Cambridge University Land Society

ALUMNI

CULS Working Paper Series

No. 03 (2016)

Information Supply and Demand in Securitized Real Estate Markets

BY

JAN-OTTO JANDL, UNIVERSITY OF FREIBURG

FRANZ FUERST, CULS FELLOW, UNIVERSITY OF CAMBRIDGE

Information Supply and Demand in Securitized Real Estate Markets

Jan-Otto Jandl[a,*], Franz Fuerst[b]

[a] *Institute of Public Finance, University of Freiburg, Bertoldstrasse 17, Freiburg, 79098, Germany*
[b] *Department of Land Economy, University of Cambridge, 19 Silver Street, Cambridge, CB3 9EP, United Kingdom*

Abstract

This study introduces a novel approach to explaining the short-term market movements of Real Estate Investment Trusts (REITs) utilizing Big Data Analytics. While literature provides several explanations for the variation of industry returns, there has been little empirical evidence of whether these patterns are attributable to the flow of information. We extract news sentiment as a proxy for information supply based on agency news and additionally investigate web search queries as an indicator for information demand. Analyzing REIT markets in the UK and US indicates a consistent pattern across both countries. While we observe a strong positive effect of news sentiment on REIT returns and stock market volatility, the effect of online search behavior is only marginal. Further, we exhibit that particularly finance-specific news sentiment measures significantly contribute to explaining NAV spreads. Hence, the application of different metrics for information supply and demand for the UK and the US yields diverse impacts on developments in securitized real estate markets.

Keywords: Behavioral Finance, Big Data Analytics, Net Asset Value, News Sentiment, Online Search Behavior

*Corresponding author
Email addresses: jan-otto.jandl@vwl.uni-freiburg.de (Jan-Otto Jandl), ff274@cam.ac.uk (Franz Fuerst)

1. Introduction

It has long been posited that animal spirits constitute an important determinant for the formation of investor beliefs [36]. Beliefs about future cash flow and investment risk that are not justified by the facts at hand are commonly regarded as investor sentiment [2]. As those beliefs constitute noise rather than information [6], trading on the basis of investor sentiment drives stock prices away from their fundamental value. The proposition that media significantly influences investor sentiment and accelerated the upswings in stock and real estate markets in the recent decade appears consistent with this notion [24].

Prior work in the field of behavioral finance stresses the relevance of psychology for investment decisions under uncertainty [34]. Modeling the mispricing process caused by irrational traders, noise trader theory emphasizes that price deviations from fundamental values might persist in the long-run [18]. Early empirical studies disclose that the variation of investor sentiment over time affects stock returns [3]. However, the estimation of sentiment remained a major challenge until recently.

Real Estate Investment Trusts (REITs) provide real estate investments which are traded in the stock market. Similar to closed-end funds, REITs are initialized through capital from external markets and invest the capital in operating assets. However, REITs are required to focus on real estate assets and have to distribute the majority of income to shareholders in order to qualify as a REIT under tax law and subsequently exhibit tax deductions. It has been observed that stock prices of REITs typically vary around their respective net asset value (NAV) per share. The NAV reflects the difference between the estimated market value of assets and the liabilities of a REIT and, thus, should correspond to the stock price. Hence, the empirical observation of NAV spreads highlights inefficiencies in REIT markets.

Prior work elaborates on various potential causes for the cyclical fluctuations of REIT share prices around their fundamental value. While empirical evidence on rational explanations such as company-specific factors yields heterogenous insights, literature increasingly stresses irrational factors as crucial factors for the existence of NAV spreads in REIT markets [51]. However, estimations of information supply and demand as decisive marekt components have not been addressed yet. Therefore, this study aims at narrowing the gap in the understanding of NAV spreads by applying Big Data Analytics. More precisely, we estimate news sentiment as a proxy for information supply and online search behavior as a proxy for information demand. The principal contribution of this study is to disclose the different effects of information in the real estate domain. While prior work incorporating text mining has extensively focused on the stock market in general [1, 24, 54, 55], literature on Real Estate Investment Trusts (REIT) has been limited in this regard. Hence, this study is motivated by the increasing potential of Big Data analytics and the lack of literature proposing a "behavioral story" [24] as possible explanation for cyclical movements in stock markets.

The remainder of this paper is structured as follows. Section 2 summarizes the relevant literature on the net asset value spread and Big Data analytics. The research methodology and sample data is described in section 3. Section 4 describes the findings for the hypothesized relationship between information and industry returns. Section 5 focuses on the relation between information variables and NAV spreads. Section 6 extends the findings for stock market volatility. Section 7 concludes.

2. Related Work

This study extends prior work on the NAV spread of REITs and the research stream of Big Data Analytics. First, we summarize the related research on the NAV spread particularly focusing on the insights for REIT markets. While literature provides heterogeneous results for most possible explanations, empirical evidence highlights that investor sentiment posits a major determinant for NAV spreads. Nevertheless, studies investigating investor sentiment mainly build upon proxies such as capital market conditions or surveys. There do not seem to exist empirical investigations on the REIT NAV spread incorporating sentiment analysis of financial news. This is even more surprising as sentiment analysis has significantly gained influence and may provide further insights into important economic problems. Hence, we additionally summarize the Big Data Analytics literature related to the instruments applied in this study.

2.1. The Net Asset Value Spread

The closed-end fund puzzle constitutes a challenging phenomenon in finance. In theory, the NAV, thus the difference between the estimated market value of the asset and the liabilities, should reflect the market value of a fund. However, closed-end funds stocks typically deviate from their per share NAV. In a seminal study, [37] divide the closed-end funds puzzle into four major determinants. While (1) new funds are observed to be traded at a premium to NAV, those (2) initial premiums tend to decrease to discounts shortly after the initial public offering (IPO). Subsequently, (3) discounts exhibit significant fluctuations over time which appear to be stationary. Finally, (4) discounts typically drop when the termination of a fund is announced. A thorough review on closed-end funds and possible explanations for the closed-end funds puzzle are provided by [20]. Analogous to closed-end funds, stock prices of REITs exhibit cyclical deviations from their respective NAV per share. Predominant solutions for explaining the third part of the puzzle, that is fluctuations in NAV spreads, can be divided into the rational and the irrational approach [51]. Rational explanations commonly argue that company-specific factors lead to the divergence between market capitalization and NAV. Single factor explanations range from company size [3, 12, 15], sectoral focus [11], to informational transparency [11]. However, thorough empirical work incorporating multiple factors is rare [3, 12, 15].

Although the presented studies provide partial evidence for the influence of company-specific factors, the results are ambiguous and rather invariant. Additionally, the time-variation in NAV spreads suggests explanation beyond unsystematic risk factors which might underline behavioral biases [25]. The irrational approach addresses an explanation beyond an idiosyncratic nature.

Accordingly, the cyclical deviations of stock prices from the respective NAV per share are argued to be due to the impact of irrational investment behavior inherent in the absence of informational efficiency in stock markets [51]. Prior studies on NAV spreads in REIT markets approximate sentiment using the average sector spread [3, 15] or multiple factors to construct a sentiment index [2, 51]. [3] are among the first to address the role of sentiment for listed property companies. They first investigate the impact of a number of company-specific factors on individual NAV spreads. Accordingly, they identify a significant positive influence of capital gains tax liabilities and company size and negative impacts of high historic returns and stock holding intensity on firm-specific NAV discounts. While capital gains tax liabilities and company size are found to increase firm-specific NAV discounts, high historic returns and the intensity of stock holdings reduce discounts. However, the explanatory power of the model more than doubles if the sector average NAV spread is additionally considered. They assume that the time variation of market sentiment highly influences NAV spreads and that the sector average NAV spread serves best as a proxy for market sentiment.

