocks are first sorted by residual news attention. The top tercile is taken to be the 'high attention' group (Panel A), and the bottom tercile is taken to be the 'low attention' group (Panel A). Within each attention group, firms are then sorted into terciles by news sentiment. Momentum portfolios are then formed within each sentiment group. Portfolios are constructed using formation and holding periods of six months, with no skip period.

sorting first on attention.

5.3.3 GFC and sub-period analysis

Our sample period contains the global financial crisis (GFC). This is particularly relevant to our analysis due to the asset universe consisting entirely of S&P 500 constituents and our investigation of long-short portfolios, which would have been disrupted by the SEC short-sale ban from 2008-09-19 to 2008-10-08. Furthermore, momentum portfolios are known to crash severely in market rebounds, following large market declines.

Before accounting for the GFC in the portfolio simulations, it is useful to take a birds-eye view of a strategy's performance throughout the sample period. We use the zero-cost momentum portfolio for illustration purposes due to its simplicity and sensitivity to the GFC and aftermath. Figure 5.2 depicts the cumulative holding period return of a momentum portfolio as a function of formation period date. Note that each point on the vertical axis represents the total holding period return to an individual portfolio, not the return on the total overlapping strategy. Since the horizontal axis represents portfolio formation date, the associated cumulative return is forward-looking. Overlayed in red is the GFC period, from the fall of Lehman Brothers (2008-09-15) to the emergence of the US out of the recession induced by the financial crisis (2009-06-30), as classified by NBER (NBER, 2018; Sinha, 2016).

Figure 5.2: This figure shows the rolling holding period return to a 6/0/6 decile momentum portfolio as a function of formation date. On the last trading day of each month, the portfolio buys (sells) the decile of stocks with the highest (lowest) six-month return, and holds positions for the following six months without rebalancing, with no skip period. As the x-axis represents formation period date, the y-axis depicting holding period returns is forward-looking. Overlayed in red is the GFC period, from the fall of Lehman Brothers, 2008-09-15, to the emergence of the US out of the recession induced by the financial crisis, 2009-06-30, as classified by NBER (NBER, 2018; Sinha, 2016)

From Figure 5.2 the sharp decline in holding period performance of momentum portfolios formed throughout the recessionary period is clear. Consistent with Daniel and Moskowitz (2016), we find that the most severe momentum drawdowns occurred in the three month period from March to May of 2009, when momentum losers significantly outperformed momentum winners. The portfolios formed close-to or during this period were the worst off; those formed in March and April lost over 150% of their equity, while the portfolio formed in February lost over 60%.

Whether or not such crashes are foreseeable and can be avoided (see Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016), the GFC is an outlier period for many of the tested portfolios within our sample. To better understand how these strategies behave *most of the time*, and to avoid having results dominated by a small number of observations, we analyse portfolio returns in different sub-periods with respect to the GFC.

To correct for results directly influenced by the GFC, we discard all data from during the short-sale ban. All positions are closed at the onset of the ban, and investment does not commence until the formation period no longer includes the ban. For the baseline strategies, this means that complete overlapping portfolio returns do not start again until almost 12 months (formation + holding period) following the end of the short-sale ban. The holding period returns to the long and short legs of the zero-cost momentum portfolio are shown in Figure 5.3, with affected or truncated holding periods represented by triangles.

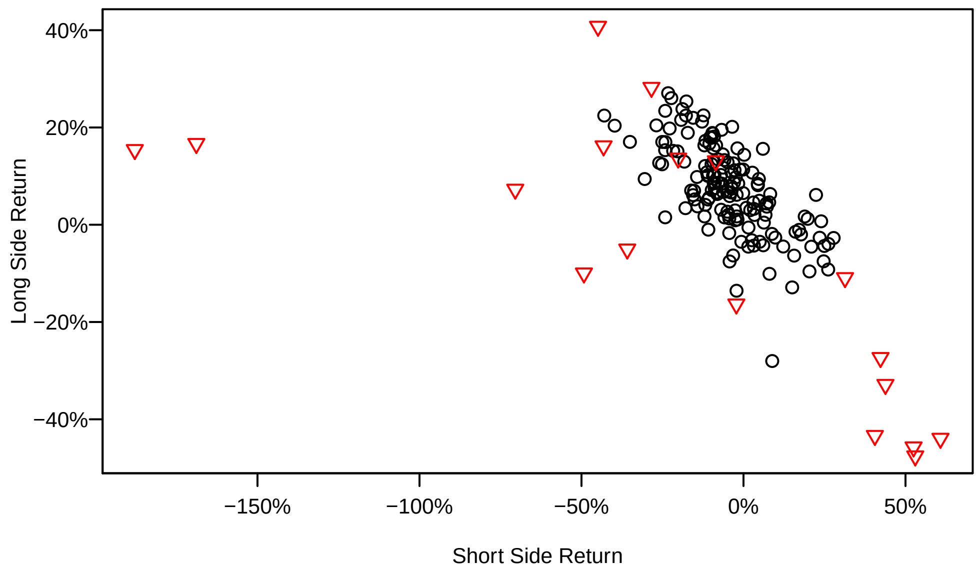

Figure 5.3: This figure shows the cumulative holding period return of the long and short legs of a decile momentum portfolio. On the last trading day of each month, the portfolio buys (sells) the decile of stocks with the highest (lowest) six-month return, and holds positions for the following six months without rebalancing, with no skip period. Holding periods truncated or influenced by the GFC are represented by triangles.

Figure 5.4 shows monthly return observations of the long and short legs of the zero-cost momentum strategy. Observations occurring during the designated GFC-affected period are again represented by triangles.

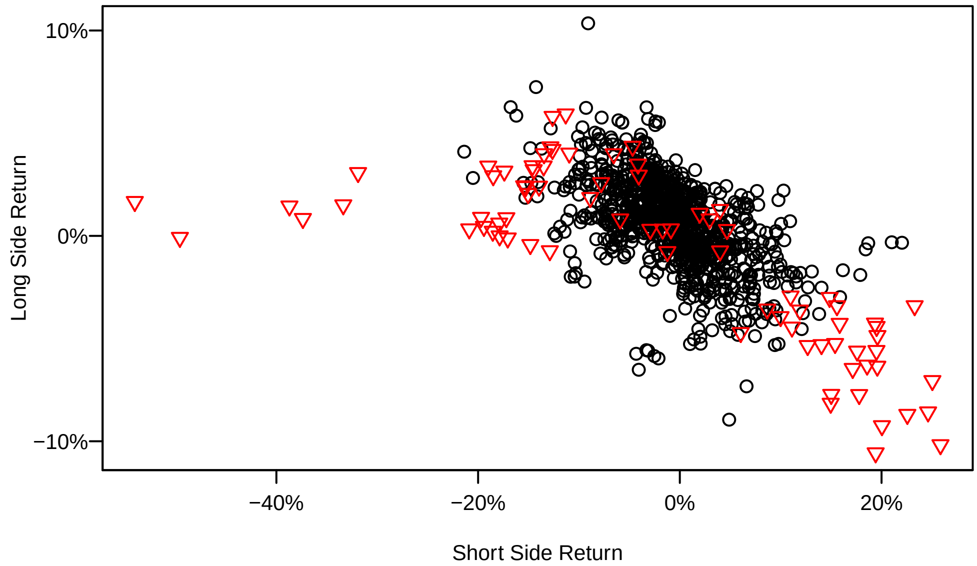

Figure 5.4: This figure shows the monthly return of the long and short legs of a decile momentum portfolio. On the last trading day of each month, the portfolio buys (sells) the decile of stocks with the highest (lowest) six-month return, and holds positions for the following six months without rebalancing, with no skip period. Observations designated to the GFC period are represented by triangles.

Table 5.15 documents mean monthly returns to long-only (*High*) and high-minus-low (*HML*) decile portfolios of stocks sorted separately by momentum, sentiment and residual attention, for pre-GFC (2003-01-01 to 2008-08-31), GFC (2009-09-01 to 2009-09-30), post-GFC (2009-10-01 to 2017-12-31), and ex-GFC (pre- and post-GFC) periods. In absence of the GFC, momentum has been the most profitable long-only portfolio, but has underperformed the high sentiment and high attention portfolios since the GFC. The high attention portfolio, which generated the lowest mean return of the long-only strategies in the ex-GFC sample, has generated the highest return since the GFC. Among the long-short strategies, momentum is the top performer, although the sentiment portfolio has performed better in the post-GFC period. Due to our relatively small sample size, the lack of statistical significance in pre-GFC and GFC periods is not surprising.

Table 5.16 presents mean monthly returns to long-only and long-short portfolios double-sorted by combinations of momentum, sentiment and residual attention. Here we have adopted a labeling methodology that hasn't been used up to this point in the text, to differentiate the sort order, sorting variable, and tercile, of the tested portfolios. *W* (*L*) represents top (bottom) tercile momentum stocks, *Pos* (*Neg*) represents top (bottom) tercile sentiment stocks, and *High* (*Low*) represents top (bottom) tercile attention stocks. The

Table 5.15: Sub-period returns to momentum, sentiment and attention portfolios

	Mom High	Sent High	Attn High	Mom HML	Sent HML	Attn HML
Pre-GFC	0.0095	0.0082	0.0061	0.0113	0.0030	0.0012
	(1.6303)	(1.6722)	(1.0537)	(1.8462)	(0.9567)	(0.3948)
GFC	-0.0198	-0.0029	0.0115	-0.0605	-0.0010	0.0285
	(-0.8605)	(-0.1094)	(0.2657)	(-1.2930)	(-0.1133)	(1.1695)
Post-GFC	0.0129**	0.0133**	0.0138**	0.0018	0.0037	0.0016
	(3.1320)	(3.3708)	(2.9397)	(0.4799)	(1.2771)	(0.7437)
Ex-GFC	0.0117***	0.0114***	0.0110**	0.0052	0.0035	0.0015
	(3.4709)	(3.7244)	(3.0128)	(1.6042)	(1.5939)	(0.8310)

This table presents raw monthly returns (t-statistics in brackets) for decile portfolios sorted separately by past return (*Mom*), news sentiment (*Sent*) and residual news coverage (*Attn*), for four subperiods within the sample. *High* represents a long-only portfololio of the top decile for the given sorting variable and *HML* represents a zero-cost portfolio that buys (sells) the top (bottom) decile for that variable. *Pre-GFC* is the period from 2003-01-01 to 2008-08-31, *GFC* is the period from 2008-09-01 to 2009-09-30, *Post-GFC* is the period from 2009-10-01 to 2017-12-31, and *Ex-GFC* includes the *Pre-* and *Post-GFC* periods.

order of the labels reflects the order in which the stocks were sorted.

Thus, the *LPos* portfolio sorts on momentum and then sentiment, and buys momentum losers (bottom tercile) with positive sentiment (top tercile); this is the lower left portfolio in Panel B of Table 5.4. Similarly, the *WHigh* portfolio sorts on momentum and then attention, and buys momentum winners (top tercile), with high attention (top tercile); this is the upper left portfolio in Panel A of Table 5.9. The *NegHigh* portfolio sorts stocks by sentiment and then attention, and then buys negative sentiment (bottom tercile) stocks with high attention; this is the lower left portfolio in Panel B of Table 5.10. The *LPos - LNeg* portfolio buys the *LPos* and sells the *LNeg* portfolio. The *WHigh - WLow* portfolio buys the *WHigh* and sells the *WLow* portfolio. The *LPos - LNeg* portfolios buys the *NegHigh* and sells the *NegLow* portfolio.

Table 5.16: Sub-period returns to double-sorted momentum, sentiment and attention portfolios

	LPos	WHigh	NegHigh	LPos -LNeg	WHigh -WLow	NegHigh -NegLow
Pre-GFC	0.0103*	0.0046	0.0083	0.0029	0.0025	0.0023
	(2.1105)	(0.9160)	(1.7703)	(1.0499)	(1.0856)	(1.1192)
GFC	-0.0108	0.0111	0.0119	0.0030	0.0025	0.0221
	(-0.4377)	(0.2835)	(0.3039)	(0.3577)	(0.5346)	(1.5926)
Post-GFC	0.0135**	0.0134**	0.0148**	0.0019	0.0017	0.0036*
	(3.3537)	(3.2690)	(3.3778)	(0.9313)	(1.3125)	(2.2503)
Ex-GFC	0.0124***	0.0102**	0.0124***	0.0023	0.0020	0.0031*
	(3.9684)	(3.1934)	(3.8061)	(1.3815)	(1.7040)	(2.4795)

This table presents raw monthly returns (t-statistics in brackets) for portfolios double-sorted by past return, news sentiment and residual news coverage, for four subperiods within the sample. Key: *W* (*L*) = high (low) momentum, *Pos* (*Neg*) = high (low) news sentiment, *High* (*Low*) = high (low) residual news coverage. I.e. *LPos* buys momentum losers with high sentiment, and *LPos-LNeg* buys (sells) momentum losers with high (low) sentiment. *WHigh* buys momentum winners with high residual news coverage, and so on. The double-sorting procedure for each portfolio is as described in Tables 5.4, 5.9, and 5.10. *Pre-GFC* is the period from 2003-01-01 to 2008-08-31, *GFC* is the period from 2008-09-01 to 2009-09-30, *Post-GFC* is the period from 2009-10-01 to 2017-12-31, and *Ex-GFC* includes the *Pre-* and *Post-GFC* periods.

5.3.4 Alternate portfolio formation procedures

Concordant news portfolios

Up to this point, the numerical values of the news sentiment scores have had no bearing on portfolio formation, as portfolios have relied only on ranks. The same is true for past return. However, it may be the case that the valence of news scores are more important for news-momentum interaction effects than the magnitude of the scores. For example, a momentum winner with negative news may have negative expected return, even if that news is among the most positive of the winners. Here we test a variation of the standard momentum strategy, in which only stocks with concordant news tone (a news tone score in the same direction of the trade position) are held. Since we are conditioning on the valence of the scores provided by Thomson Reuters, we do not center firms' formation period sentiment scores for these tests.

One approach to implementing such a strategy is to first form a standard decile momentum portfolio, then simply remove stocks from the long (short) leg of the portfolio with a numerically negative (positive) news tone score, and redistribute equity equally among the remaining positions in that leg. We label this portfolio *A1*. One might further implement an analogous time-series momentum rule: remove stocks from the long (short) leg of the portfolio with numerically negative (positive) momentum, and redistribute equity equally among the remaining positions in that leg. We label this portfolio *A2*.

A second approach is to seek replacements for the stocks removed from the long (short) leg of the portfolio by moving outside the top (bottom) momentum decile and including the next highest (lowest) ranked stock with concordant news tone. This approach compromises on momentum ranking to maintain as close as possible to the original number of stocks in each leg. We label this portfolio *B1*. One can again remove stocks from the long (short) leg with numerically negative (positive) momentum. We call this portfolio *B2*. For comparison, we also test a portfolio (*C1*) in which only stocks with concordant momentum are held and in which news tone is not considered.

Table 5.17 documents mean monthly returns for each of the portfolios described above, as well as the standard decile momentum portfolio (*Mom*). Portfolios *A1*, *A2* and *B2* each generate statistically significant returns in the ex-GFC period, indicating that filtering out positions with discordant news sentiment does lead to a superior momentum portfolio for the sample period. The relatively poor performance of portfolio *B1* suggests that this benefit is lost if discordant news stocks are replaced with those with discordant momentum, while the performance of *C1* indicates that removal of discordant momentum stocks itself has almost no effect.

Table 5.17: Subperiod returns to concordant momentum portfolios

	A1	A2	B1	B2	C1	Mom
Pre-GFC	0.0126 (1.9480)	0.0127 (1.9560)	0.0079 (1.5868)	0.0117* (2.0534)	0.0113 (1.8450)	0.0113 (1.8441)
GFC	-0.0561 (-1.0758)	-0.0430 (-0.9287)	-0.0460 (-1.0903)	-0.0480 (-1.0514)	-0.0635 (-1.3155)	-0.0595 (-1.2908)
Post-GFC	0.0077 (1.7522)	0.0077 (1.7394)	0.0041 (1.1576)	0.0058 (1.5365)	0.0019 (0.5133)	0.0018 (0.4799)
Ex-GFC	0.0095* (2.5976)	0.0095* (2.5921)	0.0055 (1.9009)	0.0079* (2.5026)	0.0053 (1.6255)	0.0052 (1.6026)

This table presents raw monthly returns (t-statistics in brackets) of six variants of zero-cost momentum portfolios, for four subperiods within the sample. *Mom* is a standard decile momentum portfolio that sorts all stocks by momentum and buys (sells) those in the highest (lowest) decile. Portfolio *A1* takes the decile momentum portfolio and removes winners (losers) with negative (positive) news sentiment. Portfolio *A2* follows the procedure of *A1*, but also removes winners (losers) with negative (positive) momentum. Portfolio *B1* takes a list of stocks sorted by momentum, and moves through the ranks from each end, discarding winners (losers) with numerically negative (positive) news sentiment until it reaches 50 stocks in each leg or runs out of stocks with concordant news. Portfolio *B2* follows the procedure of *B1*, but also removes winners (losers) with negative (positive) momentum. Portfolio *C* takes the decile momentum portfolio and removes winners (losers) with negative (positive) momentum. *Pre-GFC* is the period from 2003-01-01 to 2008-08-31, *GFC* is the period from 2008-09-01 to 2009-09-30, *Post-GFC* is the period from 2009-10-01 to 2017-12-31, and *Ex-GFC* includes the *Pre-* and *Post-GFC* periods.

Given that news sentiment is strongly positively skewed over most of the sample period (see Chapter 3), we would expect that the increased performance of the concordant portfolios is primarily due to the removal of stocks with positive news sentiment from the short side. This is confirmed by Figure 5.5, which illustrates the mean cumulative holding period return for all stocks in each leg of the standard momentum strategy compared to the subset with concordant news sentiment (i.e. the long and short legs of strategy *A1*). On average, over the entire sample period, a portfolio of momentum winners performs about as well as a portfolio of momentum losers. Excluding the GFC-affected period, momentum winners are noticeably superior, but the difference is not large. In both periods, the subset of momentum losers with negative news sentiment are shown to perform much worse, on average, than the population of momentum losers.

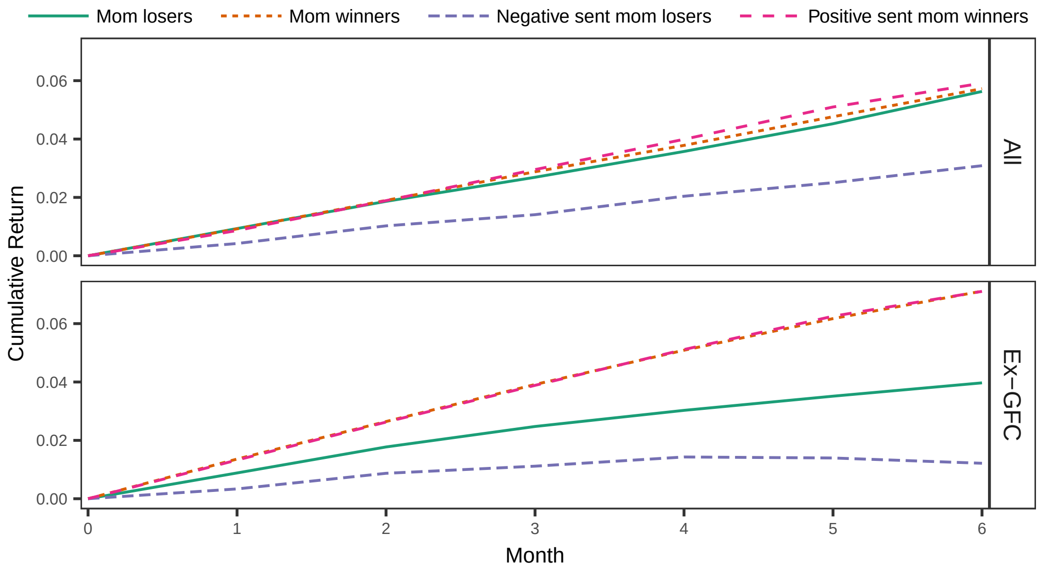

Figure 5.5: This figure shows mean buy-and-hold cumulative returns for the long and short legs of two variants of a momentum portfolio, throughout the six-month holding period. *Mom losers* are the decile of stocks with the lowest return over the previous six months. *Mom winners* are the decile of stocks with the highest return over the previous six months. *Negative sent mom losers* is the subset of *Mom losers* with numerically negative news sentiment. *Positive sent mom winners* is the subset of *Mom winners* with numerically positive news sentiment. News sentiment is measured as the mean daily sentiment score over the six-month formation period.

Despite their superior returns, the concordant portfolios, as tested, do not reflect realistic trading strategies; the number of stocks in each leg are unknown in advance, tend to be unbalanced between legs, and vary significantly through time. For example, the number of positions in each leg of portfolio *A1* throughout the sample period is shown in Figure 5.6. The average number of long (short) positions held by the portfolio is 41 (19), and reaches a minimum of 12 (1). In this regard, portfolio *B2* is only slightly more reasonable, with an average of 49 (30) long (short) positions and reaching a minimum of 2 (5). A more practical implementation of the concordant momentum strategy may select a small but fixed quantile of the lowest sentiment losers; this would concentrate holdings on average, and sacrifice some of whatever benefit lies in shorting only those stocks with numerically negative sentiment, but would avoid single-asset legs and provide a known and constant level of concentration. We test such a strategy in Section 5.3.6, but for now we continue to investigate the apparent importance of news valence within momentum losers.

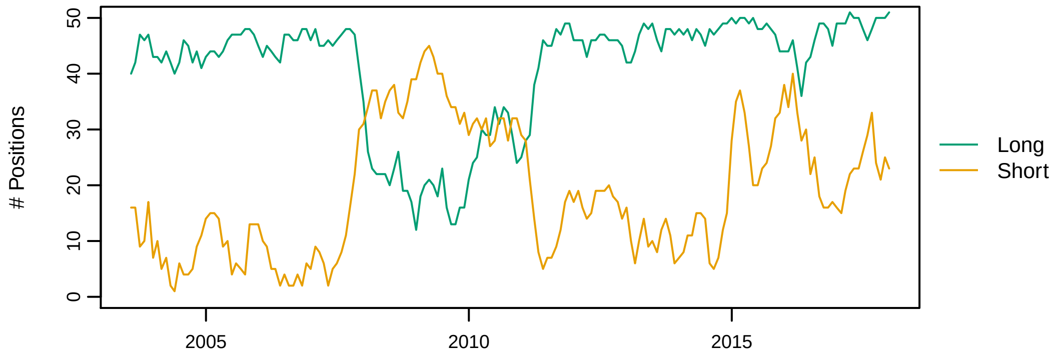

Figure 5.6: This figure shows the number of long and short positions in a portfolio that buys momentum winners with numerically positive news sentiment and sells momentum losers with numerically negative news sentiment. Momentum and news sentiment are measured over a six month formation period.

The effect obtained by removing momentum losers with numerically negative news sentiment suggests that there exists a nonlinearity in the response to news sentiment scores that is not captured by firms' relative (cross-sectional) positioning alone. However, due to the positive bias in news sentiment scores, screening for numerically negative news is likely equivalent to increasing the "precision" of the sentiment conditioning step. To investigate whether the response to negative sentiment is nonlinear, or whether we are just conditioning on a smaller portion of an approximately linear relationship, we sort the decile of momentum losers by news sentiment into a further five groups, and form a portfolio from each group. This is the same set up as the third column of Panel B, Table 5.4, except that rather than group by tercile and then tercile, we have grouped by decile (momentum losers) and quintile (the five sentiment groups).

Figure 5.7 depicts the mean monthly return for each sentiment group within the momentum losers. The relationship between mean return and news sentiment appears to be nonlinear, with all of the variance occurring at the extremes. Consistent with our previous results from the concordant portfolios, the greatest curvature is when moving into the stocks with the lowest sentiment–from group four to group five.

Yet, this does not necessarily indicate there is anything meaningful about negative scores; the difference between the second-lowest and lowest sentiment stocks may be larger, on

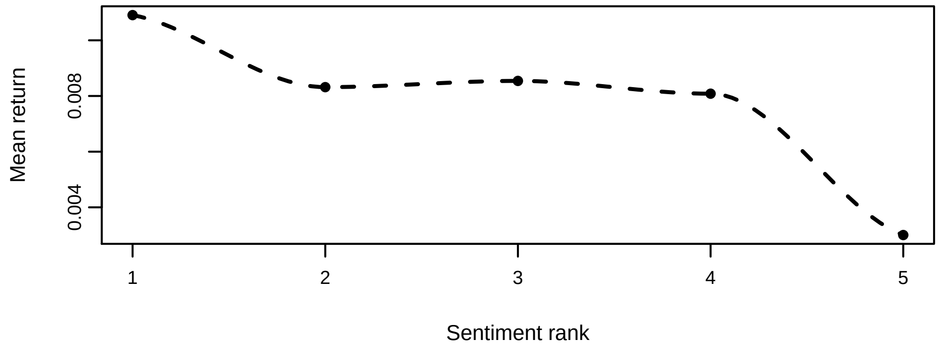

Figure 5.7: This figure depicts the performance of five sentiment quintile portfolios, formed within the decile of momentum losers, ordered from highest (1) to lowest (5) sentiment. Each month stocks are sorted by momentum and split into deciles. Within the lowest performing momentum decile (momentum losers), stocks are sorted by news sentiment and split into a further five groups. The mean monthly return of each portfolio are shown, joined by a loess smoothing function for visual aid. A six month formation period is used for momentum and sentiment, and portfolios are held for six months, with no rebalancing.

average, than the distance between other groups. Further, we do not know if moving from group four to group five actually captures the change in sign from positive to negative news sentiment. We can check these conditions by looking at the mapping from sentiment group to sentiment scores directly.

In Figure 5.8 the mean sentiment score, at the time of portfolio formation, is shown for each of the momentum loser sentiment quintiles throughout the sample period. We can see that group five does in fact capture the transition from positive to negative sentiment news items; the mean sentiment at formation is negative for the vast majority of the sample period, while that of group four is positive approximately half the time. The magnitude of the difference between group five and group four tends to be larger than the difference between the intermediate groups, but the difference between groups one and two is larger still, so this is unlikely to explain the results.

In summary, these results suggest that within our sample negative sentiment news scores attributed to firms with poor past returns contain information beyond their relative magnitude in the cross-section. This points to the potential benefits of non-arbitrary sorting procedures when relying on relative strength measures related to news sentiment, or the use of flexible learning algorithms capable of identifying such irregularities implicitly.

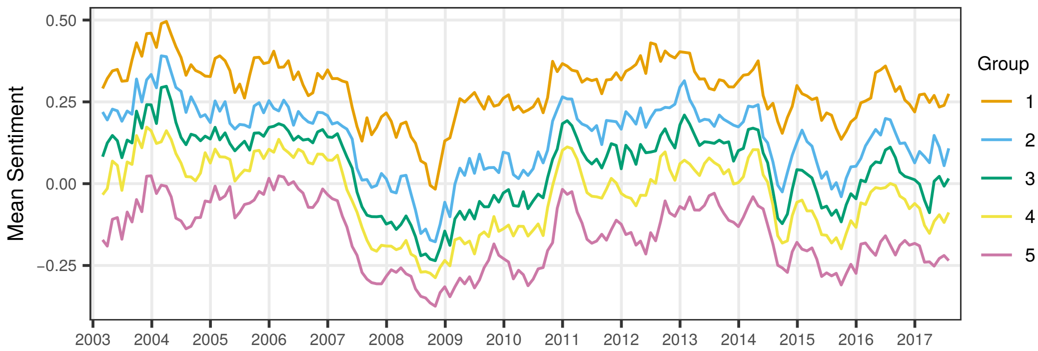

Figure 5.8: This figure shows the mean news sentiment scores, at portfolio formation, for the five momentum loser portfolios depicted in Figure 5.7. Each group, representing a quintile of stocks ranked by news sentiment from highest (group 1) to lowest (group 5), has been formed from within the decile of momentum losers. A six month formation period is used for momentum and sentiment, and portfolios are held for six months, with no rebalancing.

5.3.5 Risk Adjustments

In this section we document risk-adjusted returns for a number of key portfolios analysed in previous sections. Here we adopt the same naming convention used in Tables 5.15 and 5.16 of Section 5.3.2.

