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
October 2015

QUANTIFYING ANIMAL SPIRITS:

NEWS MEDIA AND SENTIMENT IN THE HOUSING MARKET*

Cindy K. Soo†

WORKING PAPER

October 2015

Abstract

This paper develops first measures of housing sentiment for 34 cities across the U.S. by quantifying the qualitative tone of local housing news. I find that housing media sentiment has significant predictive power for future house prices, above and beyond historically predictive factors and past returns. Sentiment leads price movements by more than two years, and is highly correlated with available survey expectations measures. The structure of the media sentiment index itself reflects a backward-looking nature consistent with extrapolative expectations. Consistent with theories of sentiment, the media sentiment index has a greater effect in markets with more minority homebuyers, more speculative investors, and across lower-priced homes. Including additional controls for subprime lending and easy credit has no impact on the magnitude of the results, but the predictive effect of sentiment is amplified in markets where more subprime loans were issued. Directly investigating the content across news articles finds that results are not driven by news stories of unobserved fundamentals.

*I am deeply grateful to Fernando Ferreira, Joe Gyourko, Olivia Mitchell, Michael Roberts, and Todd Sinai for their invaluable comments and feedback. I am also very thankful to seminar participants at Wharton, 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, Federal Reserve Banks of Boston, Chicago, and Philadelphia, and the Whitebox Advisors Graduate Student Conference for their helpful suggestions and comments. 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

Sentiment, broadly defined as the psychology behind investor beliefs, has long been posited as a determinant of asset price variation (Keynes (1936)). The potential role of sentiment is particularly important for the housing market, where financial errors can have very large consequences. Over two-thirds of households in the United States own a home and invest the majority of their portfolio in real estate (Tracy, Schneider and Chan (1999); Nakajima (2005)). The housing bust in 2006 devastated millions of homeowners across the U.S., and overwhelmed banks and financial institutions that held significant investments in mortgage-backed securities and other housing related assets. The collapse of the subprime mortgage industry quickly followed, leading the nation into its worst economic recession since the 1930s.

Shiller (2009) and others notably argued that "animal spirits" or an irrational exuberance of investors was a significant factor in the dramatic boom and bust of house prices in some markets. Testing this theory, however, is difficult without empirical measures of housing sentiment. Beliefs are not only unobservable and therefore not straightforward to quantify, but 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 discounts are naturally not available for housing (Baker and Wurgler (2006)). Survey measures that are available are limited in geographic scope. While the most recent cycle included cities that experienced unprecedented growth in house prices, there were simultaneously many markets that observed minimal or stable movements in prices (Ferreira and Gyourko (2012)). Therefore understanding the variation of price growth across local markets requires analogous city-level measures.

The goal of this paper is to provide real-time local measures of sentiment across a number of city housing markets. I construct 34 city-specific housing sentiment indices corresponding to major metropolitan areas across the U.S.. I measure housing sentiment by quantifying the qualitative tone of local housing news media coverage. This strategy is motivated by seminal literature on asset price bubbles that argues news media has a prominent relationship with sentiment through an incentive to cater to readers' preferences (Kindleberger (1978); Galbraith (1990); Shiller (2005)). This methodology is further supported by pioneering empirical work by Tetlock (2007) and others who have found media tone to be a consistent proxy for sentiment in the stock market. By analyzing local housing news articles each month, I quantify the share of the positive and negative local media tone in each market from 2000 to the end of 2013. To my knowledge, this paper contributes the first set of city measures of housing sentiment across several markets.

Housing media sentiment indices move ahead of prices at a significant lead of more than two years. Cities that experienced dramatic rises and declines in house prices are preceded by similar cycles in sentiment, while cities with milder price changes are led by more subdued or random patterns in sentiment growth. In cities with large swings in prices, sentiment appears to peak approximately some time in 2004. Though difficult to validate without comparable measures, I employ two available surveys to provide evidence that media sentiment serves as a meaningful proxy for homebuyer expectations. Though not available by city, the Michigan Survey of Consumers (SOC) directly assesses consumers' confidence in the housing market and also finds that their index peaks in early 2004, a few months ahead of a national version of my housing sentiment index. I find that for the four cities available through the Case, Shiller and Thompson (2012) survey, respondents' expectations exhibit similar trends to my media sentiment index city by city as well, with correlations between the two indices ranging from 0.7 to 0.9.

Consistent with these leading patterns, I find that housing media sentiment has robust predictive power for future house price growth. I find these effects to be both statistically significant and large in economic magnitude. This is notable as historical predictive factors have had difficulty explaining the wide swings in house prices during this period. Fundamental determinants that traditionally explained past patterns in house prices account for just a small fraction of changes post-2000 (Lai and Van Order (2010)). From 2004 to 2006, for example, house prices in Miami increased by nearly 54 percent. Observed economic fundamentals account for approximately 21 percentage points of this growth, while the media sentiment index explains an additional 32 percentage points.

These results suggest that media sentiment serves as a useful predictive factor for house prices above and beyond traditional observed variables. The positive association between the media sentiment index and future house prices is consistent with two different interpretations. One explanation is that these results capture the effects of sentiment, where over-exuberant beliefs pushed prices away from fundamentals. Examining the structure of the media sentiment index more closely does reveal a backward-looking nature over past returns consistent with behavioral theories of extrapolative expectations. An alternate interpretation is that local unobserved fundamentals simultaneously drove house prices and beliefs such that the expectations of home buyers were justified by local information. One could even argue that the leading pattern of the media sentiment index resembles a perfect forecast in some cities. The media sentiment index could instead serve as a valuable indicator of local market information that was otherwise unobserved.