Hence, the proxy for market sentiment is found to be by far the most significant determinant explaining a substantial portion of NAV spread developments. Further empirical tests stress the importance of sentiment variables. Most notably, inflation expectations and industrial confidence as economy-wide indicators for industrial and investor sentiment exhibit close correlations with average sector discounts. Hence, [3] conclude that sentiment constitutes the key determinant for NAV spreads. Likewise, [15] conduct a cross-sectional analysis and investigate the determinants of NAV spreads. Amongst a number of company-specific factors under consideration, REIT size and historical volatility are found to be the most relevant determinants. While return volatility as an indicator for larger risk commonly reduces the premium, market capitalization as a proxy for management quality and growth opportunities increases the premium. Analogous to the findings of [3], adding the common sector average premium significantly improves the explanatory power of the model [15]. Hence, a common sector effect dominates in explaining the cyclical NAV spreads.

Additional tests highlight that irrational trading behavior is particulary pronounced when NAV spreads increase. This hints at a higher portion of noise traders in the market when REIT prices are diverging from NAV. While the latter studies postulate that the average sector discount reflects some kind of economy-wide sentiment, [2] conduct a multivariate approach to explicitly quantify investor sentiment. Following a discussion on several potential measures for investor sentiment, a sentiment index is constructed based on the dividend premium, NAV spread, number and first day return of IPOs and equity share in new issues. Although the paper focuses on the stock market and does not explicitly address REIT markets, the study highlights that various capital market characteristics are potentially linked to investor sentiment. [51] are the first to include a latent "market sentiment" variable based on weekly internet polls. Analogous to previous studies, the results stress that rational economic determinants, in particular company type and stock volatility, just explain a minor portion of the divergences of REIT stock prices from their NAV. Additionally considering the latent market sentiment variable significantly increases the coefficient of determination of the structural equation model. While the prior semi-rational approaches to explaining the NAV spread provide partial evidence for the important role of sentiment, the role of media in REIT markets has not been adressed yet.

2.2. Big Data Analytics

Big Data and the advancement of Information Systems research methods have paved the way for a burgeoning stream of literature on sentiment analysis. Sentiment analysis generally includes a text corpus that is preprocessed by linguistic tools and transformed into sentiment scores using various approaches. As this method extracts subjective information from textual content, it is often referred to as sentiment analysis or opinion mining [48]. Big Data and text analytics constitute key research frontiers in IS research that induced a growing research stream on sentiment analysis [13].

Sentiment analysis in finance literature has been conducted based on various sources of information. [1] and [54] provide first insights into the dynamics between textual content and stock market behavior. [1] employ a Naive Bayes classifier and analyze the content of internet stock message boards and corresponding stock market reactions. While differences of opinion typically increase trading volume and message content helps predict volatility, the impact on stock returns is only marginal. Using a dictionary-based approach, [54] analyzes Wall Street Journal content for the years 1984 to 1999 and shows that the column-inherent sentiment is a suitable indicator for daily stock returns. Additionally, [55] investigate firm-specific news from the Walls Street Journal and Dow Jones News Service for the period between 1980 and 2004 and find that the negative content of texts forecasts stock market prices. Consequently, they conclude that "words contained in news stories are not redundant information, but instead capture otherwise hard-to-quantify aspects of firms' fundamentals" [55]. Hence, public news provides valuable information that can be utilized in trading strategies [22]. Accordingly, [29] exhibit that text classifications such as context-capturing features improve the predictive power of news content. Recent literature on daily news and stocks for the 20th century stresses media's influences on asset prices, which is especially profound during recessions [24].

Besides media content, IS research increasingly harnesses user-generated content (UGC). [7] investigate the relation between Twitter content and the Dow Jones Industrial Average (DJIA). Using established mood tracking tools, they find significant correlations between some measures for public mood and DJIA returns. Further, rather simple text processing techniques are quite powerful in extracting the public mood from large-scale Twitter feeds. [58] show how information aggregation on Wikipedia influences the voluntary disclosure behavior of public-firm management. Hence, collective information aggregation improves the informational environment in financial markets. In line with this notion, empirical evidence stresses that "wisdom of crowds" outperforms professional analysts judgements [47]. Accordingly, [27] propose a decision support system utilizing individual stock price predictions content from virtual investing communities and identify profitable trading strategies.

In addition to the variety of information sources, sentiment analysis of financial news is applied using diverse methods of content analysis. Hence, not only the increasing availability of huge amounts of information but also the burgeoning literature on advanced IS research methods increase the relevance of sentiment analysis [35]. Comparing the literature on textual sentiment in the financial domain, the most common methods of content analysis are dictionary based and machine-learning methods [35].

The dictionary-based approach uses pre-defined dictionaries to identify positive and negative words. The Harvard-IV General Inquirer Psychological Dictionary (GI) has early been developed by [53] and has extensively been used in literature [22, 54, 55]. However, the GI has been developed for the social psychology domain and therefore provides a rather general dictionary, which does not cater for the peculiarities of the finance discipline. Therefore, domain-specific lexica have been created endogenously through the corpus-based approach. Generally, the compilation of a domain-specific lexicon bases on the analysis of a large corpus of domain-specific documents by corpus-based algorithms, which typically parse the sentences and identify the associated sentiment expressions. Frequently used finance-specific dictionaries include Henry's Finance-Specific Dictionary (HE) [30, 31] and the Loughran and McDonald Financial Sentiment Dictionary (LM) [41]. The HE is based on an analysis of numerous earning press releases and provides a powerful context-specific word-list [31, 35].

[41] examine words which occur at in at least 5 % of a sample of 10-K reports published between 1994 and 2008 and compile the finance-specific LM dictionary containing 353 positive and 2,337 negative words [24, 32]. Comparisons of dictionaries emphasize that the power of finance-specific word lists is much higher than general dictionary [31, 35]. The application of dictionary-based approaches is mostly accompanied by the presumption that each word is weighted equally. Therefore, term-weighting methods have been established to remove significant bias from equal weights [32, 41].

The machine learning approach uses algorithms to extract and classify the relevant content in texts [39]. More precisely, a training set of the text corpus is classified into sentiment metrics using specific algorithms. Commonly used algorithms include Naive Bayes, Support Vector Machines and K-Nearest Neighbor [35]. Those algorithms produce classification rules that can be applied to out-of-sample text corpus. However, machine learing techniques may suffer from overfitting [52].

In order to appropriately analyze the impact of news sentiment, the attention of investors to the respective content has to be taken into account. Prior empirical studies have stressed either, the importance of the interest of investors to stock-related content and appropriate measures to quantify the latter.

Amongst the measures, Google search query data is increasingly found to impact corresponding financial markets. Accordingly, web search behavior seems to account for the attention of investors in a much more timely manner and seems to be applicable for predictive studies[17, 50]. Further, [56] point out that Google Trends data fits as a powerful measure for information demand.

We contribute to the existing literature in several ways. First, we analyze a unique data set of UK newspaper articles, which has not been analyzed in depth and arguably captures subjective information about fundamentals. Second, we apply several methods of content analysis to provide additional insights into information processing in financial markets. In addition, we simultaneously quantify information demand and information supply and treat them as distinct variables since search indices have not been found to correlate with news sentiment in the literature[16].

