

Long- and Short-Term Impact of News Messages on House Prices: A Comparative Study of Spain and the United States

Completed Research Paper

Jan-Otto Jandl

University of Freiburg

Bertoldstraße 17

79098 Freiburg, Germany

jan-otto.jandl@vwl.uni-freiburg.de

Stefan Feuerriegel

University of Freiburg

Platz der Alten Synagoge

79085 Freiburg, Germany

stefan.feuerriegel@is.uni-freiburg.de

Dirk Neumann

University of Freiburg

Platz der Alten Synagoge

79085 Freiburg, Germany

dirk.neumann@is.uni-freiburg.de

Abstract

This study adopts data mining methods to analyze the short- and long-term dynamic between news message content and property prices in Spain and the United States. We construct news sentiment indices based on various text mining methods which exhibit remarkable similarities to the respective property prices. Comparing dictionary-based and dynamic approaches, our results indicate that static methods produce the best estimators for investor sentiment in real estate markets. Using a Vector Error Correction Model framework, we analyze similarities between real estate markets both in the short- and long-run. The main finding of this study is a significant relationship between news messages and property prices in the long-run. Our results are stable, including a number of fundamental variables, and are underlined by forecast error variance decompositions and impulse response estimates, which additionally highlight the appropriateness of news sentiment as a crucial determinant for decision making.

Keywords: Decision Analytics, Text mining, Information processing, Economics of information systems, Quantitative analysis, Sentiment analysis

Introduction

Within the last decade, Spain and the United States have experienced generally unexpected price bubbles in housing markets which were widely believed to be a precursor to the subsequent Great Recession (Stiglitz 2009). Empirical and theoretical studies on price bubbles in asset markets provide various explanations. Thereof, macroeconomic fundamentals and financial market characteristics are among the most prominent rationales discussed in economic literature. Nevertheless, an increasing branch of research incorporates that the constraints of the homo economicus do not account for the emotional and cognitive biases of the homo sapiens.

According to economic theory, information can be understood as a foundation for expectations (Muth 1961). This study investigates the driving factors for real estate prices, incorporating an IS measure as a proxy for non-fundamental demand. We approximate this sentiment-driven demand through the Thomson Reuters News Archive by utilizing text mining methods. We assume that sentiment analysis on the time series of agency news can be used as an indicator for market expectations and create sentiment indices. As a result, the main contribution of this paper is to provide an additional rationale for information processing in real estate markets.

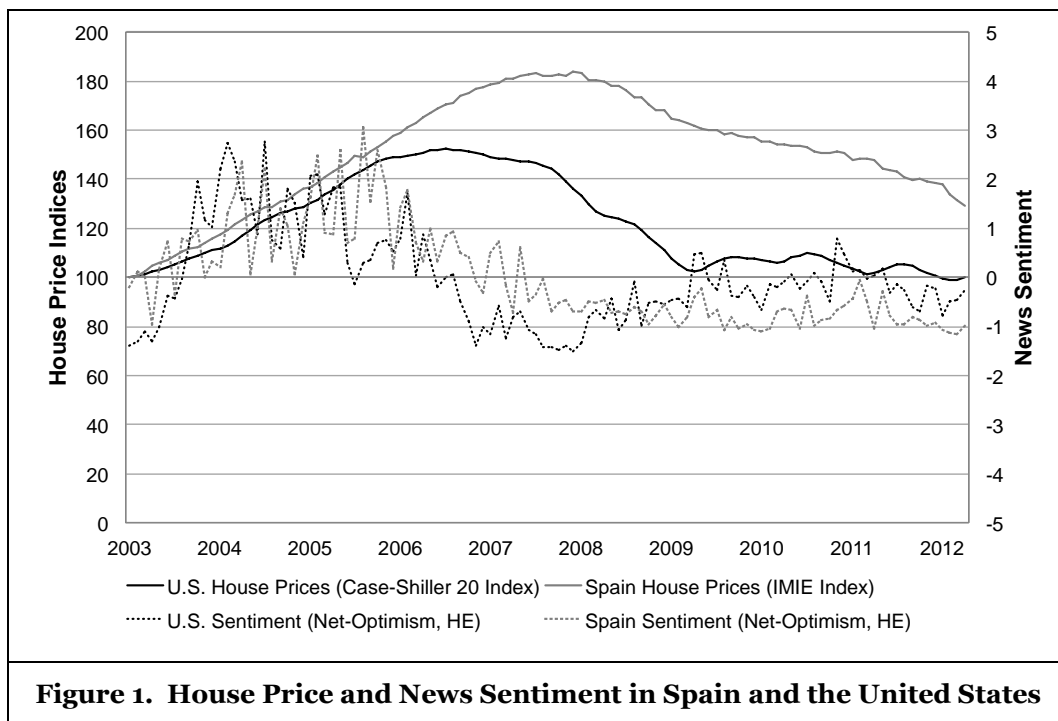


Figure 1. House Price and News Sentiment in Spain and the United States

The majority of empirical investigations on real estate markets are carried out using US data. Therefore, this study comparatively analyzes the recent real estate bubbles in the US and Spain to uncover similarities, as well as to examine the impact of news sentiment in different environments. As illustrated in Figure 1, property prices in the United States grew until mid-2006 and subsequently fell until 2009, almost returning to the average price level in 2003. Since 2009, property prices in the United States fluctuate around this level. Property prices in Spain continued to increase for almost 2 more years, reaching its climax in early 2008, with following successive declines that appear less severe than in the United States. Since its peak, the average house price decreased almost linearly, reaching the 2005 price level in May 2012.

The motivation for our research methodology lies in the empirical observation of a potential relationship between real estate prices and news sentiment, as illustrated in Figure 1. We note that increasing house prices are accompanied by positive sentiment and vice versa. This preliminary finding encouraged the construction of a sentiment index which utilizes the original sentiment measure a rate of change. The resulting sentiment indices exhibit remarkable similarities to the real estate price indices and show a high potential for information processing in property markets.

The remainder of this study proceeds as follows. Section 2 reviews the related literature. Section 3 describes the data and news sentiment indicators in detail. Empirical results are provided in section 4. The final section contains critical remarks and concludes.

Literature Review

Within the IS community, sentiment analysis has gained significant influence. The increasing relevance is underlined by a huge number of IS publications adopting sentiment analysis across various applications. Among research questions from financial markets, IS research has proven that sentiment analysis is both powerful and effective. More precisely, IS publications have contributed to relevant research questions as follows.

Various publications utilize sentiment analysis in order to investigate the relationship between company-specific news and stock returns. This strong correlation can be exploited, for example, Mittermayer (2004) and Mittermayer and Knolmayer (2006) implement an automated news categorization system with the aim of predicting intraday price trends. By selecting news from appropriate categories, a feature list is automatically created and combined with machine learning techniques to predict the direction of the stock price return. Other authors (Groth and Muntermann 2008; Groth and Muntermann 2010, 2011; Hagenau et al. 2012; Hagenau et al. 2013a; Siering 2012, 2013; Siering and Muntermann 2013) employ sentiment analysis to predict the stock price returns from company-specific news and corporate disclosures respectively. Moreover, they test their systems in trading simulations. In addition, news sentiment can be utilized to analyze how information is perceived (Siering 2013). In fact, the way in which investors and analysts absorb novel information differs. Liebmann et al. (2012) arrive at the conclusion that investors rapidly translate novel information into transactions, whereas analysts wait to respond.

Apart from predicting returns of company-specific stocks, sentiment analysis can also be used to predict changes in stock indices. Hagenau et al. (2013b) aggregate individual company-specific news to predict the movement of stock indices and, following this, test the system using a trading strategy. In addition, Tsai et al. (2010) take financial news and analyze the effects of news coverage and sentiment factors on credit ratings. Their results indicate the presence of a significant relationship between sentiment and changes in credit ratings – a better rating performance in the subsequent quarter emerges from more positive sentiment. In a similar fashion, sentiment analysis has also been extended to other sources, such as oil prices (Yu et al. 2005; Feuerriegel and Neumann 2013) and micro blogging activities in (Oh and Sheng 2011).