I subsequently perform a series of tests to explore these two competing explanations. I first exploit the textual nature of my data to directly evaluate the media index as a proxy for local information. By tracking all words across all articles, I can examine the content across articles firsthand. I then separate the articles to create a set of "media fundamentals" indices that track the positive and negative tone across articles that discuss any relevant housing market information. I find that the effect of the housing sentiment index remains extremely robust to a discussion of fundamentals, both separately and together. The only set of articles that shares some predictive power with the media sentiment index are those that discuss credit conditions. The inclusion of additional controls for subprime lending patterns and the availability of credit, however, does not affect the significance and magnitude of the predictive effect of the housing media sentiment index.

I then evaluate the sentiment interpretation of the media index by isolating a subset of articles that likely only reflect sentiment but lack any informational content. Through surveys of newspaper readership, I identify two types of articles that cater to readers who are more likely subject to sentiment. The first are those articles published over the weekend. Surveys report that weekend readership is concentrated among those who have preferences for lighter content, and thus demand more entertaining, sentiment-driven articles. The second set are those I define as "narrative" news articles. Narratives or "people-style" news are reported to increase readership among lighter readers through a human interest appeal, and consist of anecdotal stories rather than actual information. Using the detailed information in my dataset, I develop two alternative sentiment measures based on these smaller subsample of articles. Narrowing the identifying variation to these alternate measures still finds a large and significant effect of media sentiment on future house prices. Even restricting the variation to just the trend in weekend tone over and above those that occur over all weekdays does not affect the results.

In the final section, I take advantage of the availability of the cross-sectional indices to test effects that should be consistent with an interpretation of sentiment. Since arbitrage is particularly binding in the housing market, theory predicts that markets with less-informed demand should be subject to greater effects of sentiment. Shiller (2009) and others have raised concerns particularly over the disproportionate lack of access to financial advice available to minority and low-income households. Consistent with these arguments, I find that sentiment effects are significantly greater in markets with more minority mortgage applicants and among lower-priced homes. Theories on sentiment further suggest that rational traders may respond to increased sentiment in the market to anticipate increasing asset prices. Indeed, I find that the effect of sentiment is also notably larger among housing markets with a greater entrance of speculative investors. Finally, I show that while subprime lending patterns do not affect the impact of sentiment on prices, the effect of sentiment is proportionally greater in markets where subprime lending was more prevalent. Many voiced suspicions that subprime lenders targeted those most uninformed, and we would expect those subject to sentiment to be more vulnerable to taking out risky, subprime loans. These results also provide potential context for why prior studies have been unable to explain the magnitude of the boom with the ease of credit alone, and highlight potential subprime lending effects through a relationship with sentiment.

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)). At the same time, the evidence in this paper relates to a large body of work that explores determinants and consequences of the last housing boom and bust (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)). This paper also generally relates to a larger literature that explores housing price dynamics and more specifically to 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)). Finally, this paper contributes to research that links media coverage to trading activity and shows that media sentiment can be used to predict asset prices beyond stock market applications (Tetlock (2007); Tetlock, Saar-Tsechansky and Macskassy (2008); Tetlock (2011); Antweiler and Frank (2004); Barber and Loeffler

(1993); Dougal et al. (2012); Dyck and Zingales (2003); Engelberg (2008); Engelberg and Parsons (2011); Garcia (2012); Gurun and Butler (2012)).

Section 2 details how the sentiment index is calculated and validated against existing survey measures. Section 3 describes the trends in the housing media sentiment index, and reports the results of predictive regressions between the media sentiment index and prices. In Section 4, I explore the interpretation of the index as a proxy for sentiment versus unobserved fundamentals. Section 5 concludes and discusses potential avenues for future work.

2 Measuring Housing Market Sentiment with the Media

Prominent literature on bubbles and panics 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 reflect readers' expectations over assets they own. Mullainathan and Shleifer (2005) formalize these arguments by assuming readers have a disutility for news that is inconsistent with their beliefs, citing psychology literature that shows people have a tendency to favor information that confirms their priors. Gentzkow and Shapiro (2010) find empirical evidence that readers have a preference for news consistent with their political beliefs and that news outlets respond accordingly. At the same time, Shiller (2005) argues that news outlets have the power to influence reader beliefs through its chosen tone or emphasis of particular positive or negative events. Indeed, studies such as Dougal et al. (2012) have found supporting causal evidence that financial journalists can influence investor behavior and stock market returns.

Housing in particular receives heavy media attention as "a source of endless fascination for the general public, because we live in houses, we work on them every day" (Shiller (2005)). Housing is also a widely held investment by individual buyers who may be more likely subject to media slant than the typical stock market investor. Thus whether the news media influences or reflects beliefs, any positive relationship presents a unique opportunity to capture expectations in a market where alternative proxies are otherwise difficult to attain. Sentiment proxies in the stock market such as mutual fund flows, IPO first day returns, and the close-end fund discount, for example, are naturally not available for the housing market (Baker and Wurgler (2006)). Surveys such as Case and Shiller (2003) and the Michigan/Reuters Survey of Consumers provide direct assessments of home buyer expectations, but can be expensive to run and are therefore limited in frequency and/or restricted in geographic breakdown. Because news is local and recurring, however, the news media provides a potential medium through which we can quantify housing expectations both across time and city by city.

