Working Paper

Quantifying Animal Spirits: News Media and Sentiment in the Housing Market

Cindy K. Soo
Stephen M. Ross School of Business
University of Michigan

Ross School of Business Working Paper
Working Paper No. 1200
August 2013

QUANTIFYING ANIMAL SPIRITS:

NEWS MEDIA AND SENTIMENT IN THE HOUSING MARKET*

Cindy K. Soo†

University of Michigan, Ross School of Business

August 31, 2013

Abstract

Sentiment or "animal spirits" has long been posited as an important determinant of asset prices, but measures of sentiment are particularly difficult to construct for the housing market. This paper develops the first measures of sentiment across local housing markets by quantifying the positive and negative tone of housing news in local newspaper articles. I use these measures to test the role of sentiment in the run-up and crash of housing prices that instigated the great financial crisis of 2008. I find that my housing sentiment index forecasts the boom and bust pattern of house prices at a two year lead, and can predict over 70 percent of the variation in aggregate house price growth. Consistent with theories of investor sentiment, I find that my sentiment index not only predicts price variation but also patterns in trading volume. Estimated effects of sentiment are robust to an extensive list of observed controls including lagged fundamentals, lagged price growth, subprime lending patterns, and news content over typically unobserved variables. To address potential bias from latent fundamentals, I develop instruments from a subset of weekend and narrative articles that newspapers use to cater to sentiment but are plausibly exogenous to news on fundamentals. Estimates remain robust to instrumental variable estimation, suggesting bias from unobserved fundamentals is minimal.

*I am deeply grateful to Fernando Ferreira, Joe Gyourko, Olivia Mitchell, Michael Roberts, and Todd Sinai for their invaluable guidance and encouragement. I am also very thankful to Nick Barberis, Mark Duggan, Alex Edmans, Alex Gelber, Joao Gomes, Todd Gormley, Vincent Glode, Daniel Gottlieb, Philipp Illeditsch, Raimond Maurer, Greg Nini, Christian Opp, David Musto, Devin Pope, Nikolai Roussanov, Robert Stambaugh, Robert Shiller, Kent Smetters, Luke Taylor, Jeremy Tobacman, Maisy Wong, Yu Yuan, and seminar participants at the Wharton applied economics and finance seminars, UC-Berkeley, George Mason University, Harvard Business School, HEC Paris, University of Illinois at Urbana-Champaign, Miami University, University of Michigan, Michigan State, New York University, Washington University at St. Louis, and Federal Reserve Banks of Boston, Chicago, and Philadelphia for their helpful suggestions and comments. I owe special thanks to Eugene Soltes and Kumar Kesavan for their assistance on data collection. All errors are my own.

†University of Michigan, Ross School of Business, 701 Tappan Street, Ann Arbor MI 48109-1234, csoo@umich.edu

1 Introduction

From 2000-2005, U.S. home prices experienced unprecedented rapid price growth. Standard and Poor's/Case-Shiller 20 City Composite Index increased by 46 percent, with prices in some cities such as Los Angeles and Miami nearly doubling. Prices began to drop in 2006, and by 2009, the Case-Shiller Index reported price declines of over 33 percent. The housing bust devastated millions of homeowners, and foreclosures tripled to 3.9 million between 2006 and 2009. The impact of price declines rippled beyond the housing market, as banks and financial institutions held significant investments in mortgage-backed securities and other housing related assets. The collapse of the sub-prime mortgage industry quickly followed, leading the nation into its worst economic recession since the 1930s.

Standard economic explanations for house price changes have been difficult to reconcile with the wide swings during this period. Observed fundamentals that traditionally explained historical patterns in house prices, accounted for just a small fraction of house price changes post-2000 (Lai and Van Order (2010)). Many hypothesized low interest rates likely sparked an increase in housing demand, however studies such as Glaeser, Gottlieb and Gyourko (2010) find that changes in real rates explain just one-fifth of the rise in prices from 1996 to 2006. Alternatively, Shiller (2009) and others argued the "animal spirits" or irrational exuberance of investors played a significant role in the dramatic boom and bust of housing prices. While there was much discussion over the potential role of sentiment in the last housing cycle, testing this hypothesis has been difficult as sentiment measures for the housing market are particularly hard to attain. Typical sentiment proxies for the stock market such as mutual fund flows, dividend premia, and close-end fund discount are naturally not available for the housing market (Baker and Wurgler (2006)). Surveys provide direct assessments of home buyer expectations, but testing its effect on house prices is restricted due to limitations in geographic breakdown or time frequency.[1]

This paper proposes a new measure of housing market sentiment that is available across individual cities and at a monthly frequency. I measure sentiment by capturing the qualitative tone of housing news from local newspapers. Specifically, I calculate the difference between the share of

[1]For example, Case and Shiller (2003) survey consumer expectations in the housing market, but is limited to four suburban areas and an annual snapshot in time. The University of Michigan/Reuters Survey of Consumers collect data on home buyer expectations monthly, but is based on a survey of a nationally representative sample of 500 individuals and cannot be broken down geographically.

positive and negative words across local housing news articles each month. I construct local housing sentiment indices that correspond to each of the 20 city markets covered by the Case-Shiller home price index. This strategy is motivated by seminal literature on asset price bubbles that argue the news media reflects investor sentiment through an incentive to cater to readers' preferences over a particular asset (Kindleberger (1978); Galbraith (1990); Shiller (2005)). This methodology also builds on work from Tetlock (2007) and a growing number of studies that construct proxies for sentiment in the stock market with media coverage.

I find that my housing sentiment index exhibits striking patterns that forecast the boom and bust trend of housing prices from 2000 to 2011. Figure I shows that my national sentiment index increases rapidly and peaks in 2004, two years ahead the peak of national house prices in mid-2006. These patterns appear consistent across individual cities as well. Cities that experienced dramatic rises and declines in house prices are preceded by similar cycles in sentiment, whereas cities with milder price changes are led by more subdued sentiment growth. Empirically, I find that the housing sentiment index can explain over 70 percent additional variation in national house price movements above and beyond observed fundamentals. This is significant as prior studies have so far found alternative explanations to account for just a limited fraction of house price variation after 2000.[2]

To assess whether this index is a useful measure of sentiment, I validate my measure of sentiment against surveys of investor expectations in the housing market. While these surveys cannot be broken down by city, I can compare the trends in survey measures with my national sentiment index. I find that my housing sentiment index is highly correlated with housing market confidence indexes from the Survey of Consumers and the National Association of Home Builders. In particular, home buyer survey confidence also peaks in 2004, reflecting similar timing to trends in my composite index. Case, Shiller and Thompson (2012) implement annual surveys of home buyer expectations and similarly find that long term expectations peak in 2004, well ahead of house prices. These survey measures reaffirm the overall time-varying trends in my housing sentiment index and suggest the index is capturing the pattern in housing expectations.

Still, all of these measures may potentially capture variation in fundamentals. I first address

[2]For example, Zhu, Wright and He (2012) examine the role of liquidity in the housing boom, and also find that their model can account for approximately one-fifth of house price run up from 1996 to 2006.

this by controlling for an exhaustive sequence of fundamental determinants of house prices. I find that the predictive power of sentiment on house prices not only remains robust in significance, but also in magnitude. The magnitude of the estimated effect of sentiment remains highly stable to the sequential addition of observed fundamentals, suggesting that bias from unobservable factors is less likely. I further test my observed set of fundamentals by examining their explanatory power prior to 2000. The same set of controls that account for less than 10 percent of house price changes from 2000 to 2011, explain nearly 70 percent of the variation in national house price growth from 1987 to 2000. This suggests that my observed controls are not missing key fundamentals that traditionally explained house price changes.

In addition, I include observed variables that control for changes in subprime lending trends. While not considered a typical housing fundamental, subprime credit exhibited unprecedented expansion that coincided with the growth of house prices in many cities (Mian and Sufi (2009); Demyanyk and Van Hemert (2011); Goetzmann, Peng and Yen (2012)). I find that the housing sentiment index predicts house price changes above and beyond subprime lending patterns. Furthermore, I find that sentiment not only predicts house price variation but also patterns in transaction volume. This result is consistent with existing theories and empirical studies of investor sentiment (Odean (1998, 1999); Scheinkman and Xiong (2003); Barber and Odean (2000, 2008)). Interestingly, sentiment leads volume first and is followed by prices another year later. This evidence supports a hypothesis that search frictions in the housing market likely induce lags between changes in sentiment, housing transactions, and prices.

While these results are highly suggestive, the positive association between my sentiment index and house prices may still be driven by latent fundamentals. In particular, the news may report on changes in harder-to-quantify fundamentals that I do not observe. The richness of my news dataset, however, allows me to control for the content of news articles directly. Using a text search algorithm, I isolate any article that directly mentions a fundamental within its text. For example, I extract all articles that mention words related to housing fundamentals such as "taxes," "credit," or "mortgage rates." I then control for the share of positive and negative words within these fundamental new stories and find that this does not affect my results.

I then present two candidate instruments for sentiment by isolating a subset of housing news articles that cater to reader sentiment but are plausibly exogenous to news on fundamentals. The first is my measure of sentiment calculated only over housing articles published over the weekend. Weekend articles tend to cater to readers who have preferences for lighter content, and are arguably exogenous to news on fundamentals since official press releases on economic data can only occur on a weekday. The second proposed instrument is my measure of sentiment calculated only over narrative housing news articles. Narratives cater to sentiment through a human interest appeal, and are plausibly exogenous to fundamentals because they consist of anecdotal stories rather than actual information. Of course, the validity of these instruments relies on the assumption that information on fundamentals is not being reported on or somehow related through these subset of news articles. I acknowledge and test for a number of possible violations of this assumption, and find that results are consistent with the exclusion restriction. Given this, I show that the predictive power of sentiment remains robust both in significance and magnitude even after instrumenting for sentiment.

This paper provides evidence that sentiment may have a significant effect on house prices, and challenges standard explanations of the housing boom and bust that rely solely on fundamentals. The results of this paper suggest that if a fundamental drove house prices during this period, then it would also have had to drive expectations at a two year lead to prices both nationally and across cities. Furthermore, to be consistent with the empirical data, this fundamental would fail to explain prices from 1987 to 2000 but suddenly begin to drive expectations and prices differently from 2000 to 2011. This paper does not advocate that fundamentals did not play any role, but that the evidence suggests sentiment played an economically important role as well.

This paper contributes a new measure of sentiment to a larger literature that explores housing price dynamics and provides evidence that supports studies that explore the role of expectations in the housing market (Genesove and Mayer (2001); Piazzesi and Schneider (2009); Goetzmann, Peng and Yen (2012); Arce and López-Salido (2011); Burnside, Eichenbaum and Rebelo (2011); Favilukis, Ludvigson and Nieuwerburgh (2010)). The findings in this paper also relate to a large body of work that explores determinants and consequences of the last housing boom and bust, and uncovers a measure that can predict a significant fraction of variation in national house prices (Piskorski, Seru and Vig (2010); Avery and Brevoort (2010); Haughwout et al. (2011); Bhutta (2009); Bayer, Geissler and Roberts (2011); Glaeser, Gyourko and Saiz (2008); Gerardi et al. (2008); Ho and Pennington-Cross (2008)). Finally, these findings complement a number of empirical studies that attempt to quantify sentiment and provide evidence for its effect on asset prices (Edmans, Garcia and Norli (2007); Baker and Wurgler (2006, 2007); Baker, Wurgler and Yuan (2012); Baker and Stein (2004); Greenwood and Nagel (2009); Barber, Odean and Zhu (2009); Brown and Cliff (2005)).

Section 2 describes how I construct my database of newspaper articles and set of observed fundamentals. Section 3 details how the sentiment index is calculated. Section 4 and 5 present the main empirical and instrumental variable results respectively. Section 6 concludes and discusses potential avenues for future work.

2 Data Description

2.1 Newspaper Articles

My approach to measuring sentiment requires the text of newspaper articles covering the housing market. My source for news articles is Factiva.com, a comprehensive online database of newspapers.[3] Factiva categorizes its articles by subject, and provides a code that identifies articles that discuss local real estate markets. This code is determined by a propriety algorithm that remains objective across all newspapers and years. This subject code covers new and existing home sales, housing affordability indices, and housing price indices as well as supply side indicators on housing starts, building permits, housing approvals, and construction spending. Routine real estate property listings are not included. Wire-service articles are also generally excluded, as syndicated stories cannot be redistributed and typically do not appear in the Factiva database. This exclusion is actually preferable to capturing the *local* sentiment unique to each city. Wire-service articles are typically those that cover topics of more general national interest, supplied to local newspapers by large media companies such as the Associated Press. Excluding such articles ensures each city's sentiment measure is only based on news articles written by local staff writers. To that end, I also exclude any additional republished or duplicate news stories from other news outlets.[4]

I download all newspaper articles covering the housing market between January 2000 and August 2011 from the major newspaper publication in each of the following 20 cities: Atlanta, Boston,

[3] Other similar newspaper databases are Lexis Nexis and NewsBank. Factiva.com arguably has the most comprehensive coverage.

[4] I do not, however, exclude stories that are written by local staff writers but may comment on the housing market of other cities. While an article may comment on other cities, publication of these articles may be in response to a *local* interest in reading housing news. In a follow up paper, I provide evidence that suggests news mentions of other cities is a mechanism through which a contagion of sentiment is spread.

Charlotte, Chicago, Cleveland, Dallas, Denver, Detroit, Las Vegas, Los Angeles, Miami, Minneapolis, New York, Phoenix, Portland, San Francisco, San Diego, Seattle, Tampa, and Washington, D.C. I retrieve a total of 19,620 articles.

I then apply a second automated script to parse information from each article. I not only extract the text of the articles, but also useful information on the the date, headline, author, section, and copyright. My database contains each individual word of an article with its corresponding date, word position, author, and originating newspaper. My final dataset consists of a total 15,295,393 words. I then implement a final script that produces counts of positive and negative words and total words across housing articles by city and month.

Table I summarizes some descriptive statistics on the collected articles by city. Most cities have one major newspaper that dominates the news market, with the exception of Boston, Detroit, and Los Angeles, which have two. Some Associated Press articles remain in the sample, but make up less than 6 percent of the collected articles. Approximately 20 percent of the articles are found in the front or "A" section of the newspapers. Additionally, 20 percent are found in a special real estate section. Furthermore, over 30 percent of the articles are published in local news or regional editions of the newspaper. Otherwise, the majority of articles are reported in a general news or business section.

2.2 Housing Fundamentals and Additional Variables

The goal of this paper is to identify an effect of sentiment on house prices. However if housing market fundamentals also affect my news sentiment proxy, then estimating an effect of sentiment on house prices will suffer from omitted variable bias. In particular, a positive shock to fundamentals may simultaneously drive both sentiment and prices upward, biasing coefficient estimates upward. Thus, controlling for these fundamentals is key to identification. Since the true model of house prices is unknown, I apply a "kitchen sink" approach and assemble as many housing market inputs and ouputs that may account for the variation in house prices.

Rents. The "fundamental value" of an asset typically refers to its present discounted value of future cash flow. As described in the Appendix, the model assumes housing pays dividends in the form of rental services. I acquire measures of monthly rents from two sources: REIS and the Bureau of Labor Statistics (BLS). REIS provides average asking rents on rental units with common characteristics with single family homes. REIS reports monthly data on actual rental values which I normalize to match price indexes (100=January 2000). I also obtain residential rents from the Consumer Price Index Housing Survey implemented by the BLS. The BLS reports rents of primary residences as a part of the shelter component of the consumer price index. I include the BLS measure of rents as a robustness check and report the results using REIS rental indices.

Supply. I measure changes in housing supply using data on building permits and housing starts for the U.S. Census Bureau. Housing starts are the total new privately owned housing units started each month. Building permits are those authorized for new privately owned housing units in each city. I also include a measure of supply elasticity developed by Saiz (2010) with the Wharton Residential Land Use Regulatory Index (WRLURI) created by Gyourko, Saiz and Summers (2008).

Employment and Unemployment. A number of models highlight the importance of labor market variables on housing demand (Roback (1982); Rosen (1979); Nakajima (2011); Mankiw and Weil (1989)). I attain monthly employment levels and local unemployment rates by city from the BLS. I also test various measures of employment such as civilian labor force, or employment rates by particular sector, age, and industry.

Population and Income. I attain measures of income and population growth by city from the Bureau of Economic Analysis (BEA). I also use income data on loan applicants from the Home Mortgage Disclosure Act (HMDA). HMDA requires lending institutions file reports on all mortgage applications, and thus provides an exceptional profile of the pool of potential home buyers.