Methods that use the textual representation of documents to measure the positivity and negativity of the content are referred to as opinion mining or sentiment analysis. In fact, sentiment analysis can be utilized to extract subjective information from text sources, as well as to measure how market participants perceive and react to news. In this instance, one uses the observed stock price reactions following a news announcement to validate the accuracy of sentiment analysis routines. Based upon sentiment measures, one can study the relationship between news and its effect on stock markets. In addition, empirical evidence shows that a discernible relationship between news content and stock market reactions exists (Antweiler and Frank 2004; Tetlock 2007).

As sentiment analysis is applied to a wide variety of domains and text sources, research has devised various approaches to measure sentiment (Pang and Lee 2008). Within finance, recent literature surveys (Mittermayer and Knolmayer 2006; Minev et al. 2012) compare studies aimed at stock market prediction. For example, dictionary-based approaches are very frequently (Demers and Vega 2010; Henry 2008; Jegadeesh and Wu 2011; Tetlock et al. 2008) used in recent financial text mining research. These methods count the frequency of pre-defined positive and negative words from a given dictionary – producing results that are straightforward and reliable. Machine learning approaches (Antweiler and Frank 2004; Li 2010; Mittermayer and Knolmayer 2006; Schumaker and Chen 2009) offer a broad range of methods, but may suffer from overfitting (Sharma and Dey 2012).

The theoretical rationale for sentiment analysis in real estate markets is often based on the noise trader approach, which divides investors into fully rational and imperfect rational investors (DeLong et al. 1990). The investment behavior of the latter noise traders is believed to be highly influenced by sentiment. This is in line with Keynes (1936), who states that beliefs and good news might cause price bubbles. Accordingly, Shiller (2005) argues that the recent house price bubble has been caused by positive media content.

Real estate markets exhibit special characteristics. Important features are that house prices exhibit positive autocorrelation, which can be approximated by information variables and do not follow a random walk (Case and Shiller 1989). Literature on sentiment in real estate markets is scarce due to the coarseness of direct real estate returns and concentrated on Real Estate Investment Trusts (REIT) as a hybrid between direct and securitized investments. Soo (2013) is among the first to investigate the impact of regional media content directly on house prices by applying sentiment analysis to property markets and concludes that media content is a significant factor influencing regional house prices.

In contrast, our study abdicates from manually creating domain-dependent terms and incorporates various text mining methods. In addition, we base our study on objective information as represented by news announcements. Finally, the core of this study is the comparison of potential news message impact on house prices across countries that experienced house price bubbles.

Data and Descriptive Statistics

The hypothesis of this study is that the variation in real estate prices is affected by market sentiment, which we approximate using data mining methods. In particular, we determine various sentiment proxies based on news published in the *Thomson Reuters News Archive*. The hypothesis is consistent with the assumptions of Prospect Theory developed by Kahneman and Tversky (1979) that states that psychological aspects are of the utmost importance for investment decisions under uncertainty. More precisely, we assume that a change in news sentiment has a similar effect to the change of wealth in Prospect Theory on individual investment behavior.

In this section, we first elaborate on sentiment analysis methods in general and describe the various news sentiment measures investigated in this study in detail. Second, we present data on real estate prices. Third, we discuss pre-specified control variables, which are assumed to additionally influence property returns.

News Sentiment

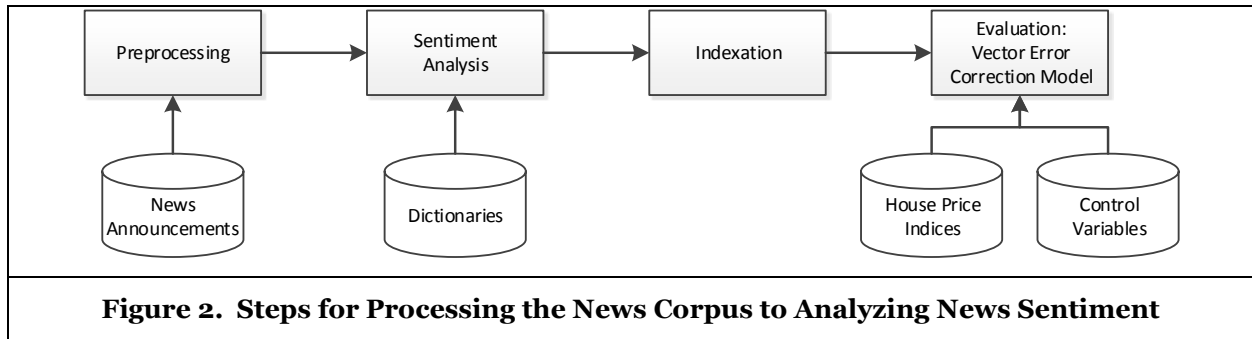
In our research framework, we process only a few data points linked to a large text basis (hundreds of announcements from a single day) along with a continuous return. Thus, we have experienced difficulties in gaining robust results with machine learning and term weighting approaches. Therefore, we focus on rule-based methods and test several metrics in combination with various dictionaries. We find that the Net-Optimism (NO) approach, along with Henry's Finance Dictionary (HE), outperforms all others. Hence, we describe the news corpus and the processes involved in sentiment analysis at first and subsequently define the Net-Optimism measure and provide descriptive statistics for the media measure.

The news corpus originates from the *Thomson Reuters News Archive* for Machine Readable News. We choose Reuters News deliberately due to the following three reasons: (1) Reuters news is third-party content and so does not originate from market participants themselves. (2) Announcements from Reuters solely contain information that is new. In comparison to announcements, articles from newspapers might be edited, perturbed, shortened and filtered by the subjective criteria of editors. (3) Articles from newspapers are normally built upon announcements from news agencies. Thus, news released from newspapers might be delayed and exhibit a time lag. Overall, choosing Reuters news provides the advantage of a more objective news source.

All announcements provided by Reuters are taken from the time span, January 1, 2003 till May 31, 2012. Based upon these, the news corpus is filtered such that we extract announcements that focus on the real-estate market. This is achieved by applying a set of filter criteria: (1) The language must be English. (2) The event type is *Story Take Overwrite* to guarantee that we do not yield an alert but the actual message. (3) We exclude announcements that contain specific words (*advisory, chronology, corrected, feature, diary, instant view, analyst view, newsmaker, corrected, refile, rpt, schedule, table, service, alert, wrapup, imbalance, update*) in their headline. In fact, special types of announcements, such as alerts or personal opinions, might have limited relevance to changes in the real-estate market and we want to exclude these. (4) We use topic codes (Reuters 2008) to select announcements (containing the label *REA* and *US* or *REA* and *ES* respectively) that deal with real estate in the United States and Spain. (5) We exclude announcements addressing changes in house prices to avoid simultaneity (Antonakis et al. 2014). (6) In order to remove white noise, we require announcements to be a total of at least 50 words.

All in all, this set of criteria filters a total of 63,731 announcements that solely address the US real-estate market and 7,279 announcements for the Spanish property market. As we perform sentiment analysis based on this news corpus, we describe the common steps involved in text mining before providing descriptive statistics.

Starting with a corpus that contains all news announcements, we proceed to filter announcements that fit our research focus. Each announcement is then subject to preprocessing which transforms the running text into machine-readable tokens. These tokens are later arranged within a term-document matrix which itself is input for the actual sentiment analysis. The final step before evaluation is aggregation, which aggregates all announcements to match the resolution of the house price data. As new house price data is released once every few months, the idea can be briefly summarized as concatenating all announcements in between to form one long document. Both preprocessing and sentiment analysis (joint with aggregation) are briefly discussed in the following sections.