My source for news articles is Factiva.com, a comprehensive online database of newspapers. Factiva categorizes its articles by subject, and provides a subject 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. Routine real estate property listings are not included. Wire-service articles (such as those by the Associated Press) are also generally excluded, as syndicated stories cannot be redistributed and typically do not appear in the Factiva database. Thus, articles in my sample are those written by local staff reporters.

I collect all news articles discussing real estate markets between January 2000 and December 2013 from the major newspaper publications of 34 cities, including the 20 cities followed by the Case-Shiller home price indices. Most cities have one major newspaper that dominates the news market, with the exception of Boston, Detroit, and Los Angeles, which have two. Newspaper sources for each city are listed in the Appendix. I extract the full text of each article, and record each individual word with its corresponding date, word position, and originating newspaper. My final dataset assembles 37,537 newspaper articles, consisting of over 29 million words.

I capture media tone through a textual analysis of the content across newspaper articles. Textual analysis is an increasingly popular methodology used to quantify the tone of financial documents (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)). I apply the most standard methodology employed by this literature, which uses a dictionary-based method to quantify the raw frequency of positive and negative words in a text. To do so, these papers typically identify words as positive or negative based on an external word list. External word lists such as those from the *Harvard IV-4 Psychological Dictionary* are preferred because they are predetermined and less vulnerable to subjectivity from the author. Recent studies have argued, however, that these general tonal lists can at times contain irrelevant words and lead to noisy measures (Tetlock, Saar-Tsechansky and Macskassy (2008)). For example, Engelberg (2008) points out that *Harvard IV-4* positive list contains word such as *company* or *shares* can unintentionally capture other effects in finance applications. Loughran and Mcdonald (2011) show 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 employ a predetermined list from the Harvard IV-4 dictionary to reduce subjectivity, but choose one that is to relevant to how the media expresses positive or negative tone over housing. Shiller (2008) suggests that the media embellishes market activity by employing superlatives that emphasize increases and upward movements. For example, my sample includes articles with headlines such as "Home Sales Skyrocket!", "Home Prices Zoom Up", or "Housing is HOT, HOT, HOT!!" Thus to capture words like *skyrocket*, *zoom,* or *hot,* I use the Harvard IV-4 lists *Increase* and *Rise* as a base set.[1] I remove any words from the original list that would result in obvious misclassifications, and then expand the remaining words with corresponding superlatives. Table 2 provides sample positive words and their corresponding word counts. Following Loughran and Mcdonald (2011), I also expand the list with inflections and tenses that retain the original meaning of each word. Word 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. [2] I repeat the above process to create negative word lists using the converse Harvard IV-4 lists *Decrease* and *Fall.*

I calculate the overall tone of housing news in each city i and period t by:

$$S_{it} = \frac{\#pos - \#neg}{\#totalwords}_{it} \tag{1}$$

i.e., the number of positive minus negative words divided by the total number of words across all housing articles in each period.[3] 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

[1] These lists can be found at http://www.wjh.harvard.edu/~inquirer/Increas.html and http://www.wjh.harvard.edu/~inquirer/rise.html. My dictionary source for synonyms is *Rogets 21st Century Thesaurus, 3rd Edition* ((2012)).

[2] Full words lists are available upon request.

[3] This calculation essentially treats all articles in one period as one long document; an alternative method is to calculate the share of positive and negative words in each individual article and then average across articles. I try both methods and do not find a difference in values.

is preceded within five words by one of these negation terms.[4] The above calculation represents the most raw and baseline index of media tone. Loughran and Mcdonald (2011) propose an alternative "term-weighted" index that also adjusts for the commonality and frequency of a word across documents. I find that this and other alternative variations are highly correlated with the baseline version above, and subsequently report all results with the baseline version above. Details on alternate versions and their correlations with the baseline index are available in the Appendix.

Validating a measures of housing sentiment is by nature challenging when beliefs are unobservable and alternative proxies are otherwise rare. Existing surveys of home buyer confidence cannot validate each measure city by city, but can be used to compare overall trends on the national level. The University of Michigan/Reuters Survey of Consumers (SOC) surveys a nationally representative sample of 500 individuals each month on their attitudes toward business and buying conditions, including those of the housing market. Specifically, the SOC asks, "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 1 plots the percentage of respondents that answered "yes" against a national version of my housing expectations index. I calculate a national index of media tone using the same weights applied to the twenty cities in Case-Shiller Composite-20 home price index. My measure of housing media tone reveals a strikingly similar pattern to the SOC survey measure. Both measures increase rapidly from 2000 to 2003, peaking in early 2004. Both fall rapidly from 2004 to 2006, dropping well below original levels of confidence in 2000. Media tone appears to generally lag survey confidence on average by two to six months, consistent with theories that the media responds to reader preferences in the housing market. Both measures rebound early 2008, peaking and declining slightly again in late 2009. Both of the increases occur before the temporary rebound of the housing market in 2009, but fall slightly again afterwards. The correlation between the SOC survey and media measure of expectations is equal to approximately 0.84.