Interest Rates. A large focus of the debate over the housing crisis has been on the role of low real interest rates and availability of easy credit. Theory shows that low interest rates should lead to increased housing demand and higher prices (Himmelberg, Mayer and Sinai (2005); Mayer and Sinai (2009); Taylor (2009)). I include measures of both real and nominal interest rates relevant to home buyers. I use the national 30-year conventional mortgage rate from the Federal Reserve Board. Following Himmelberg, Mayer and Sinai (2005), I calculate real interest rates by subtracting the Livingston Survey 10-year expected inflation rate from the 10-year Treasury bond rate. The standard user cost formula of housing suggests a 10-year rate, rather than a short-term rate, is more sensible when approximating the duration of mortgages. I also include measures of the the 10-year treasury bill rate and the 6-month London Interbank Offered Rate (LIBOR).

Subprime Lending and Leverage. Studies also hypothesize that the availability of credit should boost housing demand and prices are likely more sensitive in cities where homeowners are highly leveraged (Stein (1995); Lamont and Stein (1999)). Thus, I attain loan-to-value ratios come from a comprehensive new micro dataset provided by DataQuick, an industry data provider (Ferreira, Gyourko and Tracy (2010)). DataQuick provides detailed transaction level data on over 23 million arms length housing transaction from 1993 to 2009. Loan-to-value ratios include the total amount of mortgage debt including not only the primary but also any debt up to three loans taken to finance the home. This dataset covers transactions cover 16 cities in my sample. I also use the percent of subprime mortgages as calculated by Ferreira and Gyourko (2012). The share of subprime loans in a city is the share of loans issued by any of the top twenty subprime lenders ranked by the publication *Inside Mortgage Finance*.

Housing Prices and Volume. I measure home prices for each city from 2000 to 2011 with monthly indexes calculated by Standard & Poor's/Case-Shiller home price index. I use their composite-20 home price index to measure aggregate prices. The S&P/Case-Shiller price indices estimate price changes with repeat sales to control for the changing quality of houses being sold through time. The overall average price index over all twenty cities is 147.3, with the highest, 280.9, occurring in Miami December 2006 and the lowest hitting 67.68 in Detroit the March of 2010. The Case-Shiller Composite 20 index aggregates prices of all 20 major metropolitan areas into composite index and has a slightly higher mean of 157.2 with less variance over time. As a further robustness check, I also test quarterly home price indices calculated by the Federal Housing Finance Agency (FHFA). Since DataQuick covers transaction level data across cities, I also calculate the volume of transactions as an additional dependent variable. This dataset covers transactions for most of cities in my sample and is available monthly.

3 Measuring Housing Sentiment with News

3.1 News Media Relationship with Sentiment

This paper uses the tone of news media to capture the average level of investor sentiment in the housing market. Seminal literature on bubbles and panics commonly stress that the news media has an important relationship with investor beliefs (Kindleberger (1978); Galbraith (1990); Shiller (2005)). They argue that newspapers have a demand-side incentive to cater to reader preferences, and will spin news according to readers' opinion over assets they own. According to Shiller (2005), housing in particular receives heavy media attention being "a source of endless fascination for the general public, because we live in houses, we work on them every day."

Formal studies of media slant have found both theoretical and empirical basis to support similar arguments in the context of readers' political preferences. Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006) assume that readers have a disutility for news that is inconsistent with their beliefs, citing psychology literature that show people have a tendency to favor information that confirms their priors.[5] Gentzkow and Shapiro (2010) find empirical evidence that readers have a preference for news consistent with their beliefs and news outlets respond accordingly. For those interested, the appendix provides a formal framework that uses similar arguments to illustrate how news relates to housing sentiment and prices.

These studies assume that news reflects investor sentiment. Nonetheless, Shiller (2005) argues that news media can simultaneously fuel sentiment if readers misperceive optimism in the news for real information about fundamentals the housing market. Housing, in particular, is a widely held household investment by individual buyers. Thus the average housing investor is likely less financially sophisticated than the typical stock market investor. News slant can make it difficult for even more sophisticated readers to separate true information from sentiment and can subsequently affect trading behaviors. Indeed, additional empirical studies on political media slant find evidence that media has been able to shift public opinion and voting behavior (DellaVigna and Kaplan (2007); Gerber, Karlan and Bergan (2009)).

Thus, it is unclear whether the news media would necessarily lead or lag sentiment in the housing market. The true relationship is likely more complicated, potentially one where news media both feeds and fuel reader beliefs and preferences. While the findings of this paper raise important questions for the relationship between investor sentiment and media, the goal of this paper is to develop a robust proxy for housing sentiment. As long as media slant is positively associated with sentiment, the news media provides a way to capture an empirical measure of investor sentiment.

[5] This tendency is called confirmatory bias in the psychology literature (Lord (1979); Yariv (2002)).

3.2 Textual Analysis of News Articles

I capture news sentiment through a textual analysis of newspaper articles. Textual analysis is a increasingly popular methodology used to quantify the tone and sentiment in financial documents.[6] For example, a number of finance and accounting studies have applied textual analysis techniques to capture the tone of earnings announcements, investor chat rooms, corporate 10-K reports, IPO prospectuses, and newspaper articles (Engelberg (2008); Antweiler and Frank (2004); Li (2006); Loughran and Mcdonald (2011); Tetlock (2007); Jegadeesh and Wu (2011); Hanley and Hoberg (2010); Kothari, Li and Short (2009); Feldman and Segal (2008); Henry (2008)). Many of these papers have linked the sentiment of these documents to outcomes such as firm earnings, stock returns, and trading volume. Tetlock (2007), one of the most well known of these papers, quantifies the negative tone of the popular Wall Street Journal newspaper column "Abreast the Market." His results support the tone of news as as robust proxy for stock market sentiment.

I apply the most standard methodology employed by this literature, which quantifies the raw frequency of positive and negative words in a text. These papers typically identify words as positive or negative based on an external word list. External word lists are preferred because they are predetermined and less vulnerable to subjectivity from the author. A number of previous papers start with general positive or negative word lists provided by *Harvard IV-4 Psychological Dictionary*. Existing studies have found, however, that these general tonal lists can contain irrelevant words and lead to noisy measures (Tetlock, Saar-Tsechansky and Macskassy (2008)). For example, Engelberg (2008) points out words on the general Harvard positive list such as *company* or *shares* have limited relevance in capturing positive tone and can unintentionally capture other effects in finance applications. Indeed, several papers have specifically found limited use for the general Harvard positive list (Tetlock (2007); Engelberg (2008); Kothari, Li and Short (2009)). A recent study by Loughran and Mcdonald (2011) shows that the noise introduced by the general Harvard negative word list can also be substantial and argues that word lists should be discipline-specific to reduce measurement error.

To balance these concerns, I still use a predetermined list from the Harvard IV-4 dictionary to reduce subjectivity, but choose one that specifically reflects how the media spins excitement over

[6] Alternative labels for textual analysis are content analysis, natural language processing, or information retrieval.

asset markets. Shiller (2008) asserts that "the media weave stories around price movements, and when those movements are upward, the media tend to embellish and legitimize 'new era' stories with extra attention and detail." He argues that the media employs superlatives that emphasize price increases and upward movements. For example, a news article may describe markets as "*skyrocketing,*" "*soaring*," "*booming*" or "*heating up*." For this reason, I use the Harvard IV-4 lists *Increase* and *Rise*, words associated with increasing outlook and rising movement.[7] Nonetheless, these lists still include a few words such as *people* and *renaissance* that are clearly irrelevant and would result in obvious misclassifications. I manually remove these words, but simultaneously expand the remaining words with their dictionary synonyms.[8] For example, *skyrocket* is a synonym of *soar*, but not included in the original Harvard lists. I exclude synonyms that correspond to an alternative definition of the original word. Following Loughran and Mcdonald (2011), I also expand the list with inflections and tenses that retain the original meaning of each word. Thus counts for the root word *skyrocket*, for example, also include *skyrockets, skyrocketed,* and *skyrocketing*. The original Harvard IV-4 lists include 136 words and the expanded list, including inflections and synonyms, contains 403 words. Table II reports a sample of positive words and their corresponding word counts. I repeat the above process to create negative word lists using the converse Harvard IV-4 lists *Decrease* and *Fall*.

3.3 Calculating the Sentiment Index

Using an automated script, I generate counts of positive words by city and month. I calculate the fraction of positive words in city i and month t by simply dividing the number of positive words by the total number of words each month. The share of positive words is represented by:

$$Pos_{it} = \frac{\#positivewords}{\#totalwords}_{it} \tag{1}$$

An alternative method is to calculate the share of positive words in each individual article and then average across articles; I try both methods and they do not make a difference in values. To be conservative, I focus my analysis and report my results based on the leading text of an article. An

[7]These lists can be found at http://www.wjh.harvard.edu/∼inquirer/Increas.html and http://www.wjh.harvard.edu/∼inquirer/rise.html.

[8]My dictionary source for synonyms is *Rogets 21st Century Thesaurus, 3rd Edition* ((2012)).

article may intend to express a negative tone with the first half of its text, but contain a number of positive words in the latter half. Thus, tabulating word counts over the full text can potentially overestimate the share of positive words. Nevertheless, the share of positive words based on the full text of the articles is highly correlated with the share based on the leading text.

Still, positive words in a text may be simultaneously surrounded by a number of negative words. I address this issue by subtracting the share of negative words from the share of positive words. I define the fraction of negative words by the analogous expression:

$$Neg_{it} = \frac{\#negativewords}{\#totalwords}_{it} \tag{2}$$

and define the housing news sentiment index by:

$$S_{it} = Pos_{it} - Neg_{it} \tag{3}$$

where i and t denote the city and month respectively. I additionally adjust both negative and positive word counts for negation using the terms: *no, not, none, neither, never, nobody.* I consider a word negated if it is preceded within five words by one of these negation terms.[9] Finally, I apply a backwards 3-month moving average to smooth the series and reduce noise.[10] The window for each reporting month is based on data for that month and the preceding two months. This mirrors the same 3-month moving average used to calculate the S&P/Case-Shiller home price indices. In addition, I apply the same normalized weights used to create the Case-Shiller Composite-20 home price index to create an analogous Composite-20 housing sentiment index.

I create a number of alternate versions of the baseline index sentiment index for robustness. For example, I calculate a version of the index that uses the full, rather than just the leading, text of the articles. I also construct a version that accounts for not only the tone of news, but also the frequency of housing articles published each month. Loughran and Mcdonald (2011) also suggest a "term-weighted" index that adjusts for the commonality and frequency of a word across

[9] Loughran and Mcdonald (2011) apply the same strategy except with a preceding word distance of three words. Textual analysis studies in the computer science field use a preceding distance of five words, so I opt for the wider window.

[10] Baker, Bloom and Davis (2012) suggest a 36-month backward moving average to smooth a monthly series of an economic policy uncertainty index.

documents.[11] I find that the results remain robust to these alternative versions. Details on alternate versions and their correlations with the baseline index are available in the Appendix.

3.4 Validating Sentiment Index Patterns

Figure III plots my composite-20 housing news sentiment index with the Case-Shiller composite-20 housing price index across time. My housing news sentiment index exhibits a striking boom and bust pattern, and appears to lead the rise and fall of aggregate housing prices by more than two years. My sentiment index peaks in January 2004, while the housing price index peaks 30 months (2.5 years) later in July 2006. The lead time between my sentiment index and house prices seems striking, particulary in comparison to the stock market where sentiment predicts prices over just several days. However, the transaction process of buying a home is considerably longer than that of the stock market, and the search process for a home can actually take several months. Furthermore if news slant does feed sentiment, then this can also take some time to diffuse and spread across investors.[12]

This aggregate lead pattern is driven by similar patterns across in individual cities. Figure II plots individual sentiment indexes across time for a sample of six cities. As in the composite index, cities such as Las Vegas and Phoenix that experienced large swings in house prices were preceded by similar swings in news sentiment. Conversely, cities with more moderate increases in housing prices such as Atlanta and Minneapolis, do not appear to have clear trending patterns in news sentiment. Plots for all cities are available in Figure A.1.

One concern might be that these patterns reflect some coincidental manifestation of text across newspaper articles. While Figure II shows that the pattern of sentiment varies across cities, it is possible that the boom and bust pattern of words is common across all subjects and not necessarily specific to housing. To address this issue, I collect a random sample of articles that cover any subject or topic. I then compute a "random" sentiment index using the same methodology I used to create my housing sentiment index. If my index really reflects sentiment in the housing market, then we would not expect to see the same pattern arise from a random set of news articles.

[11] Chauvet, Gabriel and Lutz (2012) also propose a novel strategy to capture investor distress and negative sentiment using Google search queries. They find that the effects of their "housing distress index" on returns is particularly strong during the crisis.

[12] Hong and Stein (1999)model a gradual diffusion of news where only a fraction of traders receive innovations about dividends in each period.

Figure III reveals that the random index is a relatively flat line, and does not exhibit any discernible trend. This suggests that the sentiment index is at least specific to housing news.

Validating the sentiment index as a proxy for investor beliefs is naturally more challenging. By definition, beliefs are unobservable, but there exist some surveys that ask investors about the housing market. Existing survey measures are limited in frequency or geographic variation, but can be used to validate overall trends in my composite sentiment index. The Survey of Consumers (SOC) run by the University of Michigan and Reuters surveys a nationally representative sample of 500 individuals each month on their attitudes toward personal finances, business conditions, and buying conditions. One of these questions refers to the buying conditions in the housing market. Specifically, the SOC asks consumers, "Generally speaking, do you think now is a good or bad time to buy a house?" Respondents answer "yes," "no," or "do not know." Figure IV plots the percentage of respondents that answered "yes" across time. This simple question on home buyer confidence reveals a strikingly similar pattern to my composite-20 housing sentiment index. The percentage of positive home buyers also peaks well before housing prices, by more than a two year lead. Surveyed home buyer confidence actually appears to lead housing news sentiment slightly, from two to six months. This lead is consistent with a theory that news sentiment responds to consumer sentiment in the market. Interestingly, the increase in survey confidence is also followed by a similar increase in news sentiment in 2008. Both of the increases occur before the temporary rebound of the housing market in 2009, but fall again afterwards.

Case and Shiller (2003) implement even more detailed surveys of home buyer behaviors and provide more detailed perspective on investor expectations. They directly ask respondents how much they expect their house price to grow over the next ten years. Answers in 2003 revealed astonishingly high expectations; with respondents expecting prices to rise an average of 11 to 13 percent annually. Case, Shiller and Thompson (2012) recently updated these surveys each year from 2003 to 2012. Their survey covers just four suburban areas, but the similarity in timing of sentiment across the same cities in my dataset is significant. They find that long-term expectations of home buyers also peak in 2004, the same time as my sentiment index.

Panel B in Figure IV further plots my sentiment index with an index of home builder confidence constructed by the National Association of Home Builders (NAHB). The NAHB implements a monthly survey of their members, asking builders and developers to rate the current market conditions of the sale of new homes, the prospective market conditions in the next 6 months, and the expected volume of new home buyers. The NAHB index weights these answers into one index to represent an aggregate builders' opinion of housing market conditions. Figure IV shows that builder confidence index in the housing market declined significantly at similar timing to my sentiment index. Builder confidence peaks in 2005, suggesting a slight lag to home buyer confidence. My sentiment index highly correlates with survey measures of housing market confidence in both trends and timing, suggesting that news sentiment does reflect investor beliefs over the housing market. Still, both survey and news sentiment may still be driven by changes in fundamentals. I address effects from both observed and unobserved fundamentals in the following sections.

4 Does Sentiment Reflect Changes in Observed Fundamentals?

4.1 Sentiment Effects on House Price Growth

In this section I test the empirical predictions of the effect of sentiment on prices in the Appendix model and analyze whether the results reflect variation in observed fundamentals. I first test the predicted effect of sentiment on prices across time using the composite index. I approximate Equation 14 with the following estimating equation:

$$\Delta p_t = \alpha_0 + \sum_{k=0}^{K} b_k L^k \Delta s_{nt} + \gamma \Delta x_t + \delta_m + \nu_t \tag{4}$$

where a lowercase letter represents a log operator ($p_t = lnP_t$) and Δ denotes the first difference such that $\Delta p_t = lnP_t - lnP_{t-1}$. L^k is a lag operator such that lags $L^k \Delta s_n t = lnS_{n,t-k} - lnS_{n,t-k-1}$. Vector x_t controls for changes in observable fundamentals that drive housing prices over time. House price growth may generally coincide with increased home buying in particular seasons of the year (such as the summer), so I include a set of monthly fixed effects, δ_m, to control for price changes due to seasonality. I assume the error term ν_t is heteroskedastic across time and serially correlated, and calculate Newey and West (1987) standard errors that are robust to heteroskedasticity and auto-correlation up to twelve lags.