Before performing the actual sentiment analysis, several operations are involved in a preprocessing phase. The individual steps are as follows.

- **Tokenization.** Each announcement is split into sentences and single words named tokens.
- **Negations.** Negations invert the meaning of words and sentences. When encountering the word *no*, each of the subsequent three words (i.e. the object) is counted as words from the opposite dictionary. When encountering other negating terms (*rather, hardly, couldn't, wasn't, didn't, wouldn't, shouldn't, weren't, don't, doesn't, haven't, hasn't, won't, hadn't, never*), the meaning of all succeeding words is inverted (Dadvar et al. 2011). According to Tetlock (2007) and Loughran and McDonald (2011), handling negations is an inevitable prerequisite to measure positive tone.
- **Stop word removal.** Words without a deeper meaning, such as *the, is, of and so on* are named stop words and are removed. We use a list of 571 stop words (Lewis et al. 2004).
- **Synonym merging.** Synonyms, though spelled differently, convey the same meaning. Thus, approximately 150 frequent synonyms are grouped and aggregated by their meaning – a method referred to as pseudoword generation (Manning and Schütze 1999).
- **Stemming.** Stemming refers to the process of reducing inflected words to their stem (Manning and Schütze 1999). Here, we use the Porter stemming algorithm (Porter 1980).

The correlation between media measures and the corresponding stock market reaction varies strongly across different sentiment metrics. A sentiment approach that produces a reliable correlation is the Net-Optimism metric (Demers and Vega 2010). Out of these, Net-Optimism along with Henry's Finance-Specific Dictionary (Henry 2008), achieves the highest robustness and, consequently, we rely upon this approach in the following evaluation.

The Net-Optimism measure was originally developed to analyze the sentiment of a single news announcement A . We present an extended version that aggregates the sentiment of all announcements of one day t . Therefore, we estimate the Net-Optimism sentiment value $S_{NO}(t)$ incorporating the total number of words $W_{tot}(t)$, the number of negative words $W_{neg}(t)$ and the total number of positive words $W_{pos}(t)$ in the news of day t according to formula

$$S_{NO}(t) = \frac{W_{pos}(t) - W_{neg}(t)}{W_{tot}(t)}. \quad (1)$$

Thus, $S_{No}(t)$ measures the difference between the count of positive and negative words normalized by the number of total words for a the news stream of a given day t . Table 1 provides the summary statistics for various news sentiment measures as the average monthly news sentiment normalized by the annual sentiment average.

	Static Methods				Dynamic Methods		
	Tetlock (GI)	Net-Optimism (GI)	Net-Optimism (HE)	Net-Optimism (LM)	Adjusted Weighting (AW)	Bi-Normal Separation (BNS)	Tonality (TON)
First Year	2003	2003	2003	2003	2005	2005	2005
Observations	112	112	112	112	112	84	84
United States							
Minimum	-2.55	-2.36	-1.51	-3.56	-1.16	-1.07	-0.96
Maximum	2.29	1.95	2.75	2.99	9.22	2.35	2.20
Spain							
Minimum	-2.06	-1.70	-1.15	-2.46	-0.41	-2.35	-1.72
Maximum	1.67	3.10	3.05	2.47	10.30	1.19	2.18

Table 1. Summary Statistics - News Sentiment

As various indicators for news sentiment are drawn from the same kind of news, we first calculate the pairwise correlations. Table 2 reports the correlations between the levels of sentiment proxies for seven individual sentiment proxies between 2003 and 2012, including four static and three dynamic methods. The pairwise Pearson product-moment correlation, the p-value for a two-tailed test with a null of zero correlation and the number of observations are shown.

	United States							Spain						
	Static Methods				Dynamic Methods			Static Methods				Dynamic Methods		
	Tetlock (GI)	Net-O. (GI)	Net-O. (HE)	Net-O. (LM)	AW	BNS	TON	Tetlock (GI)	Net-O. (GI)	Net-O. (HE)	Net-O. (LM)	AW	BNS	TON
Tetl. (GI)	-							-						
Net-O. (GI)	-0.634							-0.947						
O. (GI)	0.000	-						0.000	-					
Net-O. (HE)	-0.214	0.622						-0.838	0.813					
O. (HE)	0.023	0.000	-					0.000	0.000	-				
Net-O. (LM)	-0.137	0.105	0.466					-0.767	0.748	0.747				
O. (LM)	0.1495	0.267	0.000	-				0.000	0.000	0.000	-			
AW	-0.113	0.029	-0.142	-0.088				-0.127	0.172	0.222	0.190			
BNS	0.236	0.758	0.134	0.353	-			0.181	0.069	0.018	0.043	-		
TON	-0.046	0.689	0.803	0.393	0.200			-0.916	0.887	0.782	0.785	0.132		
AW	0.673	0.000	0.000	0.000	0.066	-		0.000	0.000	0.000	0.000	0.228	-	
BNS	84	84	84	84	84			84	84	84	84	84		
TON	-0.028	0.664	0.814	0.385	0.204	0.996		-0.957	0.948	0.890	0.805	0.237	0.942	
AW	0.795	0.000	0.000	0.000	0.062	0.000	-	0.000	0.000	0.000	0.000	0.029	0.000	-
BNS	84	84	84	84	84	84		84	84	84	84	84	84	

Table 2. Pairwise Correlations between News Sentiment Indicators

The results indicate reasonable correlations between sentiment variables in terms of magnitude, sign and significance. While news sentiment generally tends to move together, the Tetlock (GI) measure expectedly correlates negatively with the other news sentiment indicators. This is in accordance with the nature of this metric, which is focused on negative words and can therefore be interpreted as an indicator of pessimism. Nevertheless, the Tetlock (GI) sentiment measure lacks significance for the United States. In fact, the Tetlock measure for the United States does not seem to be related to any of the other measures except Net-Optimism based on the GI and HE dictionaries. Changes in the sentiment metrics based on the GI dictionary are not closely related to changes in those based on the LM dictionary. In contrast, Spanish static sentiment methods are highly correlated in terms of impact and magnitude without exception.

The dynamic sentiment methods exhibit high correlations, with the exception of Adjusted Weighting. As magnitude and sign of correlation coefficients are not coherent, in addition to insignificant p-values, we postulate that Adjusted Weighting may not be suited as a reasonable news sentiment indicator. In conclusion, the presented news sentiment metrics seem to reflect the same tone in the corresponding news. Hence, we use various sentiment measures to check the robustness of our results. As the correlations are generally stable and significant, we expect the different sentiment methods to produce comparable results.

Sentiment Indices

The formal derivation of sentiment indices bases on the observation that the originally derived sentiment measures exhibit a remarkable similarity with changes in property prices. We calculate the indexed sentiment series $S_{DI}^I(t)$ to the base 100 in January 2003 at month t adapted from the sentiment index value of the preceding month $S_{DI}^I(t - 1)$ and add the sentiment value of the original sentiment measure for the respective month. Therefore, the sentiment index is simply calculated using an additive form of

$$S_{DI}^I(t) = S_{DI}^I(t - 1) + S_{DI}(t). \tag{2}$$

In addition to real estate (RE) indices, the derived sentiment indices for Spain and the United States based on the Net-Optimism measure and Henry’s finance-specific dictionary are illustrated in Figure 3. We observe that sentiment seems to follow a similar pattern as real estate prices. We therefore propose that these variables may have a specific relationship which we investigate using a VECM and affiliated econometric instruments.