Case, Shiller and Thompson (2012) implement surveys that provide even more detailed perspectives on investor expectations. They directly ask respondents how much they expect their house price to grow over the next ten years. Answers reveal astonishingly high expectations; with respondents expecting prices to rise an average of 11 to 13 percent each year. Survey expenses

[4] 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.

limit coverage to four suburban areas and an annual snapshot, but nonetheless provide a valuable opportunity to validate my media index on a local level for at least four cities. Panel B of Figure 1 plots the Case-Shiller survey measures against my media expectations index for San Francisco, Los Angeles, Boston, and Milwaukee. As with national trends, my media index exhibits parallel patterns to survey measures on a city level. Both my media index and survey measures of expectations in San Francisco, Los Angeles, and Boston similarly peak in 2004, and hit a low point in 2008 before rising again. Measures for Milwaukee display more moderate patterns from 2003 to 2006, both in the survey and media index. Correlations between my media index of expectations and the Case-Shiller Survey for each city range from approximately 0.7 to 0.74.

3 House Prices and Media Sentiment

3.1 Descriptive Patterns

I obtain quarterly residential home prices across my sample of 34 cities from the Federal Housing Finance Agency (FHFA). Like the Case-Shiller home price index, the FHFA home price index estimates price changes for single-family homes with repeat sales to control for the changing quality of houses being sold through time. Their estimates are based on data on repeat mortgage transactions that have been purchased or securitized by Fannie Mae or Freddie Mac. The FHFA indices cover additional cities in my sample beyond the twenty major cities followed by Standard&Poor/Case-Shiller. Both indexes, however, are highly correlated (0.87). I use the Case-Shiller Composite-20 index as a measure of aggregate prices for the U.S..

Figure 2 plots media sentiment and house prices for the composite-20 indices and a sample of four cities. As seen in Figure 1, the composite-20 sentiment index mirrors the SOC survey expectations pattern, exhibiting a striking boom and bust pattern that peaks in early to mid-2004. Aggregate house prices follow a similar rise and fall pattern, but at a lag of nearly three years. This pattern is driven by cities such as Los Angeles and Las Vegas (pictured below), which experienced a similar leading boom-bust trend in media sentiment followed by house prices. Cities with minimal price appreciation such as Atlanta and Cleveland, however, appear to have more random patterns in media sentiment from 2000-2004. While these cities did not observe large run-ups in prices, they did experience large declines in prices that seem to be somewhat preceded by patterns of decline in media sentiment as well. As Ferreira and Gyourko (2012) also document, I observe a wide variation in house price changes across cities in my sample during this period. Likewise, I find significant variation in the timing and magnitude of appreciation of sentiment index across cities as well. The full sample of plots for all 34 cities and summary descriptives are available in the Appendix.

The lead time between media sentiment and prices seems relatively large, particularly in comparison to the stock market where sentiment predicts prices and their reversals over just several days (Tetlock (2007); Baker and Wurgler (2006)). Yet both the SOC and Case-Shiller surveys also find that expectations started declining in 2004, so the pattern in the media index appears to be consistent with survey expectations. Furthermore, the SOC survey measure of housing confidence leads my media sentiment measure by two to six months, suggesting expectations move even earlier. One explanation for the large lead is that the transaction process of buying a home has a much higher fixed cost than that of the stock market, and consequently takes considerably longer for home buyers to act in response to rising optimism. The home-buying process includes several steps, from searching for mortgage lenders, qualifying for a loan, to initiating the search for a home. The actual search process for a home itself can also take several months from first search to an accepted offer. Depending on details of home inspection and financing, from first offer to final closing on a home can often take at least one month. The shortest average reported mortgage loan closing period reported by Ellie Mae, a technology company that tracks mortgage applications, is 37 days.[5] Finally, house price indices are based on publicly recorded transactions registered with the county deeds records, which are not recorded until after a transaction has finally closed.

3.2 Predicting Prices with Media Sentiment

The leading patterns across cities suggest media housing sentiment may have a predictive relationship with house prices. I explore the effect of my media measure of expectations on house prices with the following linear framework :

$$\triangle p_{it+1} = \alpha + \lambda L^k \triangle p_{it} + \beta L^k \Delta s_{it} + \gamma \triangle x_{it} + \epsilon_{it+1} \tag{2}$$

[5] See: http://www.elliemae.com/

where i denotes each city and t indicates a quarterly period. A lowercase letter represents a log operator such that ($p_t = logP_t$) and $\triangle$ denotes the first difference such that $\Delta p_{t+1} = logP_{t+1} - logP_t$. I control for past price changes in all specifications, as denoted by $\triangle p_{it}$. Predictability in house price changes has been well documented across a number of studies. Among the most well-known, Case and Shiller (1989) find significant positive serial correlation and predictability with past price growth in four markets. FHFA house price changes across my sample of 34 cities exhibit positive serial correlation with an average AR(1) coefficient equal to approximately 0.42.

Vector $\triangle x_t$ includes a set of economic variables such as rents, population, income, employment, and interest rates that have been shown to predict residential house price growth over time. The "fundamental value" of an asset typically refers to its present discounted value of future cash flow, which models of housing assume housing pays dividends in the form of rental services. I obtain city-level measures of rents from REIS.com, which provides average asking rents on rental units with common characteristics with single family homes. A number of housing studies also highlight the importance of labor market variables on housing demand (Roback (1982); Rosen (1979); Nakajima (2011); Mankiw and Weil (1989)). I attain local employment and levels and unemployment rates by city from the Bureau of Labor Statistics (BLS). The Rosen/Roback model of spatial equilibrium also suggests income as an important demand shifter. I include city measures of personal income per capita and population from the Bureau of Economic Analysis (BEA). Studies argue 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 the national 30-year conventional mortgage rate relevant to most home buyers, but also compute real interest rates following Himmelberg, Mayer and Sinai (2005) by subtracting the Livingston Survey 10-year expected inflation rate from the 10-year Treasury bond rate for robustness.