Table 5.18 documents risk-adjusted returns to long-only (*High*) and high-minus-low (*HML*) decile portfolios of stocks sorted separately by momentum, sentiment and residual attention, for the full sample (Panel A) and the full sample excluding the GFC-affected subperiod (Panel B). None of the long-only portfolios yield positive risk-adjusted return in either period, despite their strong performance in raw returns (for comparison, SPY generated a mean monthly return of 77 basis points over the complete sample period). Thus, the additional return produced by long-only portfolios can be attributed to increased exposure to market risk. On the other hand, the standard long-short momentum portfolio yields statistically significant positive alpha for both the market and three-factor models over the Ex-GFC period, and is the only univariate zero-cost portfolio to do so.

Table 5.19 presents risk-adjusted returns to long-only and long-short portfolios double-sorted by combinations of momentum, sentiment and residual attention. As in Section 5.3.3, *W* (*L*) represents top (bottom) tercile momentum stocks, *Pos* (*Neg*) represents top (bottom) tercile sentiment stocks, and *High* (*Low*) represents top (bottom) tercile attention stocks. The order of the labels reflects the order in which the stocks were sorted. The risk-adjusted results for the double-sorted portfolios are qualitatively the same as for the single-sorted portfolios; strong raw returns translating to insignificant alphas. The exception here is the portfolio of low momentum stocks with high sentiment (*LPos*), which

Table 5.18: Risk-adjusted returns to sentiment, momentum and attention

	TS Mean		CAPM		FF3		FF5	
	α	(t-stat)	α	(t-stat)	α	(t-stat)	α	(t-stat)
Panel A: All								
Sent H	0.0101**	(2.9048)	0.0015	(1.2569)	0.0014	(1.5500)	0.0018*	(1.9977)
Mom H	0.0088*	(2.4174)	0.0004	(0.2677)	0.0004	(0.2323)	0.0006	(0.3554)
Attn H	0.0108*	(2.2933)	-0.0003	(-0.1655)	0.0000	(0.0161)	0.0010	(0.5920)
Sent H-L	0.0031	(1.4729)	0.0026	(1.3253)	0.0021	(1.4527)	0.0026	(1.7349)
Mom H-L	-0.0001	(-0.0157)	0.0025	(0.5919)	0.0021	(0.4669)	0.0011	(0.2688)
Attn H-L	0.0033	(1.3222)	-0.0008	(-0.4225)	-0.0006	(-0.3181)	0.0012	(0.5501)
Panel B: Ex-GFC								
Sent H	0.0115***	(3.7266)	0.0014	(1.2906)	0.0016	(1.8183)	0.0020 *	(2.4462)
Mom H	0.0117***	(3.4711)	0.0015	(1.1499)	0.0019	(1.4870)	0.0015	(1.1446)
Attn H	0.0110**	(3.0082)	-0.0008	(-0.5915)	-0.0006	(-0.4312)	-0.0000	(-0.0284)
Sent H-L	0.0035	(1.5998)	0.0025	(1.2727)	0.0026	(1.7251)	0.0033 *	(2.0620)
Mom H-L	0.0052	(1.6026)	0.0061 *	(2.1820)	0.0063 *	(2.2337)	0.0047	(1.7915)
Attn H-L	0.0014	(0.7986)	-0.0026	(-1.9177)	-0.0025	(-1.8212)	-0.0013	(-1.0289)

This table presents raw and risk-adjusted returns (t-statistics in brackets) to portfolios sorted separately news sentiment, momentum and residual news attention. α is the intercept term in each of the factor regressions. *TS Mean* is the raw time series mean return, *CAPM* is the market model, *FF3* is the Fama and French (1993) model, and *FF5* is the Fama (2015) model. *Sent H*, *Mom H*, and *Attn H* are long-only decile portfolios that buy the top decile of stocks sorted by new sentiment, momentum and residual news coverage, respectively. *Sent H-L*, *Mom H-L*, and *Attn H-L* are their zero-cost long-short equivalents, which buy (sell) the top (bottom) decile of stocks sorted by news sentiment, momentum and residual news attention, respectively. Panel A documents results for the full sample period. Panel B documents results for the full sample, excluding observations between 2008-09-01 and 2009-09-30.

produced small, significant risk-adjusted return for each of the factor models in the Ex-GFC period.

Table 5.20 presents risk-adjusted returns to long-only and long-short portfolios triple-sorted by combinations of momentum, sentiment and residual attention. Unlike the single- and double-sorted long-only portfolios, the portfolios combining low sentiment, high attention, and positive momentum (*NegHighW* and *WNegHigh*) generate positive, statistically significant adjusted returns over both sample periods. Additionally, the zero-cost portfolio that buys low sentiment stocks with high attention and positive momentum, and sells those with low attention and low momentum (*NegHighW - NegLowL*), produces significant adjusted returns for the market and three factor models in the Ex-GFC

Table 5.19: Risk-adjusted returns to double-sorted momentum, sentiment and attention portfolios

	TS Mean		CAPM		FF3		FF5	
	α	(t-stat)	α	(t-stat)	α	(t-stat)	α	(t-stat)
Panel A: All								
LPos	0.0103**	(2.9595)	0.0017	(1.6450)	0.0018	(1.7176)	0.0020	(1.8921)
WHigh	0.0102*	(2.4456)	0.0002	(0.1638)	0.0004	(0.2616)	0.0005	(0.3524)
NegHigh	0.0122**	(2.8892)	0.0021	(1.5963)	0.0025	(1.7496)	0.0022	(1.5988)
LPos-LNeg	0.0024	(1.4361)	0.0016	(1.0652)	0.0014	(0.9915)	0.0014	(0.9569)
WHigh-WLow	0.0020	(1.7437)	0.0001	(0.0407)	0.0001	(0.1138)	0.0010	(0.8107)
NegHigh-NegLow	0.0045**	(2.7913)	0.0019	(1.4801)	0.0020	(1.4722)	0.0024	(1.5144)
Panel B: Ex-GFC								
LPos	0.0124***	(3.9667)	0.0021*	(2.0563)	0.0024*	(2.3920)	0.0022*	(2.1297)
WHigh	0.0102**	(3.1980)	-0.0002	(-0.2282)	-0.0001	(-0.0557)	0.0004	(0.3960)
NegHigh	0.0124***	(3.8038)	0.0017	(1.4863)	0.0019	(1.9572)	0.0015	(1.4659)
LPos-LNeg	0.0023	(1.3877)	0.0014	(0.8495)	0.0016	(1.0204)	0.0018	(1.1013)
WHigh-WLow	0.0019	(1.6722)	-0.0005	(-0.3709)	-0.0004	(-0.3178)	0.0003	(0.2880)
NegHigh-NegLow	0.0031*	(2.4560)	0.0007	(0.6448)	0.0007	(0.6846)	0.0008	(0.7046)

This table presents raw and risk-adjusted returns (t-statistics in brackets) for portfolios double-sorted by momentum, news sentiment and residual news attention. α is the intercept term in each of the factor regressions. *TS Mean* is the raw time series mean return, *CAPM* is the market model, *FF3* is the Fama and French (1993) model, and *FF5* is the Fama (2015) model. Key: *W* (*L*) = high (low) momentum, *Pos* (*Neg*) = high (low) news sentiment, *High* (*Low*) = high (low) residual news coverage. I.e. *LPos* buys momentum losers with high sentiment, and *LPos-LNeg* buys (sells) momentum losers with high (low) sentiment. *WHigh* buys momentum winners with high residual news coverage, and so on. The double-sorting procedure for each portfolio is as described in Tables 5.4, 5.9, and 5.10. Panel A documents results for the full sample period. Panel B documents results for the full sample, excluding observations between 2008-09-01 and 2009-09-30.

period and for the market model in the complete period. Unlike the zero-cost momentum portfolio, it also produces significant positive raw returns in both periods.

Although we have presented subperiod and risk-adjusted returns to a number of profitable double-sorted and triple-sorted portfolios, we do not believe these strategies reflect realistic trading opportunities over the sample period due to their ex-ante implausibility. In our eyes, it would not have made sense to an investor in 2003 (any more than it does now) to invest in low sentiment stocks with high momentum, or to sell momentum winners with low sentiment and low attention but high momentum, for example. We document additional tests for such portfolios only for completeness.

Not withstanding the implausibility of a number of tested portfolios, the implementation of something analogous to the concordant portfolios examined in Section 5.3.4 seems reasonable; A momentum trader, having selected their short positions based on price information alone, foregos those that are trailing positive or optimistic news stories. Or, having obtained an additional information source (news), a momentum trader concentrates their positions among those in which both “signals” (trend and news) are in agreement. This is still quite a naive strategy, since momentum winners with high sentiment perform worse than those with low sentiment, in our sample (Table 5.4).

Table 5.20: Risk-adjusted returns to triple-sorted momentum, sentiment and attention portfolios

	TS Mean		CAPM		FF3		FF5	
	α	(t-stat)	α	(t-stat)	α	(t-stat)	α	(t-stat)
Panel A: All								
NegHighW	0.0125**	(3.3008)	0.0037*	(2.2145)	0.0039*	(2.4852)	0.0033*	(2.2403)
WNegHigh	0.0122**	(3.3097)	0.0036*	(2.2029)	0.0038*	(2.3728)	0.0034*	(2.3105)
LPosHigh	0.0108*	(2.3166)	0.0000	(0.0184)	0.0003	(0.1384)	0.0010	(0.5905)
NegHighW - NegLowL	0.0052	(1.9281)	0.0047 *	(1.9659)	0.0046	(1.8965)	0.0038	(1.8064)
WNegHigh - WNegLow	0.0050**	(2.7898)	0.0031	(1.4263)	0.0032	(1.4882)	0.0036	(1.6788)
LPosHigh - LNegLow	0.0062	(1.4921)	-0.0032	(-1.5926)	-0.0029	(-1.3596)	-0.0027	(-1.4875)
Panel B: Ex-GFC								
NegHighW	0.0145***	(4.2816)	0.0043*	(2.4120)	0.0046**	(2.8035)	0.0031*	(2.1495)
WNegHigh	0.0143***	(4.3315)	0.0042*	(2.4709)	0.0045**	(2.8428)	0.0033*	(2.4096)
LPosHigh	0.0109**	(3.0789)	-0.0002	(-0.1360)	0.0001	(0.0714)	0.0012	(0.8158)
NegHighW - NegLowL	0.0068**	(2.6777)	0.0055*	(2.3101)	0.0058*	(2.4489)	0.0040	(1.9529)
WNegHigh - WNegLow	0.0048*	(2.5928)	0.0023	(1.0062)	0.0024	(1.0841)	0.0023	(1.0770)
LPosHigh - LNegLow	0.0072*	(2.0418)	-0.0032	(-1.4720)	-0.0031	(-1.4347)	-0.0028	(-1.4705)

This table presents raw and risk-adjusted returns (t-statistics in brackets) for portfolios triple-sorted by momentum, news sentiment and residual news attention. α is the intercept term in each of the factor regressions. *TS Mean* is the raw time series mean return, *CAPM* is the market model, *FF3* is the Fama and French (1993) model, and *FF5* is the Fama (2015) model. Key: *W* (*L*) = high (low) momentum, *Pos* (*Neg*) = high (low) news sentiment, *High* (*Low*) = high (low) residual news coverage. The order of keys reflects sorting order. *NegHighW* buys momentum winners (W) within the high attention tercile (High) of stocks with low news sentiment (Neg). *NegHighW - NegLowL* buys the *NegHighW* portfolio and sells momentum winners (W) within the low attention tercile (Low) of negative sentiment stocks (Neg). See Table 5.11 for details of sorting procedure. *WNegHigh* buys high attention stocks within the negative sentiment tercile of momentum winners. *WNegHigh - WNegLow* buys the *WNegHigh* portfolio and sells low attention stocks within the negative sentiment tercile of momentum winners. See Table 5.12 for details on sorting procedure. *LPosHigh* buys high attention stocks (High) within the positive sentiment tercile (Pos) of momentum losers (L). *LPosHigh - LNegLow* buys the *LPosHigh* portfolio and sells low attention stocks within the negative sentiment tercile of momentum losers. See Table 5.12 for details on sorting rocedure. Panel A documents results for the full sample period. Panel B documents results for the full sample, excluding observations between 2008-09-01 and 2009-09-30.

In addition to the concordant portfolios already presented, we also test a more practical implementation of the same idea, that buys (sells) the quintile of decile momentum winners (losers) with the highest (lowest) news sentiment. Positions are concentrated to 10-11 stocks on each leg, but this number remains constant through time. We label this portfolio is *D*. Table 5.21 documents risk-adjusted returns for concordant portfolios *A1*, *A2*, *B1*, *B2*, and *D*, for the full sample and ex-GFC periods. Each of the concordant portfolios are shown to produce statistically significant risk-adjusted returns for each of the models in the ex-GFC period.

Table 5.21: Risk-adjusted returns to concordant momentum portfolios

	TS Mean		CAPM		FF3		FF5	
	α	(t-stat)	α	(t-stat)	α	(t-stat)	α	(t-stat)
Panel A: All								
A1	0.0043	(0.8066)	0.0071	(1.5615)	0.0066	(1.3612)	0.0058	(1.2475)
A2	0.0053	(1.0757)	0.0079*	(1.9887)	0.0074	(1.7080)	0.0070	(1.7667)
B1	0.0014	(0.3339)	0.0032	(0.8930)	0.0027	(0.7055)	0.0025	(0.6679)
B2	0.0035	(0.7634)	0.0056	(1.4795)	0.0051	(1.2411)	0.0050	(1.3161)
D	0.0047	(0.9168)	0.0069	(1.6610)	0.0062	(1.4327)	0.0057	(1.3688)
Panel B: Ex-GFC								
A1	0.0095*	(2.5976)	0.0105**	(3.1471)	0.0107***	(3.3334)	0.0097**	(3.0593)
A2	0.0095*	(2.5921)	0.0105**	(3.1350)	0.0107***	(3.3195)	0.0097**	(3.0406)
B1	0.0055	(1.9009)	0.0053*	(1.9731)	0.0056*	(2.2839)	0.0054*	(2.1853)
B2	0.0079*	(2.5026)	0.0080**	(2.7440)	0.0084**	(3.1701)	0.0078**	(2.9027)
D	0.0090*	(2.0938)	0.0092**	(2.5908)	0.0093**	(3.0656)	0.0087**	(2.6604)

This table presents risk-adjusted returns (t-statistics in brackets) of four variants of zero-cost momentum portfolios, for two subperiods within the sample. Portfolio *A1* takes a decile momentum portfolio and removes winners (losers) with negative (positive) news sentiment. Portfolio *A2* follows the procedure of *A1*, but also removes winners (losers) with negative (positive) momentum. Portfolio *B1* takes a list of stocks sorted by momentum, and moves through the ranks from each end, discarding winners (losers) with numerically negative (positive) news sentiment until it reaches 50 stocks in each leg or runs out of stocks with concordant news. It then removes winners (losers) with negative (positive) momentum. Portfolio *D* takes the decile momentum portfolio and buys (sells) the fifth of winners (losers) with the most positive (negative) news sentiment. Panel A provides results from the whole sample period and Panel B provides results for the whole sample period, excluding the period from 2008-09-01 to 2009-09-30.

5.3.6 Additional tests

Unconditional sorts

The multi-way sorting procedures which have been the focus of this study do not necessarily capture the highest unconditional combinations of each variable. For example, if momentum is negatively correlated with news sentiment, then the highest momentum tercile may be the lowest in terms of sentiment. Depending on how these variables are jointly distributed over the cross-section of stocks, extremes in the unconditional combination of each variable may not occur in the extremes of the primary ranking variable—a fact related to the "mixing" versus "integrating" approaches to factor portfolio construction discussed by Fitzgibbons et al. (2017).

Table 5.22: Returns to uncondtional triple-sorted portfolios

	Distance		Summation		Product	
Decile	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
High	0.0095**	(2.8207)	0.0095**	(2.7189)	0.0096*	(2.4966)
2	0.0081*	(2.4022)	0.0087*	(2.5848)	0.0090*	(2.5588)
3	0.0101**	(3.0293)	0.0097**	(2.8944)	0.0096**	(2.8116)
4	0.0099**	(3.0172)	0.0096**	(2.9345)	0.0099**	(2.8772)
5	0.0093**	(2.8722)	0.0096**	(2.8479)	0.0099**	(2.8698)
6	0.0098**	(2.9090)	0.0094**	(2.7153)	0.0100**	(2.9686)
7	0.0092**	(2.6765)	0.0097**	(2.7778)	0.0093**	(2.7995)
8	0.0091*	(2.5104)	0.0099**	(2.7003)	0.0091**	(2.6595)
9	0.0097*	(2.4501)	0.0089*	(2.3288)	0.0085*	(2.4037)
Low	0.0069	(1.5993)	0.0067	(1.6712)	0.0069	(1.7534)
High-Low	0.0019	(0.6802)	0.0025	(0.9731)	0.0026	(1.0935)

This table presents raw monthly returns to decile portfolios ranked by news sentiment, momentum and residual news coverage. Each portfolio formation procedure sorts on each variable individually and combines individual rankings unconditionally (sorting-order invariant). *Distance* interprets individual ranks as cartesian coordinates and calculates the 'distance' from the origin ($\sqrt{rank_i^2 + rank_j^2 + rank_k^2}$). *Summation* is the sum of individual ranks ($rank_i + rank_j + rank_k$). *Product* is the product of individual ranks ($rank_i \times rank_j \times rank_k$).

This feature of the sequential sorting process is of secondary importance when the purpose is to investigate specific joint effects and conditional relationships, as is the case with the current study. Nevertheless, we test a number of jointly sorted (or integrated) portfolios designed to capture the unconditional magnitude of each variable. Table 5.22

documents raw monthly returns to decile portfolios sorted by news sentiment, momentum and residual news attention, using three different simultaneous sorting procedures. Stocks are first sorted by each variable independently so that each stock has a momentum rank, a sentiment rank and an attention rank. The three sorting procedures then differ in how these ranks are combined.

The 'distance' approach interprets each variable as a coordinate axis, and a stock's set of ranks as a coordinate in this n-dimensional space (here n=3). A stock's score is then given by its distance from the origin:

$$distance = \sqrt{rank_i^2 + rank_j^2 + rank_k^2}$$

The 'summation' score of a stock is simply the sum of its ranks:

$$summation = rank_i + rank_j + rank_k$$

Similarly, the 'product' score of a stock is given by the product of its ranks:

$$product = rank_i \times rank_j \times rank_k$$

In each of these equations, the subscripts i, j, and k refer to sorting variables.

What we are looking for is either strong monotonicity or a statistically significant return on the long-short portfolio. Yet none of the unconditional sorting procedures lead to especially different results from the conditional sorts. As expected from our previous results, the combination of low sentiment, low momentum, and low attention gives rise to the lowest performance. Beyond this, there is no discernible relationship between the joint rank scores and mean return. Table 5.23 documents returns to unconditional pairwise sorts for the three variables, using the distance method to calculate the combined score. Again, none of the pairwise combinations are seen to move together unconditionally with expected return. This is consistent with the fact that the news-informed momentum literature has so far been focused on conditional relationships.

Table 5.23: Returns to unconditional double-sorted portfolios

	Mom & Sent		Attn & Sent		Mom & Attn	
Decile	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
High	0.0088**	(2.7000)	0.0102**	(2.9205)	0.0087*	(2.3835)
2	0.0091**	(2.9601)	0.0089**	(2.6384)	0.0095**	(2.9978)
3	0.0090**	(2.9716)	0.0093**	(2.8257)	0.0086**	(2.7992)
4	0.0090**	(2.7603)	0.0095**	(2.9396)	0.0092**	(3.0460)
5	0.0092**	(2.8562)	0.0085*	(2.5267)	0.0096**	(3.1313)
6	0.0102**	(2.9658)	0.0079*	(2.2662)	0.0099**	(3.1115)
7	0.0099**	(2.8417)	0.0092*	(2.5458)	0.0097**	(2.8014)
8	0.0094*	(2.4795)	0.0103**	(2.8626)	0.0092*	(2.4464)
9	0.0094*	(2.2510)	0.0107**	(2.8080)	0.0090*	(2.1466)
Low	0.0076	(1.5474)	0.0072	(1.8387)	0.0076	(1.3939)
High-Low	-0.0007	(-0.1658)	0.0031	(1.4483)	-0.0021	(-0.1880)

This table presents raw monthly returns to decile portfolios pairwise ranked by news sentiment, momentum and residual attention. Stocks are first sorted by each variable individually. A stocks score for a given pair of variables is then calculated as the root sum of squares of the individual ranks ($\sqrt{rank_i^2 + rank_j^2}$).

Temporal weighting of news information

As outlined in Section 5.2, news sentiment scores used throughout this study are computed by first aggregating news sentiment at the daily level, and then averaging the daily observations over the formation period. For visual reference, Figure 5.9 displays mean news sentiment scores for *Apple Inc* (ticker: AAPL) computed over daily, monthly, and six monthly formation periods. The aim of the aggregation procedure is to reduce the noise embedded within the raw, high-frequency observations and capture the underlying trend of the news content. Notwithstanding the construction of the per-item and daily sentiment scores, an unweighted average is the simplest way to do this.

Yet, if news content has any relationship to future performance, it is reasonable to assume that more recent news items will bear more heavily on the trajectory of a firm than older news items—especially given that we are operating with a formation period of six months. In this section, we investigate the effect of applying a time decay our news observations when constructing portfolios.

We first obtain monthly sentiment scores for each month in the formation period by averaging the daily news scores within each month. We then take a weighted average of

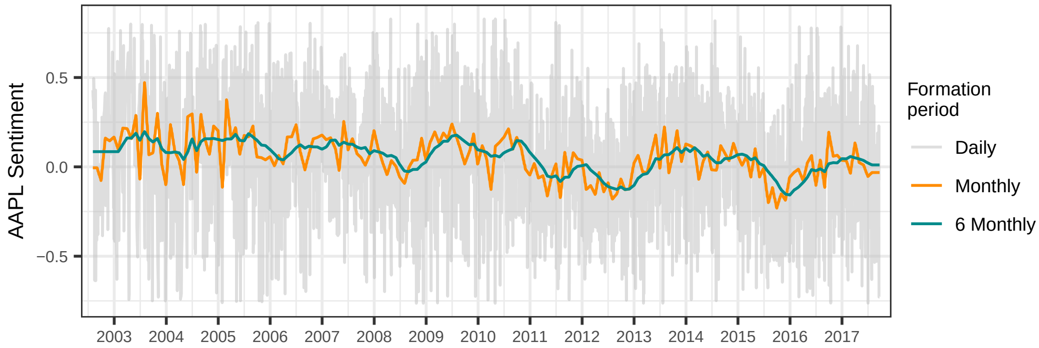

Figure 5.9: This figure shows mean news sentiment scores for *Apple Inc* (ticker: AAPL) calculated over daily, monthly and six monthly formation periods.

the monthly observations, such that the weight applied to each month in a given formation period decreases exponentially as a function of its lag-length from the current month. We use a weight ratio (*WR*) to describe the rate of this decay; a weight ratio of two implies that the sentiment score of the current month is weighted twice as heavily as the sentiment score of the oldest month in the formation period when calculating the combined score. These scores are then used to construct portfolios in the same manner as our original sentiment scores.

In order to use a weighting scheme such as temporal decay in data with missing observations, one needs to define a neutral direction in which to apply the decay. If missing data in a given formation period is excluded from the weighted average, this has the same effect as taking the missing values to be equal to zero. In the case of a positive mean variable with sample space on either side of zero (such as news sentiment), applying a decay toward zero generally translates to treating older observations as "more negative". In the context of the current study, this means that firms with only older news items will appear to have relatively low news sentiment, rather than merely dampened news sentiment. This is another reason one might consider mean- or median-centering the sentiment data, or otherwise making a deliberate decision as to what should be considered a neutral score, and whether missing values represent that neutral score. As discussed previously, we use median-centering due to our focus on cross-sectional groupings.

Table 5.24 documents mean monthly returns for sentiment decile portfolios using different weight ratios to construct the sentiment scores used for ranking. Note that a weight ratio of one is simply an equal-weighted average of each month. While the weighting schemes differ numerically, there is no qualitative significance to weighting recent observations of news sentiment more heavily than old ones. In each case, deciles one, eight and nine are the three highest performing portfolios, while decile 10 is the worst performing.

Table 5.24: Returns to news sentiment with temporal decay

	WR=1		WR=2		WR=4		WR=6	
Decile	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)	Mean	(t-stat)
Panel A: All								
High	0.0103**	(2.9530)	0.0106**	(2.9952)	0.0105**	(2.9805)	0.0104**	(2.9673)
2	0.0090**	(2.6983)	0.0088**	(2.6541)	0.0088**	(2.6661)	0.0088**	(2.7044)
3	0.0094**	(2.9312)	0.0095**	(2.9379)	0.0096**	(3.0003)	0.0099**	(3.0458)
4	0.0099**	(3.0521)	0.0098**	(3.0933)	0.0096**	(2.9927)	0.0093**	(2.9046)
5	0.0088*	(2.5949)	0.0088*	(2.5900)	0.0087*	(2.5737)	0.0087*	(2.5368)
6	0.0083*	(2.4164)	0.0079*	(2.3192)	0.0081*	(2.3984)	0.0084*	(2.5027)
7	0.0097**	(2.7849)	0.0092**	(2.6337)	0.0094**	(2.6841)	0.0093**	(2.6682)
8	0.0109**	(3.0365)	0.0115**	(3.2355)	0.0113**	(3.1529)	0.0111**	(3.0950)
9	0.0102**	(2.7300)	0.0103**	(2.7396)	0.0102**	(2.6941)	0.0104**	(2.7483)
Low	0.0080*	(2.0557)	0.0080*	(2.0621)	0.0083*	(2.1271)	0.0082*	(2.1128)
High-Low	0.0024	(1.1357)	0.0025	(1.2127)	0.0022	(1.0760)	0.0022	(1.0687)
Panel B: Ex-GFC								
High	0.0117***	(3.8209)	0.0119***	(3.8613)	0.0117***	(3.8273)	0.0117***	(3.8165)
2	0.0104***	(3.6779)	0.0102***	(3.5885)	0.0103***	(3.6281)	0.0103***	(3.6602)
3	0.0105***	(3.7795)	0.0104***	(3.7756)	0.0104***	(3.7951)	0.0107***	(3.8693)
4	0.0108***	(3.9722)	0.0109***	(4.0423)	0.0109***	(4.0351)	0.0106***	(3.9450)
5	0.0100***	(3.6283)	0.0101***	(3.6662)	0.0099***	(3.5914)	0.0098***	(3.5281)
6	0.0094***	(3.4473)	0.0091***	(3.3823)	0.0093***	(3.4497)	0.0097***	(3.6214)
7	0.0108***	(3.9944)	0.0105***	(3.7890)	0.0106***	(3.8430)	0.0106***	(3.8200)
8	0.0117***	(4.1352)	0.0122***	(4.3726)	0.0120***	(4.3424)	0.0116***	(4.2063)
9	0.0115***	(3.8760)	0.0114***	(3.8469)	0.0114***	(3.8373)	0.0116***	(3.8974)
Low	0.0090**	(2.6733)	0.0091**	(2.6876)	0.0092**	(2.7306)	0.0092**	(2.7512)
High-Low	0.0029	(1.3239)	0.0029	(1.3570)	0.0026	(1.2315)	0.0025	(1.1975)

This table presents raw monthly returns to stock portfolios sorted by news sentiment using four different termporal weighting schemes to aggregate sentiment over the formation period. The starting point for each portfolio is monthly observations of firm news sentiment, representing the average daily sentiment of all news days for the firm that month. Each portfolio then takes a weighted average of the monthly scores over the previous six months. The weight applied to each month in over given formation period decreases exponentially as a function of time. The weight ratio (\\textit{WR}) represents the rate of the temporal decay. Specifically, the weight ratio defines the ratio of the weight given to the most recent observation, to that of the oldest. Panel A documents results for the full sample. Panel B documents results for the full sample excluding the period from 2008-09-01 to 2009-09-30.