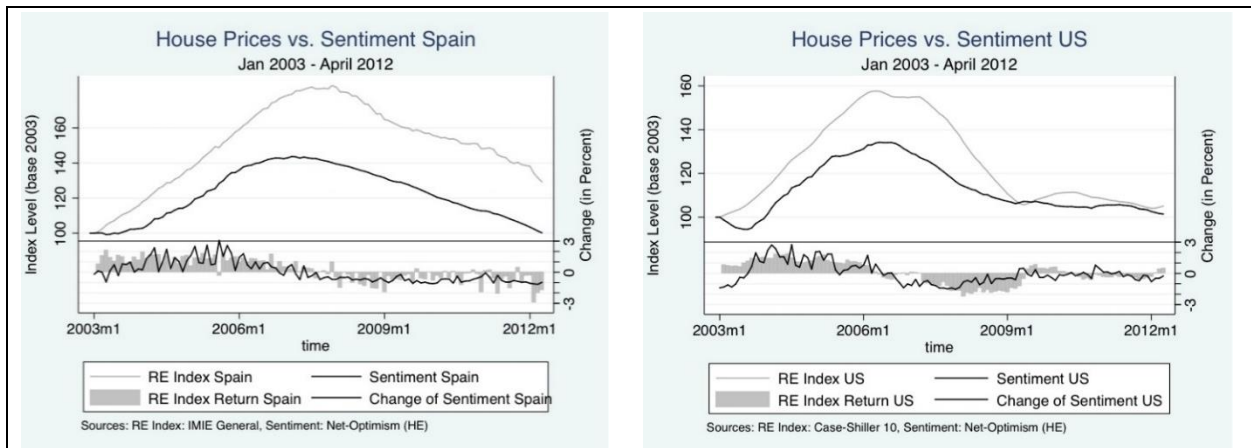


Figure 3. Sentiment and House Price Indices

Summary statistics for the determined sentiment indices are described in Table 3. A comprehensive look at the measures provides reasonable explanations for the characteristics of the indices attained. The Tetlock-measure, for example, seems to mirror the development of other dictionary-based sentiment approaches. Comparing these, we observe that Net-Optimism measures based on the GI- and LM-dictionary simultaneously peak with property prices, while Net-Optimism (HE) reaches its climax prior to real estate prices. Although a synchronous run can also be observed for the dynamic approaches, the lack of data points and unsatisfying data in the case of Adjusted Weighting led to the exclusion of dynamic sentiment metrics from further analysis.

	Static Methods				Dynamic Methods		
	Tetlock (GI) Index	Net-O. (GI) Index	Net-O. (HE) Index	Net-O. (LM) Index	AW Index	BNS Index	TON Index
First Year	2003	2003	2003	2003	2003	2005	2005
Observations	112	112	112	112	112	84	84
	United States						
Mean	88.00	123.22	112.82	97.54	105.76	119.26	119.27
Std. Dev.	6.70	10.86	11.62	5.36	4.31	11.22	11.42
Minimum	74.75	99.19	94.42	87.27	97.08	97.64	97.79
Maximum	100.00	140.72	134.24	109.01	112.84	136.11	136.38
	Spain						
Mean	72.53	126.87	122.44	121.48	102.24	120.71	121.15
Std. Dev.	16.11	15.54	14.71	13.97	3.11	11.34	11.63
Minimum	51.24	99.36	99.16	100.00	96.64	99.07	99.03
Maximum	100.48	147.87	143.66	144.08	107.86	136.14	136.24

Table 3. Summary Statistics - News Sentiment Indices

We note that the application of sentiment analysis of financial news in English for both real estate markets implicitly considers expressions in English for the property market in Spain, which may lead to a potential bias. Therefore, we take alternative adaptations of cross-linguistic sentiment analysis into account. Accordingly, sentiment analysis of financial news for the Spanish real estate market could be executed on the basis of news announcements in Spanish. In this case, the application of Spanish dictionaries is inevitable. However, pre-specified Spanish dictionaries analogous to the applied English dictionaries do not exist. As we additionally note that an analysis on the basis of different dictionaries might even deteriorate the results, we refrain from the analysis of news in Spanish. Hence, this approach limits the application to dynamic methods, which we reasonably exclude in this paper. An additional viable alternative is found to be the analysis of translated Spanish news content. An additional viable alternative cross-linguistic investigation includes the translation of Spanish news content to English and subsequent sentiment analysis. Nevertheless, prior research has pointed out that dictionary-based sentiment analysis of translated Spanish texts on the basis leads to a significant semantic loss and consequently deteriorates the results (Brooke et al. 2009). Therefore, we propose that the application of the same sentiment methods to a common news corpus seems most suitable and minimizes the linguistic bias and potential endogeneity.

Real Estate Prices

Real estate prices for the United States are based on the S&P/Case-Shiller home price indices (Case-Shiller) as well as the U.S. House Price Indices (HPI). The Case-Shiller indices are published by Standard and Poor's (S&P) and cover major Metropolitan Statistical Areas (MSA) in the United States. The indices are calculated using repeat-sales methods and are available both as a Composite 10 Index comprising 10 MSA since 1987 (CS10) and as a Composite 20 Index including 20 MSA since 2000 (CS20). Since the Case-Shiller indices use data on single-family homes that are re-sold, they are considered to measure price changes of houses given a constant level of quality. The HPI are issued by the Office of Federal Housing Finance Agency (FHFA), are determined using weighted, repeat-sales methods and are available for nine U.S. Census divisions and on a national level.

House prices for Spain are drawn from TINSA, a private company for real estate analysis and valuation, which provides various real estate price indices (IMIE) for Spain. We use the general IMIE index for Spain, as well as the IMIE index covering MSA and large cities. The IMIE index on MSA concentrates on metropolitan areas formed by municipalities, the IMIE index on capital and large cities covers around 150 Spanish cities with more than 50.000 inhabitants. The descriptive statistics for the applied series of real estate prices in terms of level data and log returns are presented in Table 4.

	Panel A: United States			Panel B: Spain		
	OFHEO (HPI)	Case-Shiller (Comp.-20)	Case-Shiller (Comp.-10)	IMIE (General)	IMIE (MSA)	IMIE (Cities)
First Year	2003	2003	2003	2003	2003	2003
Obs.	112	112	112	112	112	112
	level data					
Mean	198.85	164.76	179.12	1868.49	1863.20	1977.70
Std. Dev	17.24	24.91	28.11	287.27	278.27	320.04
Max	227.24	206.61	226.87	2283.88	2274.08	2433.35
Min	167.82	135.40	142.57	1229.40	1239.74	1270.41
	log returns					
Mean	0.09	0.02	0.05	0.23	0.21	0.22
Std. Dev	0.65	1.00	1.07	1.08	1.35	1.24
Max	1.03	1.66	1.86	2.08	2.97	2.53
Min	-1.71	-2.01	-2.23	-2.96	-3.58	-2.89

Table 4. Descriptive Statistics - Real Estate Prices

Control Variables

We include macroeconomic variables in order to cover inflation, economic growth, labor market and key interest rates. In particular, we consider consumer price indices as a proxy for inflation. We use the normalized consumer price index for Spain and the city average of all items consumer price index for the United States. Economic growth is approximated by the total industrial production indices. Unemployment rates are used as a labor market indicator. The data is seasonally adjusted and transformed into natural logarithms. In addition, we include nominal short-term and long-term key interest rates. That is, we use the 3-month yield on treasury bills as a proxy for short-term and the 10-year government bond yields as an approximator for long-term interest rates in both countries. We derive all data from Thomson Reuters Datastream. The descriptive statistics for the macroeconomic control variables are summarized in Table 5.