This set of economic "fundamentals" does an exceptional job explaining the changes in house prices prior to 2000, but has difficulty explaining the variation in the most recent cycle. The vector $\triangle x_t$ predicts nearly 70 percent of variation in composite house prices prior to 2000. Using the same set of economic variables to explain house price growth after 2000, however, explains less than 10 percent of the variation. Local economic variables explain more of the variation in prices city to city post-2000 than in the composite, but are able to explain 1.55 times more variation prior to 2000. These traditional housing factors are able to explain more variation in prices in cities that had stable house price appreciation, but account for minimal movement in cities with large swings in prices post-2000.

Thus, Equation 2 considers whether a media proxy for expectations serves as a significant predictor for house prices during this period. I first normalize my index to be positive with the same adjustment as the SOC survey measure (i.e. $pos - neg + 100$)). I then orthogonalize my measure of expectations from the observed vector of fundamentals, $\triangle x_t$. Thus, $\triangle s_t$ represents log differences in my measure of housing media sentiment, orthogonalized to observed fundamentals, and L^k is a lag operator that indicates k number of lags such that lags $L^k \Delta s_n t = lnS_{n,t-k} - lnS_{n,t-k-1}$. I impose a finite-distributed lag structure with three years of quarter periods such that $k = 12$. The parameter β then captures the the total accumulated predictive effect of expectations on house prices for each individual lag k of $L^k \triangle s_t$. Equation 2 tests the alternative hypothesis that $\beta \neq 0$ against the null that $\beta = 0$. If media sentiment simply reflects price movements or economic fundamentals already incorporated into prices, then β should not be significantly different than zero.

Table 3 presents the results from Equation 2. The first row reports the total accumulated effect of expectations, β, on the current t quarterly growth in prices. The subsequent rows breaks down the lagged effect of expectations by years. Estimates show that a one percent appreciation in accumulated lagged sentiment is positively associated with a future quarterly price appreciation of approximately 2.8 percentage points. This is significant in magnitude relative to the unconditional mean of quarterly house price appreciation of 1.97 percent. Note that these coefficient estimates represent the predicted effect of housing media sentiment above and beyond both historical housing economic variables and past prices changes. Even focusing on the effects of just one lagged year, housing sentiment and price growth nearly have a one-to-one relationship. To put magnitudes into further context, a one standard deviation increase in a one year accumulation of housing sentiment in Las Vegas predicts approximately 12 percent future quarterly price growth. The largest quarterly price growth during in my sample occurred in Las Vegas in Q2 of 2004, appreciating almost 12.5 percent.

Columns 1 through 3 of Table 3 compares t-statistics calculated based on 3 different standard error procedures. In Column 1, I assume the error term ϵ_{t+1} is heteroskedastic across time and serially correlated within city, and calculate panel Newey and West (1987) standard errors that are robust to heteroskedasticity and auto-correlation up to twelve lags. This column assumes errors are correlated within city since studies have documented little mobility in homeowners across states. Nonetheless, 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 increase my standard errors, suggesting some dependence exists across cities. Nonetheless, my estimates remain statistically significant. In column 3, I allow for further flexibility in the structure of this dependence and cluster my standard errors by time and city. Doing so increases my standard errors minimally, while estimates remain statistically significant.

3.3 Determinants of Media Sentiment

The results in Table 3 show that expectations, as proxied by media sentiment, are positively associated with future house prices. At the same time, expectations may also be influenced by past price changes. In their first survey of home buyer expectations, Case and Shiller (1988) concluded "people seem to form their expectations from past price movements rather than knowledge of fundamentals." In their updated surveys, Case, Shiller and Thompson (2012) find that home buyers' expectations appeared extremely high, projecting an appreciation of more than 10 percent per year for the next 10 years. While these expectations at first seem wildly unrealistic, Case, Shiller and Thompson (2012) note that the Case-Shiller composite-10 index had indeed appreciated nearly 11 percent per year over the last ten years from 1996 to 2006. Greenwood and Shleifer (2014) examine six different surveys of investor expectations on the stock market, and find that investor expectations do appear to extrapolate from past market returns. I explore the nature of my measure of housing expectations with an analogous linear specification in their study:

$$S_{it} = \alpha + \lambda R_{it-k} + \delta logP_{it} + \gamma logX_{it} + u_{it} \tag{3}$$

where S_{it} denotes the level of housing expectations in city i at quarter t, and R_{it-k} represents the cumulative return in city i's local housing market from period k to t. Variable p_{it} denotes the local log housing price level, and x_{it} is the same vector of economic variables from Equation 2. I report results based on standard errors clustered by time and city.

Table 4 reports the estimates from Equation 3 for four different time horizons. Columns

1-4 regress the housing media sentiment index on 6-month, 1-year, 2-year, and 3-year cumulative price growth respectively. The results show that past price appreciation predicts higher media sentiment, consistent with home buyer expectations in Case and Shiller (1988) and Case, Shiller and Thompson (2012) survey responses. Column 1 shows that the coefficient on a 1-year lagged cumulative return is equal to approximately 10.67. This implies that if returns to housing in the last 6 months increase by one standard deviation (approximately 5.4 percent), then the housing media sentiment index will increase by approximately 0.9 units. Given that the media sentiment index ranges from approximately 96 to 110, this translates to a 6 percent increase.