Taking log differences provides a convenient approximation of growth period, but also addresses concerns of nonstationarity. Serial correlation in house prices have been well documented

(Case and Shiller (1989, 1990)). Estimates will still be consistent if prices and sentiment are serially correlated, as long as this correlation weakens over time.[13] However if both prices and sentiment are nonstationary and contain unit roots, then a regression of Equation 15 could result in a significant estimate of sentiment even if the series are completely unrelated. First differencing also has an additional benefit of removing any linear time trend in price levels. For estimates to be consistent, I also impose an assumption that the error term ν_t is uncorrelated with fundamentals and both contemporaneous and lagged values of news sentiment.[14] Making this assumption is useful because it does not require that the error term be independent from future values of news sentiment. This is important because it does not rule out feedback from prices onto future values of news sentiment. In particular, newspapers may put a positive spin on news by emphasizing certain past price increases over others.

The effect of sentiment on prices is captured by the coefficients b_k. Each individual coefficient b_k represents the effect of the one-time change in sentiment growth in period $t-k$ on the equilibrium price growth in time t. Conceptually, the lagged coefficients b_k represent the lagged adjustment path of prices to sentiment.[15] As noted in the last section, Figures I reveals that composite sentiment peaks in 2004, suggesting a lag structure of nearly three years. Ultimately, I am interested in the accumulated effect of sentiment on prices, represented by the sum of the coefficients, $\sum_{k=0}^{K} b_k$ For ease of notation going forward, let $\beta = \sum_{k=0}^{K} b_k$.

Table IV tests the the hypothesis that $\beta > 0$ against the null that $H_o : \beta = 0$. If news sentiment simply reflects price movements or information about fundamentals that is already in prices, then β will not be significantly different than zero. Column (1) estimates equation 4 without any control variables. The first row reports the total accumulated effect of sentiment, β, on the current t monthly growth in prices. The subsequent rows groups the summed lagged effect of sentiment by years. The estimated coefficient describes the proportional relationship between the percentage change in lagged sentiment and prices. An estimated coefficient equal to one would

[13]In other words, to ensure that prices and sentiment are stationary and weakly dependent, weak dependence is generally defined as occurring when the correlation between observations x_t and x_{t+h} of a series approaches zero "sufficiently quickly" as $h \to \infty$.

[14]Tetlock (2007) makes a similar assumes independence given that news media has no direct relationship with returns, in constrast with concerns raised with predictor variables such as dividend yield in Stambaugh (1999).

[15]It is important to note that all estimations rely on assumptions over a particular lag structure on the data. I select this structure using a number of standard model selection criteria, but each has its acknowledged benefits and drawbacks. In addition, the lag structure restricts my estimation sample period. Since my measures for sentiment being in January 2000, my estimation evaluates prices beginning in 2003.

indicate that monthly price and lagged sentiment growth have a one-to-one relationship. Estimates show that a one percent appreciation in the sum of lagged sentiment is associated with a monthly price appreciation of approximately 0.8 percentage points. This is significant relative to the mean of monthly housing price appreciation across this period of 25 basis points.

Nonetheless, the estimated effect of sentiment may still be due to changes in fundamentals. For example, if news sentiment reports on a fundamental not yet incorporated into prices, then β may still be greater than zero but biased upwards. To address this concern, columns (1) through (6) add an increasing number of fundamental controls to the specification. I add each of the fundamental controls sequentially to test the stability of β. Column (2) controls for rental growth, column (3) adds variables for real interest rates and 30-year mortgage rates, and column (4) adds housing supply variables including new housing starts and building permits. Column (5) controls for additional labor market variables for employment, unemployment, and changing labor force, while column (6) includes controls for changing population and income. I do not present the individual coefficients for each control variable as they are not the primary interest of my analysis, but the coefficients are either generally in the right direction or not significantly different than zero. Estimates of β remains remarkably robust with the inclusion of each additional control and decline neither in significance nor magnitude. As argued by a number of previous studies, the stability of my estimates to the sequential addition of controls suggests bias from unobserved factors is less likely (Altonji, Elder and Taber (2005); Angrist and Krueger (1999)).

Figure V plots the predicted prices first using only fundamentals, and then using sentiment. The plot shows that sentiment growth is able to fit both the boom and subsequent bust of prices. In contrast, fundamentals explain a portion of the boom, but are not able to fit the subsequent bust in prices. Consistent with prior studies, observed fundamentals are not able to explain much of the variation in prices on their own. The adjusted R^2 from running a regression with fundamental controls only is 0.10.[16] Adding in lagged sentiment explains an additional 75 percent of the variation in price growth, increasing the R^2 to 0.85. From 2004 to 2006, aggregate housing prices increased by 33 percent. Observed fundamental controls account for approximately 9 percentage points, while sentiment explains an additional 24 percentage points.

[16]However, these same fundamentals were able to explain a significant variation in prices historically. As detailed in the next section, running a regression with the same fundamentals prior to this period (from 1987 to 2000) results in an adjusted R^2 of 0.69.

Column (7) adds in monthly fixed effects to control for any seasonal variation in housing prices. The magnitude of β actually increases by 10 basis points. Alternatively, the effect of sentiment could simply be capturing a linear time trend in house price changes. Column (8) shows that controlling for a simple linear time trend does reduce the magnitude of β somewhat, but estimates remain positive and significant. Further examination reveals that the coefficient estimate on the linear time trend (not shown) is negative, fitting the bust of the housing prices rather than the boom. Sentiment still largely accounts for the run-up in aggregate house prices.

Column (9) applies a specification that includes lagged measures of fundamentals. Search frictions in the housing market could also potentially affect the immediate effect of fundamentals (Wheaton (1990); Stein (1995); Krainer (2001)). Not all lags can be included due to high collinearity among fundamentals, but I select as many lags as possible with the same model selection criteria used to select the lag structure of sentiment. The effect of sentiment again remains positive, significant, and robust in magnitude. Column (10) reveals that the only variable able to drive down the magnitude of β are lagged measures of the price growth itself. This is not surprising as the predictability of house prices has been well documented (Case and Shiller (1989); Cutler, Poterba and Summers (1990)). Still, coefficient estimates of sentiment growth remain positive. In the following panel estimation, the predictive effect of sentiment remains both positive and significant beyond lagged price growth.

Still, estimations in Table IV are limited to a small number of observations ($N = 94$) and only accounts for variation in aggregate price growth. Table V utilizes the full panel dataset and tests whether sentiment has an effect on prices across cities. I estimate this effect with the following regressions:

$$\triangle p_{it} = \alpha_0 + \beta L^k \Delta s_{n,it} + \gamma \Delta x_{it} + \delta_m + c_i + \nu_{it} \tag{5}$$

where i denotes each city. In some specifications I also control for unobserved heterogeneity across cities with city dummies, c_i. I assume errors are heteroskedastic across time and serially correlated within city, and cluster Newey and West (1987) standard errors by city assuming auto-correlation up to twelve lags. I assume errors are correlated within city in my main regressions since studies have documented little mobility in homeowners across states. However, the presence of spatial correlation across my measures could severely understate calculated standard errors Foote (2007)). To address potential cross-sectional spatial dependence, I calculate Driscoll and Kraay (1998) standard errors for robustness. I find this does not affect the significance of my results, and present my main results using Newey and West (1987) standard errors.

Column (1) estimates regression 5 without any additional controls. Estimates of β are even larger in magnitude than in the aggregate specification, with an estimated coefficient for β of 1.12. Adding in fundamentals sequentially between columns (1) and (2) does not change the magnitude or significance of the results, and including all fundamentals actually increases the total effect of sentiment slightly to 1.22. The robustness of this estimates confirms the stability of β from the composite estimation, and further reduces concerns of that bias from unobserved fundamentals.[17]

Column (3) of Table V adds city fixed effects to the specification. Trading behavior in different markets may have particular characteristics that affect the differences in house price movements across different cities. Some cities may have inherently higher or lower house price levels (for example, New York may have high house prices due to particular characteristics of its location, financial center, etc.) that corresponds to innately optimistic newspapers. Transforming prices into growth terms normalizes fixed differences in house price levels across cities. Nonetheless, some markets also may also have coincidentally higher house price and news sentiment changes. Including city fixed effects removes any differences in house price appreciation due to time-invariant unobservable characteristics. The estimated effect of sentiment actually increases in magnitude after controlling for city fixed effects. This suggests that a large part of the predicted effect of sentiment can be attributed to its effect on price growth across time.

Columns (4) and (5) add month and year fixed effects. Adding just month fixed effects does not affect the results, estimates do not appear to be driven by seasonality. Including both month and year fixed effects drops the estimated coefficient by about half the magnitude. This drop in magnitude reflects the common trends in price growth across markets. The most recent boom of housing markets was notable because it was appeared to be a coordinated movement across many markets. Nonetheless, even with month and year dummies, the sentiment index still has a positive and significant predictive effect on price appreciation both statistically and economically.

[17] The number of observations between Columns (1) and (2) of Table V vary slightly since I do not have rental data for Las Vegas, but I do include Vegas when I estimate the effect of sentiment without controlling for fundamentals. Also, rental data is only available through October 2009 for most of cities. Column 1 has more observations since my sentiment indexes are available through August 2011. Some newspapers do have gaps in coverage by Factiva at various points in time, and thus are missing sentiment measures for those months.

The coefficient implies that a one percent increase in accumulated sentiment growth predicts a 0.6 percentage change in price growth (monthly). This is still large compared to the average monthly house price growth of 16 basis points across cities during this period. Column (6) alternatively controls for a linear time trend, which drives down the magnitude slightly from column (4). As in the aggregate estimates, the coefficient on the linear time trend is negative, fitting the bust of prices in many places but not the boom.

In column (7), I add lagged fundamentals and find that the magnitude of the effect declines slightly to 0.87, but is still positive and economically significant. Column (8) of Table V separately tests whether sentiment has any predictive effect from price growth above and beyond lagged prices. While the β drops to 30 basis points, the estimated effect of sentiment remains positive and significant. As in the aggregate specification, most of the explanatory power of lagged price growth comes from the first few lags (Δp_{t-1}). Lagged prices beyond the preceding year do not have much predictive power for future prices, whereas sentiment growth leads prices by more than two years.

Estimating over the whole sample period conceals whether the results are driven by the boom or bust period housing prices, or both. In columns (9) and (10), I split the sample and estimates the effect of sentiment on prices separately for each time period. Column (9) estimates equation 5 with data before July 2006, and Column (10) runs the regression with data July 2006 and afterwards. Concurrent with plots in Panel B of Figure IV, I find that sentiment predicts *both* the boom and bust of housing prices across cities. Estimated effects are positive, significant, and large in magnitude, while the magnitude of β is slightly larger for the bust than the boom. This is consistent with the observation that not all cities experienced a rise in housing prices, but a majority experienced a subsequent bust.

4.1.1 Subprime Conditions

One concern for the results in Table IV and V is that estimates could instead reflect a spurious correlation between news and the rise in the availability of credit and subprime lending patterns. The extraordinary rise in house prices from 2000-2005 was also accompanied by an unprecedented expansion of mortgage credit, particularly in the subprime market (Mian and Sufi (2009); Glaeser, Gottlieb and Gyourko (2010)). Easing lending standards and rising approval rates opened home-buying to a new set of consumers, which potentially allowed a new group of homebuyers to shift aggregate demand and drive up house price growth (Keys et al. (2010); Keys, Seru and Vig (2012); Mian, Sufi and Trebbi (2010)).[18] Mian and Sufi (2009) show that lending to subprime zip codes grew rapidly from 2002 to 2005, and sharply fell as house prices declined. Thus if news simply documents the rise and fall in subprime lending, then not controlling for these patterns may misrepresent the effect of β.

I address this possibility by including additional controls for credit and subprime lending in Table VII. Column (1) in Table VII adds controls for the changes in the six-month London Interbank Offered Rate (LIBOR). Estimations in Tables IV and V already include changes in overall the real interest rate and 30-year mortgage rate, but many adjustable-rate subprime mortgages were set at an initial fixed rate for the first two years and then indexed to changes in the LIBOR six-month rate (Mayer, Pence and Sherlund (2009); Gerardi et al. (2008)). Column (1) includes the full set of controls from column (5) in Table V, including fundamentals, lagged fundamentals, month and city fixed effects. Including changes in the 6-month LIBOR rate has no effect on the results, and the estimated effect of sentiment is still positive and significant. The estimate also remains robust in magnitude compared to estimates in column (5), Table V.

Column (2) additionally controls for the fraction of subprime mortgages and average loan-to-value ration in each city. I do not have measures for subprime lending and applicant income for Atlanta, Charlotte, Dallas, and Minneapolis. Thus, regressions in columns (2)-(5) only include data from 16 cities. Additionally, measures of subprime lending, loan-to-value, and applicant income are only available through 2008. Thus, estimations in columns 2-5 are limited to five years of data (2003-2008), and restricted to observations where both data on subprime lending and sentiment indexes are available. Nonetheless including trends of subprime lending and loan-to-value ratios does not significantly change the results. The estimated effect of sentiment on price growth declines slightly, but by less than 5 basis points. In column (5), I include additional measures of income, but specific to those reported by mortgage applicants. The effect of sentiment is again remarkably robust. β decreases slightly by 5 basis points, but remains positive and significant in magnitude.

[18]Other papers that explore subprime lending explanations and the role of mortgage securitization in the housing crisis are Bajari, Chu and Park (2008); Danis and Pennington-Cross (2008); Demyanyk and Van Hemert (2011); Gerardi et al. (2008); Goetzmann, Peng and Yen (2012); Mayer and Pence (2008); Mayer, Johnson and Faltin-Traeger (2010); Haughwout and Tracy (2009) Adelino, Gerardi and Willen (2009); Campbell, Giglio and Pathak (2011); Foote, Gerardi and Willen (2008); Mayer, Pence and Sherlund (2009); Mian and Sufi (2009); Mian, Sufi and Trebbi (2010); Piskorski, Seru and Vig (2010).

Only including additional lags of the subprime variables reduces estimates of β more substantially, but estimated effect of sentiment remains economically significant.

4.2 Sentiment Effects on Housing Trading Volume

Existing theories of sentiment also links sentiment to trading volume (Harrison and Kreps (1978); De Long et al. (1990*b*)). For example, Baker and Stein (2004) reason that when limits to arbitrage are very costly, optimistic investors are more likely to trade and drive up volume. Scheinkman and Xiong (2003) and Odean (1998) make related arguments based on overconfident investors. The model similarly provides testable empirical predictions for housing sentiment and trading volume. Equations 16 suggests a relationship between changes in sentiment and trading volume levels. Thus, I estimate the effect of sentiment on trading volume in the housing market with the following specification:

$$v_{it} = \varphi_0 + \kappa L^k \Delta s_{n,it} + \Delta x_{it} + \delta_m + c_i + \xi_{it} \tag{6}$$

where v_{it} represents the de-trended log volume of housing transactions in each month t . I measure trading volume in de-trended log levels to address concerns of nonstationarity in levels of volume in the housing market. I follow a de-trending methodology applied to volume in Campbell et al. (1993). I also control for all observed fundamentals, quarterly fixed effects, and city fixed effects, and lagged fundamentals. As in equation 4, κ represents the sum of coefficients for all lags of sentiment.

Figure VI plots the composite-20 housing sentiment index and volume of housing transactions over time. I construct a composite measure of transaction volume by aggregating the number of transactions in each city and weighting each measure with the normalized weights used to calculate the composite-20 Case-Shiller home price index. Figure VI shows that sentiment not only forecasts the pattern in prices, but also foreshadows a rise and fall in volume. Interestingly, volume appears to peak before prices. The plot shows that volume begins to drop at the end of 2005, while prices do not begin to decline until July 2006. Sentiment thus still precedes volume by approximately a 18 months (1.5 years). This pattern provides a potential explanation for the long lead in sentiment to prices. Figure VI suggests that sentiment moves first and leads to housing transactions in the following year, and this increased trading activity shows up in housing prices another year later.

Table VII presents the results for regression 6. I select a model that includes $K = 18$ lags i.e. a year and six months. Note that my volume data ends in July 2009 so that my sample period is shorter than in my estimations for prices. Columns (1)-(3) estimate the effect of sentiment on the composite-20 measure of transaction volume, and Columns (4)-(6) estimates over the panel dataset across cities. Consistent with predictions in Equation 16, the growth in sentiment has a positive association with increases in transaction volume levels. Columns (1) and (4) runs the regression with any additional controls. Sentiment growth has a positive and significant accumulated effect on trading volume both in the composite and panel data. Specifically, a one percent increase across monthly lags of sentiment growth leads to a 4.7 and 3.5 percent increase in the volume of housing transactions in the composite and panel regressions respectively.