	Panel A: United States				Panel B: Spain			
	Consumer Price Index (CPI)	Industrial Production (IP)	Short-term interest (SIR)	Long-term interest (LIR)	Consumer Price Index (CPI)	Industrial Production (IP)	Short-term interest (SIR)	Long-term interest (LIR)
Mean	206.36	93.70	1.75	4.34	94.39	113.09	2.13	4.23
Std. dev.	13.85	4.34	1.77	0.66	6.59	11.62	1.14	0.61
Max	229.18	100.82	5.01	5.31	105.70	129.40	4.35	6.20
Min	182.60	83.75	0.01	2.64	82.60	92.48	0.25	3.09
Obs	112	112	112	112	112	112	112	112

Table 5. Descriptive Statistics - Control Variables

Table 6 presents the pairwise Pearson product-moment correlations for the first differences in the control variables. In addition, the p-values and the number of observations are reported. The results indicate that serial correlation issues do not seem to be present. We calculate several collinearity diagnostic measures and use a Variance Inflation Factor (VIF) of 10 as a benchmark. As the VIF for the unemployment rates and the industrial production index in both countries are far above this benchmark, we remove these indicators.

	Panel A: United States				Panel B: Spain			
	Consumer Price Index (CPI)	Industrial Production (IP)	Short-term interest (SIR)	Long-term interest (LIR)	Consumer Price Index (CPI)	Industrial Production (IP)	Short-term interest (SIR)	Long-term interest (LIR)
CPI	-				-			
IP	0.0923	-			-0.0812	-		
SIR	0.3354	0.1877	-		0.3968	0.3572	-	
LIR	0.1284	0.0485	0.2160		0.0563	0.0001	0.1405	
	0.1793	0.0627	0.0228	-	0.5574	0.0507	0.1414	-
	111	111	111		111	111	111	
	0.2501	0.5132			0.1273	0.5974		
	0.0081				0.1832			
	111	111	111		111	111	111	

Table 6. Pairwise Correlations between Control Variables

Empirical Findings

News sentiment measures, which are derived from news about real estate fundamentals, are prone to being endogenous to those fundamentals. Therefore, we apply a model incorporating potential multiple relationships between news sentiment and fundamentals. The concept and implications of cointegration is expected to be relevant for real estate markets, as theory often suggests that variables are expected to hold some long-term association with one another. For example, rents may move together with demand-side factors in supply-constrained markets, as is the case for densely populated cities. In less constrained markets, long-term relationships may exist between economic variables and vacancy rates. In a rational capital market model, it is assumed that capital values of houses are affected by both rents and discount rates and should form a long-run relation. We expect that the long-run equilibrium is affected by such fundamentals and that the short-run relation is influenced by changes in risk premia due to investor sentiment.

Vector Error Correction Model Framework

The VECM framework allows for the identification of long-term and short-term relationships between variables, as well as the dynamic adjustment process of cointegrated variables. Using a Vector Error Correction framework, we determine if the evaluation of behavior of real estate markets can be improved by considering market sentiment in addition to macroeconomic risk factors. The Error Correction Model is based on a Vector Autoregressive (VAR) model with p lags in the form of

$$y_t^c = v^c + \sum_{i=1}^p A_i y_{t-i}^c + \epsilon_t^c, \quad (3)$$

in which y_t^c represents a $n \times 1$ vector of variables for the country c , and A_1, \dots, A_p are $n \times n$ matrices of parameters. In this model, ϵ_t is a $n \times 1$ vector of disturbances with a mean of zero and is independent and identically distributed. With at least two cointegrated variables, the VAR(p) can be rearranged in a VECM form as

$$\Delta y_t^c = v^c + \Pi y_{t-1}^c + \sum_{i=1}^{p-1} \Gamma_i^c \Delta y_{t-i}^c + \epsilon_t^c, \quad (4)$$

where the coefficients of the cointegration vectors and the error correction term are included in the matrix Π . Here, Δy_t^c denotes the $n \times 1$ vector of the first differences of the stochastic variables y_t^c and v^c a $n \times 1$ vector of the constants. The short-term dynamic is represented by the $n \times n$ matrices Γ_i^c . If the matrix Π has a reduced rank of $0 < r < n$, such that $\Pi = \alpha\beta'$ with α and β reflecting $n \times r$ matrices of rank r , β can be considered as a matrix containing the cointegrating vectors. Within this setting, parameter α reflects the speed of the adjustment vector and βy_{t-i}^c the error correction element.

The coefficients indicate the contribution of the long-term relationships in the individual equations. Replacing matrix Π in formula 3 leads to an error correction representation in the form of

$$\Delta y_t^c = v^c + \alpha\beta'y_{t-1}^c + \sum_{i=1}^{p-1} \Gamma_i^c \Delta y_{t-i}^c + \epsilon_t^c. \quad (5)$$

Cointegration Analysis

In order to apply a multivariate Vector Error Correction Model (VECM), we first perform a cointegration analysis at first. In particular, we investigate if the variables satisfy the stationarity condition by applying unit root tests and determine the optimal lag structure.

First, we determine the optimal lag structure by minimizing the value of Akaike's information criterion (AIC) and the Hannan-Quinn Information Criterion (HQIC). Second, an important precondition of the Johansen Test of Cointegration is that the variables have to be integrated of the same order. Therefore, the variables should be non-stationary at level but become stationary when converted to differences. To analyze the real estate time series in relation to macroeconomic variables, we need the regression residuals to be independent and identically distributed with a mean of zero.

Consequently, we test if the residuals are integrated of order zero, $I(0)$. As Granger (1981) states, a properly specified econometric model implies that for each model with a dependent variable of the order zero, the independent variables must also be $I(0)$ if the regression residual is to be stationary. In contrast, if the dependent variable is integrated of order 1, $I(1)$, cointegration has to be tested in order to evaluate if an error correction model is suitable. We perform augmented Dickey-Fuller tests to determine the stationary condition (Dickey and Fuller 1979). Hence, we apply unit root tests to the real estate return series and the explanatory variables. The unit root tests indicate non-stationarity at level and stationarity when considering the first difference of the variables. In addition, we validate these results by calculating Phillips and Perron test statistics, which can be interpreted as comparable statistics that are particularly robust to serial correlation and heteroscedasticity. As the variables are integrated of order one, $I(1)$, we conclude that the model is well-specified for a VEC framework.

We analyze potential cointegration relationships using the Johansen Tests for Cointegration. These tests perform maximum likelihood estimation and are used to estimate the rank and the coefficients. The equivalent likelihood-ratio test statistics λ_{trace} and λ_{max} represent the estimated eigenvalues of the reduced rank of the matrix Π and are calculated using the formulas

$$\lambda_{trace} = -T \sum_{r+1}^k \ln(1 - \lambda_i) \quad (6)$$

$$\lambda_{max} = -T \ln(1 - \lambda_i). \quad (7)$$

The test gives evidence of cointegrating equations. Hence, it evaluates the null hypothesis of no cointegration, which is the case if the log likelihood of the unconstrained model including cointegrating equations is significantly different from the log likelihood of the constrained model including no cointegrating equations. The sequential likelihood-ratio statistics are calculated starting with a test for zero cointegrating equations (that is $r=1$) and is successively determined once the cointegration rank is revealed.

Results of the Multivariate Vector Error Correction Model

First, we determine the appropriate lag structure by the minimization of the AIC and the HQIC and identify an optimal lag length of 4 months for both countries. We note that this result is contrary to the much longer lag structure identified in McCue and Kling (1994) and Schätz and Sebastian (2009). As they analyze partially different data and draw attention to other countries, we conclude that these unequivalent empirical findings reflect the fundamental differences in real estate market characteristics across countries.