5.4 Discussion

A number of promising news-informed momentum strategies have been presented in the literature, but results are currently dispersed across studies using a wide variety of news sources, investment universes, formation and holding periods, variable construction techniques and testing methodologies. A major aim of the current analysis was to benchmark and validate variants of such strategies in a highly-liquid, ex-ante-identifiable investment universe with a consistent experimental set up. As this was not intended to a replication study, our portfolios are not direct analogues of those presented in the literature but they do capture the most important features.

Specifically, we tested the performance of long-only and long-short portfolios formed using single, double, and triple-sorting procedures based on news sentiment, price momentum, and news attention. With the minor exception of the univariate sentiment portfolios, which outperformed momentum under on a risk-adjusted basis under the five-factor model (but not single or three-factor models), none of the literature-motivated strategies delivered significant excess returns within our sample. Additional tests that applied a temporal weighting scheme to sentiment did not qualitatively impact results, nor did the application of joint sorting procedures, or changes to the sorting order of the portfolios. These results were also qualitatively unaffected when controlling for the GFC, which was a particularly bad time for momentum strategies.

An unexpected feature of our results was the large variation in the performance of portfolios formed from low-sentiment (bottom tercile) stocks; conditioning on news attention and momentum within the low-sentiment subset produced the type of results we may have expected from momentum-attention portfolios selected from the complete sample. A portfolio of low sentiment (bottom tercile) stocks that bought high-attention momentum winners and sold low-attention momentum losers generated statistically significant risk-adjusted returns over the ex-GFC sample. However, we do not find this particularly informative for two reasons. First, we are bound to find regions of any feature space (momentum, sentiment, and attention in this case) that generated outperforming portfolios, ex-post. Second, we were not operating under any behavioural model that would have made conditioning on low sentiment stocks a plausible first step for identifying both winners and losers. Still, it does raise the question as to whether this relationship could have been identified in-sample and subsequently exploited out-of-sample by a data-driven approach.

In further tests, we found evidence of news sentiment being an effective screening mechanism within a traditional momentum framework; losers with positive sentiment significantly outperform those with negative sentiment which makes the removal of positive sentiment losers from the short leg of a momentum strategy profitable. The breakpoint in performance appears to correspond to the numerical sign of the aggregate sentiment score, although it is difficult to isolate the impact of sharper conditioning (more extreme sentiment) from the numerical sign change. Several variants of 'concordant' momentum portfolios that seek to remove stocks with sentiment (or sentiment and momentum) in the opposite sign of the portfolio leg were found to generate statistically significant risk-adjusted returns over the ex-GFC period. This comes at the cost of having a variable number of stocks in the portfolio or selecting a more concentrated sorting procedure up front.

Overall, our findings do not reflect the enthusiasm for news-based conditioning observed in the related literature. While our results are not inconsistent with those of such analyses (due to experimental differences), they do put into question the generalisability and robustness of the apparently profitable trading strategies described therein. That said, our results leave some room for the utility of news-based features in momentum portfolios—ex-ante plausibility of any given relationship expressed in the data is arguably irrelevant if such relationships are persistent enough to be exploited by data-driven (i.e statistical) approaches to portfolio formation, and this may be the case for the profitable low sentiment portfolios described above. Further, the use-case of removing discordant sentiment items appears to be a simple means of improving the performance of momentum portfolios and may warrant further investigation.

Chapter 6

Model-Based Portfolios

The non-parametric or "naive" news-informed portfolios tested in Chapter 5 provided little evidence of news data being a useful predictor for the cross-section of returns over the horizons of interest. Yet, it could be argued that the monotonic sorting procedures employed there and in the relevant literature represent too narrow a hypothesis space of predictive relationships—that the assumptions embedded in the tests are too restrictive. By turning to more flexible modelling approaches, such as the machine-learning techniques found in the high-frequency news analytics literature, predictive relationships may be more likely to be uncovered should they exist.

In this section, we extend the factor-sorting approach used previously by conditioning on the output of supervised statistical learning models trained over a designated subset of data. Starting with a "classical" OLS-based approach to model-informed portfolio formation before applying a wider class of algorithms such as gradient-boosted-trees and neural networks, we test the predictive utility of news information in an environment largely uninhibited by prior distributional assumptions.

We find that the combined use of news-derived features and flexible statistical learning algorithms offers only a modest increase in theoretical performance beyond a traditional momentum implementation. Measures of variable importance suggest that news is secondary to size, analyst following, and momentum in relevance for predicting future return.

6.1 Introduction

In the first analysis, we investigated the firm-level, or longitudinal, predictive capacity of news in a linear regression setting. In the second analysis, we studied the impact of news sentiment, coverage, and price momentum on the cross-section of returns. There we used the model-free or non-parametric approach, common in the factor literature, of conditioning on our information using single-, double-, and triple-sorted quantile portfolios and studying their performance.

Here we continue to focus on the cross-sectional effects of news contents and momentum, but in a setting that encompasses and extends the previous approaches. We use statistical models to learn cross-sectional relationships from the data (insofar as they exist within the chosen models' hypothesis space) and distill these relationships into predictions for each firm. We then use the factor portfolio approach to assess the efficacy and economic utility of the modelled relationships, by conditioning on these model predictions.

The use of statistical models allows us to flexibly incorporate a number of variables and their interactions in a supervised manner that cannot necessarily be captured by the relatively crude partitioning of variable space reflected by decile portfolios. Yet, by conditioning on model predictions via the factor portfolio framework, we allow any informational value of the model to be expressed nonparametrically and assessed for its economic relevance. We start with a straight forward implementation of this idea using OLS regression, as described by Haugen and Baker (1996), before moving on to more modern approaches.

6.2 In-Sample Prediction

We begin by turning to the most straight-forward (model-based) assessment of cross-sectional informativeness—the Fama-MacBeth regression (Fama and MacBeth, 1973), known outside of Finance as the mean-groups estimator (Pesaran and Smith, 1995; Pesaran, Shin, and Smith, 1999). In continuing from the previous analysis, we use six-month excess return as the dependent variable, with news sentiment, news coverage, price momentum, and all first-order interactions, as regressors. Each regressor is measured over a trailing six-month horizon.

Table 6.1 details Fama-Macbeth regression statistics with HAC corrected standard errors (Andrews, 1991). All variables were standardised each period, so that regression

Table 6.1: Fama-Macbeth Regression Statistics

Variable	Estimate	Std. Error	t value	P value
Sent	-0.0080	0.0048	-1.6681	0.0972
Mom	0.0396	0.0415	0.9541	0.3414
Coverage	-0.0076	0.0054	-1.4202	0.1574
Sent:Mom	-0.0241	0.0058	-4.1608	0.0001***
Mom:Coverage	0.0004	0.0080	0.0544	0.9567
Sent:Coverage	0.0064	0.0031	2.0966	0.0375*

This table presents Fama-Macbeth regression coefficient statistics for the predictive regression of six month excess return against news sentiment, news coverage, price momentum, and all first-order interactions. Standard errors and test statistics are based on HAC standard errors Andrews (1991). All variables were standardised to have zero mean and unit variance in each period.

coefficients can be interpreted in terms of cross-sectional standard deviation for both predictor and response variables. We can see that momentum has the large effect size of the variables, though is not statistically significant. The second-largest effect is that of the sentiment-momentum interaction variable, which is significant below the 1% level. The sentiment-coverage interaction has a small effect, significant at the 5%. Neither the momentum-coverage interaction nor any of the individual variables are statistically significant at 5% or below.

The cross-sectional regression reveals at least one significant predictor of forward excess return. However, as this represents an in-sample result it does not necessarily mean that it could have been exploited over the sample. Similarly, coefficients that vary through time with an average near zero, will not be statistically significant using the means-group estimator, but may still vary slowly enough, and with large enough magnitude, for them to be capitalised on by an investor. The out-of-sample prediction framework used by Haugen and Baker (1996), which we consider next, helps us address these concerns.

6.3 Standard Linear Approach

Before turning to what might be considered a less traditional approach to empirical finance, we begin our model-driven portfolio analysis within the context of OLS regression.

Specifically, we follow the framework set out by Haugen and Baker (1996) in developing a dynamic model for expected stock returns and using it to make return forecasts. Unlike Haugen and Baker (1996), we only use variables related to stock price momentum and news media for our forecasts.

The expected return model is essentially a rolling Fama-Macbeth regression. Each month, we perform the following OLS regression to determine the monthly payoffs (cross-sectional regression coefficients) of the independent variable ('factors'):

$$r_{j,t} = \sum_i \hat{\gamma}_{i,t} x_{j,i,t-1} + \epsilon_{j,t} \tag{6.1}$$

where

- $r_{j,t}$ = excess rate of return to stock j in month t,
- $\hat{\gamma}_{i,t}$ = regression coefficient or payoff to factor i in month t,
- $x_{j,i,t-1}$ = exposure (firm characteristic such as 6-month momentum) to factor i for stock j at the end of month $t-1$,
- $\epsilon_{j,t}$ = unexplained component of return for stock j in month t.

This provides a history of the regression coefficients ('payoff histories') for each of the factors. The payoff histories can then be used to make out-of-sample forecasts by averaging the regression coefficients observed in the previous 6 months prior to the month to be forecast:

$$E(r_{j,t}) = \sum_i \bar{\gamma}_{i,t} x_{j,i,t-1} \tag{6.2}$$

where,

- $E(r_{j,t})$ = is the expected return to stock j in month t,
- $\bar{\gamma}_{i,t} = \frac{1}{N}\sum_{k=1}^{N} \bar{\gamma}_{i,t-k}$ is the expected payoff to factor i in month t (arithmetic mean of the estimated payoff over the previous N months), $x_{j,i,t-1}$ is the exposure (firm characteristic such as 6-month momentum) to factor i for stock j at the end of month $t-1$.

We diverge from the Haugen and Baker (1996) setup by using a six-month prediction horizon for the cross-sectional regressions. This means that the expected returns model is essentially lagged by six months, rather than one month, as the dependent variable is only available for model fitting once it has been observed. I.e. the filtration at time t can only include trailing variables. In order to keep the formation period equal to that used by Haugen and Baker, we average the $N = 6$ most recent models to form the current expected return model.

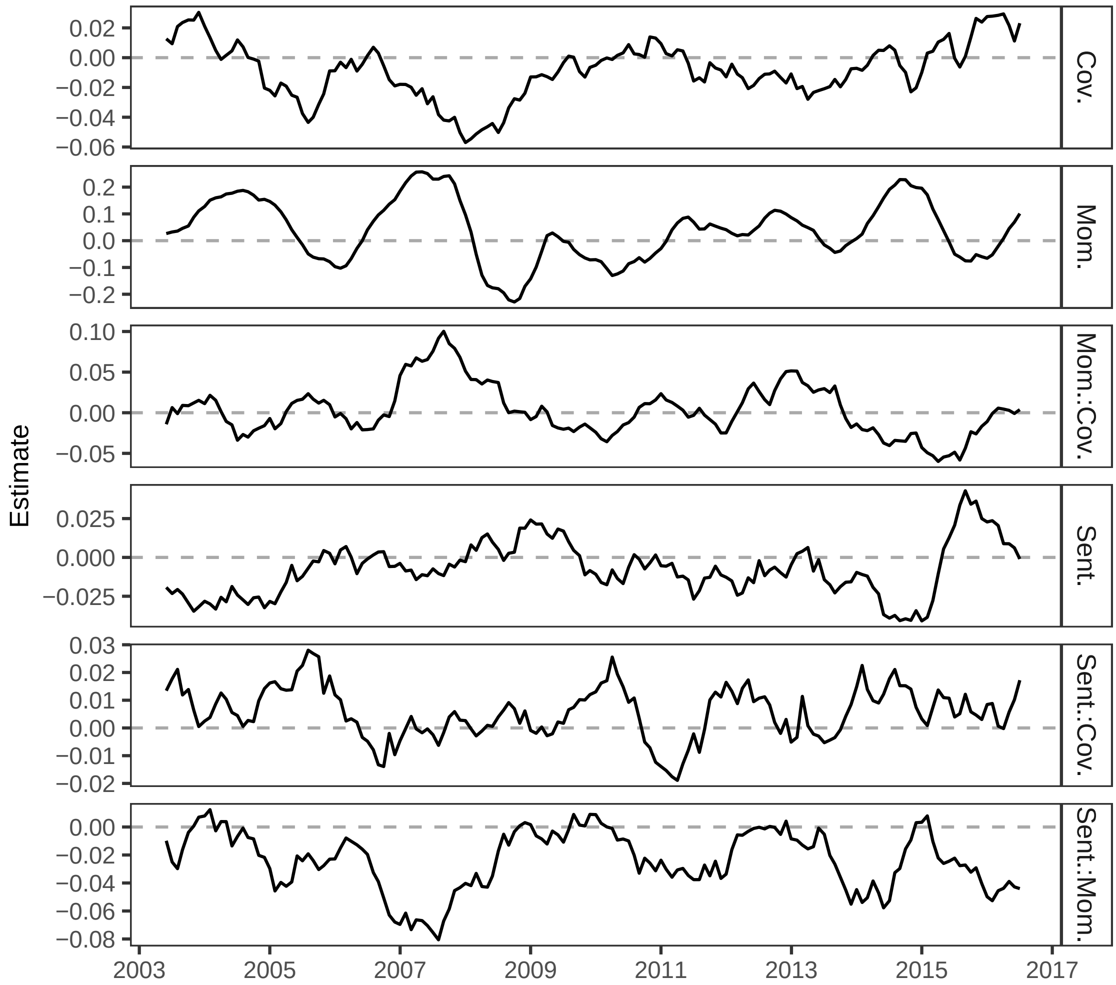

Figure 6.1: Timeseries of rolling expected return model coefficients.

We continue to use forward six month excess return as the independent variable, and six-month sentiment, six-month coverage, six-month price momentum, and all first-order interactions, as dependent variables. This is the same model used in the Fama-MacBeth regression, so the monthly coefficients are exactly those calculated in the first step of the Fama-MacBeth procedure. The timeseries of the coefficients for the expected returns model are shown in Figure 6.1. Consistent with the Fama-Macbeth regression results, the coefficient for the sentiment-momentum interaction term is the most consistent in its

Table 6.2: Expected Return Decile Portfolio Performance

	Decile										
Year	1	2	3	4	5	6	7	8	9	10	1-10
Annual Return											
2006	14.2%	16.7%	14.5%	14.9%	17.5%	16.1%	17.9%	18.1%	18.3%	13.5%	0.6%
2007	-15.1%	-6.5%	-4.3%	2.7%	1.8%	4.7%	5.4%	7.2%	7.6%	10.4%	-24%
2008	-45.5%	-34.7%	-33.5%	-30.3%	-32.9%	-34.9%	-35.4%	-41.8%	-46.2%	-55.6%	19.5%
2009	20.8%	23.2%	28.4%	33.1%	33.8%	39.1%	50.2%	54.6%	68.6%	88.4%	-66.6%
2010	22.9%	19.8%	20%	20.9%	21.2%	20.8%	19.6%	22.8%	23.3%	24.2%	-1.1%
2011	-1.8%	2.5%	3.3%	3.6%	1.6%	0.6%	-0.2%	-0.2%	-1.8%	-2.7%	1.2%
2012	15.9%	16.6%	16.8%	19.2%	17.3%	14.5%	15.2%	14.7%	17.1%	19%	-2.9%
2013	44.8%	38.9%	38.8%	36.4%	34.6%	35.4%	33.6%	32.9%	31.5%	35.8%	7.7%
2014	12.7%	13.3%	12.3%	14.6%	14.9%	14.7%	16.9%	16.2%	14.5%	13.7%	0.3%
2015	-8.4%	-1.3%	1.2%	-1.3%	1.7%	-2.4%	-0.2%	-2.9%	-3.1%	-7.5%	0.2%
2016	12.2%	9.8%	8.8%	12.4%	9.1%	13%	16.5%	19%	22.1%	27.3%	-11.4%
2017	15.7%	15.7%	18.5%	19.8%	20%	20%	16.9%	19.7%	19.9%	17.8%	-0.5%
Average returns											
	7.4%	9.5%	10.4%	12.2%	11.7%	11.8%	13.0%	13.3%	14.3%	15.4%	-6.4%
Annualised risk											
	17.6%	15.0%	14.5%	14.3%	14.6%	15.1%	15.9%	17.4%	18.9%	23.3%	18.9%

This table presents annual returns for decile portfolios formed on the basis of a rolling linear expected return model. Portfolio performance is calculated using the overlapping portfolio procedure with a holding period of six months.

margin from zero, while the momentum coefficient has the largest swings. The remaining terms stay quite close to zero for the majority of the sample.

Each month we use the expected return model to forecast the expected returns for all index constituents, and form equal-weighted decile portfolios based on these forecasts. The top decile (1) contains the high expected return stocks, while the bottom decile (10) contains the low expected return stocks. We also form a zero-cost high-minus-low portfolio from the top and bottom deciles. Positions are held for six-months without rebalancing. The strategy performance is based on the overlapping portfolio procedure discussed in the previous analysis.

Table 6.2 details the performance for each of the decile portfolios formed using the expected return model. The performance for each decile is almost perfectly opposed to the forecasts—the 'top' decile generates the worst return while the 'bottom' decile generates the highest. Although the risk of the bottom decile is also higher, the expected returns model is volatility blind (i.e. we were not forecasting risk-adjusted return), so this does not help the case for the model's performance. And, despite the higher volatility, the bottom decile still outperforms the top on a risk-adjusted basis. Increasing the model aggregation period to 12 months does not change this fact, and barely changes the overall ranking of the deciles. From these results we can conclude that the in-sample statistics do not translate into a profitable trading strategy given a straight-forward implementation of the model estimated using only trailing data.

6.3.1 Model comparison and extensions

Although the expected returns model lost money, it is still relevant to ask whether the inclusion of news sentiment and coverage added value to the model, whether it could be improved with a more flexible modeling technique, and whether any of these outperform an ex ante-defined model-free portfolio formation procedure (i.e. a traditional momentum procedure).

To asses whether the addition of news variables improved performance (within the confines of the linear regression setting), we test a linear model in line with the Haugen-Baker procedure described, but with momentum as the only regressor. The resulting model differs from the traditional momentum approach to the extent to which the momentum beta changes sign—when positive, it is identical to standard momentum, when negative, the rankings are reversed.

To assess whether the model is being limited by the linear constraint, we perform a similar procedure, but make use of a generalized additive model (GAM) (Hastie and Tibshirani, 1986; Hastie and Tibshirani, 1990) in place of the standard OLS regression. A GAM is a generalized linear model with a form something like:

$$E(Y|X_1, X_2, \dots) = \mathbf{A}^{\grave{}} + f_1(X_1) + f_2(X_2) + f_3(X_1, X_2) + \dots$$

Where Y is a response variable, f_j's are unspecified smooth (nonparametric) functions of the covariates X_i, and $\mathbf{A}^{\grave{}}$ represents any strictly parametric model components. When fitting a GAM, the basic structure of the model must be defined up front, as with standard linear regression. However, we need only specify the model in terms of 'smooth functions', with the overall shape of each function being left to the fitting procedure. GAMs can be represented using basis expansions for each smooth, each with an associated penalty controlling smoothness. In practice, this penalized likelihood maximisation problem is solved by penalized iteratively re-weighted least squares (PIRLS), with the degree of penalization (the smooothing parameters) estimated using generalized cross-validation or restricted maximum likelihood (REML) (Patterson and Thompson, 1971; Laird and Ware, 1982).

We use the GAM implementation described in Wood (2011), a very readable overview of which is given in Wood (2017). For correspondence with the linear approach, we use the same general model specification used for the OLS regression. Namely, the main effects plus first-order interaction structure described earlier. The difference being that each of these terms are now represented by smooth functions of tensors and tensor products, respectively.

Finally, for the traditional momentum strategy, the expected return model is simply each firm's six-month price momentum, standardized across firms. To recap, the four models are summarised as follows:

Nonparametric Momentum: $r_{t:t+h} \sim Mom_t$

Linear Model, Momentum: $r_{t:t+h} \sim \gamma_t \cdot Mom_t$

Linear Model, Full Specification: $r_{t:t+h} \sim \gamma_{1,t} \cdot Mom_t + \gamma_{2,t} \cdot Sent_t +$
$\gamma_{3,t} \cdot Coverage_t + \gamma_{4,t} \cdot (Sent_t \cdot Mom_t) + \gamma_{5,t} \cdot (Sent_t \cdot Coverage_t) +$
$\gamma_{4,t} \cdot (Coverage_t \cdot Mom_t)$

GAM, Full Specification: $r_{t:t+h} \sim f_{1,t}(Mom_t) + f_{2,t}(Sent_t) + f_{3,t}(Coverage_t) +$
$f_{4,t}(Sent_t, Mom_t) + f_{5,t}(Sent_t, Coverage_t) + f_{4,t}(Coverage_t, Mom_t)$

A simple way to compare the performance of each expected returns model is to compare their information coefficients throughout the sample period. The information coefficent (IC) is a widely-used practitioner-focused measure of forecast effectiveness, and is given by the cross-sectional correlation between forecasts and the realised returns (Grinold, 1989; Grinold and Kahn, 2000). We calculate IC using the Spearman rank correlation, as we are concerned with rank-order rather than the numerical accuracy of the estimates. Figure 6.2 plots the cumulative IC for each of the models. Cumulative IC, rather than the raw series, is used in illustration as it is a much easier way to visualise overall performance (Yashchin, Philips, and Stein, 1997; Philips, Yashchin, and Stein, 2003), in the same way that price is more visually intuitive than a monthly return series.

Figure 6.2: Cumulative rank information coefficient of of comparison expected returns models.

First, we can see that the naive momentum model is the only one that maintains a positive forecasting edge over the sample, while all the dynamic approaches lose any initial forecasting capacity from 2008 (GFC) onwards. The disparity between the nonparametric and parametric momentum approach implies that a trailing estimate of momentum's payoff is a poor guide to its current payoff. In the case of momentum, which has been recognised as a profitable trading strategy since at least ten years prior to the commencement of the current sample (Jegadeesh and Titman, 1993), there is justfication in stubbornly sorting stocks on past return *despite* the dynamic expected return model's evidence to the contrary. However, this justification does not exist for sentiment or coverage, hence, a rolling expected returns model is a defensible attempt to capture predictive value in the absence of strong prior assumptions.

We can also see that the addition of news variables to the linear momentum model does improve accuracy, with most of the forecasting benefit occurring post-GFC. Forecasts are improved further still by the allowance of nonlinear relationships via the GAM. One reason we might see such relative outperformance is through regularization alone; the GAM fitting procedure employs a penalty to control smoothness, therefore, if no cross-sectional relationships are observed, a penalized model could outperform by being *more* biased, i.e. closer to a constant or null model, than the linear one.

One way to see whether the flexibility of the GAM is being utilised, or is simply acting as a more regularized linear model, is to inspect the effective degrees of freedom (edf) for each of the model terms. The edf of a term will equal 1 if the model penalised the smooth term to a linear relationship and 0 if the term is penalized to exclusion, while nonlinearities are reflected by edfs greater than 1.

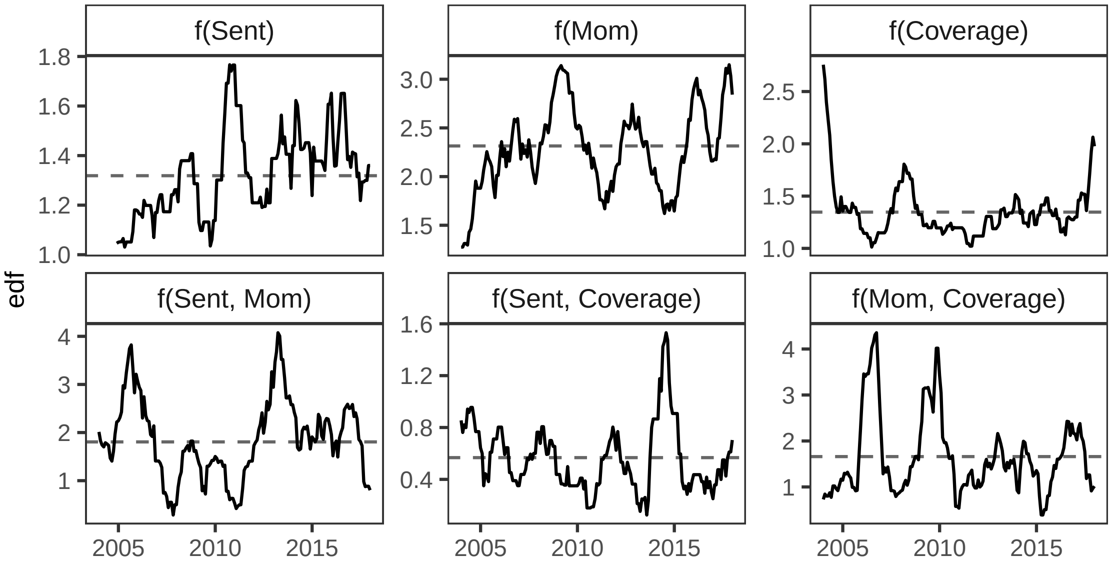

Figure 6.3: Effective degrees of freedom of the GAM expected returns model.

Figure 6.3 displays the rolling 12 month moving average edf for each of the model terms throughout the sample period. News sentiment and news coverage exhibit modest nonlinearity on average, and appear to be captured relatively well by a linear relationship, although the edf for sentiment appears to increase throughout the sample. The low edf of the sentiment-coverage interaction suggests that this term is largely redundant after the main effects have been accounted for. The relationship between forward return and momentum is consistently nonlinear from the GAM's perspective; this is consistent with the portfolio analysis results of the previous section, with most of the action occurring at the extremes. The sentiment-momentum and momentum-coverage interactions go through periods where the relationship to forward returns exhibit significant nonlinearities, though the stability of these relationships is clear. All-in-all, the GAM takes advantage of its access to greater degrees-of-freedom relative to the linear OLS model, with a mean edf of nine (compared to six for the linear model).

In this section, we found evidence of news content being a predictor of returns in the cross-section, through an in-sample Fama-MacBeth regression analysis. However, attempts to exploit these relationships in an out-of-sample context failed, with a blind momentum strategy outperforming all model-based approaches. Our results also suggested that, despite the overall failure of the forecasts, nonlinear dependencies between the variables of interest may exist, and appear to improve forecasting accuracy in a rolling-window framework.

One possible criticism of the Haugen-Baker approach is that it is too dynamic, or otherwise ill-equipped for dealing with a six-month forecast horizon, due to the six-month lag between the most recent observations of the dependent variable and deployment of the expected returns model. While it is no-doubt possible to adjust the aggregation horizon of the method and find improvements, this quickly ceases to become an out-of-sample exercise. To this point, and consistent with Haugen and Baker, we find significant sensitivity to the adjustment of such parameters. For instance, using a 36-month aggregation horizon for the expected return model, as used by Blitz, Huij, and Martens (2011), we find that the full linear model and GAM do outperform the naive momentum strategy, with the GAM still outperforming the linear model and the linear momentum-only model fairing the worst.

An alternative, and arguably more natural and robust approach, is to perform the standard train/validation/test split of the data upfront, and proceed along the usual lines of training and validating a model on the training and validation sets, and selecting a single model to test on the hold-out set. If the distribution of the training and validation sets are significantly different from the test set, the model will do poorly. However, this is a problem that comes with the territory, and we have already demonstrated the sliding window approach that is the nominal remedy for this issue.

6.4 Machine Learning Approach

In the previous section, we followed a procedure that estimated a cross-sectional model of future returns each month and averaged the N most recent of these to form the current expected returns model. For the nonparametric model used (GAM), model hyperparameters were set using a generalized cross-validation procedure (GCV) (Golub, Heath, and Wahba, 1979; Wahba, 1980) on the cross-sectional slice of the estimation month. However, this approach doesn't easily accommodate a more extensive use of machine learning.

Estimation of independent models each month limits the number of observations to be at most 500 (the number of index constituents), which for certain models is a small amount of data and may result in erroneous estimation of the prediction error (due to the slope of the training curve) and poor navigation of the bias-variance trade-off.

Accounting for time variation through model averaging is inherent in the Fama-Macbeth procedure and naturally extends to generalized linear models such as GAM. However, this is an unusual way of using the data in a machine-learning context and has no direct analogy for other classes of model, such as tree-based approaches. An argument could be made for an ensemble prediction approach that averages the predictions of independent cross-sectional models rather than the models themselves, but for large models this imposes a significant computational burden and we face the same data limitation cited above.