3. Methodology and Data

We conduct the analysis of the NAV spread phenomenon in real estate stocks analyzing information demand and supply as outlined in in Figure 1. Therefore, we describe the derivation of the news sentiment measures as a proxy for information supply at first. Following the corpus retrieval and the preprocessing phase, the resulting machine-readable tokens are arranged within a term-by-document-matrix (TDM) and serve as a starting point for sentiment analysis. Subsequently, we describe the methodology of the dictionary-based approaches applied in this study and provide the respective sample statistics. Second, we elaborate on the derivation of search volume index (SVI) as an indicator for information demand. Subsequently, we describe on the calculation and the derivation of the NAV spread data and identify potential factors which might additionally influence the cyclical behavior of NAV spreads.

That is, we present the data for prominently used control variables and provide the descriptive statistics for stock market data. Concluding, we present the econometric setup applied in this study and provide first parsimonious evidence for the relevance of information variables in real estate-related stock markets.

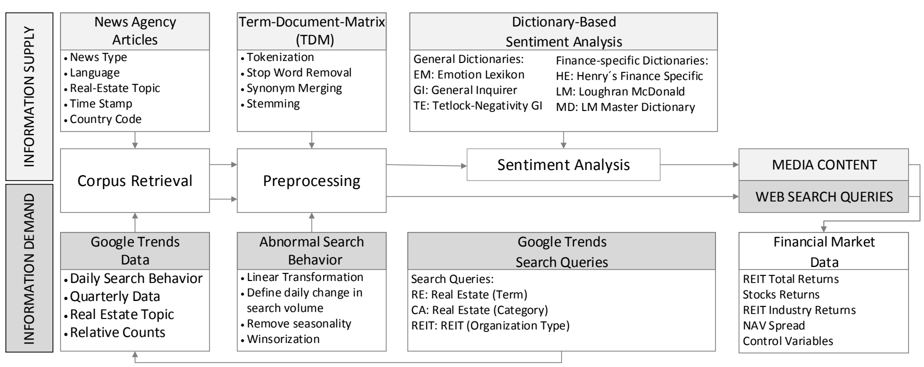

Figure 1: Research Methodology

3.1. News Sentiment

We base our analysis on a large text corpus of agency news covering country-specific newswires for the United States and the United Kingdom for the time period from January 1, 1990 until December 31, 2014. We gather the data from LexisNexis and extract real estate-related articles from news wires from Associated Press Publications including The Associated Press (National wire), The Associated Press State & Local Wire, AP Online and AP Worldstream. The Associated Press (AP) is a multinational news agency collecting worldwide third-party news on a wide range of topics. While the National as well as the State & Local Wire provide country-specific news for the United States, AP Online and AP Worldstream comprise the top national and international financial and business news from around the world. We concentrate on news in English language with direct relevance to the respective real estate markets by applying topic-specific and country-specific filters in LexisNexis.

We filter almost 80 million actual words from 157,856 newspaper articles concerned with the US real estate market. Accordingly, around 26 real estate related articles are published each day and contain more than 500 words on average. We collect 7619 news articles on the UK real estate market containing 4,788,233 words. Hence, roughly 4 news articles containing more than 625 words on average are published each day.

The sentiment analysis procedure applied in this study converts news content into quantifiable measures using a dictionary-based approach. News content is processed through tokenization, negation inversion, stop word removal, synonym merging and stemming. Subsequently, the occurrences of words are summarized within a term-by-document matrix (TDM) [43]

We split sentences using a static list of abbrevations which incorporates misleading punctuations [28] and subsequently segment the text into single words using spaces. In a second step, stop words are removed to exclude words without specific meanings [43] using a commonly used static list by [38]. During the third step, words conveying a similiar meaning are merged. We apply the procedure referred to as pseudoword generation and use around 150 synonyms from the finance domain which are aggregated into groups with similar meanings [43]. Further, closely related words are consolidated using stemming. We use the Porter stemmer [49] as described by [43] which is widely used in economic literature [26, 33]. Given all the announcements of a day, the TDM maps the text into a bag of words by considering the frequencies of each word. Annotations may be attached to specified snippets of the text such as whole documents, individual sentences and specific aspects of entities. We incorporate dictionary-based as well as machine learning techniques to quantify the tone of the news announcements. The dictionary-based approach is a simple application of text mining and is predominantly used in recent financial text mining research [24, 35, 58]. As standard psychological dictionaries typically do not refer to the financial terminology and therefore do not seem to be applicable in the finance context, we use several dictionaries in order to cater for differences in the perception of the written word. In particular, we use the Harvard-IV General Inquirer Psychosocial Dictionary (GI) [53] and additionally incorporate the Emotion Lexicon (EM) [45] and the MPQA Subjectivity Lexicon (MP) [57] to cover dictionaries with the focus on psychology. The GI has been developed in the academic fields of psychology in the 1960s to enhance computer-assisted content analyses of textual data [53]. The dictionary attaches information to words of the Harvard-IV dictionaries. From several GI dictionaries available, we use the positive and negative wordlists containing 1,915 and 2,291 words respectively. Additionally, we use Henry's Finance-Specific Dictionary (HE) [30] and Loughran and McDonald Financial Sentiment Sentiment Dictionary (LM) [41]. The HE and LM dictionaries contain finance-specific word lists have been determined endogeneously by analyzing financial news. While the HE has been developed on the basis of earning press releases, the LM originates from is based on 10-K reports. The word lists applied in this study are summarized in Table 1.

Let w_{it} be amount of words in newspaper i on date t with $p_i t^d$ representing the respective positive and $n_i t^d$ the negative word count using dictionary d.

In addition, we calculate daily measures of positive P_t^d and negative N_t^d media content reflecting the portion of positive and negative words in the news. We count positive words and normalize the measure according to

Table 1: Dictionary Sources

Lexicon	Abbr.	Positive Words	Negative Words	Source	Reference
Panel A: Psychologic Word Lists					
Emotion Lexicon	**EM**	1944	2786	Amazon's Mechanical Turk	[53]
Harvard-IV Psychological Dict.	**GI**	1684	2087	Harvard General Inquirer	[45]
Subjectivity Lexicon	**SU**	1423	2467	Amazon's Mechanical Turk	[57]
Panel B: Finance-Specific Word Lists					
Henry's Finance-Specific Dict.	**HE**	59	47	Annual press releases	[30]
Loughran and McDonald Dict.	**LM**	151	901	10-K reports from EDGAR	[41]
Loughran and McDonald Master Dict.	**MD**	152	911	10-K reports from EDGAR	[41]

$$POS_t^d = \frac{\sum p_{it}^d}{\sum w_{it}} \quad \epsilon \, [0,1] \tag{1}$$

and proceed analogously for negative words. More precisely, based on negative wordlist of the GI dictionary, [54] introduces a measure for investor sentiment defined as

$$NEG_t^d = \frac{\sum n_{it}^d}{\sum w_{it}} \quad \epsilon \, [0,1] \tag{2}$$

Hence, we present sentiment metrics for Negativity (NEG), Positivity (POS) and Optimism (OPT). The latter alternative sentiment metric incorporates positive as well as negative words from the dictionary d adopted. The Optimism score [19] is defined as the difference between the portion of positive and negative media content as

$$OPT_t^d = \frac{\sum p_{it}^d - \sum n_{it}^d}{\sum w_{it}} \quad \epsilon \, [-1,+1] \tag{3}$$

Finally, we average the variables for news sentiment to account for non-consecutive trading days as proposed by [24]. That is, we aggregate the news content available prior to market opening. While the bulk of our news data matches the trading days in our sample, we identify news content in 381 days during which the market was closed. Consequently, we define the news sentiment measure Optimism as

$$OPT_t^d = \frac{\sum_{i,s=t}^{s=t+h} p_{is}^d - \sum_{i,s=t}^{s=t+h} n_{is}^d}{\sum_{i,s=t}^{s=t+h} w_{is}}, \tag{4}$$

with h non-trading days such that $h > 0$ and respective two non-consecutive trading days t and $t + h + 1$. The Negativity and Positivitiy measures are adjusted accordingly.