λ_{trace} test				λ_{max} test			
H_0	H_A	λ_{trace}	Critical Value (5 %)	H_0	H_A	λ_{max}	Critical Value (5 %)
Panel A: Spain							
$r = 0$	$r > 0$	87.45	68.52	$r = 0$	$r = 1$	39.33	33.46
$r \leq 1$	$r > 1$	48.11	47.21	$r = 1$	$r = 2$	24.02*	27.07
$r \leq 2$	$r > 2$	24.09*	29.68	$r = 2$	$r = 3$	18.01	20.97
$r \leq 3$	$r > 3$	6.07	15.41	$r = 3$	$r = 4$	6.04	14.07
Panel B: United States							
$r = 0$	$r > 0$	88.81	68.52	$r = 0$	$r = 1$	33.81	33.46
$r \leq 1$	$r > 1$	55.00	47.21	$r = 1$	$r = 2$	24.73*	27.07
$r \leq 2$	$r > 2$	30.27	29.68	$r = 2$	$r = 3$	17.44	20.97
$r \leq 3$	$r > 3$	12.82*	15.41	$r = 3$	$r = 4$	12.80	14.07

Table 7. Johansen Tests of Cointegration for the multivariate VECM

Second, we estimate the cointegrating rank of the VECM in order to determine the number of cointegrating equations. The results of the cointegration tests are displayed in Table 7 and indicate that the real estate prices, the macroeconomic variables and the market sentiment seem to move together in the long run with multiple cointegrating vectors in each country. The trace and the maximum eigenvalue test statistics indicate that long-term dynamics are present between real estate markets and the prespecified variables. More specifically, the trace test statistics indicate the existence of three cointegrating relationships for the United States and two cointegrating relationships for Spain. In contrast, the maximum eigenvalue test suggests one cointegrating relationship in each country. As the alternative hypothesis is more precisely formulated in the latter test statistic, we continue the analysis assuming one cointegrating relationship in each country. As the variables are cointegrated and therefore have a long term association, we are able to run a VEC model.

Country	Number of Cointegrating Vectors	Real Estate Prices	Consumer Price Index	Short-term Interest Rate	News Sentiment
Spain	1	1.000	49.73 (0.000)	-60.40 (0.001)	-16.52 (0.059)
U.S.	1	1.000	-2.11 (0.000)	-11.06 (0.000)	-0.608 (0.010)

Table 8. Long-term Equilibrium Relationships

The results of the multivariate VECM for the investigation period between January 2003 and April 2012 are summarized in Table 8. As we analyze multidimensional cointegrating relationships, we use hypothesis tests to evaluate whether individual coefficients may be restricted to zero without a significant loss of information. We identify the individual factors that significantly contribute to the explanation of the market-specific equilibrium based on the tests for linear restrictions (LR test) and eliminate only a single regressor in each step. We restrict the long-term interest rate and industrial production in the United States to zero. In these cases, the information is only reported through the respective adjustment parameters.

The estimated long-term relations (β -vectors) between the real estate markets, the respective macroeconomic environment and news sentiment, as summarized in Table 8, reveal similarities in terms of magnitude, sign and significance of the coefficients. While the equilibrium relationships in Spain are determined by real estate prices, the inflation rate, short-term interest rates and news sentiment in the long run, the equilibrium in the United States is specified by real estate prices, inflation and news sentiment.

The β -vector for inflation is positive and therefore affects property prices positively in the long-term. In contrast, short-term interest rates are an indicator for liquidity, impacting property prices negatively in the long-run perspective. Our results indicate a strong positive long-term relationship between consumer prices, that is inflation, and property prices. This confirms the empirical results of Schätz and Sebastian (2009), who emphasize that this association can be interpreted as an indicator for the fact that investments in direct real estate provide opportunities to hedge against inflation.

Our empirical results reveal similarities between Spain and United States regarding the simultaneous long-term impact of news sentiment on house prices. The decomposition of the Π matrix allows for the separate evaluation of the short-run adjustment processes and the long-run relationships. The short-run adjustment processes are described by the α -vectors, which are summarized in Table 9.

Country	Error Correction	D(Real Estate Prices)	D(Consumer Price Index)	D(Short-term interest rate)	D(News Sentiment)
Spain	CE1	-0.076 (0.047)	-0.003 (0.003)	0.002 (0.002)	0.002 (0.066)
U.S.	CE1	-0.038 (0.000)	0.027 (0.071)	0.004 (0.409)	0.030 (0.016)

Table 9. Short-term Adjustment Process

We observe significantly negative coefficients of the normalized variable for the property markets in both countries, which indicates that linear combinations of the variables return to the long-term equilibrium in each case. As the average adjustment period approximately lasts $1/|\alpha|$ months, we conclude that the return to the long-term equilibrium takes one year (13 months) in Spain and two years (26 months) in the United States. The differences in the adjustment periods highlight the basic characteristics, in particular the high level of slackness of the property markets. In addition to the adjustment velocities, the impact of a disequilibrium on the respective macroeconomic environment is quite different.

A disequilibrium caused by news sentiment above its long-term equilibrium level results in a minor positive effect on short-term interest rates and impacts inflation rates and property prices negatively in the Spanish model. Interestingly, a short-term effect between news sentiment and property prices can only be determined for the United States model. We conclude that real estate prices are significantly affected by deviations in news sentiment from the long-term path. As the Spanish financial system is mainly bank-based, real estate financing depends highly on the availability and the conditions of bank loans. Therefore, we observe a highly significant coefficient for short-term interest rates as one of the main factors for the financial system in Spain while the coefficient is not significant for the United States. This result is consistent with the stronger orientation of the financial system in the United States towards capital markets.

Variance Decomposition

We gather further empirical evidence by applying forecast error variance decomposition. This instrument determines which part of the variation in property prices can be attributed to changes in the lagged explanatory variables. More precisely, the forecast error variance decomposition estimates the proportion of forecast error variance of an endogenous variable attributable to orthogonalized shocks to itself or other endogenous variables.

We expect that the forecast error variance decompositions indicate that a notable share of the long-term forecast error variance of real estate returns is explained by news sentiment and that a comparatively small share is explained by macroeconomic fundamentals. If this hypothesis holds, our analysis should indicate that the long-term influence of news on the property market is greater than the impact of the general macroeconomic environment. Nevertheless, the causality may also exist in the opposite direction. We include macroeconomic fundamentals in our model in order to eliminate any indirect effects of those factors on the co-movement between news sentiment and real estate markets. The results of the Cholesky factor error variance decomposition for both orderings are summarized in Table 10.

Months ahead	Property Prices		Consumer Price Index		Short-term Interest Rate		News Sentiment Index	
	Order 1	Order 2	Order 1	Order 2	Order 1	Order 2	Order 1	Order 2
Panel A: Spain								
1	100.0	92.97	0.00	5.18	0.00	1.67	0.00	0.16
3	83.35	71.72	2.82	9.34	13.22	18.56	0.58	0.36
6	67.80	58.91	5.94	8.69	19.17	26.90	7.07	5.48
12	37.97	38.32	24.55	22.07	10.42	13.85	24.59	25.74
24	11.77	7.78	29.24	31.09	14.02	13.88	44.95	47.23
Panel B: United States								
1	100.0	94.26	0.00	0.48	0.00	1.59	0.00	3.65
3	96.96	86.01	0.19	0.13	0.16	1.90	2.60	11.93
6	87.95	72.86	0.12	0.09	0.74	2.24	11.11	24.79
12	43.50	34.99	4.09	3.38	1.71	1.16	50.68	60.45
24	16.23	16.31	37.14	32.46	11.32	13.63	35.29	37.59

Table 10. Forecast Error Variance Decomposition

While the variance in real estate prices can be attributed to lagged property prices in the short run, the relationship weakens with lag length. In contrast, the reverse applies for the other variables, since the variance of those do not account for the variation in property prices in the short run but instead in the long run.

The variance decompositions applied for the two orderings indicate that news sentiment and the consumer price index account for a majority of the variation in the real estate series in both countries. The short-term interest rate and lagged property prices account for less than a third of the property price variation in the long run.