As in our regressions above, a primary concern is that this positive effect instead reflects positive changes from fundamentals. Thus, Columns (2)-(3) and (5)-(6) include the same set of housing fundamentals used to explain housing prices as well as month and city fixed effects. In the composite regressions, the estimated coefficient for κ remains robust to the inclusion of fundamentals in x_{it}, and further increases in magnitude after controlling for month fixed effects. In the panel regressions, including fundamentals, lagged fundamentals, month, city fixed effects reduces the magnitude of the κ in the panel regressions, but the effect of sentiment growth on volume remains positive and significant. Column (6) shows that a one percent positive appreciation in lagged sentiment leads to a 1.6 percent increase in transaction volume after controlling for lagged fundamentals. This is still well above the mean of detrended log volume (-.02). These results are consistent with empirical evidence that connects investor sentiment to trading volume (Barber and Odean (2000, 2008); Odean (1999)). The correlation between volume and prices has also been previously documented in the housing market (Stein (1995)). Genesove and Mayer (1997) provide empirical evidence that behavioral biases such as loss aversion might explain positive price-volume correlations in the housing market.

5 Does Sentiment Reflect Changes in Unobserved Fundamentals?

The previous section shows that sentiment, proxied by the tone of news, has a predictive effect for house price growth and transaction volume above and beyond a number of observed housing fundamentals. In this section I address whether this effect instead reflects effects from unobserved fundamentals. As noted in the previous section, the robustness of the estimates to the inclusion of each additional control is already strongly suggestive that bias from unobservables is less likely. Furthermore, the lead in sentiment growth to prices suggests that prices move in response to sentiment and not the reverse. One might be worried that these indexes actually overlap since Case-Shiller home price index is reported using housing transactions from previous months. However news sentiment leads prices by more than two years, and the Case-Shiller home price index is calculated over transactions from the current month and the previous two months. Even if there is some further delay in reported transactions, news sentiment peaks at such a significant year lead that it very unlikely due to some mechanical delay in the reporting of prices. Still, prediction does not eliminate the possibility that news is reporting information on unobserved fundamentals not yet incorporated into prices. Search frictions in the housing market could delay the effect of both sentiment and fundamentals on price growth.

If the housing sentiment index is affected by unobserved fundamentals, estimates of sentiment in Tables IV and V may be potentially biased. The extent of this bias depends on whether x_{it} includes the key set of fundamentals that drive house price growth. If only minor fundamentals are missing, then estimates may still be biased but only minimally. I can assess whether my observed vector x_{it} appears to miss any important housing fundamentals by testing whether it explains prices well during periods where sentiment is not suspected to be a factor. Table VIII splits the sample into two periods, pre- and post-2000, and estimates the effect on prices with fundamentals alone. If x_{it} sufficiently controls for important determinants of housing prices, then these variables should explain changes in price growth during the "pre-bubble" period, i.e. before 2000. The adjusted R^2 in column 1 shows that fundamentals explain almost 70 percent of the variation in composite housing prices before January 2000. I use the composite-10 price index since the composite-20 index is only available starting in 2000. Similar to the composite-20 index, the Case-Shiller Composite-10 home price index is a weighted average of ten major U.S. cities., which includes Boston, Chicago, Denver, Las Vegas, Los Angeles, Miami, New York, San Diego, San Francisco, and Washington, D.C. In contrast, the same fundamentals explain very little of the change in prices after 2000 with an adjusted R^2 equal to only 0.092. Columns 3 and 4 similarly show that fundamentals have greater explanatory power for housing prices across cities prior to 2000. Local fundamentals do explain at least 23 percent of the variation in prices after 2000, but are able to explain 1.55 times more prior to 2000. Fundamentals are more significant in cities that did not experience rapid growth in prices. These results suggest that if that my news sentiment index is affected by articles on unobserved fundamentals, then bias from these variables are at least minimal.

Still, the housing sentiment index may be contaminated by news reports on unobserved fundamentals. I exploit the richness of my data to isolate any articles that discuss housing fundamentals and partial out their effect directly. I identify any article that mentions words related to housing fundamentals using stem words such as "unemployment", "mortgage rates", or "taxes." Tetlock, Saar-Tsechansky and Macskassy (2008) employ a similar strategy to identify news articles that discuss firm fundamentals. The advantage of this strategy is that I can identify articles that discuss fundamentals that I both observe and do not observe. I then directly control for fraction of the positive minus negative words in these news stories that mention fundamentals in my estimations. If information on fundamentals from these articles subsequently drive prices, then controlling for words in these articles should drive down the significance and magnitude of the results in Section 4.

Table IX show that the estimated effects of sentiment on price growth remain robust to controlling for news content over fundamentals. I create individual measures of these "media fundamentals" and evaluate their effect on prices separately. I control for all lags of these measures as well as all observed controls. Columns (1) through (7) adds a control for articles discussing each housing fundamental to test the stability of β. Column (2) shows that the estimate drops after controlling for news articles discussing credit conditions, but the remains stable with the addition of remaining media fundamentals. Column 2 reports an estimated coefficient for the accumulated effect of sentiment approximately equal to 0.5, an almost one-to-two proportional relationship between lagged sentiment changes and monthly price growth. The estimated effect of positive news sentiment remains significant, positive, and large in magnitude.

5.1 Weekend and Narrative News Content

Results in Section 4 show that sentiment predicts price growth at a significant lead of more than two years, and estimated effects remain highly robust to the sequential addition of observed controls. The observed set of fundamentals explains a significant amount of variation in price growth prior to

2000, suggesting it is unlikely effects are due to a key omitted fundamental after 2000. In addition, the estimated effect does not appear to be driven by articles that discuss fundamentals in its text. To narrow the identifying variation further, I isolate two subsamples of the news articles that cater to reader sentiment but are less likely to be affected by information on fundamentals.

The first set I isolate are those articles that are published on the weekend. Weekend articles are likely correlated with sentiment because it must cater to readers who prefer content lighter in nature. Indeed, research on newspaper readership shows that lighter readers are concentrated on the weekend. The Readership Institute of Northwestern University conducted a survey of 37,000 newspaper readers in 2000 and found that readership is highest on Friday, Saturday and Sunday, driven by the greater proportion of "light" readers on the weekend. Light readers are those who spend fewer than 16 minutes reading the newspaper a week, whereas heavy readers pay attention to the news every day. Furthermore, the survey reports that these readers appear to be light readers of all news alternatives, including television news, magazines, and internet websites.[19] Thus these readers are more likely to be those who are more subject to sentiment and demand articles that cater to their preferences. This is consistent with why Saturday and Sunday editions of newspapers typically include additional sections, such as entertainment and sports, in order to draw readers who are more subject to sentiment. At the same time, weekend news articles are less likely to reflect information on fundamentals as market news tends to be reported during the business week. Furthermore, any press releases on fundamentals data can only occur on a working weekday. Thus, news stories on the weekend are more likely to be exogenous to official news reports on fundamentals. Because my dataset includes the exact date of each story, I am able to identify the exact day of the week each article is published. Thus, I create a weekend instrument that only analyzes the tone of articles that occur on Friday, Saturday, and Sunday.

Examining a smaller subset of articles allows me to run further falsification tests on the assumption that they are less likely to reflect news on fundamentals. For example, one concern might be that news releases on fundamentals are increasingly released on Friday and then reported over the weekend. If this is the case than the increase of positive or negative words on the weekend may be the result of increasing news releases concentrated at the end of the week. To test this possibility, I compile a dataset of of all the press release dates on various housing fundamentals.

[19]Survey reports can be found at http://www.readership.org/reports.asp

Specifically, I organize the schedule of press releases from the Bureau of Labor Statistics (BLS) and regional data from the Census. Table X reports the correlation of the weekend instrument with the percentage of news reports released on Friday. The first row reports the correlation of all BLS news releases and the subsequent rows reports the correlation with regional and employment releases. Column 2 reveals that the correlation with each are very low, suggesting the weekend instrument is not simply reporting news occurring on Friday. The last two rows examine the correlation with Census releases on new residential construction and sales. The weekend instrument is also uncorrelated with the percentage of these releases occurring on Friday.

Another concern might be that news on fundamentals are reported during the working week, but then summarized over the weekend. One way to address this issue is to control for the pattern of positive and negative words that occur during the weekday. If weekday articles contain information on fundamentals, then controlling for this content should address concerns that weekend content is actually a proxy or response to weekday information. I control for the fraction of positive minus negative words in weekday articles in both the first stage regressions Table XI and instrumental variable (IV) results in XII. A captured effect of sentiment is then narrowed to the differential variation between weekend and weekday news.

I create an additional instrument that from the narrative articles in my sample. A narrative article refers to one that narrates a story or account of events around particular individuals. Narrative writing is also a particular writing strategy through which newspapers can reflect sentiment and capture readers' attention. The Readership Institute Survey reports that readers have high preference for "people-centered news" or articles about local ordinary people. The study particularly encourages newspapers to increase readership through this "approach to story-writing" and finds that it is how a story is written that matters more for reader satisfaction. At the same time, narrative articles contain anecdotal stories, but tend to offer no actual data or news on fundamentals in the market. The above narrative expresses an obvious optimistic view over the housing market, but contains no actual news on any particular fundamentals. Thus, trends in news slant across narrative articles are correlated with sentiment but plausibly exogenous from any actual news on fundamentals.

I identify narrative articles by locating those that discuss individual people. I isolate any article that includes a name from name lists from the Social Security Administration (SSA) and the Census. The Social Security publishes a list of the 200 most popular first names of the 2000s. I create a list of last names with the top 1000 most frequently occurring surnames in the 2000 census. I then define an article as narrative if it discusses any of these names in its first paragraph. I exclude any articles that match a quoted statement by an individual in case these are cited statements from various experts. I then analyze the share of positive and negative words in just the identified "narrative" articles in my sample.

I then use share of positive and negative words over the smaller sample of weekend and narrative articles as instruments for my overall measure of sentiment. These instruments are only valid if they are sufficiently correlated with the housing sentiment index. I directly test the first-stage relevance between sentiment and each of my instruments with the following first-stage regression:

$$\triangle s_{n,it} = a_0 + \lambda \triangle z_{it} + \eta \triangle x_{it} + \delta_m + c_i + u_{it} \tag{7}$$

where z represents the log of the candidate instrument. Columns 1 and 2 in Table XI confirms that changes in both the weekend and narrative instruments are positively and significantly correlated with positive news sentiment. I test the strength of both instruments and report the F-statistics in bold at the bottom of Table XI. The weekend instrument is stronger than the narrative instrument, but both instruments have more than sufficient strength, with F-statistics well above the benchmark of 10.

Table 12 presents the second-stage results of instrumenting for positive news sentiment. Column (1) presents the original ordinary least squares estimates with all controls from estimating Equation 5. Columns (2) and (3) reports the results instrumenting for sentiment using the weekend and narrative index respectively. The estimated effect for sentiment on price growth remains positive, significant, and robust in magnitude. Instrumenting sentiment with the weekend instrument actually increases the magnitude of the estimated effect of sentiment on price growth substantially. While our main concern is addressing upward bias, noise from sentiment measures likely biases standard ordinary least squares estimates downward. Estimates remain robust in magnitude after instrumenting with the narrative index, though do not increase.

6 Conclusion

This paper presents evidence that sentiment has a significant effect on housing prices, particularly during the boom and bust from 2000 to 2011. While there has been much discussion and interest in the role of mass psychology or “animal spirits” in the most recent housing crisis, empirical support for this argument has been limited due to the lack of sentiment measures for the housing market. This paper provides the first measures of sentiment across local housing markets by capturing the tone of local housing news across 20 major city newspapers.

I find that sentiment forecasts the boom and bust of housing markets by a significant lead, peaking two years before house prices began to decline in 2006. Results show that sentiment growth is positively associated with future price growth, and is able to explain a significant amount of variation in the price changes above and beyond fundamentals. In particular, the housing sentiment index is able to explain an additional 70 percent of the variation in national house prices beyond observed fundamentals. Further evidence suggests these estimates are unlikely driven by latent fundamentals. Estimates are significantly robust to the inclusion of an exhaustive list of controls and remain robust to a novel instrumental variable strategy.

The findings of this paper have several potential implications. The evidence suggests that sentiment has an important effect on asset prices, and raises questions over how behavioral factors interact in economic contexts. Expectations and fundamentals likely have a more complex relationship, for example, perhaps where individuals systematically overestimate a positive shock from lower interest rates or increases in credit supply. Indeed, studies on financial literacy suggest that many investors are not able to appropriately compound interest or account for inflation (Lusardi and Mitchell (2007*b*)). Brunnermeier and Julliard (2008) find supportive evidence that particularly links money illusion to the run-up in housing prices. Furthermore, the ability of news to forecast price movements suggests measures of market sentiment may be useful indicators to monitor empirically. The central finding of this paper, however, highlights that sentiment plays an important role on aggregate economic outcomes and suggests it deserves greater attention in future work.

Appendix

A.1 Sentiment Index Robustness and Alternate Versions

Leading v. Full Text. The primary sentiment index used in this paper is the share of positive minus negative words calculated over the leading text of housing articles each city-month. I create a number of alternate versions of the baseline sentiment index for robustness. Table A.1 compares the effect of sentiment on house price growth using different versions of the housing sentiment index. Column (1) first presents the results using the baseline index, $Pos_{it} - Neg_{it}$. Column (2) similarly applies the share of positive minus negative words, but calculated using the full text of housing articles. Using the full rather than the leading text has no significant effect on the results, in precision or magnitude. The bottom panel of Table A.1 reports the correlations of each alternative with the baseline index, and shows that the full text version of the index is highly correlated with the baseline.

News Intensity. Excitement over the housing market may be evident in not only the tone of news articles, but also by how many articles cover the housing market each month. A newspaper can cater to reader sentiment through both the slant and frequency of its housing news articles. Thus to capture this dimension, I interact the baseline index with the share of housing articles published by a newspaper each month. Specifically, this version can be represented by:

$$(Pos_{it} - Neg_{it}) * \frac{\#\text{ Housing Articles}}{\#\text{ Total Articles}}_{it}$$

The share of housing articles is equal to the number of housing articles divided by the total number of news articles (in any subject) in city i and month t. Column (3) shows that this version also has no effect on the results, and is highly correlated with the baseline.

Positive v. Negative Index. Another informative robustness check is to separate the effect of positive and negative words. If the baseline index is appropriately capturing sentiment, we might expect the growth in the share of positive words to have a positive association with prices while the share of negative words should have a negative association with house prices. Indeed, columns (4) and (6) shows that the effect of just positive words is positive while negative words has an opposing negative effect. The baseline index has a greater predictive effect for house prices than just positive or negative words alone, but both still have a significant effect on house price growth individually.

Term Weighted Index. Loughran and Mcdonald (2011) also propose an index that weights each word in an article using the term-weighting formula:

$$w_{kj} = \frac{1 + logtf_{ij}}{1 + log(a)} log(\frac{N}{df_i})$$

where N represents the total number of articles in the sample, df_i, the number of articles containing at least one occurrence of the i^{th} word, tf_{ij} the raw count of the i^{th} word in the j^{th} document, and a the total number of positive words in the article. The first term accounts for the frequency of the term within each article but also applies a log transformation to attenuate the impact of high frequency words. For example, the word *soar* may appear 32,000 times in our sample while the word *skyrocket* only appears 10 times, but this does not mean *soar* is necessarily 3200 times more important than the word *skyrocket.* The second term measures the importance of the term across documents by dividing the total number of documents in the sample by the number of documents containing the particular term. Thus the word *soar* will receive a high weight based on the first term, but if it is a common word that appears in more than 90 percent of articles, then the second term will decrease the first term by more than 90 percent. I apply this weighting formula to the share of positive words and test to see this has a significant effect on the results in Column (5) of Table A.1. The results show that term-weighted share of positive words has an almost identical impact on house price growth as the non-weighted positive index.

A.2 Theoretical Framework

In this section, I present a simple theoretical framework that illustrates the potential relationship between the news media, investor sentiment, and housing prices. I specifically measure sentiment with news because prominent literature on bubbles and panics commonly stress that the news media has an important relationship with investor beliefs (Kindleberger (1978); Galbraith (1990); Shiller (2005)). They argue that newspapers have a demand-side incentive to cater to reader preferences, and will spin news according to readers' opinion over assets they own. Economic models of media slant make similar arguments in the context of readers' political preferences. Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006) assume that readers have a disutility for news that is inconsistent with their beliefs, citing psychology literature that show people have a tendency to favor information that confirms their priors.[20] Indeed, Gentzkow and Shapiro (2010) find empirical evidence that readers have a preference for news consistent with their beliefs and news outlets respond accordingly. This framework adapts models of investor sentiment (De Long et al. (1990*a*); Copeland (1976); Hong and Stein (1999)) and models of media slant (Gentzkow and Shapiro (2010); Mullainathan and Shleifer (2005)) to show how news relates to investor sentiment and asset prices.