Table 2: Sample Statistics for News Sentiment

	Mean	Std. Dev.	25%-Quan.	Med.	75%-Quan.	Mean	Std. Dev.	25%-Quan.	Med	75%-Quan.
	Panel A: United Kingdom					Panel B: United States				
Opt. (EM)	2.77	3.51	0.48	2.73	4.93	4.00	4.00	1.45	4.01	6.55
Opt. (GI)	2.63	2.00	1.44	2.63	3.81	3.23	2.45	1.78	3.19	4.65
Opt. (SU)	1.93	2.48	0.34	1.90	3.43	2.60	2.79	0.88	2.60	4.30
Opt. (HE)	0.78	1.36	0.00	0.71	1.53	0.72	1.17	0.00	0.64	1.28
Opt. (LM)	-2.02	1.75	-3.10	-1.93	-0.81	-2.00	2.08	-3.13	-1.69	-0.60
Opt. (MD)	-2.02	1.75	-3.10	-1.93	-0.81	-2.00	2.08	-3.13	-1.69	-0.60

Table 2 reports the sample statistics for the news sentiment measures. On average 2500 (12,929) words are processed daily in real estate related news articles. We divide the sentiment metrics into general psychological (EM, GI, SU) and finance-specific (HE, LM, MD) measures. We only report the difference of the proportion of positive and negative words of different dictionaries. We particularly focus on the optimism measures as a large body of prior work stresses not only the relevance of the negativity-bias, which proposes a particularly strong reaction of individuals to negative information [4], but also the importance of positive connotations in financial markets [19].

As reported, news articles on average contain more positive than negative words from the different psychological dictionaries. We note slight differences in the respective sentiment measures across both countries. The application of the finance-specific dictionaries yields diverse results which are consistent across the UK and the US. While the HE dictionary exhibits the proportion of positive words outweighing the negative words by roughly 1 percentage point, the application of the LM dictionary identifies much more negative than positive words. That is, the HE and LM measures can be regarded as distinct sentiment measures. As reported in Table 2, we exhibit a remarkably consistent pattern across the measures and both countries.

3.2. Web Search Behavior

In order to capture the attention paid by individuals to real estate-related information, we examine web search behavior on Google. Google Trends offers data including search volume indexes (SVI) on various search terms, cagegories and topics.

We examine three different real estate-related SVI. We use the SVI for the search term “Real Estate”, the SVI for the predefined category “Property” and the SVI for the organization type “REIT” as defined by Google. The SVI are transformed into daily data using the procedure proposed by [17] leading to the daily change in search volume for search term, category or organization type j in the form of

$$\Delta SVI_{j,t} = ln(SVI_{j,t}) - ln(SVI_{j,t-1}). \tag{5}$$

Table 3 provides the summary statistics the gathered SVI.

Table 3: Sample Statistics for Online Search Behavior

	Mean	Std. Dev.	25%-Quan.	Med.	75%-Quan.	Mean	Std. Dev.	25%-Quan.	Med	75%-Quan.
	Panel A: United Kingdom					Panel B: United States				
SVI (RE Terms)	35.35	17.41	20.35	27.93	45.70	53.73	19.85	34.87	47.85	72.64
SVI (RE Category)	83.65	11.90	78.08	78.69	94.57	99.55	11.74	93.25	97.61	107.56
SVI (REIT)	39.43	14.88	24.81	33.63	54.15	44.00	12.46	35.22	38.47	51.71

We follow [17] and calculate a Abnormal change in SVI (ASVI) which represents a standardized, winsorized and deseasonalized change in daily search behavior.

3.3. Financial Market

Stock market data for UK and US REITs is based on the FTSE EPRA/ NAREIT Global Real Estate Index Series data provided by the European Public Real Estate Association (EPRA) in collaboration with the Financial Times Stock Exchange (FTSE) and the National Association of Real Estate Investment Trusts (NAREIT). Information on business cycles is gathered from the National Bureau of Economic Research (NBER). Summary statistics for financial market variables are provided in Table 4.

Table 4: Sample Statistics for Financial Market Variables

	Mean	Std. Dev.	25%-Quan.	Med.	75%-Quan.	Mean	Std. Dev.	25%-Quan.	Med	75%-Quan.
	Panel A: United Kingdom					Panel B: United States				
REIT Index	1577.14	933.49	945.69	1198.59	1970.96	2078.95	1441.77	950.13	1498.68	3357.79
Stock Index	4878.75	1293.99	3850.60	5186.20	5963.80	1593.40	748.99	983.56	1663.06	2023.33
NAV Spread	-15.91	11.84	-23.56	-16.23	-7.32	4.16	12.65	-4.36	4.41	12.91

3.4. Information Flow and REIT Returns

The estimates for pairwise correlations between the information measures s highlight a significant correlation among the SVI variables. While the SVI for real estate terms and category are highly interrelated in both countries, the correlation between the SVI for REITs and real estate terms is rather low in terms of significance and magnitude. This pattern is consistent across both countries. We additionally note that both media metrics, the search query and media content variables, seem to be distinct information measures. Accordingly, we are not able to identify a significant correlation between any SVI and news sentiment.

While sentiment measures derived from psychological dictionaries are closely related, the correlations with respective measures containing finance-specific content seem rather low. The correlation matrix leads to the preliminary finding that the estimates for information demand and supply serve as distinct measures.

Table 5: Correlation Matrix

	Search Queries			Media Content					
	ASVI Terms	ASVI Category	ASVI REIT	Sent. (EM)	Sent. (GI)	Sent. (SU)	Sent. (HE)	Sent. (LM)	Sent. (MD)
				United Kingdom					
ASVI	–			0.018	0.000	-0.025	-0.022	-0.005	-0.005
Search Terms				(0.355)	(0.998)	(0.179)	(0.253)	(0.790)	(0.790)
ASVI	0.0478	–		-0.006	-0.002	0.004	-0.006	-0.015	-0.015
Category	(0.002)			(0.765)	(0.933)	(0.827)	(0.752)	(0.429)	(0.429)
ASVI	0.006	0.311	–	-0.006	-0.001	-0.008	-0.001	-0.033	-0.033
REITs	(0.729)	(0.000)		(0.771)	(0.939)	(0.690)	(0.974)	(0.079)	(0.079)
				United States					
ASVI	–			0.060	0.049	0.028	0.030	-0.029	-0.029
Search Terms				(0.002)	(0.010)	(0.143)	(0.111)	(0.121)	(0.123)
ASVI	0.468	–		-0.010	-0.007	-0.003	-0.016	-0.035	-0.035
Category	(0.000)			(0.590)	(0.725)	(0.893)	(0.411)	(0.067)	(0.067)
ASVI	0.037	0.1901	–	-0.004	0.005	0.008	0.001	0.021	0.021
REITs	(0.019)	(0.000)		(0.855)	(0.785)	(0.678)	(0.970)	(0.261)	(0.263)

Subsequently, we draw the attention to the impact of media content and search volume measures on REIT returns. We obtain real estate stock indexes as measures of investment performance and analyze each information variable. Subsequently, we investigate the value of the information measures in explaining the NAV spread phenomemon.