The results of the variance decomposition reveal that a majority of real estate price variations can be attributed to variation in news sentiment with the explanatory power increasing the lag length. Almost half of the house price variation in Spain and more than a third of real estate price variations in the United States can be attributed to news sentiment. To conclude, we observe a remarkably stable and significant explanatory power of news sentiment for the variation of property prices.

Impulse Responses

In addition to forecast error variance decomposition, we estimate impulse responses in order to determine the effect of a one-unit shock of news sentiment on the property prices over time. The impulse responses associated with one unit shocks in news sentiment are illustrated in Figure 4. Considering the direction of the responses, news sentiment always has a positive impact on real estate prices.

The left panel illustrates the impulse response of house prices to a news sentiment shock for Spain. The graph clearly highlights that the impact strengthens as the shock works through and dies out after approximately 36 months. We note that the confidence interval indicates an initial response of 5 months and an additional significant response in the period between 8 and 32 months after the initial shock in a statistical sense. In other words, a shock in news sentiment does not affect house prices immediately. The shock instead takes effect five months after the initial increase. We observe another spike in house prices as some of the feedback effects of the initial shock reverberate throughout the economy with a lag of around 8 months and lasts for another two years.

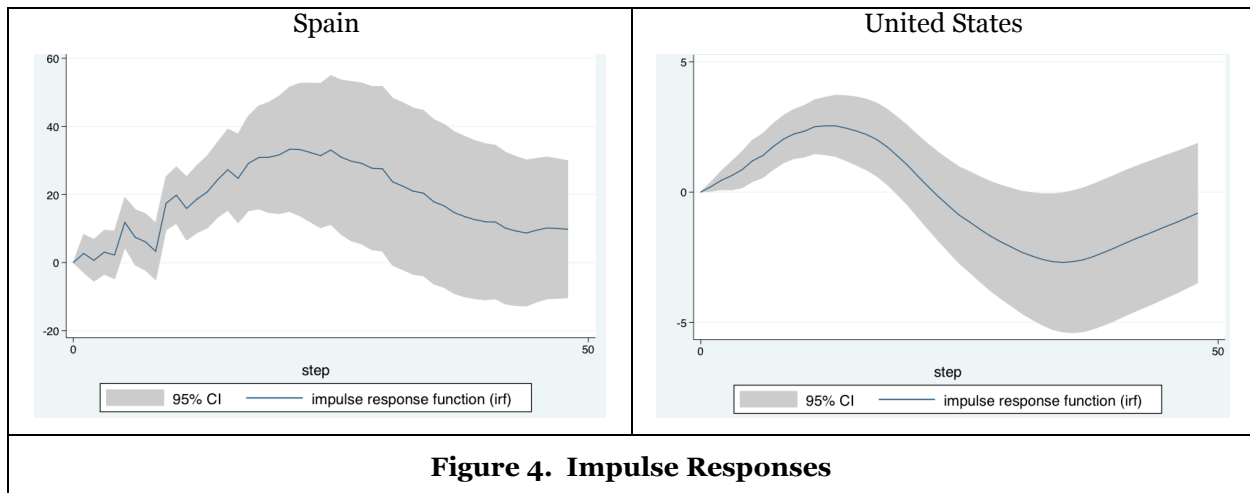


Figure 4. Impulse Responses

We also estimate impulse responses of real estate returns to unanticipated changes in news sentiment for the United States. The estimates are shown in the right panel. Similar to the responses estimated for Spain, the effect of an unexpected increase in news sentiment is positive and statistically significant. Nevertheless, the large confidence intervals which include zero indicate that after an unexpected increase in sentiment, this effect may not materialize. We conclude that as the long-term accumulated responses of the two markets are quite similar, the markets are integrated in the sense that the risk premia for various factors are the same in both markets.

Conclusion and Outlook

The house price bubbles in Spain and the United States resulted in severe consequences that have been transmitted throughout the global economy via contagion effects. Common explanations for the price developments in housing markets range from macroeconomic and financial to psychological aspects. Our study draws the attention to the role of media content in housing markets. We assume a close connection between the role of news sentiment and the real estate markets in both countries despite the obvious cultural differences. The empirical findings are in line with this main hypothesis and provide an additional rationale for the development of the housing markets by analyzing news messages from the last decade.

This study adopts data mining methods to analyze the short- and long-term dynamic between news message content and property prices in Spain and the United States. We construct news sentiment indices based on various text mining methods which exhibit remarkable similarities to the respective property prices. Comparing dictionary-based and dynamic approaches, our results indicate that static methods produce the best estimators for investor sentiment in real estate markets.

Using a Vector Error Correction Model framework, we analyze similarities between both real estate markets in the short- and long-run. The main finding of this paper is a significant relationship between news messages and property prices in the long-run. Our results are stable, including a number of fundamental variables, and are underlined by forecast error variance decompositions and impulse response estimates, which additionally highlight the appropriateness of news sentiment as a crucial determinant for long-term property price variation.

The main contribution of this paper is to explain information processing in real estate markets. We provide empirical evidence for the existence of a significant relationship between news and asset prices. While the relationship between news content and real estate prices in both countries seems evident, the adjustment process to the long-run equilibrium is different. In addition, some control variables also show a significant impact in one country while their effect in the other country is insignificant. For example, the inclusion of the short-term interest rate in the cointegrating equation of Spain can be interpreted as a key determinant in the credit channel, which is predominant in financial systems based on banking finance, such as in Spain. In contrast, the base rate does not exhibit a significant influence in markets which are financed through capital markets as in the United States.

Although the main purpose of this paper is the provision of empirical evidence for the impact of news, the model could also be used for forecasting. Hence, the accuracy of the model could be evaluated by removing data points from the series and by checking the accuracy of the out-of-sample forecast. Nevertheless, this technique requires a larger amount of data for a robust model evaluation and raises additional issues regarding the selection of a reasonable holdout period.

All in all, the presented results indicate a reasonable relation between news and house prices and contain several inspirations for future research. First, we solely concentrate on announcements of news. Therefore, a comparative study on the basis of various media sources might reveal differences in the perception of different types of news messages and the respective content. Second, the pattern of interdependency between both real estate markets and inherent potential contagion effects provide a valuable topics for future research. Third, the introduction and application of novel approaches measuring sentiment might provide additional evidence for the procession of news. However, the coarseness of data on direct house markets impedes the informative value of dynamic sentiment methods.