The most recent returns appear to have the greatest impact, though past returns have a positive and significant impact across all given columns. The magnitude of the estimated effects of lagged returns decline with further distance, where the coefficient on the one-year cumulative returns is equal to 7.3 down to 2.60 for 3 year returns. Coefficient estimates for the logged price level are not significant across all columns, indicating that the accumulation of all past returns (at least in the given sample) does not affect current housing sentiment. This result further emphasizes that the effect of lagged returns on media sentiment only holds for the most recent returns. Extending the window of returns to just 5 years ($k = 20$) in Equation 3, results in estimates that are no longer significant. These findings suggest that housing expectations, as proxied by the media, possess a backward-looking nature consistent with theories of extrapolative expectations in behavioral finance (Barberis, Shleifer and Vishny (1998); Campbell and Kyle (1993); Cutler, Poterba and Summers (1990); De Long et al. (1990a); Fuster, Laibson and Mendel (2010)). The results are also consistent with the structure of investor expectations Greenwood and Shleifer (2014) find in the stock market.

Equation 3 includes all the housing fundamentals from Equation 2, but only the housing mortgage rate has any effect of notable significance. Table 4 reports the coefficient estimates on the 30-year mortgage rate, and shows that the mortgage rate is negatively associated with media sentiment. As the mortgage rate declines, housing media sentiment increases. This is consistent with responses in the SOC that answered optimistic about housing because mortgage rates were low. As Piazzesi and Schneider (2009) point out, interest rates have historically been a major driver of housing perception and these estimates suggest that mortgage rates continued to have an impact on public perception during this sample period as well.

4 Interpretation

The above results show that housing sentiment, as captured by the media, is positively associated with future house price growth. This predictive effect is robust to known fundamental controls that have explained house prices well historically. The structure of the housing sentiment index appears to be extrapolative in nature, and peaks more than two and half years ahead of house prices on average. One interpretation of these results is that they capture the effects of sentiment, in which investor beliefs were over-optimistic and drove house prices away from fundamentals. The media sentiment index certainly does mirror patterns in the Case, Shiller and Thompson (2012) survey, where home buyer expectations look unjustifiably high and similarly peak in 2004. At the same time, however, these results could also be consistent with a story of unobserved fundamentals that were instead driving price growth. Housing markets are inherently local, and local media in particular could convey information on local fundamentals that are otherwise difficult to observe or collect. Another possibility then is that the media sentiment index represents a valuable source of unobserved information about the local housing market. Reviewing the patterns in Figure 2, particularly in markets with a defined boom and bust of prices, one could argue that the media proxy even looks like a perfect forecast, perhaps an indication expectations reflected justified local information.

The goal in this section is to provide a set of tests to explore these two interpretations empirically. The advantage of measuring housing sentiment with the news media is that we can exploit the richness of the data both across text and cross-sectional city indices. We can then use this analysis to test results that we expect to be consistent with an interpretation of sentiment versus a story of information on unobserved fundamentals.

4.1 Testing the News Content over Fundamentals

I first address the interpretation of the media sentiment index as a source of unmeasured fundamentals by examining the informational content of the articles firsthand. By keeping track of all of the text across articles, I can directly examine whether my media sentiment index is driven by articles discussing particular housing variables. Following an analogous strategy from Tetlock, Saar-Tsechansky and Macskassy (2008), I flag any news article that mentions stem words such as "rents," "population," or "taxes" that may indicate discussions of local housing market information. I then quantify the fraction of positive and negative words across these articles to create a set of "media fundamentals" analogous to my overall housing media sentiment index. For example in Table 5, "Media Rents" is an index of the positive and negative words across all local articles that discuss rents, "Media User Costs" refers to words across any articles that discuss factors enter into the user cost of housing such as property taxes and maintenance costs, and "Media Demographics" indicate any articles that discuss local population and income. Through this strategy, the news media index also presents a potential opportunity to quantify particular information in markets where fundamentals are difficult to observe.

I then additionally control for these media fundamentals in Equation 2 to see if any of these measures reduce my estimated effect of expectations. If the discussion of fundamentals from these articles drive the patterns in my main housing media sentiment index, then controlling for words in these articles should drive down the significance and magnitude of the results in Table 3. Note that I control for each for all $k = 12$ lags of each media fundamental as well as all observed controls from Equation 2. Table 5 shows that the estimated effects of media sentiment on price growth remain robust to controlling for news content over fundamentals. Columns 1 through 7 add each media control sequentially to test the stability of the coefficient estimate for the overall sentiment index, β. Coefficient estimates of β remain 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 estimates to the sequential addition of controls suggests bias from unobserved factors is less likely (Altonji, Elder and Taber (2005); Angrist and Krueger (1999)). The magnitude and precision of β actually increase after including the first media control for rents compared to the estimates in Table 3, and the magnitude continues to increase to 3.8 in column 6. This translates to a nearly four to one increase in price growth in response to accumulated housing sentiment. Column 7 shows that the effect of the overall media sentiment index slightly drops in magnitude to 3.4 after controlling for news articles discussing credit conditions, though still 0.6 percentage points higher than estimates in Table 3.