We let R_t denote the log return of the REIT indexes. We further define a dummy variable D_t which takes the value of one if and only if date t is considered to be within recession period as defined by the NBER. Prior work on real estate investments has indicated that traded real estate securities are more similar to other types of listed stocks than to the direct property market in the short run [46]. As the observed correlation between the stock market indexes FTSE 100 and S&P 500 and the respective REIT Return index is particularly strong, we use linear orthogonalization to effectively remove stock market effects from REIT returns. The residual industry effect may be regarded as the extra-market covariance [10].

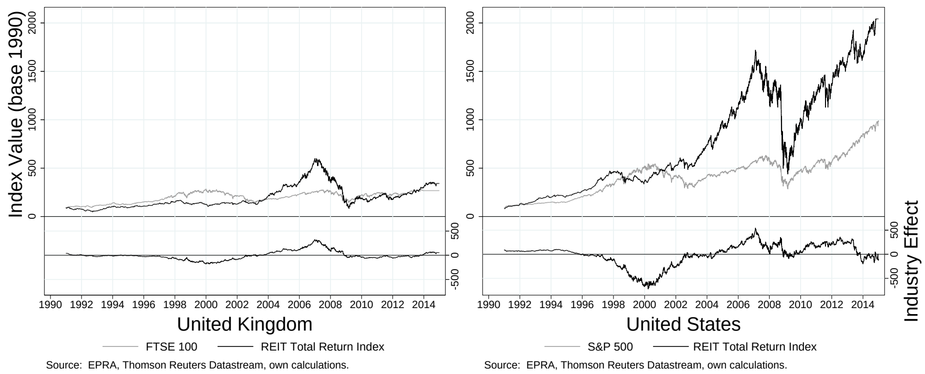

Figure 2: Stock Index, REIT Index and REIT Industry Effect.

Figure 2 illustrates the derivation of the industry effect based on the orthogonalization of REIT returns. The descriptive statistics for the return variables are provided in 6. We observe for both countries that the volatility of the REIT index is 30 to 40 basis points higher than of the respective stock market indicators. We first specify a simple model of stock returns to uncover time-series

Table 6: Sample statistics for Stock Market Returns

	Mean	Std. Dev.	25%-Quan.	Med.	75%-Quan.	Mean	Std. Dev.	25%-Quan.	Med	75%-Quan.
	Panel A: United Kingdom					Panel B: United States				
R_t^{TOT}	0.02	1.42	-0.61	0.03	0.67	0.05	1.60	-0.40	0.08	0.56
R_t^{STO}	0.02	1.14	-0.54	0.04	0.62	0.04	1.15	-0.46	0.06	0.58
R_t^{RES}	0.00	1.25	-0.65	0.01	0.65	0.00	1.20	-0.48	0.01	0.52

characteristics of the residual return series. Following [24] we apply a time series model in the form of

$$R_t^j = (1 - D_t)\beta_1^j L_s(R_t^c) + D_t\beta_2^j L_s(R_t^c) + \mu^j X_t + \epsilon_t, \tag{6}$$

with D_t denoting a dummy variable indicating a NBER recession period, L_s the lag-operator of length s, X_t a set of independent variables and ϵ_t the zero-mean error term. The set of independent variables includes X_t includes a constant term and day-of-the-week and business cycle dummies. The regression results of the parsimonious model for a maximum lag length of $s = 5$ with heteroscedasticity-consistent standard errors according to White (1980) are reported in Table 7.

Table 7: REIT Residual Return Time-Series Regression Results.

	Panel A: UK						Panel B: US				
	Expansion			Recession			Expansion			Recession	
	β_1	t-stat		β_2	t-stat		β_1	t-stat		β_2	t-stat
R_{t-1}	-0.016	-0.30		0.184	5.13		-0.402	-7.97		0.048	1.02
R_{t-2}	0.086	1.55		-0.003	-0.10		-0.065	-1.16		0.069	1.41
R_{t-3}	0.106	2.08		-0.061	-1.58		0.129	2.65		0.037	0.82
R_{t-4}	-0.016	-0.36		-0.020	-0.53		-0.022	0.48		-0.028	-0.40
	Dummy Variables UK						Dummy Variables US				
	μ_i	t-stat		μ_i	t-stat		μ_i	t-stat		μ_i	t-stat
D_t	-0.160	-1.72	X_{Thu}	0.081	1.52	D_t	-0.046	-0.41	X_{Thu}	0.050	1.01
X_{Tue}	0.086	1.63	X_{Fr}	0.071	1.35	X_{Tue}	0.113	2.15	X_{Fr}	0.149	2.85
X_{Wed}	0.162	3.03	$X_{Cons.}$	-0.064	-1.70	X_{Wed}	0.050	1.03	$X_{Cons.}$	-0.065	-1.81

The first rows of Panel A indicate statistically significant autocorrelation for REIT returns in the UK during recessions. Contrary, there is strong evidence for negative autocorrelation during expansions for REIT returns in the US which further indicate some form of mean reversion. We note that returns are significantly lower during recessions. While UK REIT returns on Monday are roughly between 8 and 16 basis points lower than returns on most other days, the day of the week effects in the US are not as pronounced in magniture ranging from 5 to 15 basis points.

We augment the time series model to include the information demand ($k = d$) and supply ($k = s$) variables in the term I_t^k according to

$$R_t^j = \beta L_S(I_t^k) + \gamma L_s(R_t^j) + \mu X_t + \epsilon_t. \tag{7}$$

The information measures are standardized such that the regression coefficients illustrate the impact of a one standard deviation shock to our news measures. Hence, the measures are converted to variables with a mean of zero and unit variance.

Table 8: Regression Results.

Lag	Search Volume Variables			Media Content Variables					
	ASVI (Terms)	*ASVI* (Cat.)	*ASVI* (REIT)	*Sent.* (EM)	*Sent.* (GI)	*Sent.* (SU)	*Sent.* (HE)	*Sent.* (LM)	*Sent.* (MD)
	Panel A: Dependent Variable: UK REIT Returns								
I^k_{t-1}	0.021	0.263	0.050	-0.001	-0.043	0.064*	0.128***	0.084**	0.084**
	(0.937)	(0.171)	(0.835)	(0.983)	(0.272)	(0.096)	(0.001)	(0.044)	(0.044)
I^k_{t-2}	-0.034	-0.171	-0.559**	-0.035	0.046	-0.031	0.018	-0.041	-0.040
	(0.908)	(0.372)	(0.023)	(0.377)	(0.243)	(0.422)	(0.636)	(0.361)	(0.365)
I^k_{t-3}	-0.283	0.003	-0.234	-0.052	-0.061	-0.069*	-0.105**	0.097**	0.097**
	(0.281)	(0.987)	(0.326)	(0.187)	(0.120)	(0.073)	(0.007)	(0.029)	(0.029)
	Panel B: Dependent Variable: US REIT Returns								
I^k_{t-1}	-0.357	-0.416*	-0.088	0.009	0.024	0.046	0.076*	0.070*	0.070*
	(0.127)	(0.060)	(0.712)	(0.805)	(0.523)	(0.217)	(0.056)	(0.066)	(0.067)
I^k_{t-2}	0.076	0.110	-0.371	-0.041	0.002	-0.046	-0.055	0.016	0.016
	(0.754)	(0.632)	(0.151)	(0.297)	(0.967)	(0.235)	(0.171)	(0.685)	(0.683)
I^k_{t-3}	0.174	0.191	-0.260	-0.093**	-0.084**	-0.059	-0.090**	-0.049	-0.049
	(0.452)	(0.387)	(0.273)	(0.018)	(0.031)	(0.126)	(0.027)	(0.217)	(0.218)