Agents: I assume there are two types of agents in the economy: fully rational traders and imperfectly rational optimists that have a preference for news that confirms their priors. Agents are otherwise identical in utility maximization and risk aversion parameters. In each period t, the fraction of optimistic traders are present in the economy each period at measure μ_t, and fully rational agents are present in the economy at measure $(1 - \mu_t)$. All agents have constant absolute risk aversion where γ denotes the common coefficient of risk aversion. Thus, the allocation to the risky asset is unaffected by the accumulation of wealth. For simplicity, I assume there is no consumption decision, no labor supply decision, and no bequest. The resources agents have to invest are completely exogenous. In each period, agents choose an optimal allocation of housing, X_t , to maximize the following:

$$\max_{H_t} E[-e^{-2\gamma W_{t+1}}]$$

subject to the budget constraint:

$$W_{t+1} = W_t(1 + r_f(1-\tau)) + X_t[P_{t+1} + D_{t+1} - P_t(\delta_t + m_t + (1-\tau_t)(1 + r_f + \pi_t)]$$

where W_t represents wealth in period t. Agents allocate wealth between a risk-free asset that guarantees a risk-free rate of $r_f > 0$ each period and a risky asset of housing that pays dividends, D_t, in the form of housing services each period. Housing is in supply quantity Q_t each period, and the risk-free asset is in perfectly elastic supply. The price of housing stock is denoted by P_t. I assume housing depreciates at rate δ_t, requires maintenance and repairs at a fraction of house value m_t, and incurs property tax liabilities at rate π_t. Furthermore, all investors must pay a marginal

[20] This tendency is called confirmatory bias in the psychology literature (Lord (1979); Yariv (2002)).

income tax of τ_t, but may deduct property taxes from taxable income and otherwise borrow or lend at the risk-free rate r_f. This represents the user cost of housing as formalized by Poterba (1984). For ease of notation going forward, let $\omega_t = \delta_t + m_t + (1-\tau_t)(1+r_f+\pi_t)$.

Maximizing expected utility over X_t yields the following optimal demand function for housing:[21]

$$X_t = \frac{EP_{t+1} + D_{t+1} - P_t\omega_t}{2\gamma E\sigma^2_{P_{t+1}}}. \tag{8}$$

Since this is just a linear demand function, for simplicity let the above be represented by:[22]

$$X_t = \alpha_t - \omega P_t \tag{9}$$

Rational traders demand housing according to equation (1), but I assume optimists overestimate the expected price of housing relative to rational traders by an additional positive parameter θ.[23] Thus relative to rational traders, optimists shift their demand curves upward by an additional θ.

$$X_t^{Opt} = \alpha_t + \theta - \omega P_t \tag{10}$$

Newspapers: I also assume that optimistic investors have a preference for news that confirms their positive beliefs. Gentzkow and Shapiro (2007) model this preference by assuming readers have a quadratic disutility for news that conflicts with their priors, and derive an equation for newspaper readership approximately equal to $a - (S_n - S_i)^2$ where a is a constant, S_{nt} is slant reported by newspaper n, and S_{it} is the overall level of sentiment in city i and period t. In this framework, the overall level of sentiment in the economy is equal to the fraction of optimists, μ_t, multiplied by their level of optimism, θ. Thus, $S_{it} = \mu_t\theta$, and the optimal level of news slant that maximizes a newspaper's readership is equal to:

[21] With normally distributed returns, maximizing the above is the same as maximizing mean-variance utility. I rewrite the agents problem such that they maximize the following expected utility each period: $EU = E[W_{t+1}] - \gamma\sigma^2_{W_{t+1}} = W_t(1+r_f)(1-\tau_t) + X_t[E_tP_{t+1} + D_{t+1} - \omega_tP_t] - X_t\gamma E_t\sigma^2_{P_{t+1}}$, where $\sigma^2_{W_{t+1}}$ is the one-period ahead variance of wealth and σ^2_{Pt+1} is the one period ahead variance of price. This follows the set up in De Long et al. (1990*a*).

[22] where $\alpha = \frac{EP_{t+1}+D_{t+1}}{2\gamma E\sigma^2_{P_{t+1}}}$ and $\omega = \frac{\omega_t}{2\gamma E\sigma^2_{P_{t+1}}}$.

[23] Conversely, this framework could also apply to a set of pessimists who underestimate the expected price of housing by a negative parameter θ.

$$S^*_{nt} = S_{it} = \mu_t \theta \tag{11}$$

Thus news slant, or the sentiment in news, directly reflects the overall level of reader sentiment.

Equilibrium Price: Given the presence of μ_t optimists and $(1-\mu_t)$ rational traders, equilibrium is characterized by setting demand equal to supply, $(1-\mu_t)(\alpha - \omega P_t) + \mu_t(a + \theta - \omega P_t) = Q_t$. Thus the equilibrium price equals:

$$P_t = \frac{(\alpha_t + \mu_t \theta - Q_t)}{\omega} \tag{12}$$

Equation 12 reveals that investor sentiment has a positive association with prices ($\frac{dP_t)}{d\mu_t \theta} > 0$). Using equation 11, we can rewrite equation 12 in terms of news sentiment:

$$P_t = \frac{(\alpha_t + S^*_{nt} - Q_t)}{\omega} \tag{13}$$

Then the price change from t to $t+1$ can be expressed by:

$$\triangle P_{t+1} = \frac{1}{\omega}[(\triangle \alpha_{t+1}) + (\triangle S^*_{nt+1}) - (\triangle Q_{t+1})] \tag{14}$$

where $\triangle P_{t+1} = P_{t+1} - P_t$. Thus Equation (7) predicts that changes in news sentiment $(\triangle S^*_{nt+1})$ are positively associated with changes in prices (P_{t+1}). Positive fundamentals such as dividends, D_t, will also drive prices up, while increasing costs and housing stock will have dampening effect on prices. If there are no optimists in the market $(\mu_t = 0)$ or sentiment remains unchanged, then prices will equal $P_t = \frac{(\alpha - Q_t)}{\omega}$ and are only moved by changes in fundamentals and rational expectations in α, β, and Q_t.

Examining the effect of sentiment in the housing market allows me to analyze not only the time-varying effects of sentiment but also the cross-sectional effect of sentiment across different local housing markets. Let $\triangle P_{it} = P_{it} - P_{it-1}$ be the change in prices in city i and $\triangle P_{jt}$ represent the changing prices in city j. The difference in house price changes across cities can be written as:

$$\triangle P_{it} - \triangle P_{jt} = \frac{1}{\omega}[(\triangle \alpha_{it} - \triangle \alpha_{jt}) + (\triangle S^*_{i,nt} - \triangle S^*_{j,nt}) - (\triangle Q_{it} - \triangle Q_{jt})] \tag{15}$$

Equation 15 shows that if the price increase from $t-1$ to t is greater in city i than in city j, then this is due to either a greater increase in components in $\triangle \alpha_{it}$ or in investor sentiment (proxied by news sentiment $\triangle S^*_{i,nt}$).

Trading Volume. Increasing sentiment driven by the rising demand from optimists in the economy has further implications for trading volume in each housing market. Suppose the fraction of optimists increases from t to $t+1$ such that $\mu_{t+1} > \mu_t$. Trading volume, V_{t+1}, is then equal to the additional demand for housing from the fraction of optimists period to period:[24]

$$\begin{aligned} V_{t+1} &= \mu_{t+1} X_{t+1}^{Opt} - \mu_t X_t^{Opt} \\ &= \frac{1}{\omega}(S_{nt+1} - S_{nt})(\alpha - Q) \end{aligned} \tag{16}$$

Equation 16 illustrates that as sentiment increases, trading volume will be pushed upward. The greater the demand from optimists is relative to the previous period, the greater the volume of trades. This framework predicts that positive changes in sentiment should lead to increases in trading volume.

Lagged Effect. The above framework assumes that news only reflects investor sentiment. However, Shiller (2005) argues that news media can simultaneously fuel sentiment if readers misperceive optimism in the news for real information about fundamentals the housing market. Housing, in particular, is a widely held household investment by individual buyers. Thus the average housing investor is likely less financially sophisticated than the typical stock market investor. Survey evidence shows that a majority of Americans do suffer from surprisingly low levels of financial literacy (Lusardi and Mitchell (2007*a*,*b*)). Even more sophisticated investors may find it difficult to process quantitative data on market fundamentals. Indeed, Engelberg (2008) provides empirical evidence from earnings announcements that qualitative information on positive fundamentals is especially difficult to process. News slant can make it difficult for readers to separate true information from sentiment, and can subsequently affect trading behaviors. Empirical studies on political media slant show that the media has been able to shift public opinion and voting behavior (DellaVigna and Kaplan (2007); Gerber, Karlan and Bergan (2009)). Engelberg and Parsons (2011) show that different local media coverage of the stock market drives different trading outcomes across markets. If this is the case, then news sentiment in period t can also drive investor sentiment in future periods, $\mu_{t+1}\theta$, and prices would be positively associated with both contemporaneous and lagged values of news sentiment, $S_{nt}$and S_{nt-k}.

[24]I assume that α and Q stay constant here to make the effect of sentiment clear.

Furthermore, this framework also assumes that transactions in the housing market are immediate and costless. The transaction process of buying a home is by no means immediate, and the search process for a home can actually take several months. Thus there can be several lags between a change in sentiment and its effect on prices, and potentially no contemporaneous effect at all. If news slant does feed sentiment, then this can also take some time to diffuse and spread across investors.[25] Thus I consider the effect of both contemporaneous and lagged effects of sentiment in my empirical estimations.

[25] Hong and Stein (1999)model a gradual diffusion of news where only a fraction of traders receive innovations about dividends in each period.

References

Adelino, Manuel, Kristopher Gerardi, and Paul S. Willen. 2009. "Why Don't Lenders Renegotiate More Home Mortgages? Redefaults, Self-Cures, and Securitization." Federal Reserve Bank of Atlanta Working Paper 2009-17.

Altonji, J.G., T.E. Elder, and C.R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1): 151–184.

Angrist, J.D., and A.B. Krueger. 1999. "Empirical Strategies in Labor Economics." *Handbook of Labor Economics*, 3: 1277–1366.

Antweiler, Werner, and Murray Z. Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance*, 59(3): pp. 1259–1294.

Arce, Óscar, and David López-Salido. 2011. "Housing Bubbles." *American Economic Journal: Macroeconomics*, 3(1): 212–41.

Avery, Robert, and Kenneth Brevoort. 2010. "The Subprime Crisis: How Much Did Lender Regulation Matter." Division of Research and Statistics. Board of Governors of the Federal Reserve System.

Bajari, Patrick, Chenghuan Sean Chu, and Minjung Park. 2008. "An Empirical Model of Subprime Mortgage Default From 2000 to 2007." National Bureau of Economic Research Working Paper 14625.

Baker, Malcolm, and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *The Journal of Finance*, 61(4): pp. 1645–1680.

Baker, Malcolm, and Jeffrey Wurgler. 2007. "Investor Sentiment in the Stock Market." *The Journal of Economic Perspectives*, 21(2): pp. 129–151.

Baker, Malcolm, and Jeremy C Stein. 2004. "Market Liquidity as a Sentiment Indicator." *Journal of Financial Markets*, 7(3): 271 – 299.

Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan. 2012. "Global, Local, and Contagious Investor Sentiment." *Journal of Financial Economics*, 104(2): 272 – 287.

Baker, S.R., N. Bloom, and S.J. Davis. 2012. "Measuring Economic Policy Uncertainty." *manuscript, Stanford University.*

Barber, Brad M., and Terrance Odean. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *The Journal of Finance*, 55(2): pp. 773–806.

Barber, Brad M., and Terrance Odean. 2008. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *The Review of Financial Studies*, 21(2): pp. 785–818.

Barber, Brad M., Terrance Odean, and Ning Zhu. 2009. "Do Retail Trades Move Markets?" *The Review of Financial Studies*, 22(1): pp. 151–186.

Bayer, Patrick, Christopher Geissler, and James W. Roberts. 2011. "Speculators and Middlemen: The Role of Flippers in the Housing Market." National Bureau of Economic Research Working Paper 16784.

Bhutta, N. 2009. "Regression discontinuity estimates of the effects of the GSE act of 1992." Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board.

Brown, Gregory W., and Michael T. Cliff. 2005. "Investor Sentiment and Asset Valuation." *The Journal of Business*, 78(2): 405–440.

Brunnermeier, M.K., and C. Julliard. 2008. "Money Illusion and Housing Frenzies." *Review of Financial Studies*, 21(1): 135–180.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 2011. "Understanding Booms and Busts in Housing Markets." National Bureau of Economic Research Working Paper 16734.

Campbell, John Y., Sanford J. Grossman, and Jiang Wang. 1993. "Trading Volume and Serial Correlation in Stock Returns." *The Quarterly Journal of Economics*, 108(4): pp. 905–939.

Campbell, John Y., Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices." *The American Economic Review*, 101(5): 2108–31.

Case, Karl E., and Robert J. Shiller. 1989. "The Efficiency of the Market for Single-Family Homes." *The American Economic Review*, 79(1): pp. 125–137.

Case, Karl E., and Robert J. Shiller. 1990. "Forecasting Prices and Excess Returns in the Housing Market." *Real Estate Economics*, 18(3): 253–273.

Case, Karl E., and Robert J. Shiller. 2003. "Is There a Bubble in the Housing Market?" *Brookings Papers on Economic Activity*, 2003(2): pp. 299–342.

Case, K.E., R.J. Shiller, and A. Thompson. 2012. "What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity*.

Chauvet, Marcelle, Stuart Gabriel, and Chandler Lutz. 2012. "Fear and Loathing in the Housing Market: Evidence from Search Query Data." Available at SSRN: http://ssrn.com/abstract=2148769.

Copeland, Thomas E. 1976. "A Model of Asset Trading Under the Assumption of Sequential Information Arrival." *The Journal of Finance*, 31(4): pp. 1149–1168.

Cutler, David M., James M. Poterba, and Lawrence H. Summers. 1990. "Speculative Dynamics and the Role of Feedback Traders." *The American Economic Review*, 80(2): pp. 63–68.

Danis, Michelle A., and Anthony Pennington-Cross. 2008. "The Delinquency of Subprime Mortgages." *Journal of Economics and Business*, 60(1-2): 67–90.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics*, 122(3): 1187–1234.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990*a*. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98(4): pp. 703–738.

De Long, J Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990*b*. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *The Journal of Finance*, 45(2).

Demyanyk, Yuliya, and Otto Van Hemert. 2011. "Understanding the Subprime Mortgage Crisis." *Review of Financial Studies*, 24(6): 1848–1880.

Driscoll, John C, and Aart C Kraay. 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics*, 80(4): 549–560.

Edmans, Alex, Diego Garcia, and Ayvind Norli. 2007. "Sports Sentiment and Stock Returns." *The Journal of Finance*, 62(4): 1967–1998.

Engelberg, Joseph. 2008. "Costly Information Processing: Evidence from Earnings Announcements."

Engelberg, Joseph E., and Christopher A. Parsons. 2011. "The Causal Impact of Media in Financial Markets." *The Journal of Finance*, 66(1): 67–97.

Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh. 2010. "The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk-Sharing in General Equilibrium." National Bureau of Economic Research Working Paper 15988.

Feldman, Ronen, Govindaraj Suresh Livnat Joshua, and Benjamin Segal. 2008. "The Incremental Information Content of Tone Change in Management Discussion and Analysis." SSRN Working Paper.

Ferreira, Fernando, and Joseph Gyourko. 2012. "Heterogeneity in Neighborhood-Level Price Growth in the United States, 1993-2009." *The American Economic Review*, 102(3): 134–40.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy. 2010. "Housing busts and household mobility." *Journal of Urban Economics*, 68(1): 34–45.

Foote, Christopher L. 2007. "Space and Time in Macroeconomic Panel Data: Young Workers and State-Level Unemployment Revisited." Federal Reserve Bank of Boston Working Paper 07-10.

Foote, Christopher Lee, Kristopher Gerardi, and Paul S. Willen. 2008. "Negative equity and foreclosure: Theory and evidence." *Journal of Urban Economics*, 64(2): 234–245.

Galbraith, John. 1990. *A Short History of Financial Euphoria.* New York:Viking Press.

Genesove, David, and Christopher J. Mayer. 1997. "Equity and Time to Sale in the Real Estate Market." *The American Economic Review*, 87(3): pp. 255–269.

Genesove, David, and Christopher Mayer. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *The Quarterly Journal of Economics*, 116(4): pp. 1233–1260.

Gentzkow, M., and J.M. Shapiro. 2007. "What Drives Media Slant? Evidence from US Daily Newspapers." National Bureau of Economic Research Working Paper 12707.

Gentzkow, Matthew, and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy*, 114(2): pp. 280–316.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica*, 78(1): pp. 35–71.