Credit Conditions. The estimated effect of housing media sentiment remains significant, positive, and large in magnitude after accounting for an extensive set of fundamental articles. The only media fundamental to reduce the magnitude somewhat in Table 5 is the fraction of positive and negative words across articles discussing credit conditions. While not a typical housing fundamental historically, an additional new factor largely debated during the housing crisis was the availability of easy credit. Mian and Sufi (2011) show that 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)).[6]

While the results in Table 5 suggest the predictive effects of sentiment are not driven by a news discussion of credit conditions, I collect additional credit variables to further control for subprime lending and easy credit during this period. I first include the 6-month London Interbank Offered Rate (LIBOR). Estimations in Table 3 already include changes in the 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)). I also collect loan-to-value ratios and subprime lending variables from a proprietary dataset provided by DataQuick, an industry data provider (Ferreira, Gyourko and Tracy (2010)). DataQuick collects detailed transaction level data on over 23 million housing transactions from 1993 to 2012, and provides full coverage for 23 cities in my sample. Thus, the sample size is smaller, but still provides a substantial number of observations for estimation ($N = 762$). 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. I calculate the measure of percent of subprime mortgages following 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*.

I re-estimate Equation 2 for the sample with additional controls for credit lending. Since the sample size is smaller than my original estimations, I present the results separately in Table 6.

[6] 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).

Column 1 also first provides a baseline estimate for the smaller sample of cities for comparison. As in Table 3, column 1 includes controls for past price changes and the vector of housing fundamentals X_t. Estimated effects are larger in magnitude for the smaller sample with a coefficient estimate for lagged media sentiment of 4.8. Column 2 additionally controls for the changes in the six-month LIBOR rate and shows estimates for β does not waver in significance or magnitude. Column 3 adds controls for lagged local loan-to-value ratios, and estimates remain similarly robust. As Glaeser, Gottlieb and Gyourko (2010) point out, it is difficult for loan-to-value (LTV) ratios to explain price changes as LTVs did not change much across time during this period. Column 4 shows that estimates remain significant after controlling for lagged subprime lending, but remains robust in magnitude at 3.4.

Consistent with Mian and Sufi (2011), I observe that the percent of subprime lending across cities rises and falls contemporaneously with house prices, similarly peaking in early to mid 2006.[7] Thus as with prices, the housing sentiment media index appears to peak before patterns of subprime lending as well. Ferreira and Gyourko (2011) find that the share of subprime lending actually appears to lag the boom in house prices, and note that these patterns may suggest that lending may have responded to rather than influenced price movements. Mian and Sufi (2009) do not find evidence that suggests high price expectations enabled lenders to expand credit to subprime borrowers, however, while Agarwal et al. (2012) finds evidence of positive sentiment effects on overall loan approvals.

The results in this section suggest that patterns in the housing media sentiment index are not driven by news content on fundamentals. While controls are subject to those fundamentals we opt to include, the informational content we can choose explored through the text is unlimited. Nonetheless, given the set I include in the above results, the predictive effects of the media sentiment index does not appear driven by mentions of any particular fundamental across articles. The only news articles that appear to have some effect on the magnitude of the media index coefficients are those that include discussions of credit, however including additional controls for subprime lending does not diminish the predictive effect of the media index on prices.

[7] Interestingly, I find that the fraction of subprime lending also booms and busts in cities such as Denver where prices remained relatively stable, peaking at more than 20 percent in early 2006.

4.2 Weekend and Narrative Sentiment Indices

The previous section examines the media index as a proxy for unobserved fundamentals by directly examining the informational content of the news articles. Conversely, the following section directly explores the sentiment interpretation of the media index by isolating news articles likely to only reflect sentiment and unlikely to contain any informational content. I extract two subsamples of news articles that fall into this category. The first set are those articles that are published on the weekend. The Readership Institute of Northwestern University conducted a 2000 survey of 37,000 newspaper readers and found that readership is highest on the weekend, driven by the greater proportion of "light" readers.[8] Light readers are defined as those who spend fewer than 16 minutes reading the newspaper a week. These readers also prefer content that is lighter in informational content and entertaining in style. Thus, weekend articles are those more likely subject to sentiment due to the profile of reader preferences it must cater to in order to maximize its readership.

Because my dataset tracks the date of each news article, I am able to identify the exact day of the week each article is published. Using this information, I define an alternative "weekend" sentiment index:

$$S_{it}^{Weekend} = (\frac{\#pos - \#neg}{\#totalwords}_{it})^{Weekend} \tag{4}$$

I employ the same identical methodology used to calculate my overall sentiment index, but restrict my analysis to the tone of articles that occur on a Saturday or Sunday. Not only are weekend news articles more likely to reflect sentiment, they are also less likely to reflect information on fundamentals. Market news tends to be reported during the business week, and official press releases on housing data only occur on a working weekday. Nevertheless while news stories on the weekend are less likely to be contaminated by news on fundamentals, weekend editions could still contain articles that summarize information that is reported during the working week. To address this issue, I calculate an analogous index of the share of positive and negative words that occur on a weekday, and control for these patterns directly in my estimations that I discuss further below.

The second set of articles I extract are those I define as "narrative" news articles in my sample. A narrative article refers to one that narrates a story or account of events around particular individuals. The 2000 Readership Institute Survey reports that readers have high preference for

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

"people-centered news." These articles contain anecdotal stories, but tend to offer no actual news or information on the market. Narrative articles can even provide misinformation, by supporting a particular positive or negative sentiment with one individual story. The study finds that the "light" readers prefer articles that are easier to read, and newspapers that are successful at satisfying these readers are those that publish more "feature-style" stories. Thus similar to the weekend, narrative articles are also those more likely driven by sentiment and entertainment in order to capture reader attention.