The results of the vector autoregression model (VAR) are reported in Table 8 and indicate a positive relation between news content and subsequent stock returns. Accordingly, a shock to news sentiment indicators increases UK REIT returns by 6-12 basis points. The results further indicate mean reversion of news sentiment with increasing lag length. We observe that the Optimism (HE) metric provides the strongest results. Correspondingly, results for the United States indicate that Optimism (HE) outperforms all other sentiment measures. A shock to the sentiment measures in the US moves returns by roughly 7 basis points and is overall less significant. This might be due to the fact that the UK REIT market is relatively small compared to its US counterpart. Further, we observe that the ASVI has a negative impact on REIT returns with a diverse lag structure in both countries. As expected, the ASVI for REITs outperforms the web search measures in the UK. More precisely, the remaining ASVI metrics are insignificant. The ASVI for REITs significantly impacts returns at lag 2 with around 60 basis points which is consistent with the findings of [21]. In contrast, the only significant ASVI for the United States is ASVI (Category). The measure negatively impacts returns with around 40 basis points following a one-standard deviation shock.

We conclude that news sentiment measures based on financial dictionaries, namely (HE) and (LM) seem most suitable for securitized real estate markets. Further, search volume development exhibits significant explanatory power if predefined filter for REITs or the real estate category are applied. However, the impact of Google search behavior generally seems less significant than the relevance of news sentiment.

3.5. Information Flow and REIT NAV Spreads

This paper was motivated by the preliminary finding of particularly strong correlations between news sentiment measures and the NAV spreads in UK and US REIT markets. The empirical evidence so far has been based on standard linear regression models. Hence, the econometric results reveal the approximate average relationship between the described information measures and REIT market returns. That is, optimistic news content has been found to increase while abnormal online search behavior tends to decrease REIT returns. In the following we aim at investigating potential variations in the relationship across the conditional distribution of NAV spreads. More precisely, we apply quantile regressions which allows us to reveal the impact of the information measures on the entire distribution of NAV spreads.

The historical NAV spread for UK and US REITs is illustrated in Figure 3. The average spread for UK REITS is -15.8 percent and commonly varies between -23 and 7 percent during the sample period. US REITs exhibit an average premium of 4.1 percent with common spreads between 4 and 13 percent. The variation of spreads is comparatively high during the NBER recession period from December 3rd, 2007 until July 1st, 2009. During this time, discounts

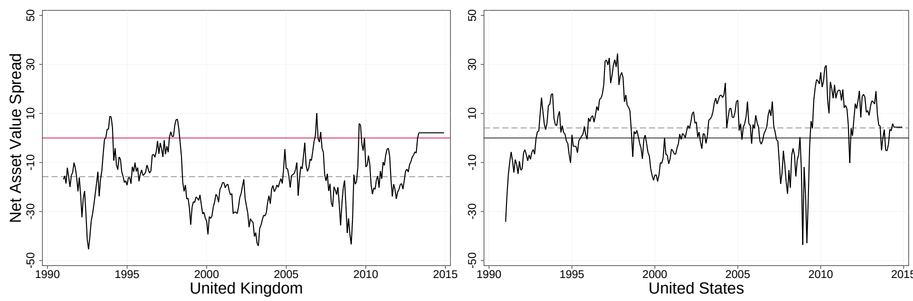

Figure 3: Development of Net Asset Value Spreads in UK and US REIT Markets

fell short of roughly -45 percent in both countries, reaching a local maximum already a year after the end of the recession period.

We incorporate the 10-year government bond yield and the 3-month treasury bill rate as control variables [10, 8]. The descriptive statistics for the exogenous variables are provided in Table 9.

Table 9: Sample statistics for Covariates

	Mean	Std. Dev.	25%-Quan.	Med.	75%-Quan.	Mean	Std. Dev.	25%-Quan.	Med	75%-Quan.
	Panel A: United Kingdom					Panel B: United States				
Government Bond Yield (10y)	4.89	1.76	3.76	4.72	5.43	4.96	1.79	3.71	4.82	6.26
Treasury Bills (3m)	4.79	3.36	3.42	4.9	5.93	3.03	2.26	0.78	3.22	4.97

The rationale for the application of quantile regression is the conditional distribution of NAV spreads across both counries. Figure 4 plots the dependent varialbes by quantiles and highlights that NAV spreads are almost symetrically distributed with extreme vaules in the lower and upper quantiles. Therefore, we cater to investigate the impact of information measures across different quantiles in particular focusing on the 10% and the 90%-quantiles.

The quantile regression methodology minimizes the nondifferentiable function based on the simplex method for each quantile q in the form of

$$Q(\beta_q) = \sum_{i:y_i \geq x'_i}^{N} q|y_i - x'_i\beta_q| + \sum_{i:y_i < x'_i}^{N} (1-q)|y_i - x'_i\beta_q| \tag{8}$$

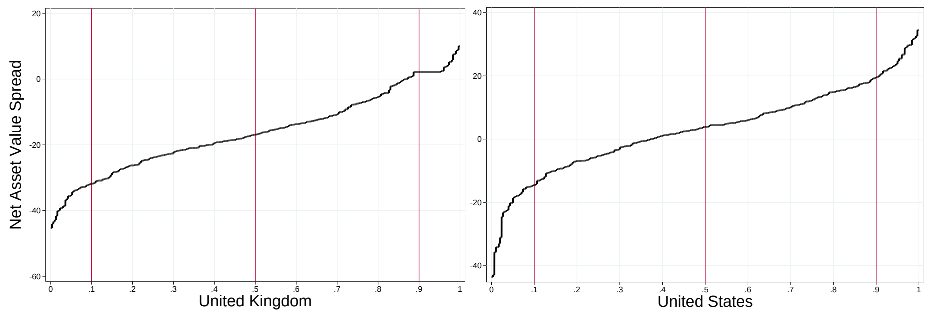

Figure 4: Conditional Distribution Function of NAV Spreads

and has additional advantages. First, quantile regressions are not to be prone to non-normal errors and outliers as ordinary linear regression models and are further invariant to monotonistic transformations.