Gerardi, Kristopher, Andreas Lehnert, Shane M. Sherlund, and Paul Willen. 2008. "Making Sense of the Subprime Crisis." *Brookings Papers on Economic Activity*, 2008: pp. 69–145.

Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics*, 1(2): 35–52.

Glaeser, Edward L., Joseph Gyourko, and Albert Saiz. 2008. "Housing supply and housing bubbles." *Journal of Urban Economics*, 64(2): 198 – 217.

Glaeser, Edward L., Joshua D. Gottlieb, and Joseph Gyourko. 2010. "Can Cheap Credit Explain the Housing Boom?" National Bureau of Economic Research Working Paper 16230.

Goetzmann, William, Liang Peng, and Jacqueline Yen. 2012. "The Subprime Crisis and House Price Appreciation." *The Journal of Real Estate Finance and Economics*, 44(1): 36–66.

Greenwood, R., and S. Nagel. 2009. "Inexperienced Investors and Bubbles." *Journal of Financial Economics*, 93(2): 239–258.

Gyourko, Joseph E., Albert Saiz, and Anita A. Summers. 2008. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index." *Urban Studies*, 45(3): 693–729.

Hanley, Kathleen Weiss, and Gerard Hoberg. 2010. "The Information Content of IPO Prospectuses." *The Review of Financial Studies*, 23(7): pp. 2821–2864.

Harrison, J.M., and D.M. Kreps. 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *The Quarterly Journal of Economics*, 92(2): 323–336.

Haughwout, Andrew F., Donghoon Lee, Joseph S. Tracy, and Wilbert Van der Klaauw. 2011. "Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis." FRB of New York Staff Report Working Paper 514.

Haughwout, Andrew F., Okah Ebiere, and Joseph S. Tracy. 2009. "Second Chances: Subprime Mortgage Modification and Re-Default." FRB of New York Staff Report 417.

Henry, Elaine. 2008. "Are Investors Influenced By How Earnings Press Releases Are Written?" *Journal of Business Communication*, 45(4): 363–407.

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. 2005. "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions." *The Journal of Economic Perspectives*, 19(4): pp. 67–92.

Ho, Giang, and Anthony Pennington-Cross. 2008. "Predatory Lending Laws and the Cost of Credit." *Real Estate Economics*, 36(2): 175–211.

Hong, Harrison, and Jeremy C. Stein. 1999. "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets." *The Journal of Finance*, 54(6): pp. 2143–2184.

Jegadeesh, N., and D. Wu. 2011. "Word Power: A New Approach for Content Analysis." AFA 2012 Chicago Meetings Paper.

Keys, Benjamin J., Amit Seru, and Vikrant Vig. 2012. "Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets." *Review of Financial Studies*, 25(7): 2071–2108.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *The Quarterly Journal of Economics*, 125(1): pp. 307–362.

Kindleberger, Charles P. 1978. *Manias, Panics, and Crashes: A History of Financial Crises.* . First ed., John Wiley and Sons, Inc.

Kothari, SP, X. Li, and J.E. Short. 2009. "The Effect of Disclosures by Management, Analysts, and Business Press on Cost of Capital, Return Volatility, and Analyst Forecasts: A Study Using Content Analysis." *The Accounting Review*, 84(5): 1639–1670.

Krainer, John. 2001. "A Theory of Liquidity in Residential Real Estate Markets." *Journal of Urban Economics*, 49(1): 32 – 53.

Lai, Rose N., and Robert A. Van Order. 2010. "Momentum and House Price Growth in the United States: Anatomy of a Bubble." *Real Estate Economics*, 38(4): 753–773.

Lamont, Owen, and Jeremy C. Stein. 1999. "Leverage and House-Price Dynamics in U.S. Cities." *The RAND Journal of Economics*, 30(3): pp. 498–514.

Li, Feng. 2006. "Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?"

Lord, Charles G.; Ross, Lee; Lepper Mark R. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology*, 37(11): 2098–2109.

Loughran, Tim, and Bill Mcdonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance*, 66(1): 35–65.

Lusardi, Annamaria, and Olivia S. Mitchell. 2007*a*. "Baby Boomer Retirement Security: The Roles of Planning, Financial literacy, and Housing Wealth." *Journal of Monetary Economics*, 54(1): 205–224.

Lusardi, Annamaria, and Olivia S Mitchell. 2007*b*. "Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education." *Business Economics*, 42(1): 35–44.

Mankiw, N.Gregory, and David N. Weil. 1989. "The Baby Boom, the Baby Bust, and the Housing Market." *Regional Science and Urban Economics*, 19(2): 235 – 258.

Mayer, Christopher, and Karen Pence. 2008. "Subprime Mortgages: What, Where and to Whom?" Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board 2008-29, Washington, D.C.

Mayer, Christopher, and Todd Sinai. 2009. "U.S. House Price Dynamics and Behavioral Finance." *Policy Making Insights from Behavioral Economics*, , ed. Christopher L. Foote, Lorenz Goette and Stephan Meier, Chapter 5. Boston, MA:Federal Reserve Bank of Boston.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives*, 23(1): 27–50.

Mayer, Christopher, Kathleen W. Johnson, and Oliver Faltin-Traeger. 2010. "Issuer Credit Quality and the Price of Asset Backed Securities." *The American Economic Review*, 100(2): pp. 501–5.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2010. "The Political Economy of the US Mortgage Default Crisis." *The American Economic Review*, 100(5): 1967–98.

Mian, Atif, and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *The Quarterly Journal of Economics*, 124(4): pp. 1449–1496.

Mullainathan, Sendhil, and Andrei Shleifer. 2005. "The Market for News." *The American Economic Review*, 95(4): pp. 1031–1053.

Nakajima, Makoto. 2011. "Understanding House-Price Dynamics." *Business Review*, , (Q2): 20–28.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3): pp. 703–708.

Odean, Terrance. 1998. "Volume, Volatility, Price, and Profit When All Traders Are above Average." *The Journal of Finance*, 53(6): pp. 1887–1934.

Odean, Terrance. 1999. "Do Investors Trade Too Much?" *The American Economic Review*, 89(5): pp. 1279–1298.

Piazzesi, Monika, and Martin Schneider. 2009. "Momentum Traders in the Housing Market: Survey Evidence and a Search Model." *The American Economic Review*, pp. 406–11.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig. 2010. "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis." *Journal of Financial Economics*, 97(3): 369 – 397.

Poterba, James M. 1984. "Tax Subsidies to Owner-Occupied Housing: An Asset Market Approach." *The Quarterly Journal of Economics*, 99(4): 729–52.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy*, 90(6): pp. 1257–1278.

Roget's 21st Century Thesaurus, Third Edition. 2012. *Roget's 21st Century Thesaurus, Third Edition.*

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In *In Current Issues in Urban Economics.* , ed. Peter Mieszkowski and Mahlon Straszheim. Baltimore:Johns Hopkins Univerity Press.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *The Quarterly Journal of Economics*, 125(3): 1253–1296.

Scheinkman, José A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy*, 111(6): pp. 1183–1220.

Shiller, Robert J. 2005. *Irrational Exuberance.* Princeton, NJ:Princeton University Press.

Shiller, Robert J. 2008. *The Subprime Solution.* Princeton, NJ:Princeton University Press.

Shiller, Robert J. 2009. *Animal Spirits.* Princeton, NJ:Princeton University Press.

Stambaugh, Robert F. 1999. "Predictive regressions." *Journal of Financial Economics*, 54(3): 375–421.

Stein, Jeremy C. 1995. "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects." *The Quarterly Journal of Economics*, 110(2): 379–406.

Taylor, John B. 2009. *Getting Off Track: How Government Actions and Interventions Caused, Prolonged, and Worsened the Financial Crisis.* Stanford:Hoover Institution Press.

Tetlock, Paul 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, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. "More than Words: Quantifying Language to Measure Firms' Fundamentals." *The Journal of Finance*, 63(3): pp. 1437–1467.

Wheaton, William C. 1990. "Vacancy, Search, and Prices in a Housing Market Matching Model." *Journal of Political Economy*, 98(6): pp. 1270–1292.

Yariv, Leeat. 2002. "I'll See It when I Believe It: A Simple Model of Cognitive Consistency." *Cowles Foundation Discussion Paper No. 1352.*

Zhu, Yu, Randall Wright, and Chao He. 2012. "Housing and Liquidity." Society for Economic Dynamics.

Table I: Descriptive Statistics for Newspaper Housing Articles

	Newspaper Publication	# Articles	AP	A-section	Real Estate	Local	General
All Cities		19,620	6%	19%	20%	28%	45%
Atlanta	The Atlanta Journal-Constitution	647	0	24	13	29	60
Boston	Boston Herald/Boston Globe	966	3	23	15	24	43
Charlotte	The Observer	556	14	23	28	17	33
Chicago	Chicago Tribune	1,965	8	79	66	13	27
Cleveland	The Plain Dealer	303	1	18	13	20	62
DC	The Washington Post	1,171	6	13	38	27	24
Dallas	The Dallas Morning News	1,294	0	3	0	74	22
Denver	The Denver Post	432	1	13	0	11	83
Detroit	Detroit News/Detroit Free Press	624	5	48	23	10	55
LA	LA Times/LA Daily News	3,579	5	17	14	18	69
Las Vegas	Las Vegas Review-Journal	588	0	15	0	4	92
Miami	The Miami Herald	678	7	27	11	14	51
Minneapolis	Star Tribune	625	1	17	20	1	79
NYC	New York Times	1,372	4	19	33	17	42
Phoenix	The Arizona Republic	1,921	0	19	5	52	29
Portland	The Oregonian	509	2	18	16	35	42
San Diego	The San Diego Union-Tribune	1,086	7	14	26	16	52
San Francisco	The San Francisco Chronicle	530	0	27	8	8	81
Seattle	The Seattle Times	398	29	25	36	5	59
Tampa	Tampa Tribune	376	0	30	2	43	41

Note: Table 1 lists each city, its corresponding newspaper, and descriptive statistics for my sample of housing news articles. My source for housing news articles is Factiva.com, which provides a subject code to identify articles that cover housing market news. My sample covers articles from January 2000 to August 2011. "AP" lists the percent of articles that are credited to the Associated Press. "A-section" refers to the percent of articles located in the front or "A" section of the newspaper. "Real Estate" is the percent of articles that were published in a special real estate section of the newspaper. "Local News" refers to those articles listed in the metropolitan or any specific regional news section of the newspaper. Most of the articles are found in a general news or business news section of the newspaper. It is possible for one article to show up in more than one category. For example, if an article is in the real estate section of the regional edition of the newspaper than it would show up in both columns 6 and 7. Thus, the percents will not necessarily add up to 100 percent for each city.

Table II: Sample Positive Words and Word Counts

word	% of Total PosWord Count	Freq.
BOOM	3.24	959
BOOST	1.17	348
BRIGHT	0.36	106
EXCEED	0.33	98
EXTEND	0.52	154
GOOD	2.29	678
GREAT	0.69	203
HEAT	3.1	917
HOPE	0.69	205
JUMP	2.67	790
LEAP	0.49	145
POSITIVE	0.3	89
SHOOT	0.44	130
SIZZLE	0.48	143
SKYROCKET	0.34	101
SOAR	2.23	660
SPIKE	0.32	96
SPRINGING	0.49	145
STRONG	2.4	711
SURGE	1.91	565

Note: This base list of positive words are from the word lists *Increas* and *Rise* word lists in the Harvard IV-4 Psychological Dictionary. I use these lists to maintain the objectivity of a predetermined list, but also reflect how the media spins excitement over asset markets. Shiller (2008) in particular argues that the media expresses a positive slant through superlatives that emphasize price increases and upward movements. I then expand the original word list with synonyms, alternate tenses, and inflections. I also eliminate obvious misclassifications. The original Harvard list consisted of 136 words while the extended Inc-NEW list contains 403 words. This table presents a sample of words and their corresponding word counts.

Table III: Summary Statistics – Sentiment, Prices, Volume, and Fundamentals

	Obs.	Mean	Std.Dev.	Min	Max
Housing Sentiment Indexes :					
Composite-20	139	1.093	1.716	-3.152	4.435
Cities	2515	1.102	2.606	-10.355	9.979
Case-Shiller Housing Price Indexes:					
Composite-20	138	154.451	31.077	100.000	206.520
Cities	2760	144.543	40.856	64.030	280.870
Volume of Housing Transactions:					
Composite-20	114	5326.172	1580.046	2538.817	9797.987
City	2046	4584.942	2762.715	160.000	21809.000
Fundamentals:					
Real Interest Rate	138	1.833	0.812	-0.010	4.020
30-yr Mortgage Rate	138	6.087	0.933	4.230	8.520
LIBOR 6-month Rate	138	3.001	2.006	0.400	7.000
Rental Index	2242	113.317	12.255	89.661	154.958
Unemployment Rate	2760	6.074	2.461	2.100	16.600
Employment (Thousands)	2760	2161.150	1917.453	158.500	8757.600
Housing Starts	2760	1259.820	1099.021	49.000	6291.000
Building Permits	2760	1894.101	1534.131	57.000	20802.000
Log Population	2400	15.201	0.635	14.109	16.764
Log Income	2400	18.882	0.699	17.554	20.755
Average Loan-to-Value	1872	0.743	0.099	0.331	0.882
Share of Subprime Lending (in Amt)	1872	12.937	7.247	0.000	34.963
Log Loan Applicant Income	1728	11.450	0.265	10.939	12.192

Note: Housing sentiment indices in this table are the difference between the share of positive and negative words each city-month ($pos - neg/total$), see Section 3 for full details on how the index is calculated. Data for the sentiment indices go through July 2011, Case-Shiller home prices are reported with a two-month lag so are available through June 2011, volume of housing transactions are provided by DataQuick through June 2009, and rent is available from REIS through October 2009. Composite-20 versions of the housing sentiment index and transaction volume are calculated using the same normalized weights used to calculate the Composite-20 Case-Shiller index. There are some gaps in newspaper coverage in the data, thus data for housing sentiment indices are not completely balanced. The index can only be calculated for months where newspaper coverage is available in the data, thus some cities are missing sentiment index data in months where the newspaper was not covered by Factiva. Details on the sources of the housing fundamentals are available in Section 2.

Table IV: Sentiment Predicts National House Price Appreciation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sum of Lagged Sentiment	0.756***	0.763***	0.760***	0.760***	0.760***	0.763***	0.838***	0.696***	0.932***	0.120
	(0.080)	(0.076)	(0.074)	(0.077)	(0.079)	(0.078)	(0.125)	(0.163)	(0.206)	(0.073)
Year 1 Lags ($L^1 + ... + L^{12}$)	0.213***	0.184***	0.189***	0.189***	0.186***	0.185***	0.193***	0.169***	0.273***	0.021
	(0.018)	(0.026)	(0.025)	(0.026)	(0.024)	(0.025)	(0.029)	(0.039)	(0.071)	(0.021)
Year 2 Lags ($L^{13} + ... + L^{24}$)	0.101***	0.087***	0.089***	0.089***	0.088***	0.087***	0.095***	0.100***	0.164***	0.012
	(0.030)	(0.027)	(0.024)	(0.024)	(0.023)	(0.024)	(0.025)	(0.028)	(0.042)	(0.011)
Year 3 Lags ($L^{25} + ... + L^{43}$)	0.442***	0.493***	0.482***	0.482***	0.486***	0.491***	0.550***	0.427***	0.495**	0.087
	(0.051)	(0.049)	(0.051)	(0.053)	(0.056)	(0.058)	(0.103)	(0.139)	(0.176)	(0.052)
Rents	.	✓	✓	✓	✓	✓	✓	✓	✓	✓
Interest Rate Variables	.	.	✓	✓	✓	✓	✓	✓	✓	✓
Labor Market Variables	.	.	.	✓	✓	✓	✓	✓	✓	✓
Housing Supply	.	.	.	.	✓	✓	✓	✓	✓	✓
Population and Income	.	.	.	.	.	✓	✓	✓	✓	✓
Month Fixed Effects	.	.	.	.	.	.	✓	✓	✓	✓
Linear Time Trend	.	.	.	.	.	.	.	✓	.	.
Lagged Fundamentals	.	.	.	.	.	.	.	.	✓	.
Lagged Price Growth	.	.	.	.	.	.	.	.	.	✓
Observations	94	94	94	94	94	94	94	94	94	94
Adjusted R^2	0.854	0.863	0.866	0.86	0.858	0.852	0.835	0.838	0.841	0.974

Note: * 10% significance, ** 5% level, *** 1% level. Newey and West (1987) standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. L^k denotes the lag $t-k$. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The rows below break down total sum of the monthly lags of sentiment by lagged years. "Year 1 Lags" equals the sum of lagged sentiment from L^1 to L^{12}, "Year 2 Lags" is the sum of lags L^{13} to L^{24}, "Year 3 Lags" is the sum from lags L^{25} to L^{43}.. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. The lag structure is chosen through a joint F-test. Including additional lags after L^{43} does not affect the results. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing.