References

- Antonakis, J., Bendahan, S., Jacquart, P., and Rafael, L. 2014. "Causality and Endogeneity: Problems and Solutions," in *The Oxford Handbook of Leadership and Organizations*. In Press, D. V. Day (ed.), Oxford, UK: Oxford University Press.
- Antweiler, W., and Frank, M. Z. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *The Journal of Finance* (59:3), pp. 1259–1294.
- Brooke, J., Tofiloski, M., and Taboada, M. 2009. "Cross-Linguistic Sentiment Analysis: From English to Spanish," *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria, September 14-16, 2009, pp. 50–54.
- Case, K. E., and Shiller, R. J. 1989. "The Efficiency of the Market for Single-Family Homes.," *The American Economic Review* (79:1), pp. 125–137.
- Dadvar, M., Hauff, C., and de Jong, F. 2011. "Scope of Negation Detection in Sentiment Analysis," in *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR 2011)*, Amsterdam, Netherlands. February 04, 2011, Amsterdam, Netherlands, pp. 16–20.
- DeLong, J. B., Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann 1990. "Noise Trader Risk in Financial Markets," *Journal of Political Economy* (98:4), pp. 703–738.
- Demers, E. A., and Vega, C. 2010. "Soft Information in Earnings Announcements: News or Noise?" *SSRN Electronic Journal*.
- Dickey, D. A., and Fuller, W. A. 1979. "Distribution of estimators for time series regression with a unit root," *Journal of the American Statistical Association* (74), pp. 417–431.
- Feuerriegel, S., and Neumann, D. 2013. "News or Noise? How News Drives Commodity Prices," in *Proceedings of the International Conference on Information Systems (ICIS 2013)*, Milan, Italy. December 15-18, 2013, Association for Information Systems.
- Granger, C. W. J. 1981. "Some properties of time series data and their use in econometric model specification," *Journal of Economics and Statistics* (48), pp. 121–130.
- Groth, S. S., and Muntermann, J. 2008. "A Text Mining Approach to Support Intraday Financial Decision-Making," in *Proceedings of the 14th Americas Conference on Information Systems (AMCIS 2008): Learning from the past & charting the future of the discipline*, A. R. Montazemi, and I. Benbasat (eds.), Toronto, Ontario, Canada. August 14-17 2008, Association for Information Systems.
- Groth, S. S., and Muntermann, J. 2010. "Discovering Intraday Market Risk Exposures in Unstructured Data Sources: The Case of Corporate Disclosures," in *43rd Hawaii International Conference on System Sciences (HICSS)*, Honolulu, Hawaii, USA. January 5-8, 2010, pp. 1–10.
- Groth, S. S., and Muntermann, J. 2011. "An intraday market risk management approach based on textual analysis," *Decision Support Systems* (50:4), pp. 680–691.
- Hagenau, M., Hauser, M., Liebmann, M., and Neumann, D. 2013a. "Reading All the News at the Same Time: Predicting Mid-term Stock Price Developments Based on News Momentum," in *46th Hawaii International Conference on System Sciences (HICSS)*, Wailea, Hawaii, USA. January 7-10, pp. 1279–1288.
- Hagenau, M., Korczak, A., and Neumann, D. 2012. "Buy on bad news, sell on good news: How insider trading analysis can benefit from textual analysis of corporate disclosures," *Workshop on Information Systems and Economics (WISE 2012)*, Orlando, Florida, USA.

- Hagenau, M., Liebmann, M., and Neumann, D. 2013b. "Automated news reading: Stock Price Prediction based on Financial News Using Context-Capturing Features," *Decision Support Systems* .
- Henry, E. 2008. "Are Investors Influenced By How Earnings Press Releases Are Written?" *Journal of Business Communication* (45:4), pp. 363–407.
- Jegadeesh, N., and Wu, D. 2011. "Word Power: A New Approach for Content Analysis," *SSRN Electronic Journal* .
- Kahneman, D., and Tversky, A. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* (47:2), pp. 263–292.
- Keynes, J. M. 1936. *The General Theory of Employment, Interest and Money*, London.
- Lewis, D. D., Yang, Y., Rose, T. G., and Li, F. 2004. "RCV1: A New Benchmark Collection for Text Categorization Research," *Journal of Machine Learning Research* (5), pp. 361–397.
- Li, F. 2010. "The Information Content of Forward-Looking Statements in Corporate Filings: A Naïve Bayesian Machine Learning Approach," *Journal of Accounting Research* (48:5), pp. 1049–1102.
- Liebmann, M., Hagenau, M., and Neumann, D. 2012. "Information Processing in Electronic Markets: Measuring Subjective Interpretation Using Sentiment Analysis," in *Proceedings of the International Conference on Information Systems (ICIS 2012)*, Orlando, USA. December 16-19, 2012, Association for Information Systems.
- Loughran, T., and McDonald, B. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance* (66:1), pp. 35–65.
- Manning, C. D., and Schütze, H. 1999. *Foundations of statistical natural language processing*, Cambridge, Mass: MIT Press.
- McCue, T. E., and Kling, J. L. 1994. "Real estate returns and the macroeconomy. some empirical evidence from real estate investment trust data," *Journal of Real Estate Research* (9), pp. 277–287.
- Minev, M., Schommer, C., and Grammatikos, T. 2012. "News and stock markets: A survey on abnormal returns and prediction models," *Technical Report*, University of Luxembourg, Luxembourg.
- Mittermayer, M.-A. 2004. "Forecasting Intraday stock price trends with text mining techniques," in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, R. H. Sprague (ed.), Island of Hawaii, Big Island, HI, USA. January 5-8, 2004, Los Alamitos, Calif.: IEEE Computer Soc.
- Mittermayer, M.-A., and Knolmayer, G. F. 2006. "Text Mining Systems for Market Response to News: A Survey," *Working Paper* 184, Institute of Information Systems, University of Bern, Bern, Switzerland.
- Muth, J. F. 1961. "Rational Expectations and the Theory of Price Movements," *Econometrica* (29:3), pp. 315–355.
- Oh, C., and Sheng, O. R. L. 2011. "Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement," in *Proceedings of the International Conference on Information Systems (ICIS 2011)*, D. F. Galletta, and T.-P. Liang (eds.), Shanghai, China. December 4-7, 2011.
- Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *FNT in Information Retrieval (Foundations and Trends in Information Retrieval)* (2:1–2), pp. 1–135.
- Porter, M. F. 1980. "An algorithm for suffix stripping," *Program* (14:3), pp. 130–137.
- Reuters 2008. "Reuters NewsScope Archive v2.0: User Guide,"
- Schätz, A., and Sebastian, S. 2009. "The links between property and the economy – evidence from the British and German markets," *Journal of Property Research* (26:2), pp. 171–191.
- Schumaker, R. P., and Chen, H. 2009. "Textual analysis of stock market prediction using breaking financial news," *ACM Transactions on Information Systems* (27:2), pp. 1–19.
- Sharma, A., and Dey, S. 2012. "A comparative study of feature selection and machine learning techniques for sentiment analysis," in *Proceedings of the 2012 Research in Applied Computation Symposium (RACS 2012)*, Y. Cho, and A. S. I. G. o. A. Computing (eds.), San Antonio, Texas. October 23-26, 2012, New York, NY: ACM, pp. 1–7.
- Shiller, R. J. 2005. *Irrational Exuberance*: Princeton University Press.
- Siering, M. 2012. "'Boom' or 'Ruin' -- Does It Make a Difference? Using Text Mining and Sentiment Analysis to Support Intraday Investment Decisions," in *45th Hawaii International Conference on System Sciences (HICSS)*, Maui, Hawaii, USA. January 04-January 07, pp. 1050–1059.
- Siering, M. 2013. "Investigating the Impact of Media Sentiment and Investor Attention on Financial Markets," in *Enterprise Applications and Services in the Finance Industry*, W. Aalst, J. Mylopoulos, M. Rosemann, M. J. Shaw, C. Szyperski, F. A. Rabhi, and P. Gomber (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3–19.

- Siering, M., and Muntermann, J. 2013. "The Role of Misbehavior in Efficient Financial Markets: Implications for Financial Decision Support," in *Enterprise Applications and Services in the Finance Industry*, W. Aalst, J. Mylopoulos, M. Rosemann, M. J. Shaw, C. Szyperski, F. A. Rabhi, and P. Gomber (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 42–58.
- Soo, C. 2013. "Quantifying Animal Spirits: News Media and Sentiment in the Housing Market," The Wharton School, University of Pennsylvania.
- Stiglitz, J. E. 2009. "The global crisis, social protection and jobs," *International labour review* (1), pp. 1–12.
- Tetlock, P. C. 2007. "Giving content to investor sentiment: The role of media in the stock market," *The Journal of Finance* (62:3), pp. 1139–1168.
- Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. 2008. "More than words: Quantifying Language to Measure Firms' Fundamentals," *The Journal of Finance* (63:3), pp. 1437–1467.
- Yu, L., Wang, S., and Lai, K. K. 2005. "A Rough-Set-Refined Text Mining Approach for Crude Oil Market Tendency Forecasting," *International Journal of Knowledge and Systems Sciences* (2:1), pp. 33–46.