Table 10: Quantile Regression Results

	United Kingdom						United States					
	OLS	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)	OLS	Q(0.10)	Q(0.25)	Q(0.50)	Q(0.75)	Q(0.90)
Government	2.08***	4.95***	3.03	3.32***	2.92**	-0.16	0.09	0.73	-0.75	-1.02	-0.01	1.68
Bond (10y)	(0.000)	(0.000)	(0.002)***	(0.000)	(0.032)	(0.870)	(0.919)	(0.742)	(0.425)	(0.275)	(0.991)	(0.455)
Treasury	-2.33***	-2.74***	-1.84**	-2.37***	-3.28***	-0.44	-1.94***	-2.53	-2.09***	-2.00***	-2.38**	-1.78
Bill (3m)	(0.000)	(0.001)	(0.010)	(0.000)	(0.002)	(0.552)	(0.004)	(0.134)	(0.004)	(0.005)	(0.014)	(0.297)
Sentiment	7.34***	8.60***	8.66***	8.14***	5.41	0.16	7.17***	5.83	5.16**	5.59**	6.14*	3.74
Opt. (HE)	(0.000)	(0.002)	(0.000)	(0.000)	(0.112)	(0.948)	(0.001)	(0.303)	(0.032)	(0.020)	(0.059)	(0.512)
Cons.	-24.22***	-43.22***	-30.77***	-24.22***	-10.66**	2.72	8.52***	-6.83	6.52**	14.20***	17.55***	15.60**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.018)	(0.406)	(0.004)	(0.346)	(0.034)	(0.000)	(0.000)	(0.034)
N	251	251	251	251	251	251	297	297	297	297	297	297

p-statistics in parantheses, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

The results of the quantile regression are reported in Table 10 and additionally illustrated in Figure 5. While we observe a constant effect of news content and online behavior on NAV spreads in the US across all quantiles, the results stress the relevance of media during extreme NAV discounts in the UK. That is, in times of extreme NAV discounts, media content has a significantly different effect than the average OLS estimate suggests. Further applying the Breusch-Pagan Test for heteroscedasticity justifies the usage of quantile regression technique.

All in all, the empirical results stress different impact along the conditional distribution function of NAV spreads in both countries. While the estimates indicate no significant differences to the OLS estimates for the US, we observe significant differences in the estimates in the lower quantiles hinting at a changing impact of information along the conditional distribution of NAV spreads in the UK. We note that the empirical findings are in line with the main insights for REIT stock returns. Therefore, our sample data supports the notion that the development of the UK REIT market is highly influenced by news content.

4. Conclusion

This paper investigates the impact of information supply and demand in financial markets. We analyze huge amounts of news articles and online search behavior to unravel insights into challenging financial frictions. The study is motivated by the lack of literature proposing a "behavioral story" [24] as a possible explanation for developments of Real Estate Investment Trust (REIT) [51]. As common explanations for the NAV spread are not homogenous and research lacks empirical investigations on investor sentiment, we utilize sentiment analysis to gather a proxy for sentiment based on information supplied by the media and further analyze web search behavior to estimate an indicator for sentiment based on information demanded by individual investors.

We construct sentiment measures and apply several well-known approaches from finance-specific literature. Analyzing news from The Associated Press for two different countries, we suppose to capture valuable information which is not inherent in financial market data. The first finding is that the application of different dictionaries yields distinct measures of news sentiment. We further analyze online search behavior from Google. User-generated content (UGC) revealing the attention of individuals to certain topics has long been lacked an appropriate database. The search volume index on real estate related terms, category and REITs exhibit different interrelations with corresponding stock market developments. Consequently, we apply time-series models which exhibit increased explanatory including the information variables described. Although the empirical evidence provides consistent results across both countries studies, the informaion measures in the UK exhibit much higher relation in terms of magnitude and significance. In addition to REIT returns, we analyze a challenging phenomenon in REIT markets which has often been attributed to behavioral biases. Hence, we collect several control variables which have previously been found to have predictive power for real estate stocks. While the inclusion of information supply more than doubles the explanatory power of the benchmark model, information demand does not exhibit a significant relation to NAV spreads. Further, we analyze the impact of news on different quantiles of NAV spreads. The empirical results stress a constant effect of news content on NAV spreads across all quantiles in the US and highlight the relevance of media during extreme NAV discounts in the UK. Therefore, the empirical evidence supports the notion that the development of the UK REIT market is highly influenced by news content. This is in line with the notion that behavioral biases yield to frictions in asset markets [24].

However, the dataset and methodologies applied additionally provide potential for future research. In particular, the interaction between self-revealed attention to information and information content seems a challenging research path which might lead to further insights into the characteristics of stock markets.

References

[1] Antweiler, W. and M. Z. Frank (2004): "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *The Journal of Finance*, 59, 1259–1294.

[2] Baker, S. A. and J. Wurgler (2007): "Investor Sentiment in the Stock Market," *Journal of Economic Perspectives*, 21, 129–151.

[3] Barkham, R. J. and C. W. R. Ward (1999): "Investor Sentiment and Noise Traders: Discount to NAV in Listed Property Companies in the UK," *Journal of Real Estate Research*, 18, 291–312.

[4] Baumeister, R. F., E. Bratslavsky, C. Finkenauer, and K. D. Vohs (2001): "Bad is stronger than good," *Review of General Psychology*, 5, 323–370.

[5] Bendler, J., S. Wagner, T. Brandt, and D. Neumann (2014): "Taming Uncertainty in Big Data," *Business & Information Systems Engineering*, 6, 279–288.

[6] Black, F. (1986): "Noise," *Journal of Finance*, 41, 529–543.

[7] Bollen, J., H. Mao, and X.-J. Zeng (2011): "Twitter mood predicts the stock market," *Journal of Computational Science*, 2, 1–8.

[8] Boudry, W. I., J. G. Kallberg, and C. H. Liu (2010): "An Analysis of REIT Security Issuance Decisions," *Real Estate Economics*, 38, 91–120.

[9] Brent, W., L. Kelly, D. Lindesy, and R. M. Price (2011): "NAV And Risk: The Case Of REITs," *Journal of Business & Economics Research*, 9, 31–42.

[10] Brooks, C. and S. Tsolacos (2003): "International evidence on the predictability of returns to securitized real estate assets: econometric models versus neural networks," *Journal of Property Research*, 20, 133–155.

[11] Brounen, D. and P. M. A. Eichholtz (2001): "Capital Structure Theory: Evidence from European Property Companies' Capital Offerings," *Real Estate Economics*, 29, 615–632.

[12] Capozza, D. R. and S. Lee (1995): "Property Type, Size and REIT Value," *The journal of real estate research*.

[13] Chen, H., R. H. L. Chiang, and V. C. Storey (2012): "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly*, 36, 1165–1188.

[14] Choi, H. and H. Varian (2012): "Predicting the Present with Google Trends," *The Economic Record*, 88, 2–9.

[15] Clayton, J. and G. MacKinnon (2000): "Explaining the Discount to NAV in REIT Pricing: Noise or Information? Working Paper No. 258268," *SSRN Working Paper*, 1–43.

[16] Da, Z., J. Engelberg, and P. Gao (2011): "In Search of Attention," *The Journal of Finance*, 66, 1461–1499.

[17] ——— (2014): "The Sum of All FEARS Investor Sentiment and Asset Prices," *Review of Financial Studies*, 27, 1–33.

[18] De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990): "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98, 703–738.

[19] Demers, E. and C. Vega (2008): *Soft Information in Earnings Announcements: News or Noise?*, vol. no. 951 of *International finance discussion papers*, Washington and D.C: Federal Reserve Board.

[20] Dimson, E. and C. Minio-Kozerski (1999): "Closed-End Funds: A Survey," *Financial Markets, Institutions & Instruments*, 8, 1–41.

[21] Drake, M. S., D. T. Roulstone, and J. R. Thornock (2012): "Investor Information Demand: Evidence from Google Searches Around Earnings Announcements," *Journal of Accounting Research*, 50, 1001–1040.

[22] Engelberg, J. E., A. V. Reed, and M. C. Ringgenberg (2012): "How are shorts informed?" *Journal of Financial Economics*, 105, 260–278.

[23] Fama, E. F. and K. R. French (1988): "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246–273.

[24] García, D. (2013): "Sentiment during Recessions," *The Journal of Finance*, 68, 1267–1300.