Table V: Sentiment Predicts City House Price Appreciation (Panel)

									Boom	Bust
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sum of Lagged Sentiment	1.120***	1.223***	1.342***	1.276***	0.670***	1.015***	0.846***	0.266***	0.814***	1.035***
	(0.120)	(0.131)	(0.127)	(0.125)	(0.170)	(0.171)	(0.144)	(0.048)	(0.309)	(0.153)
Year 1 Lags ($L^1 + ... + L^{12}$)	0.248***	0.282***	0.323***	0.294***	0.156***	0.257***	0.200***	0.068***	0.289***	0.206***
	(0.034)	(0.042)	(0.039)	(0.037)	(0.035)	(0.039)	(0.041)	(0.013)	(0.085)	(0.036)
Year 2 Lags ($L^{13} + ... + L^{24}$)	0.311***	0.440***	0.479***	0.462***	0.226***	0.366***	0.314***	0.093***	0.249**	0.371***
	(0.041)	(0.050)	(0.048)	(0.047)	(0.058)	(0.065)	(0.051)	(0.018)	(0.112)	(0.056)
Year 3 Lags ($L^{25} + ... + L^{43}$)	0.560***	0.501***	0.540***	0.521***	0.287***	0.392***	0.332***	0.105***	0.275*	0.458***
	(0.065)	(0.083)	(0.070)	(0.069)	(0.098)	(0.089)	(0.085)	(0.024)	(0.144)	(0.083)
Fundamentals	.	✓	✓	✓	✓	✓	✓	✓	✓	✓
City Fixed Effects	.	.	✓	✓	✓	✓	✓	✓	✓	✓
Month Fixed Effects	.	.	.	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	.	.	.	.	✓	.	.	.	.	.
Linear Time Trend	.	.	.	.	.	✓	.	.	.	.
Lagged Fundamentals	.	.	.	.	.	.	✓	.	.	.
Lagged Price Growth	.	.	.	.	.	.	.	✓	.	.
Observations	1564	1106	1106	1106	1106	1106	1106	1106	399	707
Adjusted R^2	0.287	0.466	0.517	0.582	0.634	0.589	0.669	0.863	0.471	0.576

Note: * 10% significance, ** 5% level, *** 1% level. This table estimates the effect of sentiment across cities. The number of observations decline from columns (1) to (2) because fundamentals are only available through 2009 while sentiment indices are available through July 2011. Newey and West (1987) standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. L^k denotes the lag $t-k$. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The rows below break down total sum of the monthly lags of sentiment by lagged years. "Year 1 Lags" equals the sum of lagged sentiment from L^1 to L^{12}, "Year 2 Lags" is the sum of lags L^{13} to L^{24}, "Year 3 Lags" is the sum from lags L^{25} to L^{43}.. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing.

Table VI: Sentiment Predicts City House Prices Beyond Subprime Lending Trends

	(1)	(2)	(3)	(4)	(5)
Sum of Lagged Sentiment	0.830***	0.778***	0.769***	0.692***	0.421***
	(0.148)	(0.162)	(0.161)	(0.154)	(0.135)
Year 1 Lags ($L^1 + ... + L^{12}$)	0.193***	0.213***	0.212***	0.224***	0.144***
	(0.043)	(0.049)	(0.049)	(0.045)	(0.035)
Year 2 Lags ($L^{13} + ... + L^{24}$)	0.304***	0.290***	0.287***	0.263***	0.169***
	(0.052)	(0.057)	(0.057)	(0.051)	(0.049)
Year 3 Lags ($L^{25} + ... + L^{43}$)	0.333***	0.275***	0.270***	0.204**	0.108
	(0.088)	(0.098)	(0.097)	(0.097)	(0.08)
Month Fixed Effects	✓	✓	✓	✓	✓
City Fixed Effects	✓	✓	✓	✓	✓
Fundamentals	✓	✓	✓	✓	✓
Lagged Fundamentals	✓	✓	✓	✓	✓
LIBOR 6-month rate	✓	✓	✓	✓	✓
% Subprime Loans	.	✓	✓	✓	✓
Loan-To-Value	.		✓	✓	✓
Loan Applicant Income	.		.	✓	✓
Lagged Credit Variables	.		.	.	✓
Observations	1106	876	876	771	771
Adjusted R^2	0.667	0.707	0.709	0.735	0.793

Note: * 10% significance, ** 5% level, *** 1% level. This table estimates the effect of sentiment across cities. The number of observations decline from columns (1) to (2) because data for % of subprime loans are only available for 16 cities in the sample and only through September 2009, observations further decline because loan applicant income from the HMDA database are only available through 2008.Newey and West (1987) standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. L^k denotes the lag $t-k$. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The rows below break down total sum of the monthly lags of sentiment by lagged years. "Year 1 Lags" equals the sum of lagged sentiment from L^1 to L^{12}, "Year 2 Lags" is the sum of lags L^{13} to L^{24}, "Year 3 Lags" is the sum from lags L^{25} to L^{43}.. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. The lag structure is chosen through a standard joint F-test. Including additional lags after L^{43} does not affect the results. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing.

Table VII: Sentiment Predicts the Volume of Housing Transactions

	Composite			Panel		
	(1)	(2)	(3)	(4)	(5)	(6)
Sum of Lagged Sentiment	4.674***	4.909***	5.531***	3.658***	2.709***	1.479**
	(1.189)	(1.337)	(1.915)	(0.678)	(0.668)	(0.594)
Year 1 Lags ($L^1 + ... + L^{12}$)	3.555***	3.938***	4.13**	2.957***	2.355***	1.381***
	(0.890)	(1.012)	(1.681)	(0.556)	(0.532)	(0.486)
Year 2 Lags ($L^{13} + ... + L^{18}$)	1.119**	0.971	1.401	0.701***	0.354*	0.098
	(0.505)	(0.874)	(0.893)	(0.208)	(0.206)	(0.153)
Rents	.	✓	✓	.	✓	✓
Interest Rate Variables	.	✓	✓	.	✓	✓
Labor Market Variables	.	✓	✓	.	✓	✓
Housing Supply	.	✓	✓	.	✓	✓
Population and Income	.	✓	✓	.	✓	✓
Month Fixed Effects	.	✓	✓	.	✓	✓
City Fixed Effects	.	.	.	.	✓	✓
Lagged Fundamentals	.	.	✓	.	.	✓
Observations	96	96	96	1578	1481	1437
Adjusted R^2	0.430	0.613	0.595	0.068	0.261	0.507

Note: * 10% significance, ** 5% level, *** 1% level. Newey-West standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. This tables estimates the effect of sentiment on detrended log volume. I use detrended log volume to address non stationarity concerns, and detrend volume following Campbell, Grossman and Wang (1993). Specifically, I subtract the one year backward moving average. Newey and West (1987) standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. L^k denotes the lag $t-k$. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The rows below break down total sum of the monthly lags of sentiment by lagged years. "Year 1 Lags" equals the sum of lagged sentiment from L^1 to L^{12}, "Year 2 Lags" is the sum of lags L^{13} to L^{24}, "Year 3 Lags" is the sum from lags L^{25} to L^{43}.. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. The lag structure is chosen through a standard joint F-test. Including additional lags after L^{43} does not affect the results. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing.

Table VIII: Explanatory Power of Observed Fundamentals Pre- and Post-2000

	Composite-10		Panel	
	Pre-2000	Post-2000	Pre-2000	Post-2000
	(1)	(2)	(4)	(5)
Rents	1.424***	0.373	0.840***	0.365**
	(0.166)	(0.704)	(0.110)	(0.179)
Interest Rate Variables	✓	✓	✓	✓
Labor Market Variables	✓	✓	✓	✓
Housing Supply	✓	✓	✓	✓
Population and Income	✓	✓	✓	✓
Month Fixed Effects	✓	✓	✓	✓
City Fixed Effects	.	.	✓	✓
Observations	119	118	2136	2241
Adjusted R^2	**0.693**	**0.092**	**0.363**	**0.234**

Note: * 10% significance, ** 5% level, *** 1% level. Newey-West standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. L^k denotes the lag $t-k$. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing. This table shows that they key set of fundamentals explain prices much better prior to the suspected bubble period, post-2000. For example, the R^2 in column 1 shows that the key set of fundamentals is able to explain nearly 70 percent of the variation in aggregate price growth prior to 2000. After 2000, however, this same set of fundamentals explains very little of the variation in price growth with an adjusted $R^2 = 0.09$. This suggests that the main set of results at least incorporate the key set of fundamentals that typically explain housing price growth, and that price movements post-2000 must be due to some other variable. Thus, sentiment estimates in the main results are less likely driven by bias from an unobserved fundamental.

Table IX: Is Sentiment Driven By News Stories on Unobserved Fundamentals?

	Dep Var: Housing Price Growth, t=monthly						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sum of Lagged Sentiment	0.826***	0.540**	0.636***	0.572***	0.589***	0.530**	0.581**
	(0.148)	(0.212)	(0.208)	(0.21)	(0.215)	(0.256)	(0.279)
Media Rents	✓	✓	✓	✓	✓	✓	✓
Media Credit Conditions	.	✓	✓	✓	✓	✓	✓
Media Labor Market Conditions	.	.	✓	✓	✓	✓	✓
Media Housing Supply	.	.	.	✓	✓	✓	✓
Media User Costs	.	.	.	.	✓	✓	✓
Media Demographics	.	.	.	.	.	✓	✓
Media Local GDP & Inflation	.	.	.	.	.	.	✓
Month Fixed Effects	✓	✓	✓	✓	✓	✓	✓
City Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Fundamentals	✓	✓	✓	✓	✓	✓	✓
Lagged Fundamentals	✓	✓	✓	✓	✓	✓	✓
Observations	1094	1094	1093	1093	1093	1093	1093
Adjusted R^2	0.678	0.683	0.717	0.717	0.717	0.718	0.723

Note: * 10% significance, ** 5% level, *** 1% level. Newey-West standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing. This table directly controls for news content over fundamentals by identifying any news article that mentions a particular fundamental in its text. The variable "Media Rents", for example, is the share of positive minus negative words in any articles that mention any word related to "rents" in its text. This table shows that controlling for articles that mention fundamentals has minimal effect on the estimated effect of sentiment on house prices.

Table X: Correlation of Weekend Instrument with Friday News Releases

% of Releases on Friday	Correlation with Weekend Instrument
All BLS	0.07
Any Metro or Regional	-0.01
County Employment	-0.04
Regional Employment	-0.05
Metro Area Employment	0.00
CPI	-0.02
PPI	0.14
New Residential Construction	-0.02
New Residential Sales	-0.01

Note: This table test for a possible violation of the exclusion restriction for the weekend instrument. The validity of the weekend instrument relies on the assumption that no news on fundamentals is being released over the weekend. One possible violation of this assumption is that news is increasingly released on Friday and therefore reported over the weekend. I put together a database of the schedule of economic data releases from the BLS and the Census. This table shows that the fraction released on Friday is uncorrelated with the share of positive minus negative words over the weekend. The first column lists the types of press releases, including all releases by the Bureau of Labor Statistics, any release on metropolitan or regional specific fundamentals, release on employment, measures of inflation, and housing specific fundamentals from the Census. The second column reports the simple correlation between the fraction of these releases that occur on Friday with the weekend instrument.

TABLE XI: WEEKEND AND NARRATIVE INSTRUMENTS FOR SENTIMENT, FIRST-STAGE

	Dep Var: Sentiment Growth, t=monthly	
	Weekend	Narrative
Instrument	0.458***	0.208***
	(0.100)	(0.031)
Weekday News Tone	✓	.
Rents	✓	✓
Interest Rate Variables	✓	✓
Labor Market Variables	✓	✓
Housing Supply Variables	✓	✓
Population and Income	✓	✓
Month Fixed Effects	✓	✓
City Fixed Effects	✓	✓
Lagged Fundamentals	✓	✓
F-statistic	**233.776**	**46.089**
Observations	1856	1856
Adjusted R^2	0.663	0.108

Note: * 10% significance, ** 5% level, *** 1% level. Newey-West standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing. This table reports the first-stage estimates of sentiment on the weekend and narrative instruments. The bottom panel reports the F-statistic for the instruments in bold to test for instrument strength. Both instruments are sufficiently relevant to the housing sentiment index, with F-statistics well above the benchmark rule of 10.

Table XII: Predicting Price Growth Using Positive Sentiment, IV Results

	OLS	Weekend IV	Narrative IV
	(1)	(2)	(3)
Sum of Lagged Sentiment	0.837***	1.247***	0.805**
	(0.096)	(0.217)	(0.382)
Year 1 Lags (L1+...+L12)	0.18***	0.305	0.01
	(0.035)	(0.187)	(0.304)
Year 2 Lags (L13+...+L24)	0.294***	0.500***	0.47**
	(0.039)	(0.153)	(0.215)
Year 3 Lags (L25+...+L43)	0.363***	0.441***	0.326
	(0.042)	(0.091)	(0.201)
Weekday News Tone	.	✓	.
Month Fixed Effects	✓	✓	✓
City Fixed Effects	✓	✓	✓
Fundamentals	✓	✓	✓
Lagged Fundamentals	✓	✓	✓
Observations	1106	1106	1106
Adjusted R^2	0.669	0.647	0.648

Note: * 10% significance, ** 5% level, *** 1% level. Newey-West standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing. This table presents the original OLS estimates in column (1), and the instrumental variable estimates using the weekend and narrative instruments in columns (2) and (3) respectively. The estimated effect of sentiment remains robust to both instrumental variable strategies, suggesting bias from unobserved factors in the original estimates are less likely.

Table A.1: Comparing Effect Of Alternative Sentiment Indices

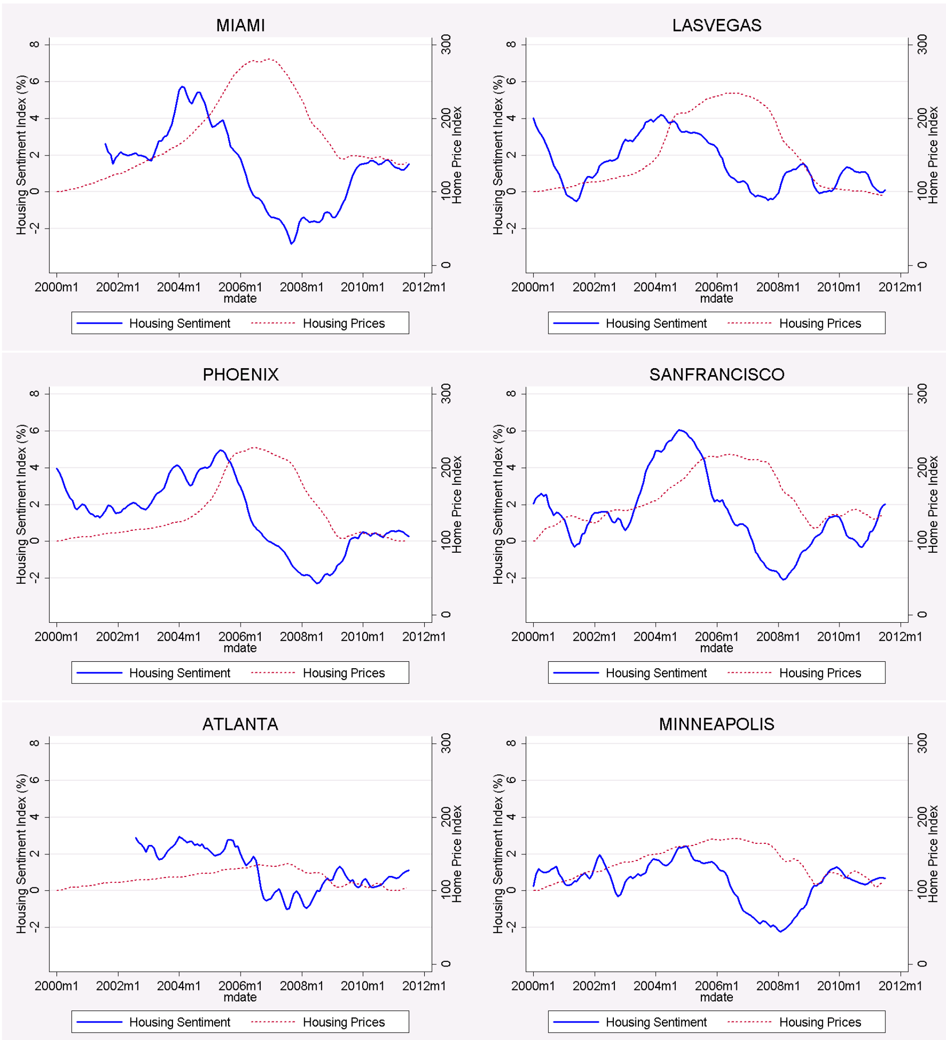
	Dep Var: Housing Price Growth, t=monthly					
	(1)	(2)	(3)	(4)	(5)	(6)
	Pos-Neg *(baseline)*	Pos-Neg *(full text)*	Pos-Neg* *% housing articles*	Positive	Positive *(term-weighted)*	Negative
Sum of Lagged Sentiment	0.846***	0.803***	0.802***	0.264***	0.277**	-0.349***
	(0.144)	(0.137)	(0.149)	(0.075)	(0.108)	(0.069)
Fundamentals	✓	✓	✓	✓	✓	✓
City Fixed Effects	✓	✓	✓	✓	✓	✓
Month Fixed Effects	✓	✓	✓	✓	✓	✓
Lagged Fundamentals	✓	✓	✓	✓	✓	✓
Correlation with *baseline*	1.00	0.784	0.655	0.674	0.510	-0.666
Observations	1106	1106	1106	1106	1106	1106
Adjusted R^2	0.669	0.694	0.696	0.624	0.615	0.662

Note: * 10% significance, ** 5% level, *** 1% level. Newey-West standard errors that are robust to heteroskedasticity and auto-correlation up to 12 lags are in parentheses. Sum of Lagged Sentiment sums all the coefficient estimates of current and lagged sentiment growth together. The corresponding standard errors for the linear combination of estimates are reported in parentheses below. Estimates of lagged logged sentiment measure the impact of a one percent increase in the monthly growth of sentiment on the monthly growth in prices i.e. monthly capital appreciation on housing. This table compares the effect of alternate versions of the sentiment index on house prices and shows that results are qualitatively the same. "Pos-Neg" represents the difference between the share of positive and negative words in the leading text of housing articles each city-month. Pos-Neg *(full text)* is the same index calculated over the full text of the articles. Column (3) adds another dimension of sentiment by interacting the baseline index with the fraction of all newspaper articles that cover the housing market. This index accounts for both the tone of newspaper articles and the number of articles published on housing. Columns (4)-(6) considers the effect of positive and negative sentiment separately. Column (4) uses just the share of positive words. Column (5) calculates a "term-weighted" positive index, which add weights for the commonality and frequency of a word across documents (Loughran and Mcdonald (2011)). "Negative" is the share of negative words across articles each month.