Further, the degree of model averaging in the previous approach, i.e. the choice of N in the above description, was ultimately ad-hoc[1]. In this section, models will be trained across a multi-period block of data, and hence time-variation of parameters within the training set will be accounted for implicitly. This does not guarantee (or even imply) better results, but it does remove a decision variable to which results are apparently sensitive, since the size of the training and validation set is dictated by the requirement to leave enough data for adequate out-of-sample testing. It also allows us to confidently apply a wider range of statistical models, as the data used to train and validate any particular model has increased.

6.4.1 Methodology

Overview

Our general procedure is to train a set of statistical models on a subsample of the data (all data prior to January 2012), with model evaluation and hyperparameter tuning performed within this period through cross-validation.

The performance of these models is then compared on the basis of in-sample statistics, and the best model is chosen for out-of-sample assessment (January 2012 - December 2017) against the benchmarks. The out-of-sample assessment is a portfolio-based analysis in which the chosen model makes rolling return predictions that are then used to construct decile portfolios.

Predictor Variables and Data Preparation

Our analysis utilises a relatively small set of available features:

- 6 month price momentum
- 12 month price momentum
- 6 month news sentiment
- 12 month news sentiment

[1] In the current analysis our choice was not ad-hoc; we followed the decisions made in the literature, but we say the choice was *ultimately* ad-hoc, as the estimation horizon used in the studies we followed was not founded on any particular conceptual or empirical claim.

- 6 month news coverage
- 12 month news coverage
- market capitalisation
- analyst following

This is the same set of variables used for the previous approach, but with the addition of the 12-month variants of news sentiment, news coverage and momentum, and market capitalisation and analyst following. The variable set is expanded as we now have much more data being used to train any given model. Further, the models tested in the first section required ex-ante specification of any interaction terms. Including all first-order interactions of the expanded set of variables increases the number of model terms from six to 36, making analysis unwieldy[2].

As before, we take the natural logarithm of news volume and market capitalisation to improve the distributional characteristics of these variables.

As we are concerned with cross-sectional forecasts and not market-timing, we standardise each variable by period (i.e. cross-sectional z-score). The median, rather than mean, was used for centering. Unlike the mean, the median is robust to outliers and leaves an equal number of companies either side of zero, which is intuitively appealing in the context of factor portfolio construction. This scaling results in most observations being bound between plus and minus three. Standardised values falling further than four standard deviations from the median for each period were binned at (plus or minus) four, in order to mitigate the impact of extreme values on model training and performance statistics.

For observations in which the recorded news volume was zero over the respective formation horizon, news sentiment was filled with the medians for that variable and date. Initially, a -1/1 dummy variable was added to indicate whether a company had a news volume of zero and had hence been median-filled. However, early tests indicated that this variable had no explanatory power so it was removed for simplicity.

[2]However, the question could still be posed "if the additional variables warrant inclusion ex-ante at all, what effect may they have had on the conclusions of the first analysis?". As a compromise, we performed the original analysis with with a model that included the additional predictors, but not their interactions. This kept the model specification relatively simple while still accounting for any benefit of the news variables that could be attributed to them being a proxy for market cap or analyst following. We found no qualitative difference from the results discussed, with the rank-order of all models remaining the same. However, the GAM ended with positive cumulative IC near that of the naive momentum strategy. As such, our conclusions from the first section do not change, but we now have additional evidence that the GAM was not providing value merely through regularization.

To further remove temporal variation in the cross-sectional distribution of each variable, additional tests were conducted using variable rankings instead of raw values, prior to standardisation. Monthly data was used for all model fitting and hyperparameter selection.

Learning Models

The following classes of model were considered on the training/validation set:

- Linear Penalized regression
- Penalized regression with polynomial features
- Gradient-boosted trees (GBM)
- Densely connected, feed-forward neural network
- KNN regression

Additional details on the training and hyper-parameter selection of each class are presented along with the their respective results. The learning task was framed as a regression problem, with the prediction target being standardised six-month forward return.

Model Evaluation and Comparison

Hyperparameter tuning for each model class was based on cross-validated information coefficient (IC). The information coefficient of a factor (or forecast in this case) is given by the average cross-sectional correlation of the factor with the target (forward return) in each period. Compared to RMSE, information coefficient more closely reflects what we care about in practice when constructing factor portfolios, namely that predictions in each period correspond to holding period returns.

As adjacent observations are highly overlapping (monthly sampling with 6 month forecast horizon), cross validation was performed using a variation of *hv*-block sampling (Racine, 2000). In this procedure, each validation fold is a contiguous block, and temporally overlapping observations between the training and validation set are removed. Further details regarding this choice are provided in Appendix 6.7.

Having set hyperparameters for each model class (e.g. regularization term in penalized regression, number of trees in GBM, and so on) using IC, the different model classes were compared. For this comparison, a broader range of statistics were used. These include: the distribution of predicted values, the balanced accuracy of winner and loser deciles, the distribution of predicted deciles among realised deciles, the realised return distribution of predicted winners and losers, and in-sample backtest results.

6.5 Model Training and Evaluation

6.5.1 Penalised Linear Regression

Three different penalised regression models were tested:

- Lasso Regression
- Ridge Regression
- ElasticNet Regression ($\alpha = 0.5$)

For each of these models the regularisation parameter was varied using 100 different values, logarithmically spaced, between 10^{-3} (less regularisation) and $10^{1.5}$ (more regularization).

Figure 6.4 shows the cross-validation estimate of IC for each model, as a function of regularization parameter. It can be seen that the IC of all tested models is negative, which suggests that we should take the *worst* model and reverse the scores. More seriously though, it suggests that the linear relationships that hold, on average, in one part of the data, do not generalise.

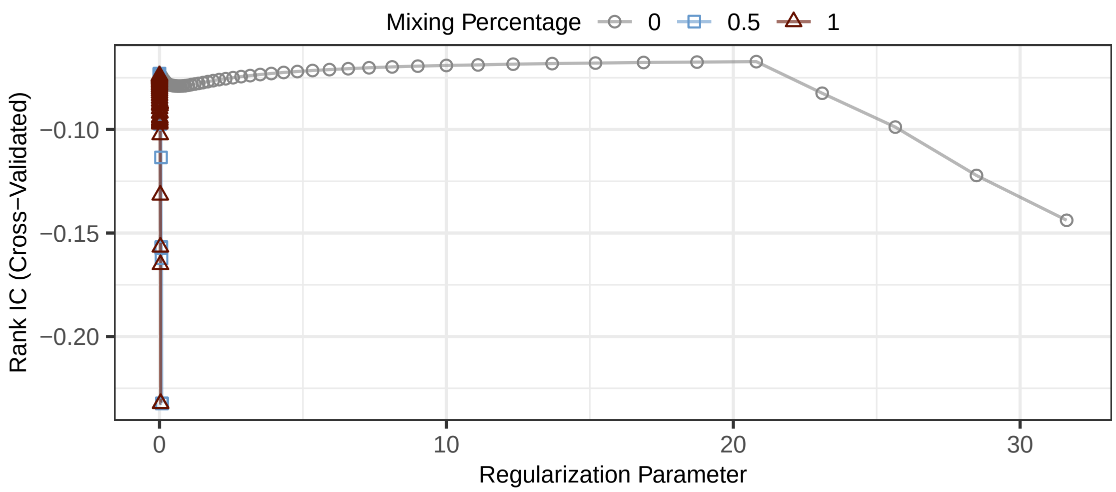

Figure 6.4: Cross-validated rank IC estimates for penalised linear regression models.

This is not only the case for performance in terms of IC. Of all linear models tested, the one with the lowest RMSE was the mean model; all coefficients are zero except for the intercept term representing the unconditional mean of the target variable.

6.5.2 Penalized Polynomial Regression

Here the three penalised regression models listed above were tested, but with the addition of second-order polynomial transformations of all original variables, including interaction terms. The regularisation parameter was varied using 100 different values, logarithmically spaced, between 10^-3 and $10^1.8$.

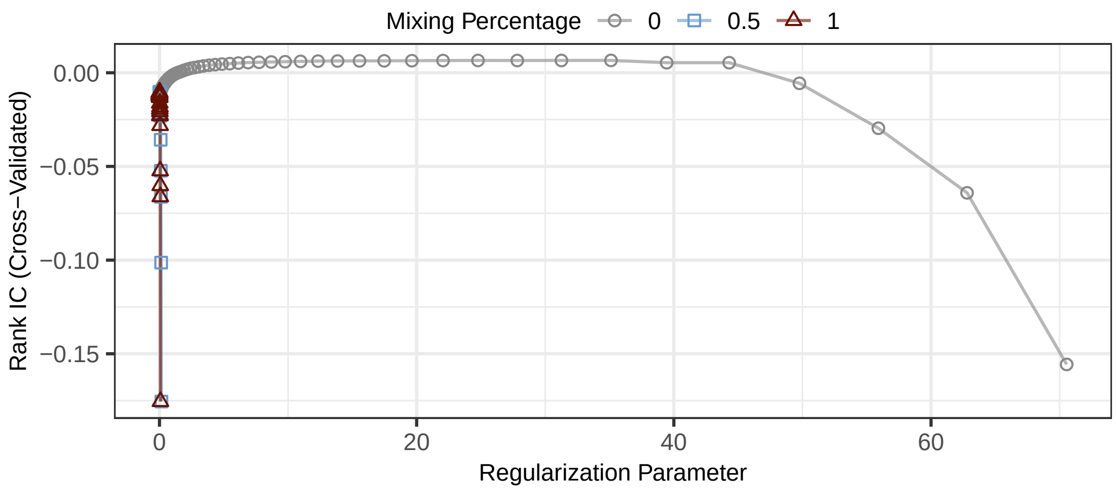

Figure 6.5: Cross-validated rank IC estimates for penalised polynomial regression models.

Figure 6.5 shows the cross-validation curves for the polynomial regression models. Again, most hyperparameter choices result in a negative IC, with the best performing model achieving only a marginally positive IC.

6.5.3 KNN Regression

The tuning parameter for the KNN algorithm, which directly controls the bias and variance of the resulting model, is the number of the neighbours (k) used to make the prediction for each observation. As the number of neighbours increases, so does the model's bias—approaching the mean model as $k \to n$.

The KNN cross-validation curve is shown in Figure 6.6. The IC estimates for the KNN models are only marginally positive, with highest IC being the model with $k = 25$.

6.5.4 GBM

The primary hyperparameters for GBMs are the number of boosting iterations (number of trees), maximum interaction depth of each tree, and shrinkage (regularization).

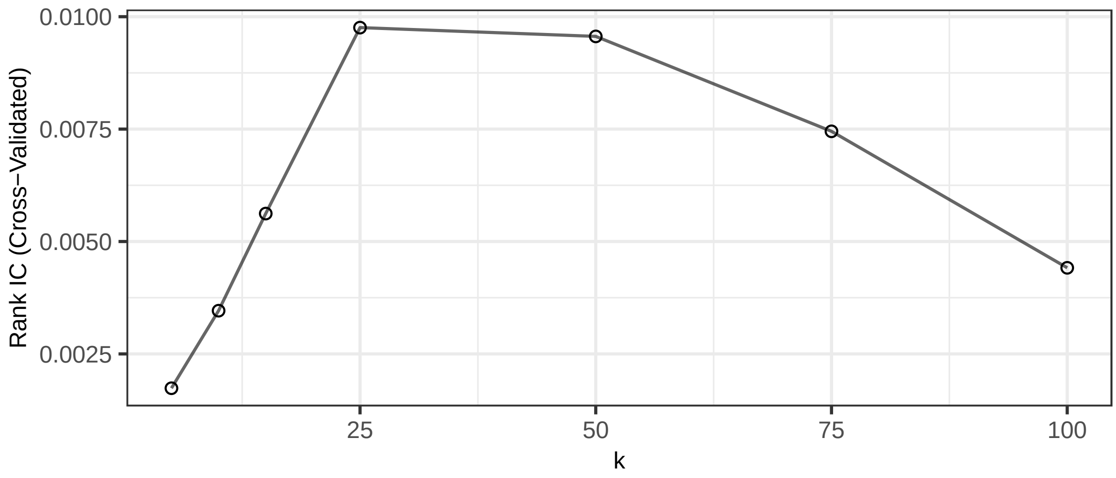

Figure 6.6: Cross-validated rank IC estimates for KNN regression models.

The value chosen for interaction depth should reflect the level of dominant interactions in the data. While this is generally not known, in most situations this tends to be low, and an interaction depth between three and seven works well, with little improvement to be gained beyond this range (Friedman, 2002).

As the shrinkage parameter (ν) controls the learning rate of the boosting procedure, there is a trade-off between ν and the number of trees (Friedman, 2001); smaller values of ν lead to a larger number of boosting operations for the same trading risk. Empirically, the best strategy has been to set ν to be small ($\nu < 0.1$) and choose the number of trees by early stopping (Friedman, 2001; Friedman, Hastie, and Tibshirani, 2009).

We apply a grid-search with $\nu = \{0.01, 0.05, 0.1\}$, interaction depth $= \{1, 2, 4, 6\}$, and no. trees $= \{100, 500, 1500, 2500, 3500, 5000\}$. The subsampling rate for stochastic gradient boosting (Friedman, 2002) was set at 0.5, which is a typical recommended value (Friedman, Hastie, and Tibshirani, 2009). The minimum number of observations per node was taken to be 10—the default for the R implementation (Greenwell et al., 2019).

The model giving the highest IC, still only marginally positive, had 5000 trees, an interaction depth of four, and shrinkage of 0.05. The results of the cross-validation grid-search procedure are illustrated in Figure 6.7.

Ranking Model The procedure described above was also used to select a GBM model using the rank of feature and target ranks, rather than their normalised values. Sacrificing the dispersion properties of the data (since the features and target were already standardised each period) can further assist to prevent over-fitting and focus on quantile-based relationships.

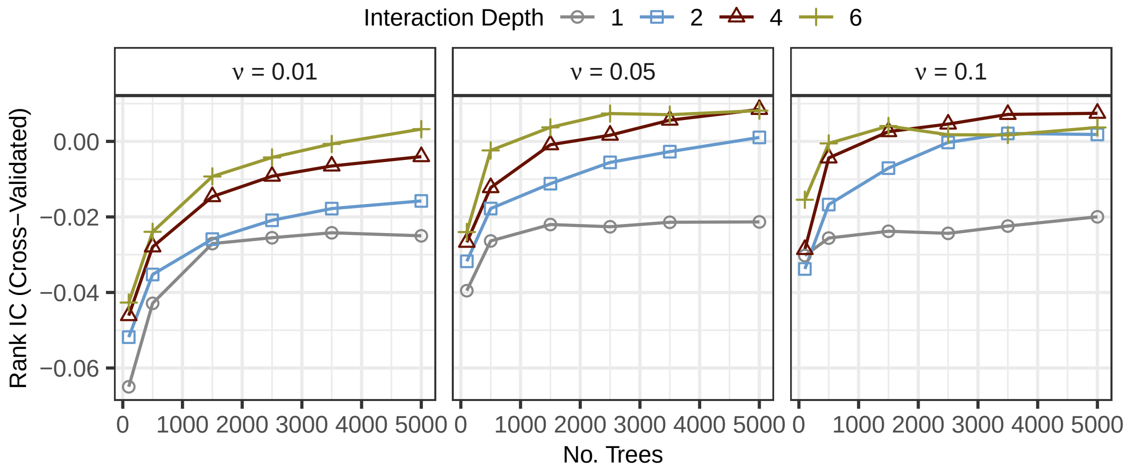

Figure 6.7: Cross-validated rank IC estimates for GBM regression models.

The best performing GBM model with ranked data used 5000 trees, an interaction depth of four, and a shrinkage of 0.10. The results of the cross-validation grid-search procedure for the rank-based GBM model are shown in Figure 6.8.

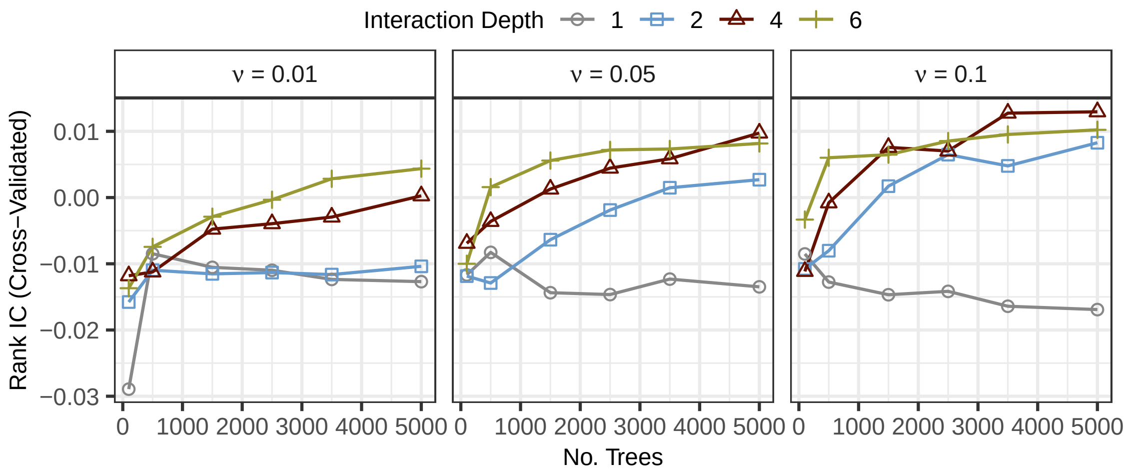

Figure 6.8: Cross-validated rank IC estimates for GBM regression models with ranked inputs.

6.5.5 ANN

The ANN architectures tested were single layer feed-forward networks of five, eight, 16, and 32 hidden units, with a relu activation function used for the hidden layer. All models were trained with mini batch gradient descent with an adaptive learning rate using the adam algorithm (Kingma and Ba, 2014). Models were tested with and with out dropout applied to the hidden layer (50% dropout rate) and early stopping (Prechelt, 1998) was used to set the number of epochs used for the final model. Each model was trained with and without PCA transformation of the features (within each fold). Mini batch sizes of 32, 64, 128, and 256 were also tested.

The highest rank IC was achieved by the five unit architecture with dropout, however, due to the very low variance of the model, the predictions could not reliably be split into deciles (due to repeated predictions), rendering it unsuitable for portfolio application. The next best model was the eight unit architecture, with dropout, trained using a mini batch size of 128 and using PCA transformed inputs.

As with the other classes of model, the best performing ANN only achieved a very modest cross-validated IC—0.021 in this case. Figure 6.9 shows the cross-validated rank IC, averaged over the 128 and 256 mini batch sizes, for three training epochs, with and without dropout, and with PCA transformed inputs.

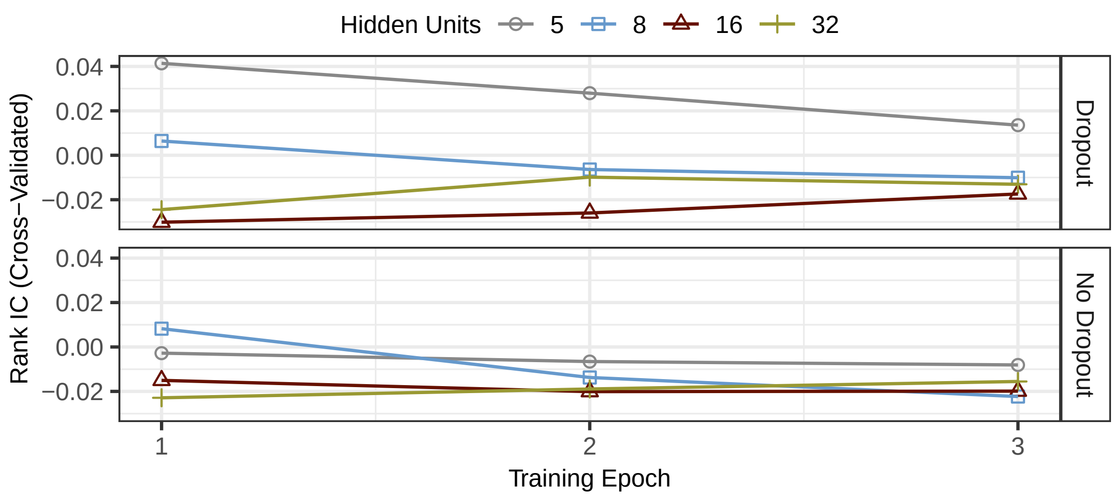

Figure 6.9: Cross-validated rank IC estimates for ANN models using PCA transformed features, averaged over mini batch sizes of 128 and 256.

6.5.6 Variable Importance

A useful feature of GBMs is they allow straight-forward estimation of the relative influence of individual input features in the variation of the prediction over the joint input variable distribution, i.e. a measure of an input's relative importance.

As described in Friedman (2001), relative influence can be estimated as the average improvement in the split criterion (MSE in this case), across all splits in all trees, made by each variable. The variables with the largest improvement in the split criterion are considered the most important.

Figure 6.10 displays the relative influence of feature variables in the GBM and rank GBM models. Relative importance for each model has been scaled such that the total relative influence adds to 100.

Market capitalisation is the most influential predictor of forward return in both models. This is likely to be a result of both its effect as a risk factor in itself (e.g. SMB) and its moderation of other variables, such as news coverage, which are largely explained firm size and analyst coverage (which is also in the top four for both models) (see for example, Hillert, Jacobs, and Mueller, 2014).

In the standard GBM model the next two of the top four most influential predictors are the 12- and six-month momentum. Whereas 12-month news coverage and sentiment occupy these positions under the rank model. This may suggest that useful information is lost when considering only the rank of momentum returns in lieu of the distribution, or that news-based variables are relatively more meaningful when "de-noised" through ranking. Interestingly, under both models the six-month news variables are less influential than their 12-month counterparts.

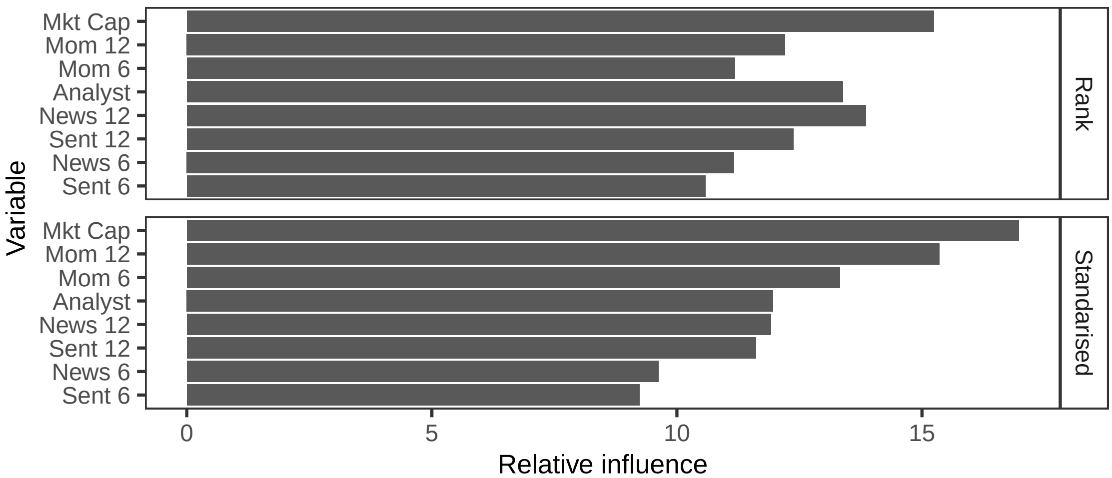

Figure 6.10: Relative influence of input variables in GBM models

6.5.7 Model Comparison

Firstly, it is worth noting the distribution of performance across the models for each of the cross-validation periods, illustrated in Figure 6.11. For instance, it is apparent that the performance of all models varies considerably across the folds—median (mean) rank IC is negative for three (four) out of the ten folds. This points to significant time-variation in the relationships between the predictors and forward return throughout the sample period. Additionally, it can be seen that the eighth fold corresponds to particularly poor performance for all models. This is the fold for which the validation period spanned 2009-08-31 to 2010-05-28, capturing the aggressive rebound immediately following the GFC.

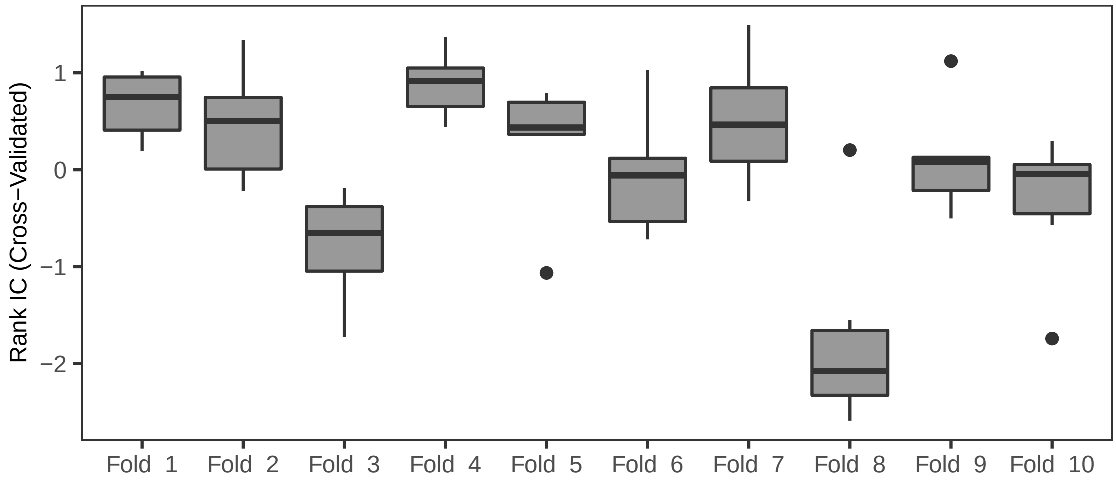

Figure 6.11: Cross-validated rank IC estimates for all models across training-validation folds.

Having chosen the hyperparameters for each class of model, a final model must be selected for the out-of-sample analysis. As we have already tested a large number of models against the validation set(s) and optimised for rank IC, we use another metric to choose between the models.

While the IC is calculated over the entire cross-section of assets and predictions each period, what we rely on most in our portfolio implementation is that the highest (lowest) quantile predictions correspond to the highest (lowest) quantile of forward returns; we are insensitive to the prediction-observation correspondence in the intermediate quantiles.

To reflect this, in each period of the validation set, we subtract the average forward return of the decile of assets with the lowest predicted return from the average forward return of the decile of assets with the highest predicted returns, and average the result over all periods. Like all other cross-validated metrics, this gives a score for each validation fold which are then averaged to give the final score for each model. We refer to this metric as the "decile difference" and it is analogous to an imprecise long-short backtest over the validation set.

Variation in the decile difference between models is largely explained by rank IC. We introduce decile difference mainly to guard against another layer of over-fitting while being slightly more precise about the desired predictive characteristics of our final model.

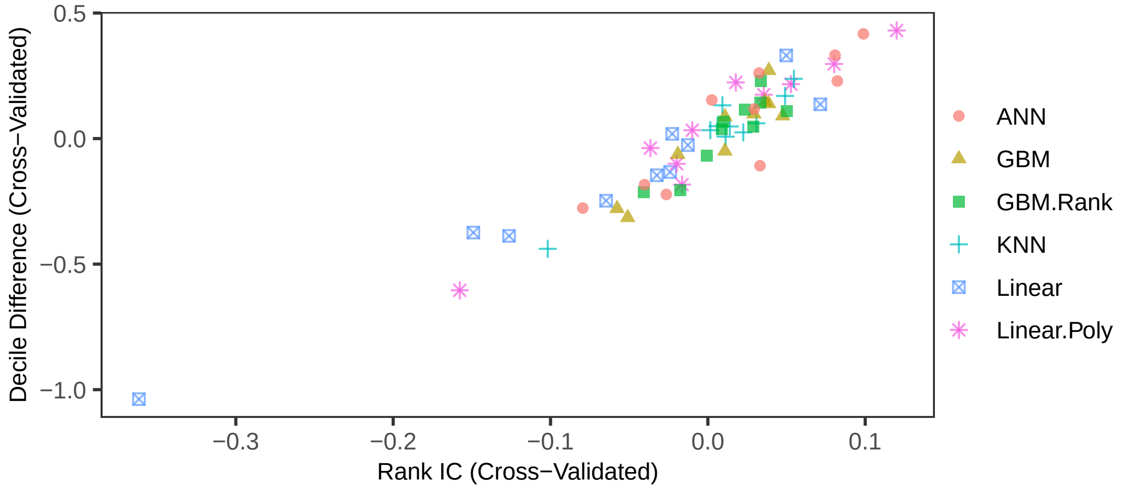

Figure 6.12: Cross-validated decile difference versus rank IC estimates for all models across training-validation folds.