[25] Gentry, W. M., C. M. Jones, and C. J. Mayer (2004): "Do Stock Prices really reflect Fundamental Values? The Case of REITs," *NBER Working Paper Series*, 1–38.

[26] Gentzkow, M. and J. M. Shapiro (2010): "What Drives Media Slant? Evidence From U.S. Daily Newspapers," *Econometrica*, 78, 35–71.

[27] Gottschlich, J. and O. Hinz (2014): "A decision support system for stock investment recommendations using collective wisdom," *Decision Support Systems*, 59, 52–62.

[28] Grefenstette, G. and P. Tapanainen (1994): *What is a word, what is a sentence ? Problems of Tokenisation*, vol. 004 of *RXRC Grenoble*, Grenoble: Rank Xerox Research Centre.

[29] Hagenau, M., M. Liebmann, and D. Neumann (2013): "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems*, 55, 685–697.

[30] Henry, E. (2008): "Are Investors Influenced By How Earnings Press Releases Are Written?" *Journal of Business Communication*, 45, 363–407.

[31] Henry, E. and A. J. Leone (2009): "Measuring Qualitative Information in Capital Markets Research," *SSRN Working Paper*.

[32] Jegadeesh, N. and Di Wu (2013): "Word power: A new approach for content analysis," *Journal of Financial Economics*, 110, 712–729.

[33] Jensen, J., S. Naidu, E. Kaplan, L. Wilse-Samson, D. Gergen, M. Zuckerman, and A. Spirling (2012): "Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech," *Brookings Papers on Economic Activity*, Fall, 1–81.

[34] Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

[35] Kearney, C. and S. Liu (2014): "Textual sentiment in finance: A survey of methods and models," *International Review of Financial Analysis*, 33, 171–185.

[36] Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money*, Cambridge: Macmillan Cambridge University Press.

[37] Lee, C. M. C., A. Shleifer, and R. H. Thaler (1991): "Investor Sentiment and the Closed-End Fund Puzzle," *The Journal of Finance*.

[38] Lewis, D. D., Y. Yang, T. G. Rose, and F. Li (2004): "RCV1: A New Banchmark Collection for Text Categorization Research," *Journal of Machine Learning Research*, 5, 361–397.

[39] Li, F. (2010): "The Information Content of Forward-Looking Statements in Corporate Filings-A Naïve Bayesian Machine Learning Approach," *Journal of Accounting Research*, 48, 1049–1102.

[40] Ling, D. C. and A. Naranjo (1997): "Êconomic Risk Factors and Commercial Real Estate Returns," *Journal of Real Estate Finance and Economics*, 15, 283–307.

[41] LOUGHRAN, T. AND B. MCDONALD (2011): "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, 66, 35–65.

[42] LOVINS, J. B. (1968): "Development of a Stemming Algorithm," *Mechanical Translation and Computational Linguistics*, 11, 22–31.

[43] MANNING, C. D. AND H. SCHÜTZE (2003): *Foundations of statistical natural language processing*, Cambridge and Mass: MIT Press, 6. print. with corr ed.

[44] MODIGLIANI, F. AND M. H. MILLER (1958): "The Cost of Capital, Corporation Finance and the Theory of Investment," *The American Economic Review*, 48, 261–297.

[45] MOHAMMAD, S. M. AND P. D. TURNEY (2010): "Emotions Evoked by CommonWords and Phrases: Using Mechanical Turk to Create an Emotion Lexicon," *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, LA, California.

[46] MORAWSKI, J., H. REHKUGLER, AND R. FÜSS (2008): "The nature of listed real estate companies: property or equity market?" *Financial Markets and Portfolio Management*, 22, 101–126.

[47] NOFER, M. AND O. HINZ (2014): "Are crowds on the internet wiser than experts? The case of a stock prediction community," *Journal of Business Economics*, 84, 303–338.

[48] PANG, B. AND L. LEE (2008): "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, 2, 1–135.

[49] PORTER, M. F. (1980): "An algorithm for suffix stripping," *Program*, 14, 130–137.

[50] Preis, T., H. S. Moat, and H. E. Stanley (2013): "Quantifying Trading Behavior in Financial Markets Using Google Trends," *Scientific Reports*, 3.

[51] Rehkugler, H., F. Schindler, and R. Zajonz (2012): "The Net Asset Value and Stock Prices of European Real Estate Companies: Explaining Net Asset Value Spreads by an Empirical Model," *Zeitschrift für Betriebswirtschaft*, 82, 53–77.

[52] Sharma, A. and S. Dey (2012): "A Comparative Study of Feature Selection and Machine Learning Techniques for Sentiment Analysis," *Proceedings of the 2012 ACM Research in Applied Computation Symposium, New York*, 1–7.

[53] Stone, P. J. (1966): *The general inquirer A computer approach to content analysis*, Cambridge and Mass: M.I.T. Press.

[54] Tetlock, P. C. (2007): "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *The Journal of Finance*, 62, 1139–1168.

[55] Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy (2008): "More than Words: Quantifying Language to Measure Firms' Fundamentals," *The Journal of Finance*, 63, 1437–1467.

[56] Vlastakis, N. and R. Markellos (2012): "Information Demand and Stock Market Volatility," *Journal of Banking and Finance*, 36, 1808–1821.

[57] Wilson, T., J. Wiebe, and P. Hoffmann (2005): "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proceedings of the Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, Vancouver, 347–354.

[58] Xu, S. X. and X. M. Zhang (2013): "Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investor Reaction," *MIS Quarterly*, 37, 1043–1068.

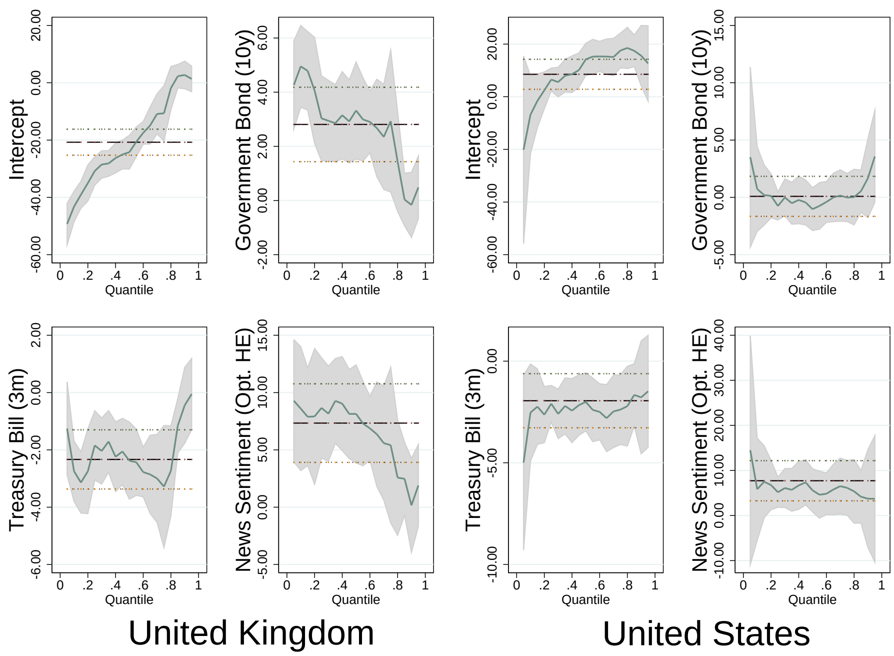

Figure 5: Quantile Regression Results