Figure I: Composite-20 Housing Sentiment and Case-Shiller Home Price Index

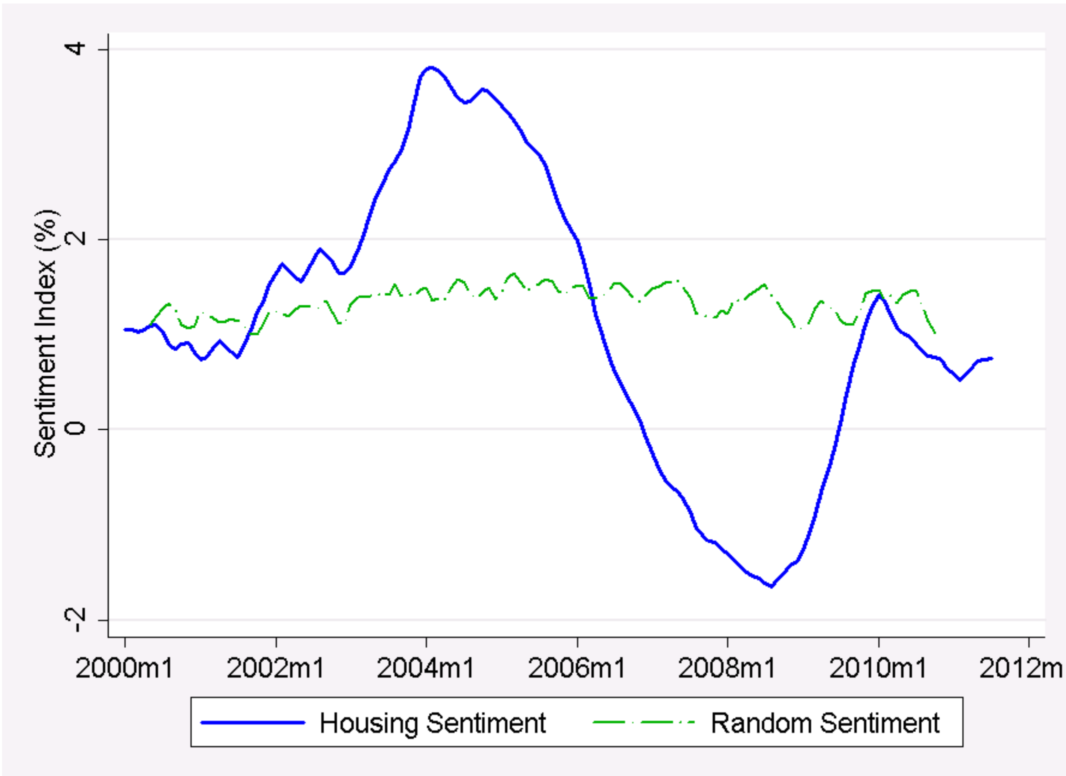
Note: This figure plots the composite-20 sentiment index and the composite-20 Case-Shiller housing price index. Lines are smoothed for seasonal variation and noise with a 6-month backward and forward moving average. Housing prices and sentiment are calculated using a 3-month backward moving average in empirical estimations.

Figure II: Housing Sentiment and Case-Shiller Home Price Indexes by City



Note: Figure 2 plots the housing sentiment index and housing price indexes for individual cities. Lines are smoothed for seasonal variation and noise with a 6-month backward and forward moving average. Housing prices and sentiment are calculated using a 3-month backward moving average in empirical estimations.

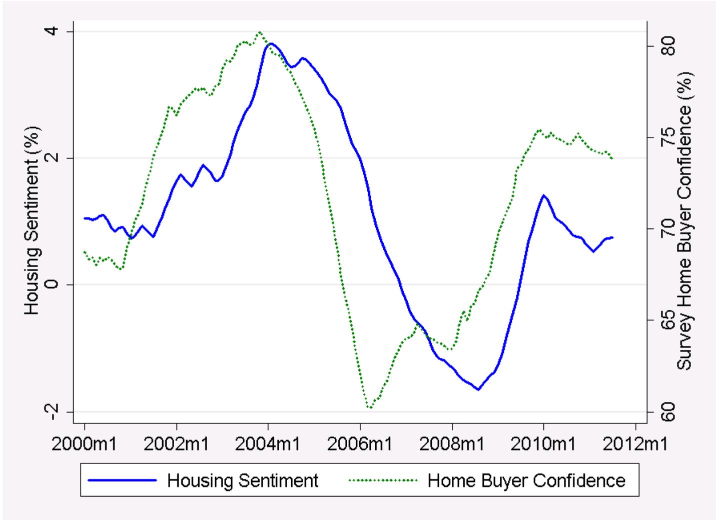
Figure III: Random Sentiment Placebo Test



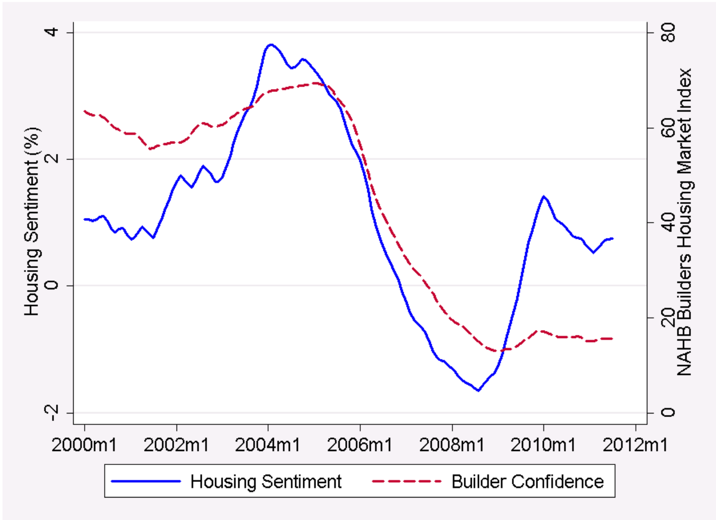
Note: Figure 3 presents evidence that the pattern of positive minus negative words is specific to housing articles. "Housing Sentiment" is the share of positive minus negative words calculated over newspaper articles that cover the housing market. "Random" is the share of positive minus negative words across a random sample of articles of any subject each city-month. As seen in the plot, random sentiment generally remains relatively flat and does not exhibit the same boom and bust pattern as housing sentiment. Lines are smoothed for seasonal variation and noise with a 6-month backward and forward moving average.

Figure IV: Validating Sentiment Against Surveys of Housing Market Confidence

Panel A. Housing Sentiment Index and Survey of Consumers Home Buyer Confidence

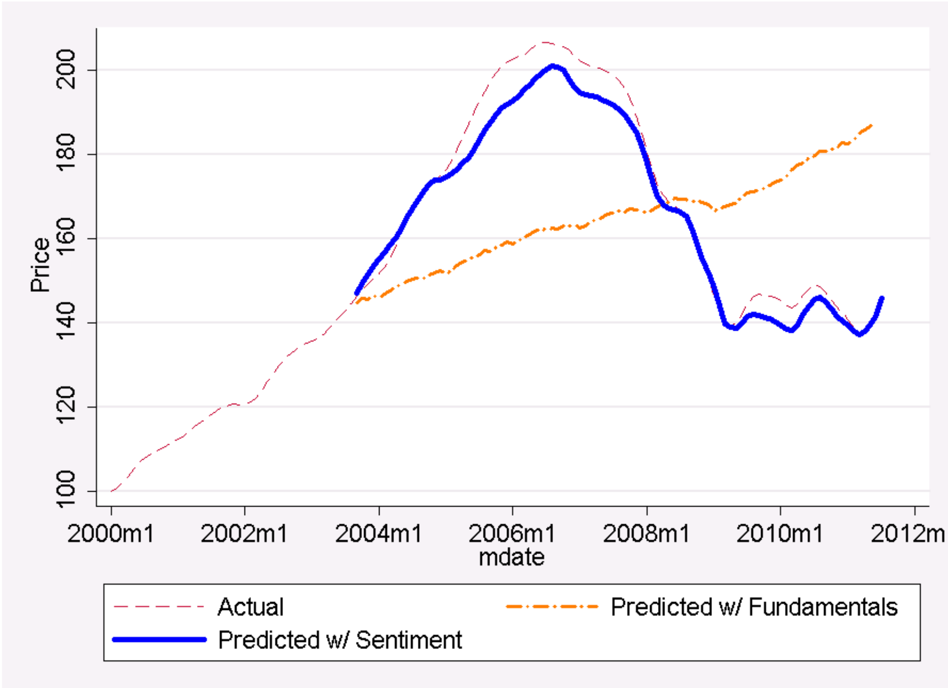


Panel B. Housing Sentiment Index and National Association of Home Builders Confidence Index



Note: Panel A plots the composite-20 housing sentiment index with a national survey of home buyer confidence. The Survey of Consumers surveys a nationally representative sample of 500 consumers and asks whether they think it is a good time to buy a home. Consumers answer "Yes/No/Don't Know." The green dashed line represents the percentage of those surveyed who answered "Yes." Panel B plots the composite housing sentiment index with a national survey of members of home builder confidence. The National Association Home Builders asks members of their association each month to rate the current market conditions of the sale of new homes, the prospective market conditions in the next 6 months, and the expected volume of new home buyers. The NAHB index weights these answers into one index to represent an aggregate builders' opinion of housing market conditions. The timing the sentiment index coincides with survey measures of confidence, suggesting that it is reflecting investor beliefs over the housing market. Lines are smoothed for seasonal variation and noise with a 6-month backward and forward moving average.

Figure V: Predicting House Price Growth with Sentiment Index v. Fundamentals

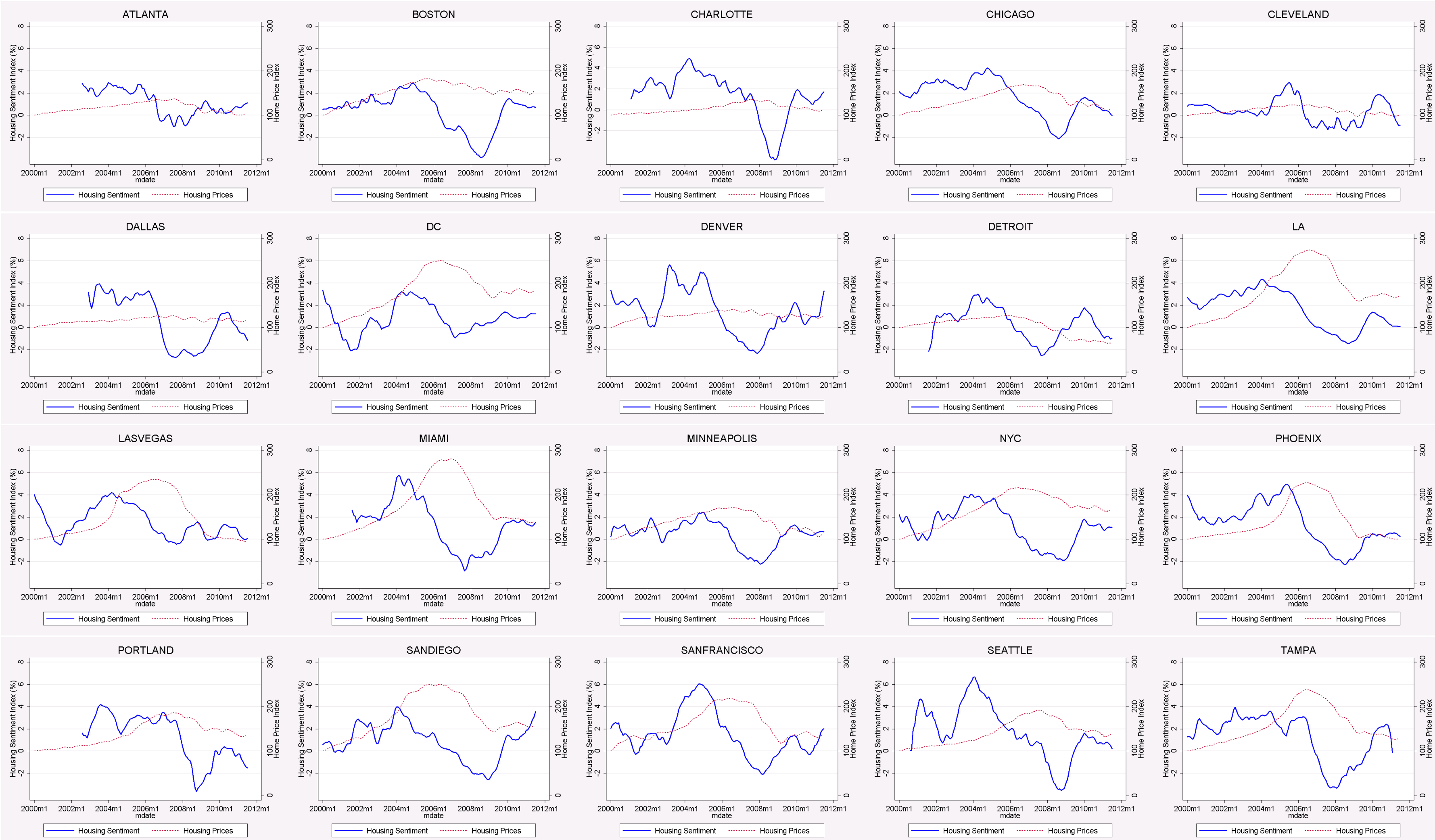


Note: Figure 7 plots observed composite-20 prices and predicted prices. The dashed line represents prices predicted with contemporaneous. fundamentals alone. The solid line plots prices predicted with positive sentiment only. The picture illustrates that sentiment can explain a significant variation in prices. More importantly, sentiment fits the prediction to the timing of the boom and bust, whereas fundamentals only predict a linear projection of prices.

Figure VI: Composite-20 Housing Sentiment Index and Transaction Volume

Note: Figure 8 plots a composite-20 volume of housing transactions and my housing sentiment index. Data for transaction volume comes from DataQuick. I calculate a composite-20 measure of volume using the same weights used to create the Case-Shiller Composite-20 Home price Index. Lines are smoothed for seasonal variation and noise with a 6-month backward and forward moving average.

Figure A.1: Housing Sentiment Index and Housing Prices, By City



Note: This figure plots the housing sentiment index and housing price indexes for each of the sample 20 cities. Lines are smoothed for seasonal variation and noise with a 6-month backward and forward moving average. Housing sentiment index equals the share of positive minus negative words across housing newspaper articles in each city-month.