The cross-validated performance of each model in terms of RMSE, MAE, Rank IC and Decile Difference, are detailed in Table 6.3. The RMSE and MAE values for the GBM Rank model have been excluded, as the target was of different scale (being rank) to the other models. ANN was the best performing model in terms of the Decile Difference and Rank IC, and so was chosen for the out-of-sample testing.

Table 6.3: Cross-Validated Model Performance

Model	Rank IC	Decile Difference	RMSE	MAE
ANN	0.021	0.072	0.962	0.746
GBM	0.008	0.013	1.005	0.755
GBM.Rank	0.013	0.026	-	-
KNN	0.010	0.033	0.991	0.751
Linear	-0.067	-0.187	0.955	0.723
Linear.Poly	0.007	0.045	0.955	0.723

This table presents cross-validated performance metrics for each of the models trained on the training/validation set. *Rank IC* is the average cross-sectional rank correlation between predictions and targets, *Decile Difference* is the average difference between the mean target value corresponding to the highest prediction decile, and those in the lowest prediction decile.

6.6 Out-of-Sample Results

Table 6.4 documents the risk-adjusted performance of decile portfolios sorted by ANN-predicted forward return, for the training, test and full sample periods. Risk-adjusted returns are comparable across both periods, suggesting that we did not over-fit the model to the training data. In particular, the risk-adjusted performance of the high-minus-low portfolio is essentially the same in both periods, although neither is statistically significant.

Figures 6.14 and 6.14 illustrate the performance of the high-minus-low and high ANN portfolios in terms of cumulative log excess return. The return of high-minus-low portfolio is measured against the risk-free rate, while the return of the high portfolio is measured against the market. Cumulative summation plots, such as those of excess return or information ratio, provide a means to easily visually detect changes in performance, via changes in the slope of the cumulative sum. Consistent with the tabulated results, there does not appear to be any significant changes in performance between the training and test periods for either of the portfolios.

Table 6.5 includes the excess return and risk-adjusted performance for each of the ANN decile portfolios over the test period. The top-decile portfolio (long-only) generates statistically significant risk-adjusted return for the Fama and French (1993) three-factor and the Fama (2015) five-factor risk models. The high-minus-low portfolio generates statistically significant risk-adjusted returns under the five-factor model.

Table 6.4: Returns to ANN portfolios: Train, Test, and Full samples

	Train		Test		Full	
Decile	α	(t-stat)	α	(t-stat)	α	(t-stat)
High	0.0033*	(2.0345)	0.0020	(1.3994)	0.0019	(1.8055)
2	0.0016	(1.1717)	0.0007	(0.6381)	0.0003	(0.2838)
3	0.0010	(0.7300)	0.0005	(0.4854)	0.0000	(0.0239)
4	0.0013	(0.6269)	0.0003	(0.3371)	-0.0002	(-0.1687)
5	0.0013	(1.4871)	-0.0001	(-0.0624)	0.0004	(0.6078)
6	0.0011	(1.6159)	-0.0000	(-0.0134)	0.0002	(0.3907)
7	0.0012	(1.4270)	-0.0004	(-0.5012)	0.0001	(0.1837)
8	0.0008	(1.0165)	-0.0006	(-0.6384)	-0.0002	(-0.3347)
9	0.0003	(0.2979)	-0.0006	(-0.6681)	-0.0008	(-1.0822)
Low	-0.0015	(-1.1087)	-0.0011	(-0.7461)	-0.0024*	(-2.2112)
High-Low	0.0024	(1.1935)	0.0029	(1.2942)	0.0028*	(1.9771)
Mom High	0.0004	(0.1628)	0.0019	(1.2645)	0.0001	(0.0608)
Mom Low	-0.0012	(-0.2828)	-0.0043	(-1.3095)	-0.0043	(-1.7072)
Mom High-Low	-0.0022	(-0.2957)	0.0056	(1.4402)	0.0023	(0.5424)
Sent High	0.0032*	(2.0732)	0.0007	(0.4803)	0.0012	(1.0478)
Sent Low	0.0006	(0.3888)	-0.0052	(-1.8537)	-0.0022	(-1.4172)
Sent High-Low	0.0008	(0.3896)	0.0054	(1.6549)	0.0024	(1.2408)

This table presents risk-adjusted (CAPM) monthly returns for equal-weighted portfolios sorted by ANN model predictions. Portfolios were constructed using formation and holding periods of six months, with no skip period. *Train* is the period over which the model was trained (2004-01-30 to 2011-12-30), *Test* is the out-of-sample period (2012-07-31 to 2017-12-29), and *Full* is the full sample period (2004-01-30 to 2017-12-29. Returns to the High, Low and High-minus-Low Momentum ((Mom)) and Sentiment ((Sent)) portfolios are provided for comparison.

Table 6.5 also includes excess and risk adjusted returns for the the high, low, and high-minus-low portfolios for univariate momentum and sentiment sorts. Broadly, the ANN portfolios perform favourably against the momentum and sentiment portfolios, except under the three-factor model, for which the low, and high-minus-low sentiment portfolios are statistically significant at the 5% level while neither of the corresponding ANN or momentum portfolios are.

Tables 6.6, 6.7, and 6.8 document the factor loadings of the high, low, and high-minus-low ANN portfolios, respectively, over the training and test periods.

The high ANN prediction portfolio loads significantly on the market, HML, and SMB

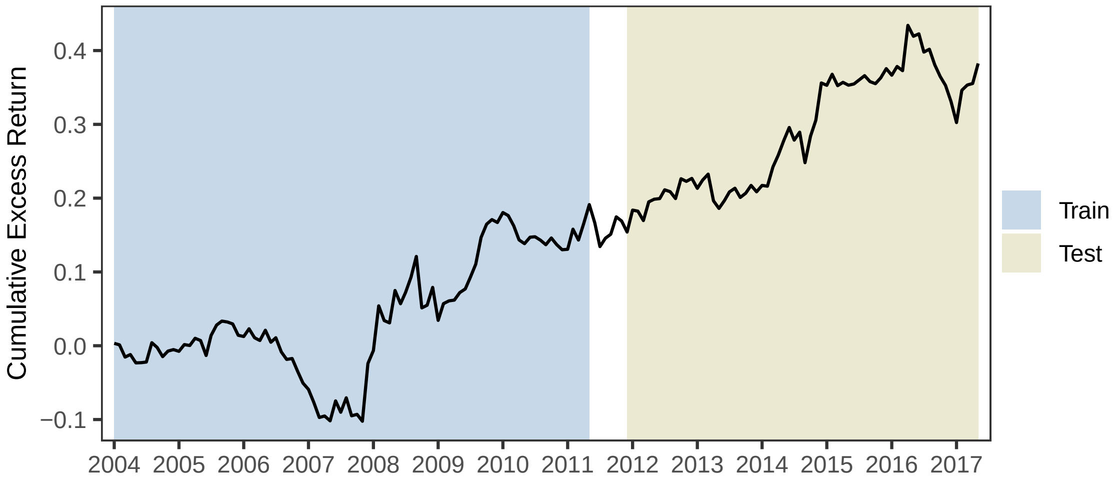

Figure 6.13: Cumulative logarithmic excess return of the ANN 'zero-cost' high-minus-low portfolio.

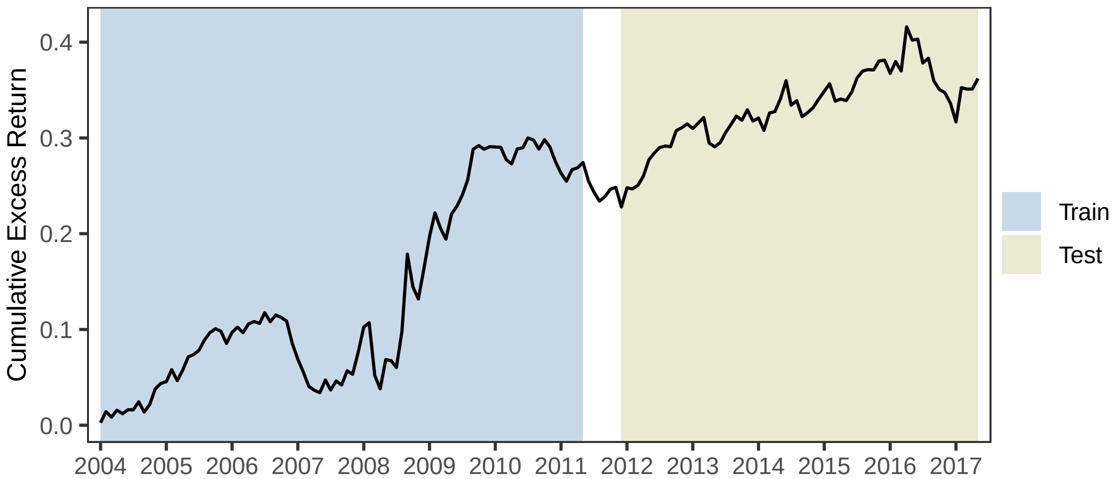

Figure 6.14: Cumulative logarithmic excess return of the long-only high ANN portfolio.

factors under the three-factor model in both training and test sets. Under the five-factor model, the high ANN prediction portfolio loads significantly on the same three factors during the training period, but loses exposure to the HML factor in the testing period, loading more significantly on the CMA factor instead.

The low ANN prediction portfolio has relatively less factor exposure than the high portfolio, its only significant loading being market beta. This is true across both sample periods. The high-minus-low ANN prediction portfolio has a similar factor profile to the long-only high portfolio over the training sample, but loses most of its factor exposure in the test sample. The only significant loading of the high-minus-low portfolio in the test sample, across any model, is the SMB factor (and alpha under the five-factor model).

Table 6.5: Returns to ANN portfolios: Test Sample

	TS Mean		CAPM		FF3		FF5	
Decile	α	(t-stat)	α	(t-stat)	α	(t-stat)	α	(t-stat)
High	0.0143***	(3.8055)	0.0020	(1.3994)	0.0029*	(2.2806)	0.0032**	(3.0453)
2	0.0134***	(3.6811)	0.0007	(0.6381)	0.0014	(1.4353)	0.0015	(1.6616)
3	0.0137***	(3.6512)	0.0005	(0.4854)	0.0010	(1.1081)	0.0012	(1.2845)
4	0.0131***	(3.6295)	0.0003	(0.3371)	0.0007	(0.7669)	0.0008	(1.1089)
5	0.0123***	(3.5564)	-0.0001	(-0.0624)	0.0001	(0.1781)	0.0003	(0.3417)
6	0.0121***	(3.5478)	-0.0000	(-0.0134)	-0.0001	(-0.1228)	0.0001	(0.1344)
7	0.0120***	(3.4638)	-0.0004	(-0.5012)	-0.0006	(-0.7177)	-0.0004	(-0.5433)
8	0.0118**	(3.4219)	-0.0006	(-0.6384)	-0.0007	(-0.8806)	-0.0006	(-0.7725)
9	0.0115**	(3.3587)	-0.0006	(-0.6681)	-0.0008	(-0.8716)	-0.0008	(-0.8379)
Low	0.0107**	(3.0887)	-0.0011	(-0.7461)	-0.0014	(-0.9438)	-0.0012	(-0.8821)
High-Low	0.0035	(1.4755)	0.0029	(1.2942)	0.0040	(1.8694)	0.0042*	(2.0015)
Mom High	0.0140***	(3.6914)	0.0019	(1.2645)	0.0025	(1.7336)	0.0028*	(2.0955)
Mom Low	0.0106	(1.9540)	-0.0043	(-1.3095)	-0.0043	(-1.4230)	-0.0038	(-1.3473)
Mom High-Low	0.0034	(0.7206)	0.0056	(1.4402)	0.0062	(1.7204)	0.0060	(1.6772)
Sent High	0.0128***	(3.6114)	0.0007	(0.4803)	0.0010	(0.7979)	0.0010	(0.8505)
Sent Low	0.0089	(1.7436)	-0.0052	(-1.8537)	-0.0052*	(-2.0123)	-0.0046	(-1.8375)
Sent High-Low	0.0039	(0.9744)	0.0054	(1.6549)	0.0057*	(2.1897)	0.0051	(1.8198)

This table presents raw and risk-adjusted excess monthly returns for equal-weighted portfolios sorted by ANN predictions, over the out-of-sample (test) period (2012-07-31 to 2017-12-29). α is the intercept term in each of the factor regressions. *TS Mean* is the excess time series mean return (return above risk-free rate), *CAPM* is the market model, *FF3* is the Fama and French (1993) model, and *FF5* is the Fama (2015) model. Portfolios are constructed using formation and holding periods of six months, with no skip period. Returns to the High, Low and High-minus-Low Momentum ((Mom)) and Sentiment ((Sent)) portfolios are provided for comparison.

Table 6.6: Sub-period factor regressions the high ANN prediction portfolio

	α	MKT	HML	SMB	RMW	CMA
Panel 1: Training Sample						
TS	0.0064					
	(1.0517)					
CAPM	0.0033*	1.1685***				
	(2.0345)	(27.4171)				
FF3	0.0030*	1.0305***	0.2394**	0.3187***		
	(2.4313)	(28.9691)	(2.8764)	(4.0915)		
FF5	0.0027	1.0433***	0.2330**	0.3209***	0.0615	0.0339
	(1.7950)	(22.1930)	(2.8792)	(4.0256)	(0.6169)	(0.3609)
Panel 2: Test Sample						
TS	0.0143***					
	(3.8055)					
CAPM	0.0020	0.9827***				
	(1.3994)	(13.1768)				
FF3	0.0029*	0.9131***	0.1090*	0.2878***		
	(2.2806)	(18.5661)	(2.0257)	(3.8928)		
FF5	0.0032**	0.9241***	-0.0222	0.3087***	0.0606	0.3200*
	(3.0453)	(20.0184)	(-0.3373)	(4.1794)	(0.7446)	(2.5172)

This table presents raw and risk-adjusted returns (t-statistics in brackets) to the decile of stocks with the highest ANN prediction. α is the intercept term in each of the factor regressions. *TS Mean* is the excess time series mean return (return above risk-free rate), *CAPM* is the market model, *FF3* is the Fama and French (1993) model, and *FF5* is the Fama (2015) model. *MKT* represents the market return minus the risk-free rate. *SMB* represents the small-minus-big factor. *SMB* represents the high-minus-low factor. *RMW* represents the robust-minus-weak factor. *CMA* represents the conservative-minus-agressive factor. *Training Sample* is the period from 2004-01-30 to 2011-12-30, and *Test Sample* is the period from 2012-07-31 to 2017-12-29.

Table 6.7: Sub-period factor regressions the low ANN prediction portfolio

	α	MKT	HML	SMB	RMW	CMA
Panel 1: Training Sample						
TS	0.0017					
	(0.2644)					
CAPM	-0.0015	1.2460***				
	(-1.1087)	(35.0838)				
FF3	-0.0016	1.2432***	-0.0606	0.0579		
	(-1.1382)	(20.5672)	(-0.4168)	(0.7537)		
FF5	-0.0015	1.2338***	-0.0358	0.0626	-0.0198	-0.1215
	(-0.8382)	(16.2826)	(-0.2350)	(0.7731)	(-0.1158)	(-0.9958)
Panel 2: Test Sample						
TS	0.0107**					
	(3.0887)					
CAPM	-0.0011	0.9527***				
	(-0.7461)	(37.1112)				
FF3	-0.0014	0.9611***	0.1314	-0.0516		
	(-0.9438)	(39.4574)	(1.3966)	(-0.8651)		
FF5	-0.0012	0.9663***	0.0650	-0.0631	-0.0395	0.1690
	(-0.8821)	(41.7893)	(0.7138)	(-0.9063)	(-0.5399)	(1.9042)

This table presents raw and risk-adjusted returns (t-statistics in brackets) to the decile of stocks with the lowest ANN prediction. α is the intercept term in each of the factor regressions. *TS Mean* is the excess time series mean return (return above risk-free rate), *CAPM* is the market model, *FF3* is the Fama and French (1993) model, and *FF5* is the Fama (2015) model. *MKT* represents the market return minus the risk-free rate. *SMB* represents the small-minus-big factor. *SMB* represents the high-minus-low factor. *RMW* represents the robust-minus-weak factor. *CMA* represents the conservative-minus-agressive factor. *Training Sample* is the period from 2004-01-30 to 2011-12-30, and *Test Sample* is the period from 2012-07-31 to 2017-12-29.

Table 6.8: Sub-period factor regressions the high-minus-low ANN prediction portfolio

	α	MKT	HML	SMB	RMW	CMA
Panel 1: Training Sample						
TS	0.0023					
	(1.0925)					
CAPM	0.0024	-0.0420				
	(1.1935)	(-1.2329)				
FF3	0.0022	-0.1740***	0.2713*	0.2713*		
	(1.3513)	(-4.8070)	(2.4090)	(2.4860)		
FF5	0.0012	-0.1354**	0.2405*	0.2744*	0.1705	0.1568
	(0.5570)	(-2.6014)	(2.0497)	(2.5344)	(0.8747)	(1.0194)
Panel 2: Test Sample						
TS	0.0035					
	(1.4755)					
CAPM	0.0029	0.0478				
	(1.2942)	(0.6066)				
FF3	0.0040	-0.0312	-0.0204	0.3435**		
	(1.8694)	(-0.5827)	(-0.1628)	(3.2622)		
FF5	0.0042*	-0.0243	-0.1006	0.3714**	0.0854	0.1908
	(2.0015)	(-0.4707)	(-0.8440)	(3.0292)	(0.7237)	(1.1243)

This table presents raw and risk-adjusted returns (t-statistics in brackets) to the decile of stocks with the highest ANN prediction. α is the intercept term in each of the factor regressions. *TS Mean* is the excess time series mean return (return above risk-free rate), *CAPM* is the market model, *FF3* is the Fama and French (1993) model, and *FF5* is the Fama (2015) model. *MKT* represents the market return minus the risk-free rate. *SMB* represents the small-minus-big factor. *SMB* represents the high-minus-low factor. *RMW* represents the robust-minus-weak factor. *CMA* represents the conservative-minus-agressive factor. *Training Sample* is the period from 2004-01-30 to 2011-12-30, and *Test Sample* is the period from 2012-07-31 to 2017-12-29.

6.7 Discussion

In this analysis, we examined model-driven portfolio formation procedures based on news sentiment, news coverage, and price momentum. We tested the hypothesis that news-informed portfolios using flexible statistical models outperform the simple news and momentum portfolios explored in the previous chapter. Our findings suggest that news sentiment and coverage do not offer a simple means of improving price momentum stategies, even when used in conjunction with modeling procedures that can flexibly accommodate many forms of nonlinear dependencies.

We began with a traditional in-sample Fama and MacBeth (1973) predictive regression analysis of six-month forward return that included news sentiment, news coverage, price momentum, and their first order interactions as explanatory variables. The regression results revealed statistically significant predictive power in sentiment-momentum and sentiment-coverage interactions.

We next employed a rolling linear regression framework, as set out by Haugen and Baker (1996), to test whether these predictive effects were exploitable out-of-sample. The performance of decile portfolios formed on the basis of the linear expected returns model demonstrated that this was not the case—if the expected return model had any predictive capacity at all, it appeared to be in the wrong direction.

Using the same rolling regression procedure, we then compared the news-informed linear regression model to three alternatives; a traditional momentum approach, a linear regression model using only momentum, and a news-informed GAM model. Using cumulative IC as a guide to model efficacy, we found that the inclusion of news variables improved performance beyond the momentum-only regression model and further improvement resulted from nonlinearities captured by the GAM. Yet, all statistical models faired worse than the traditional momentum strategy, which was the only strategy to end the sample with a positive cumulative IC.

We then moved to a testing procedure amenable to a larger class of statistical models; the Haugen and Baker (1996) style analysis, while natural from a linear regression perspective, represents an unusual and somewhat limiting use of data when a wider array of models are being considered. We adopted the common approach to statistical prediction in which the dataset is initially split into training/validation, and testing sets, with model training and tuning being performed on the former and testing on the latter. The main differences from the previous approach being that the training/validation window is static (and dictated by the size of the testing sample), and training is not constrained to monthly strata.

The best performing model from the training/validation set, a single layer neural network, was then tested out-of-sample using the overlapping portfolio procedure. The model performed well out-of-sample, with the long-only portfolio delivering statistically significant alpha under Fama-French three- and five-factor models. However, excess return was not statistically significant for the single-factor model (CAPM), and performance was only very modestly better than that of the naive momentum strategy. The long-short model portfolio only delivered statistically significant alpha under the five-factor model and was comparable to (but less than) the naive long-short sentiment strategy under most models.

The long-only model portfolio loaded significantly on HML, and SMB style factors during the training sample, and on SMB, and CMA style factors during the test sample. The long-short model portfolio did not load significantly on any style factors in either sample.

In summary, while there appears to be some durable performance benefit incorporating news-based variables into model-driven momentum portfolios, the gains are unlikely to outweigh the additional costs and complexity for many practitioners.

Appendix A: Choice of Sampling Technique for Cross-Validation

6.7.1 Introduction

For any model intended to be applied in practice, generalisation error, that is its prediction efficacy on new data, is paramount. A crucial step in model selection and evaluation is therefore estimation of this quantity. As terminology in this area can be confusing, it is useful to define some quantities of interest up front.

Training error is the average loss over the training sample

$$\overline{\text{err}} = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}(X))$$

Generalization error or *test error* is the prediction error of a model $\hat{f}(X)$, produced on training set $\mathcal{T}$, over an independent test sample

$$\mathrm{Err}_{\mathcal{T}} = E\left[L(Y, \hat{f}(X))|\mathcal{T}\right]$$

where $L(\cdot)$ is the loss function and x, Y are feature and target variables, respectively. Although estimation of $\mathrm{Err}_{\mathcal{T}}$ is our goal, most methods estimate the *expected prediction error*.

$$\mathrm{Err} = E\left[L(Y, \hat{f}(X))\right] = E[\mathrm{Err}_{\mathcal{T}}],$$

i.e. the expected generalization error over all training sets. As discussed by Friedman, Hastie, and Tibshirani (2009), estimating the conditional quantity $\mathrm{Err}_{\mathcal{T}}$ effectively does not seem possible given only the information in the training set.

Ideally, we would split our data into three parts: a training set, a validation set, and a test set. In this situation, the training set is used to fit the models, the validation set is used to estimate the prediction error for model selection, and the test set is used to assess the generalization error of the final model.

In the current case, we have insufficient data to usefully split it into three parts and so we must approximate the independent validation step by efficient sample re-use. For this we will use cross-validation, which is likely the most widely used approach for estimating the prediction error Err.

K-fold cross validation begins by splitting the data into K equal-sized parts. For each of these parts, the model is trained on the other K-1 parts of the data, and the prediction error is then calculated on the part not used for training. The average of the K estimates of the prediction error is then the cross-validation estimate of prediction error and model tuning parameters are selected so as to minimise this value.

$K = 5$ or $K = 10$ are generally recommended as a good compromise between the bias and variance of the cross-validation estimate of prediction error (Breiman and Spector, 1992; Kohavi, 1995). We proceed with choice of $K = 5$ due to the lower computational burden.

A complication in the current case is that our data consists of highly overlapping observations. Due to our formation and prediction horizons spanning 12 months each, temporally adjacent observations for a given stock will contain information that differs by only 1 out of the 365 days that comprise the observation window. If observations from the training and validation folds are overlapping, overfit predictions will appear unrealistically good. This is why cross-validation is explicitly based on the premise of using independent data for the training and test samples.

As a result, we have to be deliberate in how we select our training and validation samples, and we will see that the usual approach of unconstrained random sampling is unsuitable.

To visualise different sampling procedures, we will use a dummy grid of 90 periods and 30 tickers, which reflects approximately the same ticker/period ratio as our actual training set.

6.7.2 Random Sampling

As alluded, the typical approach to K-fold cross-validation is to select folds by randomly shuffling the data. Figure 6.15 shows the locations of validation and training observations for one (of 5) hypothetical cross-validation folds using unconstrained random sampling.

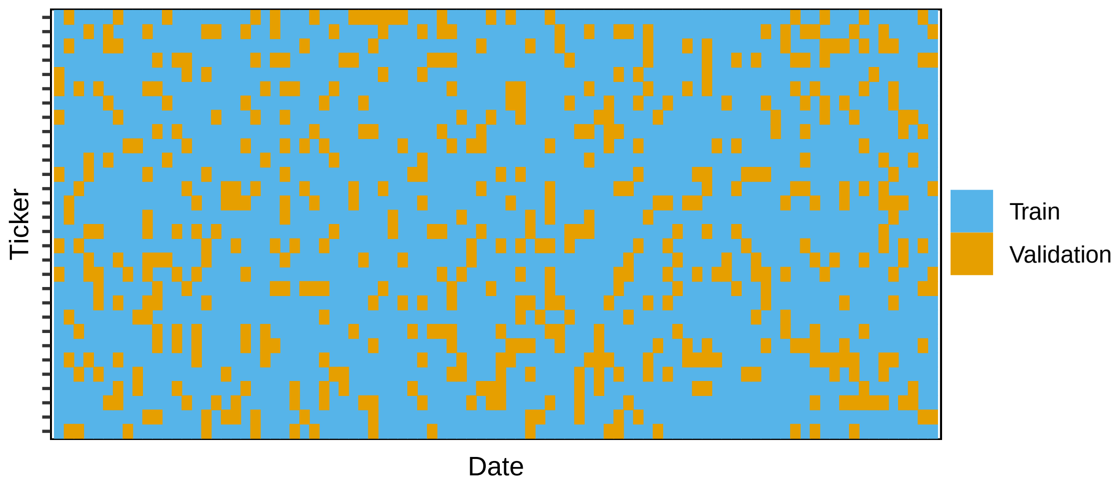

Figure 6.15: K-fold cross-validation with unconstrained random sampling as applied to panel data.

If temporally adjacent (x-axis) observations for a given ticker (y-axis) contain a significant amount of overlap, it is easy to see that there will be very little new or "unseen" data in the validation set. This means that our estimate of prediction error will behave like training error, which consistently decreases with model complexity.

The prediction error estimated using this sampling approach will therefore lead us to select models of increasing complexity (variance), as overfitting is not being penalized.

This can be verified by applying the random sampling cross-validation procedure to the training of a kNN regression model using a subset of our actual training data. The effective degrees of freedom of a kNN model is approximately N/k, where N is the number of training observations and k is the number of 'neighbours' used, giving us a direct measure of model variance to compare.

Figure 6.16 shows the 5-fold cross-validation error and test set error for a kNN regression model with $k = 1, 10, 20, 50, 100$. The training and cross validation sample includes observations, sampled weekly, up to and including 2016-12-31. The test set consists of observations, sampled weekly, from 2010-01-01 to 2011-12-31. Both of these are subsets of our actual training set.

The cross-validation estimate decreases montonically with increasing model complexity—as expected, while the test error *increases* monotonically with model complexity. Note that both the error (RMSE) and model degrees of freedom have been normalised to the 0-1 interval for clarity.

In the following sections, we will go on to discuss some intuitive alternatives to unconstrained sampling for dealing with dependent observations.

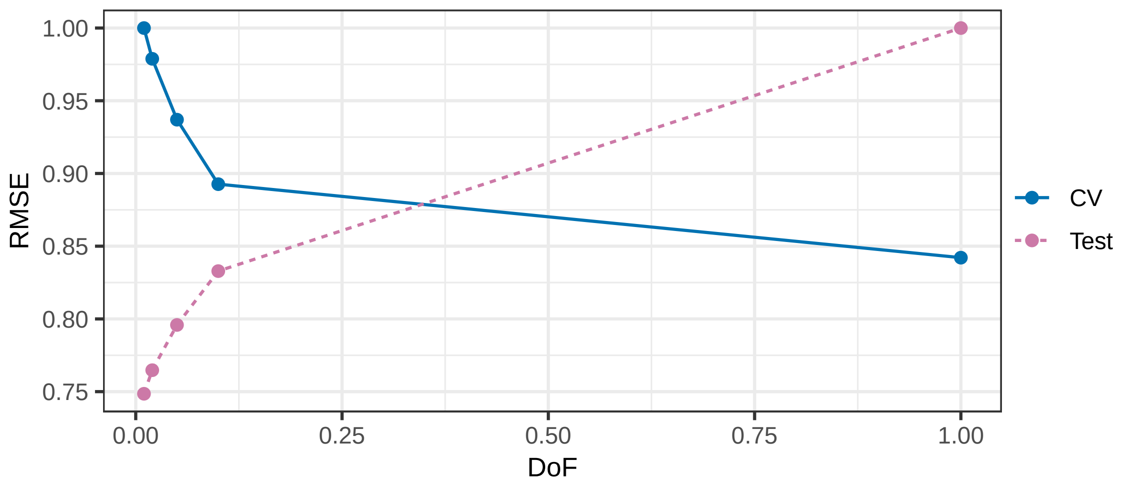

Figure 6.16: Expected prediction error and realised test error versus model complexity for kNN regression with unconstrained K-fold cross-validation.

6.7.3 Randomly Selected Unique Observation Sets

The dependency issue arises because our sampling frequency is higher than our prediction horizon, so an obvious solution may be to remove intermediate (overlapping) observations from the data set. However, this just equates to limiting sampling frequency to the horizon of the target.

With long outcome horizons, as is the current case, this leads to a very coarse data set, and is generally not recommended (De Prado, 2018). Furthermore, the starting point for such an approach, such as the last day of the month, is arbitrary. Unless we are prepared to throw-away the majority of our data, we must find alternative ways to sample from it.

One approach is to sample randomly, but restrict subsequent draws to observations that do not overlap with previous draws. This results in a set of non-overlapping (or partially-overlapping within a predefined range) observations that can then be broken into training and validation sets. To make use of the full data set (in expectation) and/or to obtain the desired number of folds, this procedure must be repeated.

Figure 6.17 shows one observation set generated in this way on the dummy data. In this case, observations for a given ticker were constrained to be at least five periods apart (top panel) or 11 periods apart (bottom panel). Note that the sample has not further been split into training and validation folds.

It is apparent that we lose a majority of our original data set in generating the non-overlapping set. This means that we require many more cross-validation folds to use the entire data set in expectation. In addition to being computationally expensive, performing cross-validation using only a small fraction of the data for each fold may give a poor estimate of the prediciton error, depending on the slope of the training curve at that sample size (Friedman, Hastie, and Tibshirani, 2009).

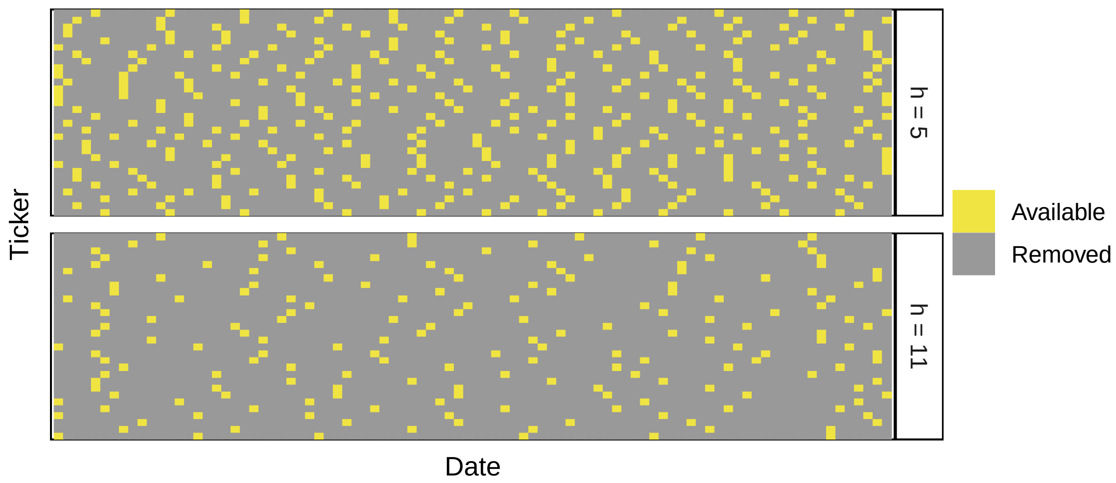

Figure 6.17: Random sampling of panel data with subsequent draws restricted by minimum temporal distance, h.

6.7.4 Random Validation, Restricted Train

As we are primarily concerned with overlapping observations *between* training and validation sets, we can relax the sampling procedure to allow overlapping observations *within* training and validation sets.

One way to create folds with these characteristics is to select the validation set using random sampling then remove the h overlapping observations preceding and following the observations in the test set. In the single series (i.e. not panel data) context, this is known as *modified* or *non-dependent* cross-validation (Chu and Marron, 1991; Bergmeir and Benítez, 2012). Figure 6.18 shows this method applied to the dummy sample, with $h = 1$ and $h = 11$.

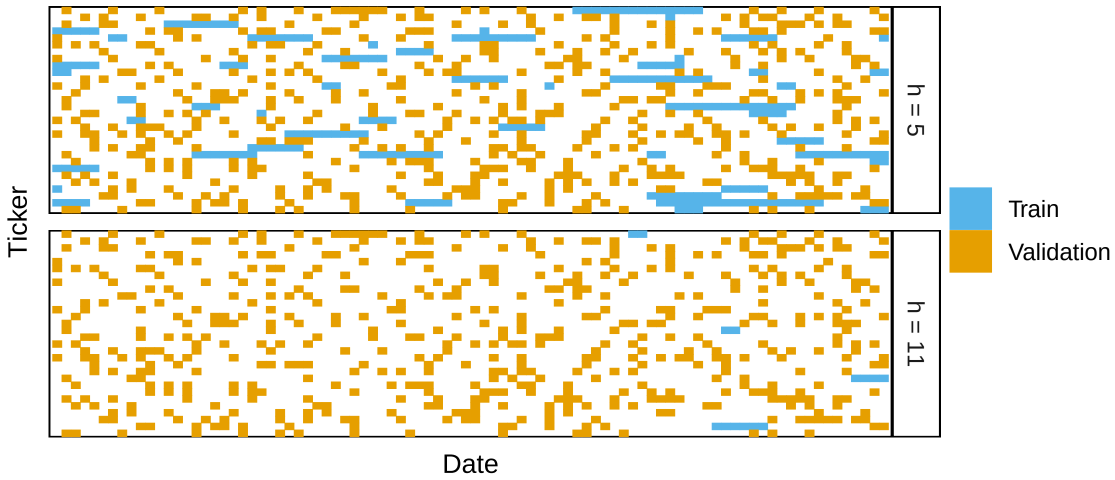

Figure 6.18: Random sampling of validation set, and overlapping observations, within h periods, removed from test set.

While this method offers an improvement for small degrees of separation (h), we can see that very few observations are left for the training set using the parameters specified. The visualisation illustrates that training data are often unable to find a place within large regions of the dataset sparsely populated by validation observations. By randomly selecting the entire validation set first, we are unable to retain (even approximately) the desired $(K - 1)/1$ training/validation ratio.

6.7.5 Iterative Restrictive Sampling

We can vary the previous approach by iteratively selecting training and validation and observations that conform to our separation requirements. That is, we alternatively select training and validation observations in a $(k - 1)/1$ ratio, with each draw restricted to be at least h periods away from observations already added to the other (i.e. training or validation) set. This way we can sample in a $(k - 1)/1$ ratio and maintain the desired train-validation split.

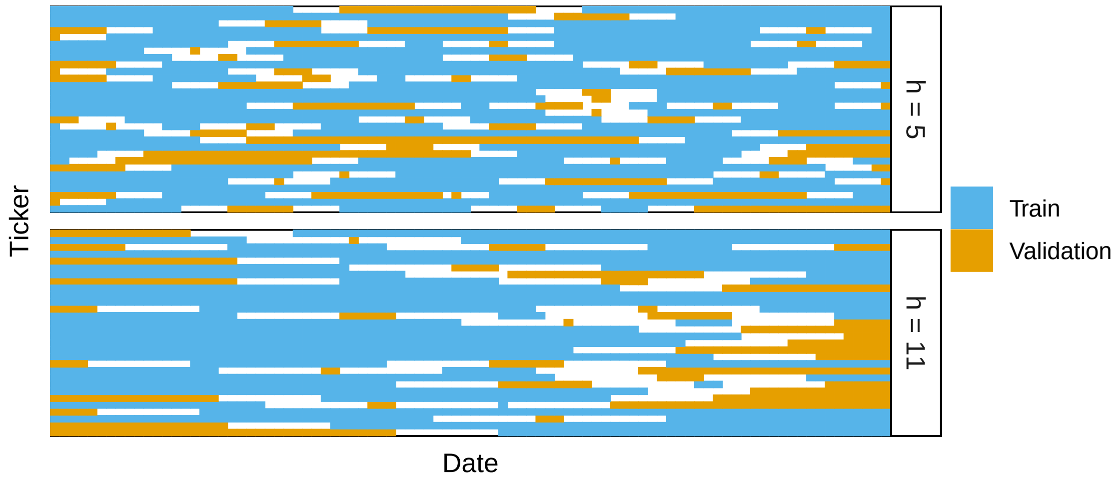

Figure 6.19: Iterative, random sampling of validation and test sets, with subsequent draws restricted to be at least h periods apart from the other (training or validation) set.

This method works reasonably well on the dummy data, as most observations are used in each split and the desired train/test ratio is retained. However, like all the methods discussed so far, as the horizon of overlap increases relative to sampling frequency, it becomes more computationally expensive and less data-efficient.

6.7.6 hv-Block Sampling

A simpler and more scalable way to approach this type of sampling (allowing overlap within but not between sets) is to select the validation set as a block, and remove training observations within h periods. Taking v as the length of the validation set, this is similar to the hv-block cross-validation procedure described by Racine (2000).

By selecting the validation set in temporal blocks, the overlapping data to be removed is reduced to a contiguous buffer either side of the validation set. In the case of our actual training set, this means that two years of data are removed for each fold, which is approximately 25% of the entire training set. The hv-block sampling procedure is illustrated in Figure 6.20 for five-fold cross-validation, with $h = 11$.

6.7.7 Group Sampling

The final option we will consider, which again allows overlapping observations within, but not between, test and validation folds, is to sample observations by group. This exploits the fact that our primary concern is the structural temporal dependency introduced

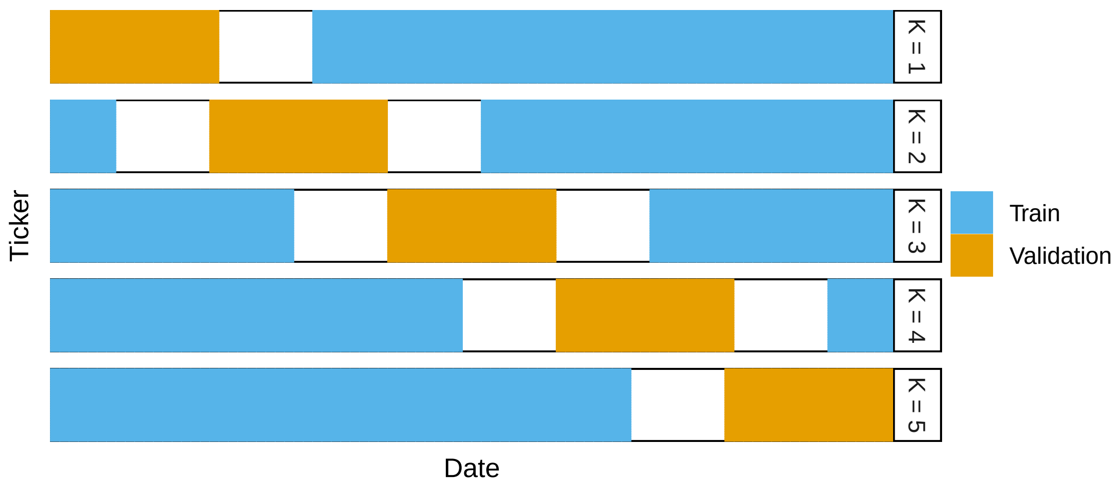

Figure 6.20: *hv*-block sampling ($h = 11$) for five-fold cross-validation, in which validation sets are selected in non-overlapping contiguous blocks and training sets are the remaining data after removal of observations within h periods of the validation set for that fold.

.

through analysis of a target horizon longer than our data sampling rate, rather than any incidental cross-sectional relationship between assets. Group sampling is the most data efficient that maintains (within-asset) temporal separation between training and validation sets, as it allows all the data to be used.

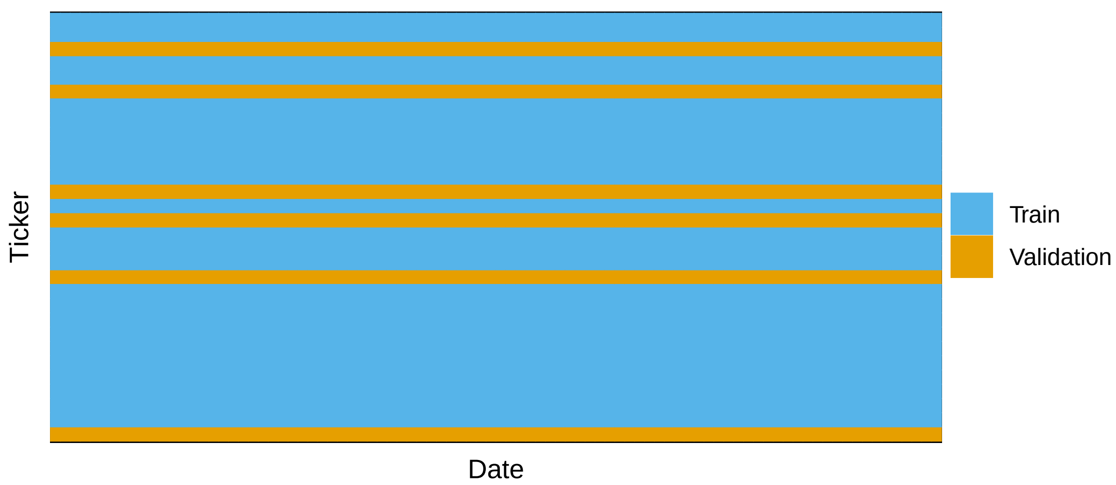

Figure 6.21: Group sampling for five-fold cross-validation, in which training and validation sets are partitioned by randomly selected groups (tickers).

.

If the relationship between features and predictors is not stable through time (non-stationary joint distribution), but is stable across assets, then this type of sampling may lead to optimistic estimates of prediction error.

Unfortunately, with scarce data and overlapping observations there is no perfect solution.

6.7.8 Block versus Group Sampling

Here we compare cross-validation results between the group sampling and *hv*-block sampling methods described above, with the aim of selecting the appropriate number of neighbours (k) for a kNN regression model. Ten cross-validation folds were used for the *hv*-block sampling and five folds were used for the group sampling. This keeps the number of observations in the training set comparable between the two methods (due to the unusable data in each *hv*-block fold).

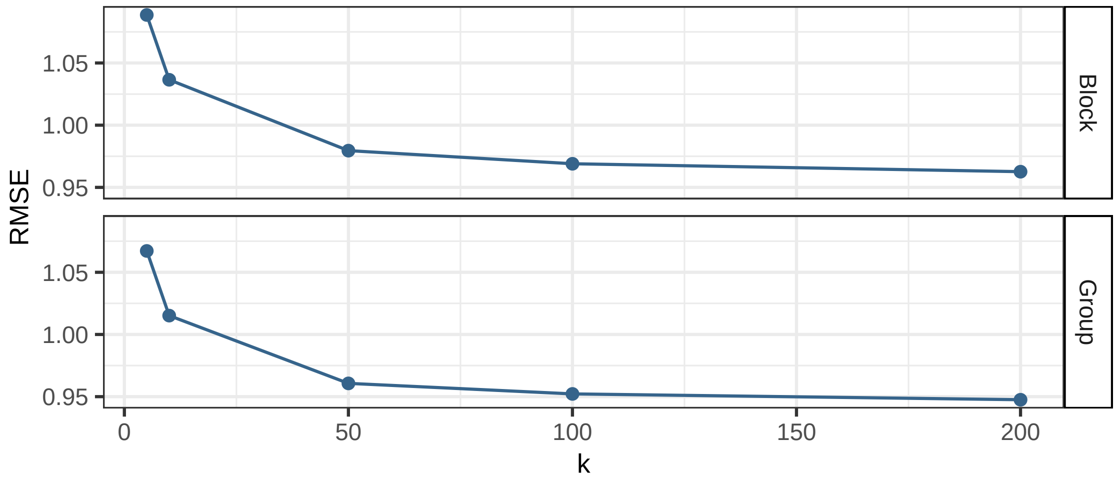

Figure 6.22: Comparison of cross-validation curves for kNN regression model using group and *hv*-block sampling approaches.

.

Figure 6.22 shows the CV curve for the two different sampling methods. The two approaches paint the same picture regarding model complexity and expected prediction error, with both models suggesting a high-bias, low-variance fit.

The block sampling procedure results in a slightly more pessimistic estimate of the prediction error, and more closely represents the data stratification we would encounter in practice. Unlike group sampling, block sampling also allows us to have temporally distinct in-sample and out-of-sample results.

Chapter 7

Conclusions and Future Work

7.1 Summary of Research

This thesis provides new evidence on the utility of news analytics in quantitative equity portfolio management. In particular, the research examines three broad approaches to extracting the potential informational edge provided by news information in the context of momentum-style equity strategies. That is, factor portfolios with holding periods on the order of 6 to 12 months.

All studies employ a proprietary Thomson-Reuters dataset of 5.3 million news items relating to historical constituents of the S&P 500 stock index from 2003 to 2018. The focus on historical constituents of the S&P 500 and use of an institutional-grade data source places our results in an empirical context relevant to the liquidity and mandate constraints often faced by practitioners while reflecting an economically relevant market strata of interest to academics.

The first study investigates the longitudinal predictive capacity of news sentiment and news coverage through a series of firm-level predictive regressions. Firm-level predictability is an important mechanism through which news content could potentially add value to portfolio construction, that may be missed in typical cross-sectional conditioning procedures. We find no supporting evidence that either news content or news volume are useful predictors of future return at the firm-level. Rates of statistical significance for predictive model coefficients were comparable to those of random covariates, and the majority of effect sizes were economically irrelevant. In relaxing the assumption of exogeneity of news, the only persistent relationships observed in the data were endogenous. The most important contribution of this study is that it is the first (to the best of my knowledge) to examine firm-level predictability at horizons of one-month or more with plausible news-based regressors.

The second study examines the performance of investment strategies employing nonparametric conditioning on news history and stock price momentum using single, double, and triple-sorted decile portfolios. The primary analysis technique for this study is the calendar-time overlapping portfolio approach of Jegadeesh and Titman (1993). Tests focus on strategies motivated by the prior literature on news-informed momentum-style portfolios that have not been tested within a stringent investment universe such as the S&P 500, or subject to practical rebalancing constraints. The results of this analysis offer little evidential support for the utility of news analytics in momentum-style or momentum-enhanced portfolios. Plausibly ex-ante-identifiable strategies, such as those motivated by the literature, failed to generate risk-adjusted excess returns, even after controlling for the GFC. Additional tests revealed that removing positive-sentiment stocks from the short leg of a traditional momentum strategy leads to significant risk-adjusted returns over the ex-GFC period. This is potentially a practical and economically relevant use-case for news sentiment in momentum portfolios, though its generalisability requires further support. The main contribution of this study is the benchmarking of numerous news-informed factor portfolios in an experimentally-consistent and practitioner-relevant setting.

The third study considers model-driven approaches to portfolio formation with news and price momentum inputs. Unlike the second study in this thesis and the major papers in the field, which condition on inputs using "naive" sorting procedures, the focus here is on using statistical models to learn in-sample relationships which are then deployed over out-of-sample horizons. We begin with linear and nonlinear implementations of the sliding window approach employed by Haugen and Baker (1996) before moving onto the fixed-sample train/validation/test approach to out-of-sample forecasting common in the machine-learning literature. We find that the combined use of news-derived features and flexible statistical learning algorithms offers only a modest increase in performance beyond a traditional momentum implementation. Measures of variable importance suggest that news is secondary to size, analyst following, and momentum in relevance for predicting future return.

7.2 Limitations and Future Research

This work scrutinises the economic relevance of news analytics for quantitative equity portfolios operating under momentum-style investment constraints. I believe the approaches we have taken to do this are statistically sound, conceptually defensible, and represent a genuine advancement of the literature. However, this is by no means an exhaustive attempt at extracting an informational edge from news media in the context of portfolio management, and there are a number of avenues left to explore.

A caveat to the findings of this thesis is the fact that the smallest firms within the S&P 500 still have large market capitalisations, and so the selected asset universe neglects small firms. While this was largely a deliberate decision so as to ensure large-capacity investability and economic relevance for any profitable strategies reported, it does meaningfully limit the applicability of this thesis for investment managers who are able to invest in smaller capitalisation firms. Indeed, previous findings suggest that small, less liquid stocks are more sensitive to news flow than larger stocks, so further research across different stock samples may still yield practical low-frequency trading opportunities not present in the S&P 500.

Additionally, our analysis ultimately sticks to a relatively small number of predictors, even when using statistical learning algorithms that could handle many more. It is plausible that a purely data-driven approach, in which a large set of news features are filtered and ensembled using advanced feature construction and selection techniques, could produce results with economic gains meaningfully larger than those presented in the current work.

For example, researchers who find this a compelling avenue to investigate may begin by constructing permutations of common of technical signals over a range of plausible formation periods using sentiment, coverage, and other news-based features, in addition to the traditional variants based on prices. As many of these features will be highly correlated, and therefore redundant, a model capable of regularization and selection, such as lasso, could be used to filter the predictor set. To avoid overfitting what is likely to be a time-varying and largely random structure of relative variable importance, a grouped selection procedure (e.g. grouped lasso) could be employed to identify the *types* of features that were important in the training set.

Comparing the stability of this supervised feature set across markets of the same asset class may identify an expanded set of predictors to be deployed in the validation set. Subsequent model selection and testing basically in line with the methodology used in this thesis would suffice for finalising and evaluating such an approach. There are many different ways to come at a more purely data-orientated style of investigation, but we hope this conveys the flavour of analysis being suggested.

It is also possible that a move in the other direction, that of greater human judgment, could capitalise on aspects of news analytics missed by our work. For example, an experienced equity analyst, being a consumer of financial news and actively involved in equity markets at the firm level, could plausibly identify components of news analytics and construct an associated mental model of news-market interaction that is not subsumed by the dominating behavioural theories in the literature or be likely to be identified by data-driven approaches (due to enormity of the hypothesis space for feature construction).

Based on personal discussions with equity analysts and portfolio managers, they are often acutely aware of news cycles and market responses, particularly around earnings announcements. This includes an awareness of the (approximate) impact of quantitative funds operating in the same market that trade on information such as analyst upgrades and corporate announcements. While most of the market response to news appears, both anecdotally and empirically, to be short lived (1 day or less), there may be lower frequency dynamics that active market participants are aware of that are yet to be rigorously tested and documented in the literature.

Those interested in this avenue of research might begin with "foundational" exploration using granular features of firm and news data to identify pockets of predictability. This may include studying the response of different industries, to different categorisations of news, from different outlets, and at different distances from reporting season. While some of these have been investigated in a piecemeal way across the literature, there is plenty of scope for a patient, foundational analysis of modern news analytics data tied to causal hypotheses of the relevant market mechanics.

Another way in which this thesis leaves room for future research is in its focus on momentum strategies. In the case of the current work, momentum was at the intersection of empirical plausibility (due to existing empirical work), conceptual plausibility (due to the role of public information in major behavioural theories of momentum), and suggested formation and holding periods that could be adequately incorporated into the available dataset (contrary to the value factor, which appears to operate at a much lower frequency). These facts do not suggest that focus on other factors would not have been worthwhile. To the best of my knowledge, there are no published attempts to enhance value, quality, or low-volatility strategies using news analytics. As such, incorporation of news features into an expanded set of factor portfolios appears to be fertile ground for research.

A further strand of research that has received very little academic attention, but seems to be a realistic use-case for news analytics in quantitative portfolio management, is tactical asset allocation—i.e. dynamic allocations across markets, often via indexed exposures. The current literature suggests aggregate news sentiment is useful for market timing within a given market, but to the best of my knowledge there are no studies aimed at practical implementations of cross-market allocations.

One could test an implementation of a tactical asset allocation strategy such as risk parity, that holds a basket of markets within each asset class. Rather than holding an equal-weighted, cap-weighted or risk-weighted exposure within each asset class, the strategy could invest only in the subset of markets with the highest sentiment trend (or other news-based signal). Comparing this to analogous strategies using price trend and momentum would be a relatively straight-forward analysis highly relevant to asset allocators. Many extensions to this idea are also possible, such as augmentation with statistical models and expansion of the factor universe used to tilt exposures. Since news analytics can be stratified by asset class, its efficacy for tilts across asset classes is also ripe for investigation.

Finally, a number of studies report a stronger relationship between news and volatility than news and returns, and there are lots of examples in the literature of news-informed volatility models. However, the research on news data and *comovement* is scarce. From conversations with fund managers, it is apparent that network effects through shared supply lines, customer bases, and other economic sensitivities, are increasingly being used in lieu of formal industry groupings for purposes of risk monitoring and position sizing. It is plausible that the role of news media in these types of network identification and risk-forecasting problems is under-documented in the literature, and may warrant further investigation.

Bibliography

Admati, Anat R (1985). “A noisy rational expectations equilibrium for multi-asset securities markets”. In: *Econometrica* 53.3, pp. 629–657.

Admati, Anat R and Paul Pfleiderer (1988). “A theory of intraday patterns: Volume and price variability”. In: *The Review of Financial Studies* 1.1, pp. 3–40.

Aggarwal, Charu C and ChengXiang Zhai (2012). “A survey of text classification algorithms”. In: *Mining Text Data*. Ed. by Charu C Aggarwal and ChengXiang Zhai. Springer Science & Business Media. Chap. 6, pp. 163–222.

Ahmad, Khurshid, JingGuang Han, Elaine Hutson, Colm Kearney, and Sha Liu (2016). “Media-expressed negative tone and firm-level stock returns”. In: *Journal of Corporate Finance* 37, 152–172.

Ahmed, Mansoor, Anirudh Sriram, and Sanjay Singh (2020). “Short Term Firm-Specific Stock Forecasting with BDI Framework”. In: *Computational Economics* 55.3, 745–778.

Akerlof, George A. (Aug. 1970). “The Market for "Lemons": Quality Uncertainty and the Market Mechanism”. In: *Quarterly Journal of Economics* 84 (3), pp. 488–500.

Allen, David E, Michael McAleer, and Abhay K Singh (2017). “An entropy-based analysis of the relationship between the Dow Jones Index and the TRNA Sentiment series”. In: *Applied Economics* 49.7, pp. 677–692.

Amihud, Yakov (2002). “Illiquidity and stock returns: cross-section and time-series effects”. In: *Journal of Financial Markets* 5.1, pp. 31–56.

Ammann, Manuel, Roman Frey, and Michael Verhofen (2014). “Do Newspaper Articles Predict Aggregate Stock Returns?” In: *Journal of Behavioral Finance* 15.3, 195–213.

Andrews, Donald WK (1991). “Heteroskedasticity and autocorrelation consistent covariance matrix estimation”. In: *Econometrica*, pp. 817–858.

Antweiler, Werner and Murray Z Frank (2004). “Is all that talk just noise? The information content of internet stock message boards”. In: *The Journal of Finance* 59.3, pp. 1259–1294.

Aria, Massimo and Corrado Cuccurullo (2017). "bibliometrix: An R-tool for comprehensive science mapping analysis". In: *Journal of Informetrics* 11.4, pp. 959–975.

Audrino, Francesco, Fabio Sigrist, and Daniele Ballinari (APR-JUN 2020). "The impact of sentiment and attention measures on stock market volatility". In: *International Journal of Forecasting* 36.2, 334–357.

Baker, Malcolm and Jeffrey Wurgler (2006). "Investor sentiment and the cross-section of stock returns". In: *The Journal of Finance* 61.4, pp. 1645–1680.

– (2007). "Investor sentiment in the stock market". In: *The Journal of Economic Perspectives* 21.2, pp. 129–151.

Banerjee, Snehal and Ilan Kremer (2010). "Disagreement and Learning: Dynamic Patterns of Trade". In: *The Journal of Finance* 65 (4), pp. 1269–1302.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998). "A model of investor sentiment". In: *Journal of Financial Economics* 49.3, pp. 163–187.

Barroso, Pedro and Pedro Santa-Clara (2015). "Momentum has its moments". In: *Journal of Financial Economics* 116.1, pp. 111–120.

Bender, Oliver, Franz Josef Och, and Hermann Ney (2003). "Maximum entropy models for named entity recognition". In: *Proceedings of the seventh conference on natural language learning at HLT-NAACL*. Vol. 4. Association for Computational Linguistics, pp. 148–151.

Benello, Julian, Andrew W Mackie, and James A Anderson (1989). "Syntactic category disambiguation with neural networks". In: *Computer Speech & Language* 3.3, pp. 203–217.

Bergmeir, Christoph and José M. Benítez (2012). "On the use of cross-validation for time series predictor evaluation". English. In: *Information Sciences* 191, pp. 192–213.

Bianconi, Marcelo, Xiaxin Hua, and Chih Ming Tan (JUL 2015). "Determinants of systemic risk and information dissemination". In: *International Review of Economics & Finance* 38, 352–368.

Bikel, Daniel M, Scott Miller, Richard Schwartz, and Ralph Weischedel (1997). "Nymble: a high-performance learning name-finder". In: *Proceedings of the fifth conference on applied natural language processing*. Association for Computational Linguistics, pp. 194–201.

Black, Fischer (1986). "Noise". In: *The Journal of Finance* 41.3, pp. 528–543.

BlackRock (2021). *What is Factor Investing?* URL: `https://www.blackrock.com/us/individual/investment-ideas/what-is-factor-investing` (visited on 05/14/2021).

Blitz, David, Joop Huij, and Martin Martens (2011). "Residual momentum". In: *Journal of Empirical Finance* 18.3, pp. 506–521.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008.

Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011). "Twitter mood predicts the stock market". In: *Journal of Computational Science* 2.1, pp. 1–8.

Börner, Katy, Chaomei Chen, and Kevin W Boyack (2003). "Visualizing knowledge domains". In: *Annual Review of Information Science and Technology* 37.1, pp. 179–255.

Borovkova, Svetlana and Diego Mahakena (JUL 3 2015). "News, volatility and jumps: the case of natural gas futures". In: *Quantitative Finance* 15.7, 1217–1242.

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, and Matthew Richardson (2013). *Which News Moves Stock Prices? A Textual Analysis*. NBER Working Papers 18725. National Bureau of Economic Research, Inc.

Boudoukh, Jacob, Ronen Israel, and Matthew Richardson (2019). "Long-horizon predictability: A cautionary tale". In: *Financial Analysts Journal* 75.1, pp. 17–30.

Boudoukh, Jacob and Matthew Richardson (1994). "The Statistics of Long-Horizon Regressions Revisited". In: *Mathematical Finance* 4.2, pp. 103–119.

Boulton, Thomas J and Marcus V Braga-Alves (2010). "The skinny on the 2008 naked short-sale restrictions". In: *Journal of Financial Markets* 13.4, pp. 397–421.

Box, George EP and David R Cox (1964). "An analysis of transformations". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 26.2, pp. 211–243.

Breiman, Leo and Philip Spector (1992). "Submodel Selection and Evaluation in Regression. The X-Random Case". In: *International Statistical Review / Revue Internationale de Statistique* 60.3, pp. 291–319.

Breusch, Trevor S and Adrian R Pagan (1979). "A simple test for heteroscedasticity and random coefficient variation". In: *Econometrica*, pp. 1287–1294.

Britten-Jones, Mark, Anthony Neuberger, and Ingmar Nolte (2011). "Improved inference in regression with overlapping observations". In: *Journal of Business Finance & Accounting* 38.5-6, pp. 657–683.

Brown, Gregory W and Michael T Cliff (2005). "Investor sentiment and asset valuation". In: *The Journal of Business* 78.2, pp. 405–440.

Brown, Gregory W. and Michael T. Cliff (2004). "Investor sentiment and the near-term stock market". In: *Journal of Empirical Finance* 11.1, pp. 1–27. ISSN: 0927-5398.

Butler, Matthew and Vlado Keselj (2009). "Financial Forecasting Using Character N-Gram Analysis and Readability Scores of Annual Reports." In: *Canadian Conference on AI*. Springer, pp. 39–51.

Cahan, Rochester H., Steven F. Cahan, Tan (Charlene) Lee, and Nhut H. Nguyen (2017). "Media Content, Accounting Quality, and Liquidity Volatility". In: *European Accounting Review* 26.1, 1–25.

Cahill, Daniel, Marvin Wee, and Joey W. Yang (SEP 2017). "Media sentiment and trading strategies of different types of traders". In: *Pacific-Basin Finance Journal* 44, 160–172.

Cahlik, Tomas (2000). "Comparison of the maps of science". In: *Scientometrics* 49.3, pp. 373–387.

Callon, Michel, Jean-Pierre Courtial, and Francoise Laville (1991). "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry". In: *Scientometrics* 22.1, pp. 155–205.

Calomiris, Charles W. and Harry Mamaysky (2019). "How news and its context drive risk and returns around the world". In: *Journal of Financial Economics* 133.2, 299–336.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2011). "Robust inference with multiway clustering". In: *Journal of Business & Economic Statistics* 29.2, pp. 238–249.

Cameron, A Colin and Douglas L Miller (2015). "A practitioner's guide to cluster-robust inference". In: *Journal of Human Resources* 50.2, pp. 317–372.

Campbell, John Y and Albert S Kyle (1993). "Smart money, noise trading and stock price behaviour". In: *The Review of Economic Studies* 60.1, pp. 1–34.

Campbell, John Y. and Robert J. Shiller (1988). "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors". In: *Review of Financial Studies* 1.3, 195–228.

Cao, Jie, Bing Han, and Qinghai Wang (2017). "Institutional investment constraints and stock prices". In: *Journal of Financial and Quantitative Analysis* 52.2, pp. 465–489.

Carhart, Mark M (1997). "On persistence in mutual fund performance". In: *The Journal of Finance* 52.1, pp. 57–82.

Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto, and P Krishna Gummadi (2010). "Measuring user influence in twitter: The million follower fallacy." In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, p. 30.

Chan, Felix, Robert B. Durand, Joyce Khuu, and Lee A. Smales (SEP 2017). "The Validity of Investor Sentiment Proxies". In: *International Review of Finance* 17.3, 473–477.

Chan, Samuel WK and Mickey WC Chong (2017). "Sentiment analysis in financial texts". In: *Decision Support Systems* 94, pp. 53–64.

Chan, Wesley S (2003). "Stock price reaction to news and no-news: drift and reversal after headlines". In: *Journal of Financial Economics* 70.2, pp. 223–260.

Charniak, Eugene (1996). *Statistical language learning*. MIT press.

Checkley, M. S., D. Arion Higon, and H. Alles (2017). "The hasty wisdom of the mob: How market sentiment predicts stock market behavior". In: *Expert Systems with Applications* 77, 256–263.

Chen, Mu-Yen, Chien-Hsiang Liao, and Ren-Pao Hsieh (2019). "Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach". In: *Computers in Human Behavior* 101, 402–408.

Chincarini, Ludwig B and Daehwan Kim (2010). *Quantitative equity portfolio management: An Active Approach to Portfolio Construction and Management*. McGraw Hill Professional.

Christiano, Lawrence J (2012). "Christopher A. Sims and vector autoregressions". In: *The Scandinavian Journal of Economics* 114.4, pp. 1082–1104.

Chu, C.-K. and J. S. Marron (1991). "Comparison of Two Bandwidth Selectors with Dependent Errors". In: *The Annals of Statistics* 19.4, pp. 1906–1918.

Clark, Peter K (1973). "A subordinated stochastic process model with finite variance for speculative prices". In: *Econometrica*, pp. 135–155.

Clements, Adam E. and Neda Todorova (2016). "Information Flow, Trading Activity and Commodity Futures Volatility". In: *Journal of Futures Markets* 36.1, 88–104.

Cobo, Manuel J, Antonio Gabriel López-Herrera, Enrique Herrera-Viedma, and Francisco Herrera (2011). "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field". In: *Journal of Informetrics* 5.1, pp. 146–166.

Cochrane, John H (1991). "A critique of the application of unit root tests". In: *Journal of Economic Dynamics and Control* 15.2, pp. 275–284.

Coqueret, Guillaume (2020). "Stock-specific sentiment and return predictability". In: *Quantitative Finance* 20.9, pp. 1531–1551.

Coulter, Neal, Ira Monarch, and Suresh Konda (1998). "Software engineering as seen through its research literature: A study in co-word analysis". In: *Journal of the American Society for Information Science* 49.13, pp. 1206–1223.

Crain, Steven P, Ke Zhou, Shuang-Hong Yang, and Hongyuan Zha (2012). "Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond". In: *Mining text data*. Ed. by Charu C Aggarwal and ChengXiang Zhai. Springer. Chap. 5, pp. 129–161.

Cribari-Neto, Francisco and Wilton Bernardino da Silva (2011). "A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model". In: *Advances in Statistical Analysis* 95.2, pp. 129–146.

Cui, Xin, Daniel Lam, and Arun Verma (2017). *Embedded Value in Bloomberg News and Social Sentiment Data*. URL: `http://b.bloomberg.com/reg-EDFSentimentWhitepaper?src=blog` (visited on 09/07/2017).

Daelemans, Walter, Jakub Zavrel, Peter Berck, and Steven Gillis (1996). "MBT: A memory-based part of speech tagger-generator". In: *WVLC*. arXiv preprint cmp-lg/9607012.

Daniel, K, D Hirshleifer, and A Subrahmanyam (1998). "Investor psychology and security market under- and overreactions". In: *The Journal of Finance* 53.6, 1839–1885.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers (1997). "Measuring mutual fund performance with characteristic-based benchmarks". In: *The Journal of Finance* 52.3, pp. 1035–1058.

Daniel, Kent and Tobias J Moskowitz (2016). "Momentum crashes". In: *Journal of Financial Economics* 122.2, pp. 221–247.

Das, S, A Martinez-Jerez, and P Tufano (2005). "eInformation: A clinical study of investor discussion and sentiment". In: *Financial Management* 34.3, 103–137.

Das, Sanjiv R. and Mike Y. Chen (2007). "Yahoo! for Amazon: Sentiment extraction from small talk on the web". In: *Management Science* 53.9, 1375–1388.

Davis, Angela K, Jeremy M Piger, and Lisa M Sedor (2012). "Beyond the numbers: Measuring the information content of earnings press release language". In: *Contemporary Accounting Research* 29.3, pp. 845–868.

De Bondt, Werner F. M. and Richard Thaler (July 1985). "Does the Stock Market Overreact?" In: *The Journal of Finance* 40 (3), pp. 793–805.

De Jong, Kenneth (Oct. 1988). "Learning with genetic algorithms: An overview". In: *Machine Learning* 3.2, pp. 121–138.

De Prado, Marcos Lopez (2018). *Advances in Financial Machine Learning*. John Wiley & Sons.

DeJong, David N, John C Nankervis, N Eugene Savin, and Charles H Whiteman (1992). "The power problems of unit root test in time series with autoregressive errors". In: *Journal of Econometrics* 53.1-3, pp. 323–343.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann (1990). "Noise Trader Risk in Financial Markets". In: *Journal of Political Economy* 98.4, pp. 703–738.

Dewally, Michael (2003). "Internet investment advice: Investing with a rock of salt". In: *Financial Analysts Journal* 59.4, pp. 65–77.

Dougal, Casey, Joseph Engelberg, Diego Garcia, and Christopher A Parsons (2012). "Journalists and the stock market". In: *Review of Financial Studies* 25.3, pp. 639–679.

Driscoll, John C and Aart C Kraay (1998). "Consistent covariance matrix estimation with spatially dependent panel data". In: *Review of Economics and Statistics* 80.4, pp. 549–560.

Durbin, James and Geoffrey S Watson (1950a). "Testing for serial correlation in least squares regression: I". In: *Biometrika* 37.3/4, pp. 409–428.

– (1950b). "Testing for serial correlation in least squares regression: II". In: *Biometrika* 38, pp. 159–177.

– (1971). "Testing for serial correlation in least squares regression. III". In: *Biometrika* 58.1, pp. 1–19.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara (2002). "Is Information Risk a Determinant of Asset Returns?" In: *The Journal of Finance* 57.5, pp. 2185–2221.

Easley, David and Maureen O'Hara (2004). "Information and the cost of capital". In: *The Journal of Finance* 59.4, pp. 1553–1583.

Edgerton, David and Ghazi Shukur (1999). "Testing autocorrelation in a system perspective testing autocorrelation". In: *Econometric Reviews* 18.4, pp. 343–386.

Engelberg, Joseph E. and Christopher A. Parsons (2011). "The Causal Impact of Media in Financial Markets". In: *The Journal of Finance* 66.1, pp. 67–97. ISSN: 1540-6261.

Engle, Robert F and Victor K Ng (1993). "Measuring and testing the impact of news on volatility". In: *The Journal of Finance* 48.5, pp. 1749–1778.

Engle, Robert F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4, pp. 987–1007.

Engle, Robert F., Martin Klint Hansen, and Asger Lunde (Dec. 2012). *And Now, The Rest of the News: Volatility and Firm Specific News Arrival*. CREATES Research Paper 2012-56. Department of Economics and Business Economics, Aarhus University. URL: `https://ideas.repec.org/p/aah/create/2012-56.html`.

Epps, Thomas W and Mary Lee Epps (1976). "The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis". In: *Econometrica*, pp. 305–321.

Fabozzi, Frank J, Sergio M Focardi, and Petter N Kolm (2010). *Quantitative equity investing: Techniques and strategies*. John Wiley & Sons.

Fama, Eugene F (1998). "Market efficiency, long-term returns, and behavioral finance1". In: *Journal of Financial Economics* 49.3, pp. 283–306.

Fama, Eugene F and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics* 33.1, pp. 3–56.

Fama, Eugene F and James D MacBeth (1973). "Risk, return, and equilibrium: Empirical tests". In: *Journal of Political Economy* 81.3, pp. 607–636.

Fama Eugene F.; French, Kenneth R. (Apr. 2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116 (1).

Fang, Lily and Joel Peress (2009). "Media coverage and the cross-section of stock returns". In: *The Journal of Finance* 64.5, pp. 2023–2052.

Feldman, Ronen, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal (2010). "Management's tone change, post earnings announcement drift and accruals". In: *Review of Accounting Studies* 15.4, pp. 915–953.

Ferguson, Nicky J., Dennis Philip, Herbert Y. T. Lam, and Jie Michael Guo (2015). "Media Content and Stock Returns: The Predictive Power of Press." In: *Multinational Finance Journal* 19.1, pp. 1–31.

Feuerriegel, Stefan and Helmut Prendinger (2016). "News-based trading strategies". In: *Decision Support Systems* 90, pp. 65–74.

Finkel, Jenny Rose, Trond Grenager, and Christopher Manning (2005). "Incorporating non-local information into information extraction systems by gibbs sampling". In: *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, pp. 363–370.

Fisher, Ronald A (1936). "The use of multiple measurements in taxonomic problems". In: *Annals of Human Genetics* 7.2, pp. 179–188.

Fitzgibbons, Shaun, Jacques Friedman, Lukasz Pomorski, and Laura Serban (2017). "Long-Only Style Investing: Don't Just Mix, Integrate". In: *The Journal of Investing* 26.4, pp. 153–164.

Foote, Christopher L. (2007). *Space and time in macroeconomic panel data: young workers and state-level unemployment revisited*. Working Papers 07-10. Federal Reserve Bank of Boston.

Forman, George (2003). "An extensive empirical study of feature selection metrics for text classification". In: *Journal of Machine Learning Research* 3.Mar, pp. 1289–1305.

Foster, F Douglas and Subramanian Viswanathan (Oct. 1990). "A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Securities Markets". In: *Review of Financial Studies* 3 (4), pp. 593–624.

– (Jan. 1993a). "The Effect of Public Information and Competition on Trading Volume and Price Volatility". In: *Review of Financial Studies* 6 (1), pp. 23–56.

– (1993b). "Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models". In: *The Journal of Finance* 48.1, pp. 187–211.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2009). *The elements of statistical learning*. 2nd ed. Springer series in statistics New York.

Friedman, Jerome H (2001). "Greedy function approximation: a gradient boosting machine". In: *Annals of Statistics*, pp. 1189–1232.

– (2002). "Stochastic gradient boosting". In: *Computational Statistics & Data Analysis* 38.4, pp. 367–378.

Fung, Gabriel Pui Cheong, Jeffrey Xu Yu, and Wai Lam (2003). "Automatic stock trend prediction by real time news". In: *Advances in Data Mining and Modeling*. World Scientific, pp. 48–59.

Gabrovšek, Peter, Darko Aleksovski, Igor Mozetič, and Miha Grčar (2017). "Twitter sentiment around the Earnings Announcement events". In: *PloS One* 12.2, e0173151.

Gan, Baoqing, Vitali Alexeev, Ron Bird, and Danny Yeung (2020). "Sensitivity to sentiment: News vs social media". In: *International Review of Financial Analysis* 67.

Garcia, A. (2016). "Global financial indices and twitter sentiment: A random matrix theory approach". In: *Physica A-Statistical Mechanics and its Applications* 461, 509–522.

Garcia, Diego (2013). "Sentiment during Recessions". In: *Journal of Finance* 68.3, 1267–1300.

Garfield, Eugene (Apr. 2004). "Historiographic Mapping of Knowledge Domains Literature". In: *Journal of Information Science* 30 (2), pp. 119–145.

Garfield, Eugene, A. I. Pudovkin, and V. S. Istomin (2003). "Why do we need algorithmic historiography?" In: *Journal of the American Society for Information Science and Technology* 54 (5), pp. 400–412.

Geva, Tomer and Jacob Zahavi (2014). "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news". In: *Decision Support Systems* 57, pp. 212–223.

Glasserman, Paul and Harry Mamaysky (2019). "Does Unusual News Forecast Market Stress?" In: *Journal of Financial and Quantitative Analysis* 54.5, 1937–1974.

Golub, Gene H, Michael Heath, and Grace Wahba (1979). "Generalized cross-validation as a method for choosing a good ridge parameter". In: *Technometrics* 21.2, pp. 215–223.

Granger, Clive WJ (1969). "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica*, pp. 424–438.

Greene, William H (2005). *Econometric Analysis*. 5th ed. Prentice Hall.

Greenwell, Brandon, Bradley Boehmke, Jay Cunningham, and GBM Developers (2019). *gbm: Generalized Boosted Regression Models*. R package version 2.1.5. URL: `https://CRAN.R-project.org/package=gbm`.

Grefenstette, Gregory and Pasi Tapanainen (1994). "What is a word, what is a sentence? Problems of Tokenisation". In: *The 3rd International Conference on Computational Lexicography*. Budapest, pp. 79–87.

Griffin, John M, Nicholas H Hirschey, and Patrick J Kelly (2011). "How important is the financial media in global markets?" In: *The Review of Financial Studies* 24.12, pp. 3941–3992.

Griffith, John, Mohammad Najand, and Jiancheng Shen (JAN 2 2020). "Emotions in the Stock Market". In: *Journal of Behavioral Finance* 21.1, 42–56.

Grinold, Richard C and Ronald N Kahn (2000). *Active portfolio management*. McGraw Hill New York.

Grinold, Richard C. (1989). "The fundamental law of active management". In: *The Journal of Portfolio Management* 15.3, pp. 30–37. ISSN: 0095-4918.

Groß-Klußmann, Axel and Nikolaus Hautsch (2011). "When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions". In: *Journal of Empirical Finance* 18.2, 321–340.

Grossman, Sanford J and Joseph E Stiglitz (1976). "Information and competitive price systems". In: *The American Economic Review* 66.2, pp. 246–253.

Gutierrez, Roberto C and Eric K Kelley (2008). "The Long-Lasting Momentum in Weekly Returns". In: *The Journal of Finance* 63.1, pp. 415–447.

Hagenau, Michael, Michael Liebmann, and Dirk Neumann (2013). "Automated news reading: Stock price prediction based on financial news using context-capturing features". In: *Decision Support Systems* 55.3, pp. 685–697.

Hamilton, James Douglas (2014). *Time series analysis*. Princeton university press.

Hanna, Alan J., John D. Turner, and Clive B. Walker (2020). "News media and investor sentiment during bull and bear markets". In: *European Journal of Finance*.

Hansen, Lars Peter and Robert J Hodrick (1980). "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis". In: *Journal of Political Economy* 88.5, pp. 829–853.

Hansen, Lars Peter and Kenneth J Singleton (1982). "Generalized instrumental variables estimation of nonlinear rational expectations models". In: *Econometrica*, pp. 1269–1286.

Harris, Zellig S (1954). "Distributional structure". In: *Word* 10.2-3, pp. 146–162.

Hart, Roderick P (2000). *DICTION 5.0: the text analysis program*. Austin, TX: Digitex.

Hastie, TJ and R Tibshirani (1986). "Generalized Additive Models (with discussion)". In: *Statistical Science* 1, pp. 297–318.

– (1990). *Generalized Additive Models*. Chapman and Hall.

Haugen, Robert A and Nardin L Baker (1996). "Commonality in the determinants of expected stock returns". In: *Journal of Financial Economics* 41.3, pp. 401–439.

He, Hua and Jiang Wang (WIN 1995). "Differential Information and Dynamic Behavior of Stock Trading Volume". In: *Review of Financial Studies* 8 (4), pp. 919–972.

Hendershott, Terrence, Dmitry Livdan, and Norman Schuerhoff (2015). "Are institutions informed about news?" In: *Journal of Financial Economics* 117.2, 249–287.

Henry, Elaine (2008). "Are investors influenced by how earnings press releases are written?" In: *The Journal of Business Communication* 45.4, pp. 363–407.

Henry, Elaine and Andrew J. Leone (2016). "Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone". In: *The Accounting Review* 91.1, pp. 153–178.

Heston, Steven L and Nitish Ranjan Sinha (2017). "News vs. Sentiment: Predicting Stock Returns from News Stories". In: *Financial Analysts Journal* 73.3, pp. 1–17.

Hillert, Alexander, Heiko Jacobs, and Sebastian Mueller (2018). "Journalist disagreement". In: *Journal of Financial Markets* 41, 57–76.

– (2014). "Media Makes Momentum". In: *Review of Financial Studies* 27.12, 3467–3501.

Hirshleifer, David and Siew Hong Teoh (2003). "Limited attention, information disclosure, and financial reporting". In: *Journal of Accounting and Economics* 36 (1-3), pp. 0–386.

Hjalmarsson, Erik (2006). "New methods for inference in long-run predictive regressions". In: *FRB International Finance Discussion Paper* 853.

Ho, Kin-Yip, Yanlin Shi, and Zhaoyong Zhang (2013). "How does news sentiment impact asset volatility? Evidence from long memory and regime-switching approaches". In: *North American Journal of Economics and Finance* 26.SI, 436–456.

Hodrick, Robert J (1992). "Dividend yields and expected stock returns: Alternative procedures for inference and measurement". In: *The Review of Financial Studies* 5.3, pp. 357–386.

Hofmann, Thomas (2001). "Unsupervised learning by probabilistic latent semantic analysis". In: *Machine Learning* 42.1, pp. 177–196.

Holden, Craig W. and Avanidhar Subrahmanyam (Mar. 1992). "Long-Lived Private Information and Imperfect Competition". In: *The Journal of Finance* 47 (1), pp. 247–270.

Hong, Harrison and Jeremy C. Stein (1999). "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets". In: *The Journal of Finance* 54.6, pp. 2143–2184.

Huang, Chenn-Jung, Jia-Jian Liao, Dian-Xiu Yang, Tun-Yu Chang, and Yun-Cheng Luo (2010). "Realization of a news dissemination agent based on weighted association rules and text mining techniques". In: *Expert Systems with Applications* 37.9, pp. 6409–6413.

Huang, Dashan and Guofu Zhou (2017). "Upper bounds on return predictability". In: *Journal of Financial and Quantitative Analysis* 52.2, pp. 401–425.

Huberman, Gur and Tomer Regev (2001). "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar". In: *The Journal of Finance* 56.1, pp. 387–396.

Huynh, Thanh D and Daniel R Smith (July 2017). "Stock Price Reaction to News: The Joint Effect of Tone and Attention on Momentum". In: *Journal of Behavioral Finance*, pp. 1–25.

Hyndman, Rob J. and Yeasmin Khandakar (2008). "Automatic Time Series Forecasting: The forecast Package for R". In: *Journal of Statistical Software, Articles* 27.3, pp. 1–22.

Jaynes, Edwin T (2003). *Probability theory: The logic of science*. Cambridge university press.

Jegadeesh, Narasimhan and Sheridan Titman (1993). "Returns to buying winners and selling losers: Implications for stock market efficiency". In: *The Journal of Finance* 48.1, pp. 65–91.

Jiang, Jing (2012). "Information extraction from text". In: *Mining text data*. Ed. by Charu C. Aggarwal and ChengXiang Zhai. Springer, pp. 11–41.

Joachims, Thorsten (1998). "Text Categorization with Support Vector Machines: Learning with Many Relevant Features". In: *Proceedings of the 10th European Conference on Machine Learning*. ECML'98. Chemnitz, Germany: Springer-Verlag, pp. 137–142. ISBN: 3-540-64417-2, 978-3-540-64417-0.

Johnman, Mark, Bruce James Vanstone, and Adrian Gepp (2018). "Predicting FTSE 100 returns and volatility using sentiment analysis". In: *Accounting & Finance* 58.1, SI, 253–274.

Jurafsky, Daniel and James H. Martin (2009). *Speech and Language Processing*. Accessed: 2017-08-04.

– (2016). *Speech and Language Processing*. URL: `https://web.stanford.edu/~jurafsky/slp3/` (visited on 08/04/2017).

Kandel, Eugene and Neil D. Pearson (Aug. 1995). "Differential Interpretation of Public Signals and Trade in Speculative Markets". In: *Journal of Political Economy* 103 (4), pp. 831–872.

Karagozoglu, Ahmet K and Frank J Fabozzi (2017). "Volatility Wisdom of Social Media Crowds". In: *The Journal of Portfolio Management* 43.2, pp. 136–151.

Kearney, Colm and Sha Liu (2014). "Textual sentiment in finance: A survey of methods and models". In: *International Review of Financial Analysis* 33, 171–185.

Kelly, Stephen and Khurshid Ahmad (2018). "Estimating the impact of domain-specific news sentiment on financial assets". In: *Knowledge-Based Systems* 150, 116–126.

Kendall, Maurice G (1954). "Note on bias in the estimation of autocorrelation". In: *Biometrika* 41.3-4, pp. 403–404.

Khuu, Joyce, Robert B. Durand, and Lee A. Smales (DEC 2016). "Melancholia and Japanese stock returns-2003 to 2012". In: *Pacific-Basin Finance Journal* 40.B, SI, 424–437.

Kim, Soon-Ho and Dongcheol Kim (NOV 2014). "Investor sentiment from internet message postings and the predictability of stock returns". In: *Journal of Economic Behavior & Organization* 107.B, SI, 708–729.

Kingma, Diederik P and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Kohavi, Ron (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection". In: Morgan Kaufmann, pp. 1137–1143.

Kothari, SP, Xu Li, and James E Short (2009). "The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis". In: *The Accounting Review* 84.5, pp. 1639–1670.

Kraussl, Roman and Elizaveta Mirgorodskaya (2017). "Media, sentiment and market performance in the long run". In: *European Journal of Finance* 23.11, 1059–1082.

Kravet, Todd and Volkan Muslu (2013). "Textual risk disclosures and investors' risk perceptions". In: *Review of Accounting Studies* 18.4, pp. 1088–1122.

Kroujiline, Dimitri, Maxim Gusev, Dmitry Ushanov, Sergey V. Sharov, and Boris Govorkov (NOV 2016). "Forecasting stock market returns over multiple time horizons". In: *Quantitative Finance* 16.11, 1695–1712.

Kumar, Alok and Charles Lee (2006). "Retail investor sentiment and return comovements". In: *The Journal of Finance* 61.5, pp. 2451–2486.

Kyle, Albert S (1985). "Continuous auctions and insider trading". In: *Econometrica*, pp. 1315–1335.

Laird, Nan M and James H Ware (1982). "Random-effects models for longitudinal data". In: *Biometrics*, pp. 963–974.

Leinweber, David and Jacob Sisk (2011). "Event-driven trading and the "new news"". In: *The Journal of Portfolio Management* 38.1, pp. 110–124.

Leung, Henry and Thai Ton (2015). "The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks". In: *Journal of Banking & Finance* 55, pp. 37–55.

Lewis, David D (1992). "An evaluation of phrasal and clustered representations on a text categorization task". In: *Proceedings of the 15th annual international*

ACM SIGIR conference on research and development in information retrieval. ACM, pp. 37–50.

Li, Feng (2006). “Do stock market investors understand the risk sentiment of corporate annual reports?” Working paper.

– (2010a). “Textual analysis of corporate disclosures: A survey of the literature”. In: *Journal of Accounting Literature* 29, p. 143.

– (2010b). “The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach”. In: *Journal of Accounting Research* 48.5, pp. 1049–1102.

Li, Qing, TieJun Wang, Ping Li, Ling Liu, Qixu Gong, and Yuanzhu Chen (2014). “The effect of news and public mood on stock movements”. In: *Information Sciences* 278, 826–840.

Li, Xiaodong, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng (2014). “News impact on stock price return via sentiment analysis”. In: *Knowledge-Based Systems* 69.SI, 14–23.

Liu, Bing and Lei Zhang (2012). “A Survey of Opinion Mining and Sentiment Analysis”. In: *Mining Text Data*. Ed. by Charu C Aggarwal and ChengXiang Zhai. Springer Science & Business Media. Chap. 13.

Long, J Scott and Laurie H Ervin (2000). “Using heteroscedasticity consistent standard errors in the linear regression model”. In: *The American Statistician* 54.3, pp. 217–224.

Loughran, Tim and Bill McDonald (2011). “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks”. In: *The Journal of Finance* 66.1, pp. 35–65.

– (2015). “The use of word lists in textual analysis”. In: *Journal of Behavioral Finance* 16.1, pp. 1–11.

– (2016). “Textual Analysis in Accounting and Finance: A Survey”. In: *Journal of Accounting Research* 54.4, pp. 1187–1230.

Loughran, Tim, Bill McDonald, and Hayong Yun (2009). “A Wolf in Sheep’s Clothing: The Use of Ethics-Related Terms in 10-K Reports”. In: *Journal of Business Ethics* 89.1, pp. 39–49.

Lovins, Julie Beth (1968). “Development of a stemming algorithm”. In: *Mech. Translat. & Comp. Linguistics* 11.1-2, pp. 22–31.

Lu, Yang-Cheng and Yu-Chen Wei (2013). “The Chinese News Sentiment Around Earnings Announcements”. In: *Romanian Journal of Economic Forecasting* 16.3, 44–58.

Lütkepohl, Helmut (2006). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Al-Maadid, Alanoud, Guglielmo Maria Caporale, Fabio Spagnolo, and Nicola Spagnolo (2020). "The impact of business and political news on the GCC stock markets". In: *Research in International Business and Finance* 52.

Malkiel, Burton G (1977). "The Valuation of Closed-End Investment-Company Shares". In: *The Journal of Finance* 32.3, pp. 847–859.

Manela, Asaf and Alan Moreira (2017). "News implied volatility and disaster concerns". In: *Journal of Financial Economics* 123.1, 137–162.

Mankiw, N Gregory and Matthew D Shapiro (1986). "Do we reject too often?: Small sample properties of tests of rational expectations models". In: *Economics Letters* 20.2, pp. 139–145.

Mankiw, N. Gregory, David Romer, and Matthew D. Shapiro (1985). "An Unbiased Reexamination of Stock Market Volatility". In: *The Journal of Finance* 40.3, pp. 677–687.

Manning, Christopher D and Hinrich Schütze (1999). *Foundations of statistical natural language processing*. Vol. 999. MIT Press.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze (2008). *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press.

Maragoudakis, Manolis and Dimitrios Serpanos (2016). "Exploiting Financial News and Social Media Opinions for Stock Market Analysis using MCMC Bayesian Inference". In: *Computational Economics* 47.4, pp. 589–622.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy (2014). "Sentiment analysis algorithms and applications: A survey". In: *Ain Shams Engineering Journal* 5.4, pp. 1093–1113. ISSN: 2090-4479.

Merton, Robert C (1987). "A simple model of capital market equilibrium with incomplete information". In: *The Journal of Finance* 42.3, pp. 483–510.

Miangah, Tayebeh Mosavi (2014). "Finite-State Technology in Natural Language Processing". In: *4th International Conference on Information Technology Management, Communication and Computer Science, At Tehran, Iran*.

Mikheev, Andrei (1998). "Feature lattices for maximum entropy modelling". In: *Proceedings of the 17th international conference on computational linguistics-Volume 2*. Association for Computational Linguistics, pp. 848–854.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Mikolov, Tomáš, Wen-tau Yih, and Geoffrey Zweig (2013). "Linguistic regularities in continuous space word representations". In: *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pp. 746–751.

Milgrom, Paul and Nancy Stokey (1982). "Information, trade and common knowledge". In: *Journal of Economic Theory* 26.1, pp. 17–27.

Millo, Giovanni (2017). "Robust Standard Error Estimators for Panel Models: A Unifying Approach". In: *Journal of Statistical Software, Articles* 82.3, pp. 1–27.

Miner, Gary (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.

Minh, Dang Lien, Abolghasem Sadeghi-Niaraki, Huynh Duc Huy, Kyungbok Min, and Hyeonjoon Moon (2018). "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network". In: *IEEE Access* 6, 55392–55404.

Mitchell, Mark L and J Harold Mulherin (1994). "The impact of public information on the stock market". In: *The Journal of Finance* 49.3, pp. 923–950.

Mitra, Gautam and Leela Mitra (2011). *The handbook of news analytics in finance*. Vol. 596. John Wiley & Sons.

Mitra, Leela, Gautam Mitra, and Dan Dibartolomeo (2009). "Equity portfolio risk estimation using market information and sentiment". In: *Quantitative Finance* 9.8, pp. 887–895.

Myskova, Renata, Petr Hajek, and Vladimir Olej (2018). "Predicting Abnormal Stock Return Volatility Using Textual Analysis of News - A Meta-Learning Approach". In: *Amfiteatru Economic* 20.47, 185–201.

Nadeau, David and Satoshi Sekine (2007). "A survey of named entity recognition and classification". In: *Lingvisticae Investigationes* 30.1, pp. 3–26.

Narayan, Paresh Kumar (2019). "Can stale oil price news predict stock returns?" In: *Energy Economics* 83, 430–444.

Narayan, Paresh Kumar and Deepa Bannigidadmath (APR 2017). "Does Financial News Predict Stock Returns? New Evidence from Islamic and Non-Islamic Stocks". In: *Pacific-Basin Finance Journal* 42, 24–45.

Nardo, Michela, Marco Petracco-Giudici, and Minás Naltsidis (2016). "Walking down wall street with a tablet: A survey of stock market predictions using the web". In: *Journal of Economic Surveys* 30.2, pp. 356–369.

Nassirtoussi, Arman Khadjeh, Saeed Aghabozorgi, Ying Wah Teh, and David Chek Ling Ngo (2015). "Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment". In: *Expert Systems with Applications* 42.1, 306–324.

Nassirtoussi, Arman Khadjeh, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo (2014). "Text mining for market prediction: A systematic review". In: *Expert Systems with Applications* 41.16, pp. 7653–7670.

NBER (2018). *Business Cycle Dating Committee, National Bureau of Economic Research*. URL: `https://www.nber.org/cycles/sept2010.html` (visited on 12/09/2018).

Newey, Whitney K and Kenneth D West (1994). "Automatic lag selection in covariance matrix estimation". In: *The Review of Economic Studies* 61.4, pp. 631–653.

– (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Nguyen, Thien Hai, Kiyoaki Shirai, and Julien Velcin (2015). "Sentiment analysis on social media for stock movement prediction". In: *Expert Systems with Applications* 42.24, 9603–9611.

Niederhoffer, Victor (1971). "The analysis of world events and stock prices". In: *The Journal of Business* 44.2, pp. 193–219.

Nofer, Michael and Oliver Hinz (2015). "Using Twitter to Predict the Stock Market Where is the Mood Effect?" In: *Business & Information Systems Engineering* 57.4, 229–242.

Nooijen, Steven J. and Simon A. Broda (2016). "Predicting Equity Markets with Digital Online Media Sentiment: Evidence from Markov-switching Models". In: *Journal of Behavioral Finance* 17.4, 321–335.

Nuij, Wijnand, Viorel Milea, Frederik Hogenboom, Flavius Frasincar, and Uzay Kaymak (APR 2014). "An Automated Framework for Incorporating News into Stock Trading Strategies". In: *IEEE Transactions on Knowledge and Data Engineering* 26.4, 823–835.

Ohta, Tomoko, Yuka Tateisi, and Jin-Dong Kim (2002). "The GENIA corpus: An annotated research abstract corpus in molecular biology domain". In: *Proceed-*

ings of the second international conference on Human Language Technology Research. Morgan Kaufmann Publishers Inc., pp. 82–86.

Palmer, David D and Marti A Hearst (1994). "Adaptive sentence boundary disambiguation". In: *Proceedings of the fourth conference on applied natural language processing*. Association for Computational Linguistics, pp. 78–83.

– (1997). "Adaptive multilingual sentence boundary disambiguation". In: *Computational Linguistics* 23.2, pp. 241–267.

Pang, Bo, Lillian Lee, et al. (2008). "Opinion mining and sentiment analysis". In: *Foundations and Trends® in Information Retrieval* 2.1–2, pp. 1–135.

Pástor, L'uboš and Robert F Stambaugh (2003). "Liquidity risk and expected stock returns". In: *Journal of Political Economy* 111.3, pp. 642–685.

Patterson, H Desmond and Robin Thompson (1971). "Recovery of inter-block information when block sizes are unequal". In: *Biometrika* 58.3, pp. 545–554.

Peng, Lin and Wei Xiong (2006). "Investor attention, overconfidence and category learning". In: *Journal of Financial Economics* 80 (3), pp. 563–602.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Pesaran, M Hashem, Yongcheol Shin, and Ron P Smith (1999). "Pooled mean group estimation of dynamic heterogeneous panels". In: *Journal of the American Statistical Association* 94.446, pp. 621–634.

Pesaran, M Hashem and Ron Smith (1995). "Estimating long-run relationships from dynamic heterogeneous panels". In: *Journal of Econometrics* 68.1, pp. 79–113.

Petersen, Mitchell A (2009). "Estimating standard errors in finance panel data sets: Comparing approaches". In: *The Review of Financial Studies* 22.1, pp. 435–480.

Pfaff, Bernhard (2008). "VAR, SVAR and SVEC Models: Implementation Within R Package vars". In: *Journal of Statistical Software, Articles* 27.4, pp. 1–32.

Philips, Thomas K., Emmanuel Yashchin, and David M. Stein (2003). "Using Statistical Process Control to Monitor Active Managers". In: *The Journal of Portfolio Management* 30.1, pp. 86–94.

Pineiro-Chousa, Juan, Marcos Vizcaino-Gonzalez, and Ada Maria Perez-Pico (2017). "Influence of Social Media over the Stock Market". In: *Psychology & Marketing* 34.1, 101–108.

Porter, Martin F (1980). "An algorithm for suffix stripping". In: *Program* 14.3, pp. 130–137.

Prechelt, Lutz (1998). "Early stopping-but when?" In: *Neural Networks: Tricks of the trade*. Springer, pp. 55–69.

Price, S McKay, James S Doran, David R Peterson, and Barbara A Bliss (2012). "Earnings conference calls and stock returns: The incremental informativeness of textual tone". In: *Journal of Banking & Finance* 36.4, pp. 992–1011.

Purda, Lynnette and David Skillicorn (2015). "Accounting variables, deception, and a bag of words: assessing the tools of fraud detection". In: *Contemporary Accounting Research* 32.3, pp. 1193–1223.

Pyo, Dong-Jin and Jungho Kim (2019). "News media sentiment and asset prices in Korea: text-mining approach". In: *Asia-Pacific Journal of Accounting & Economics*.

Qian, Edward E, Ronald H Hua, and Eric H Sorensen (2007). *Quantitative equity portfolio management: modern techniques and applications*. CRC Press.

Racine, Jeff (2000). "Consistent cross-validatory model-selection for dependent data: hv-block cross-validation". In: *Journal of Econometrics* 99.1, pp. 39–61. ISSN: 0304-4076.

Ranco, Gabriele, Darko Aleksovski, Guido Caldarelli, Miha Grcar, and Igor Mozetic (SEP 21 2015). "The Effects of Twitter Sentiment on Stock Price Returns". In: *PloS One* 10.9.

Ravi, Kumar and Vadlamani Ravi (2015). "A survey on opinion mining and sentiment analysis: tasks, approaches and applications". In: *Knowledge-Based Systems* 89, pp. 14–46.

Reynar, Jeffrey C and Adwait Ratnaparkhi (1997). "A maximum entropy approach to identifying sentence boundaries". In: *Proceedings of the fifth conference on applied natural language processing*. Association for Computational Linguistics, pp. 16–19.

Richardson, Matthew and Tom Smith (1991). "Tests of financial models in the presence of overlapping observations". In: *The Review of Financial Studies* 4.2, pp. 227–254.

Riley, Michael D (1989). "Some applications of tree-based modelling to speech and language". In: *Proceedings of the workshop on Speech and Natural Language*. Association for Computational Linguistics, pp. 339–352.

Ross, Stephen A (2009). *Neoclassical finance*. Princeton University Press.

Sabherwal, Sanjiv, Salil K Sarkar, and Ying Zhang (2011). "Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news". In: *Journal of Business Finance & Accounting* 38.9-10, pp. 1209–1237.

Sadik, Zryan A., Paresh M. Date, and Gautam Mitra (2019[a]). "News augmented GARCH(1,1) model for volatility prediction". In: *IMA Journal of Management Mathematics* 30.2, 165–185.

– (APR 2019[b]). "News augmented GARCH(1,1) model for volatility prediction". In: *IMA Journal of Management Mathematics* 30.2, 165–185.

Schmeling, Maik (2007). "Institutional and individual sentiment: Smart money and noise trader risk?" In: *International Journal of Forecasting* 23.1, pp. 127–145.

– (2009). "Investor sentiment and stock returns: Some international evidence". In: *Journal of Empirical Finance* 16.3, pp. 394–408.

Schmid, Helmut (1994). "Probabilistic part-of-speech tagging using decision trees". In: *International conference on new methods in language processing*, pp. 1–9.

Schumaker, Robert P and Hsinchun Chen (2009). "Textual analysis of stock market prediction using breaking financial news: The AZFin text system". In: *ACM Transactions on Information Systems (TOIS)* 27.2, p. 12.

Schumaker, Robert P, Yulei Zhang, Chun-Neng Huang, and Hsinchun Chen (2012). "Evaluating sentiment in financial news articles". In: *Decision Support Systems* 53.3, pp. 458–464.

Sebastiani, F (2002). "Machine learning in automated text categorization". In: *ACM Computing Surveys* 34.1, 1–47.

Seng, Jia-Lang and Hsiao-Fang Yang (2017). "The association between stock price volatility and financial news - a sentiment analysis approach". In: *Kybernetes* 46.8, 1341–1365. ISSN: 0368-492X.

Serrano-Guerrero, Jesus, Jose A Olivas, Francisco P Romero, and Enrique Herrera-Viedma (2015). "Sentiment analysis: A review and comparative analysis of web services". In: *Information Sciences* 311, pp. 18–38.

Shi, Yanlin, Kin-Yip Ho, and Wai-Man Liu (2016). "Public news arrival and the idiosyncratic volatility puzzle". In: *Journal of Empirical Finance* 37, 159–172.

Shiller, RJ (1981). "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" In: *American Economic Review* 71.3, 421–436.

Shiller, Robert J (1984). "Stock prices and social dynamics". In: *Brookings Papers on Economic Activity* 1984.2, pp. 457–510.

Shleifer, Andrei and Robert W Vishny (1997). "The limits of arbitrage". In: *The Journal of Finance* 52.1, pp. 35–55.

Shumway, Tyler (1997). "The Delisting Bias in CRSP Data". In: *The Journal of Finance* 52.1, pp. 327–340.

Shumway, Tyler and Vincent A. Warther (1999). "The Delisting Bias in CRSP's Nasdaq Data and Its Implications for the Size Effect". In: *The Journal of Finance* 54.6, pp. 2361–2379.

Siganos, Antonios, Evangelos Vagenas-Nanos, and Patrick Verwijmeren (2014). "Facebook's daily sentiment and international stock markets". In: *Journal of Economic Behavior & Organization* 107, pp. 730–743.

Silla, Carlos and Celso Kaestner (2004). "An analysis of sentence boundary detection systems for English and Portuguese documents". In: *Computational Linguistics and Intelligent Text Processing*, pp. 135–141.

Sims, Christopher A (1980). "Macroeconomics and reality". In: *Econometrica*, pp. 1–48.

Sinha, Nitish Ranjan (2016). "Underreaction to news in the US stock market". In: *Quarterly Journal of Finance* 6.02, p. 1650005.

Smales, Lee A. (2015[a]). "Asymmetric volatility response to news sentiment in gold futures". In: *Journal of International Financial Markets Institutions & Money* 34, 161–172.

– (2015[b]). "Time-variation in the impact of news sentiment". In: *International Review of Financial Analysis* 37, 40–50.

– (2016). "Time-varying relationship of news sentiment, implied volatility and stock returns". In: *Applied Economics* 48.51, 4942–4960.

Song, Qiang, Anqi Liu, and Steve Y Yang (June 2017). "Stock portfolio selection using learning-to-rank algorithms with news sentiment". In: *Neurocomputing*, pp. 20–28.

Souza, Thársis Tuani Pinto, Olga Kolchyna, Philip C Treleaven, and Tomaso Aste (2016). "Twitter sentiment analysis applied to finance: A case study in the retail industry". In: *The Handbook of Sentiment Analysis in Finance*. Ed. by Gautam Mitra, Yu Xiang, and En Tsao. Albury Books. Chap. 23.

Sprenger, Timm O., Andranik Tumasjan, Philipp G. Sandner, and Isabell M. Welpe (2014). "Tweets and Trades: the Information Content of Stock Microblogs". In: *European Financial Management* 20.5, 926–957.

Stambaugh, Robert F (1993). "Estimating conditional expectations when volatility fluctuates". In: *NBER Working Paper* t0140.

– (1999). "Predictive regressions". In: *Journal of Financial Economics* 54.3, pp. 375–421.

Stanford NLP Group (2017). *Stanford Tokenizer*. URL: `https://nlp.stanford.edu/software/tokenizer.html` (visited on 08/08/2017).

Stone, Philip J, Dexter C Dunphy, and Marshall S Smith (1966). "The general inquirer: A computer approach to content analysis." In:

Strauß, Nadine, Rens Vliegenthart, and Piet Verhoeven (2016). "Lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands". In: *Public Relations Review* 42.4, pp. 548–555.

Subrahmanyam, Avanidhar (1991). "Risk aversion, market liquidity, and price efficiency". In: *The Review of Financial Studies* 4.3, pp. 417–441.

Summers, Lawrence H. (1986). "Does the Stock Market Rationally Reflect Fundamental Values?" In: *The Journal of Finance* 41.3, pp. 591–601.

Sun, Andrew, Michael Lachanski, and Frank J. Fabozzi (2016). "Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction". In: *International Review of Financial Analysis* 48, 272–281.

Tauchen, George E and Mark Pitts (1983). "The price variability-volume relationship on speculative markets". In: *Econometrica*, pp. 485–505.

Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy (2008). "More than words: Quantifying language to measure firms' fundamentals". In: *The Journal of Finance* 63.3, pp. 1437–1467.

Tetlock, Paul C. (Sept. 2010). "Does Public Financial News Resolve Asymmetric Information?" In: *Review of Financial Studies* 23 (9), pp. 3520–3557.

– (2011). "All the News That's Fit to Reprint: Do Investors React to Stale Information?" In: *Review of Financial Studies* 24.5, 1481–1512.

– (2007). "Giving content to investor sentiment: The role of media in the stock market". In: *The Journal of Finance* 62.3, 1139–1168.

Thompson, Samuel B (2011). "Simple formulas for standard errors that cluster by both firm and time". In: *Journal of Financial Economics* 99.1, pp. 1–10.

Thomson Reuters (2015). *Exploit Signals in News for Quantitative Strategies and Systematic Trading: Turning Unstructured Data into Actionable Insights to Maximize Returns and Spot Event Risk*. URL: `https://financial.thomsonreuters.com/content/dam/openweb/documents/pdf/financial/exploit-signals-in-news-for-quantitative-strategies.pdf` (visited on 09/09/2017).

Tomanek, Katrin, Joachim Wermter, and Udo Hahn (2007). "Sentence and token splitting based on conditional random fields". In: *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, pp. 49–57.

Tumarkin, Robert and Robert F Whitelaw (2001). "News or noise? Internet postings and stock prices". In: *Financial Analysts Journal* 57.3, pp. 41–51.

Uhl, Matthias W (2014). "Reuters Sentiment and Stock Returns". In: *Journal of Behavioral Finance* 15.4, 287–298.

Uhl, Matthias W, Mads Pedersen, and Oliver Malitius (2015). "What's in the News? Using News Sentiment Momentum for Tactical Asset Allocation". In: *Journal of Portfolio Management* 41.2, 100–112.

Uhl, Matthias W. (2018). "Volatility Aversion in the Options Market Based on News Sentiment". In: *Journal of Derivatives* 25.4, 24–35.

Valkanov, Rossen (2003). "Long-horizon regressions: theoretical results and applications". In: *Journal of Financial Economics* 68.2, pp. 201–232.

Vanstone, Bruce J and Tobias Hahn (2017). "Australian momentum: performance, capacity and the GFC effect". In: *Accounting & Finance* 57.1, pp. 261–287.

Vanstone, Bruce James, Adrian Gepp, and Geoff Harris (2019). "Do news and sentiment play a role in stock price prediction?" In: *Applied Intelligence* 49.11, 3815–3820.

Veronesi, Pietro (1999). "Stock market overreactions to bad news in good times: a rational expectations equilibrium model". In: *The Review of Financial Studies* 12.5, pp. 975–1007.

Wahba, Grace (1980). "Spline bases, regularization, and generalized cross validation for solving approximation problems with large quantities of noisy data". In: *Proceedings of the International Conference on Approximation theory in Honour of George Lorenz*. Austin, Texas. Academic Press, pp. 905–912.

Wang, Jiang (Feb. 1994). "A Model of Competitive Stock Trading Volume". In: *Journal of Political Economy* 102 (1), pp. 127–168.

Wang, Kuei-Yuan, Chien-Kuo Chen, and Hsiao-Chi Wei (2015). "Media Coverage and Stock Return in the Taiwan Stock Market". In: *Acta Oeconomica* 65.2, 35–53.

Wei, Yu-Chen, Yang-Cheng Lu, Jen-Nan Chen, and Yen-Ju Hsu (2017). "Informativeness of the market news sentiment in the Taiwan stock market". In: *North American Journal of Economics and Finance* 39, 158–181.

White, Halbert (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". In: *Econometrica*, pp. 817–838.

Wiley (2021). *How to Clear Permissions for a Thesis or Dissertation*. URL: `https://www.wiley.com/network/researchers/latest-content/how-to-clear-permissions-for-a-thesis-or-dissertation` (visited on 07/01/2021).

Wood, S. N. (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models”. In: *Journal of the Royal Statistical Society (B)* 73.1, pp. 3–36.

– (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.

Wu, Chen-Hui and Chan-Jane Lin (JUN 2017). “The impact of media coverage on investor trading behavior and stock returns”. In: *Pacific-Basin Finance Journal* 43, 151–172.

Wüthrich, Beat, D Permunetilleke, Steven Leung, W Lam, Vincent Cho, and J Zhang (1998). “Daily prediction of major stock indices from textual www data”. In: *HKIE Transactions* 5.3, pp. 151–156.

Wysocki, Peter (1999). “Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards”. Working paper.

Xiong, Guiyang and Sundar Bharadwaj (DEC 2013). “Asymmetric Roles of Advertising and Marketing Capability in Financial Returns to News: Turning Bad into Good and Good into Great”. In: *Journal of Marketing Research* 50.6, 706–724.

Yang, Steve Y, Sheung Yin Kevin Mo, Anqi Liu, and Andrei A Kirilenko (June 2017). “Genetic programming optimization for a sentiment feedback strength based trading strategy”. In: *Neurocomputing*, S0925231217311086–.

Yang, Steve Y, Qiang Song, Sheung Yin K Mo, Kaushik Datta, and Anil Deane (2015). “The impact of abnormal news sentiment on financial markets”. In: *Journal of Business and Economics* 6.10, pp. 1682–1694.

Yang, Steve Y., Sheung Yin Kevin Mo, and Anqi Liu (2015). “Twitter financial community sentiment and its predictive relationship to stock market movement”. In: *Quantitative Finance* 15.10, SI, 1637–1656.

Yashchin, Emmanuel, Thomas K Philips, and David M Stein (1997). “Monitoring active portfolios using statistical process control”. In: *Computational Approaches to Economic Problems*. Springer, pp. 193–205.

Yerkes, Rustin (2011). “Regulatory Trading Restrictions, Overvaluation, and Insider Selling”. In: *The Journal of Trading* 7.3, pp. 85–102.

Yu, Yang, Wenjing Duan, and Qing Cao (2013). “The impact of social and conventional media on firm equity value: A sentiment analysis approach”. In: *Decision Support Systems* 55.4, SI, 919–926.

Zeileis, Achim (2004). “Econometric Computing with HC and HAC Covariance Matrix Estimators”. In: *Journal of Statistical Software, Articles* 11.10, pp. 1–17.

Zhang, Junni L, Wolfgang K Härdle, Cathy Y Chen, and Elisabeth Bommes (2016). "Distillation of news flow into analysis of stock reactions". In: *Journal of Business & Economic Statistics* 34.4, pp. 547–563.

Zhang, Xi, Yunjia Zhang, Senzhang Wang, Yuntao Yao, Binxing Fang, and Philip S. Yu (2018). "Improving stock market prediction via heterogeneous information fusion". In: *Knowledge-Based Systems* 143, 236–247.

Zhang, Ying, Peggy E Swanson, and Wikrom Prombutr (2012). "Measuring effects on stock returns of sentiment indexes created from stock message boards". In: *Journal of Financial Research* 35.1, pp. 79–114.

Zhang, Yongjie, Weixin Song, Dehua Shen, and Wei Zhang (AUG 2016). "Market reaction to internet news: Information diffusion and price pressure". In: *Economic Modelling* 56, 43–49. ISSN: 0264-9993.

Zheludev, Ilya, Robert Smith, and Tomaso Aste (2014). "When can social media lead financial markets?" In: *Scientific Reports* 4, p. 4213.

Zhou, Guofu (2010). "How much stock return predictability can we expect from an asset pricing model?" In: *Economics Letters* 108.2, pp. 184–186.

Zhou, Zhongbao, Meng Gao, Qing Liu, and Helu Xiao (2020). "Forecasting stock price movements with multiple data sources: Evidence from stock market in China". In: *Physica A-Statistical Mechanics and its Applications* 542.