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 housing experts. I then analyze the tone across only the identified narrative articles in my sample, and calculate an analogous narrative sentiment index, $S_{it}^{Narrative}$, to my weekend-only index above.

Using these alternative sentiment measures, I again estimate equation 5 with the weekend and narrative indices respectively. The framework and all fundamental controls remain the same, with just the overall sentiment index replaced. The tests in the previous section included the set of media fundamentals to see whether informational articles in fact drove the variation behind the estimated effect of sentiment on prices. Conversely, the estimations in this section instead tests whether narrowing the identifying variation to a smaller subsample of sentiment-driven articles still maintain the predictive effect of sentiment from baseline results. Estimations using the weekend sentiment index also additionally include controls for the tone across weekday news. If weekend articles simply summarize information discussed during the week, then we should not expect to find a weekend sentiment effect distinct from weekday tone. In these estimations, the identifying variation is narrowed even further to just the changes in the tone across weekend articles above and beyond the patterns already observed throughout the business week.

Table 7 presents the results of estimating an effect of sentiment with just the weekend and narrative indices. Column 1 first presents the baseline results from Table 3 using the overall sentiment index for comparison. As in the Table 3, the first row reports the total effect of lagged media sentiment, β, on future quarterly appreciation in house prices. Column 2 reports the results from estimating equation 5 using the weekend sentiment index. Because estimations in column 2 also include controls for weekday tone, the estimated coefficient on the weekend sentiment index, β^{Wknd}, represents the differential effect of sentiment of weekend over weekday news. Results show that the estimated effect of sentiment remains significant and large in magnitude. The magnitude of the estimated coefficient declines slightly from 2.8 percentage points to 2.6, but is not statistically different from baseline estimations. Column 3 reports the estimates using just the narrative sentiment index. Estimates for narrative sentiment, $\beta^{Narrative}$, again remain robust in both statistical significance and magnitude. Coefficients in column 3 are even slightly larger in magnitude than the weekend estimation results, but again are not statistically different than the baseline results. Thus as previous results show that media reports on fundamentals do not drive out the effect of sentiment, estimates restricted to to narrowed sentiment measures continue to show a significant and large effect of sentiment on house prices.

4.3 Cross-Sectional Effects of Housing Media Sentiment

The preceding sections take advantage of the textual nature of the data, which allows for a direct analysis of the content across articles. Measuring sentiment with news media further provides an avenue to create indices across several markets in real time, creating an opportunity to explore sentiment effects cross-sectionally. Given the city-wide variation in house price changes during this period, this is particularly useful when exploring the relationship between sentiment and prices in the housing market. With access to city-specific indices, we can explore whether prices and sentiment vary systematically according to any city-level traits. The cross-sectional nature of the data thus allows us to test whether cross-sectional differences exist based on characteristics that we would expect to be consistent with a theory of sentiment.

Minority Home Buyers. Baker and Wurgler (2006) highlight two channels through which theory predicts sentiment has cross-sectional effects on prices: 1) where demand is less informed and 2) where arbitrage constraints are particularly binding. Since arbitrage constraints are completely binding in the housing market, this suggests potential cross-sectional effects lie in differences in informed demand for housing among buyers. For example, Shiller (2008) raises concerns to the disproportionate lack of access to adequate financial advice available to minority groups that may lead to investment decisions based on minimal or biased information. Indeed in a comprehensive survey of financial literacy, Lusardi and Mitchell (2007) find that financial knowledge and planning are at lowest levels among Hispanic and Black respondents. The Home Mortgage Disclosure Act (HMDA) requires lending institutions to file reports on all mortgage applications, and thus provides an opportunity to test cross-sectional effects based on the demographic profile of the pool of potential home buyers.

Following Ferreira and Gyourko (2012), I define a "% minority" variable based on the fraction of African-Americans and Hispanics loan applicants coded by the HMDA dataset. I calculate the average fraction of minority loan applicants across the sample period for each city, and then divide the 34 cities in my sample into two equal groups based on whether the city contains a low or high fraction of potential minority homebuyers. I then estimate the following equation to test whether sentiment effects differ across group:

$$\triangle p_{it+1} = \alpha + \lambda L^k \triangle p_{it} + \beta L^k \Delta s_{it} + High + \beta_{High} L^k \Delta s_{it} \times High + \gamma \triangle x_{it} + \epsilon_{it+1} \qquad (5)$$

Equation 5 examines the same predictive relationship between prices and media sentiment as in Equation 2, but now explores additional interactions between the fraction of potential minority homebuyers and lagged sentiment. $High$ is an indicator variable for a city with a high fraction of minority buyers, and is interacted with all included lags of media sentiment. Since the *Low* group is omitted from the regression, β now represents the baseline effect of sentiment for cities with the lowest fraction of minority homebuyers. The parameter β_{High} then captures the additional sentiment effect of being in the high group. If there are no cross-sectional differences across cities, the coefficient β_{High} should not be significantly different than zero. If we presume the pool of buyers with low access to financial advice to be more likely subject to sentiment, then we would expect sentiment effects to be larger in cities with a larger fraction of minority buyers. In other words, a significant coefficient $\beta_{High} > 0$ would indicate a cross-sectional effect consistent with a theory of sentiment.

Column 1 of Table 7 reports the coefficient estimates β in the first row and β_{High} in the row directly below. The results reveal that sentiment effects do appear to be greater in cities with a higher fraction of minority loan applicants. Baseline estimates for the low group indicate a one percent appreciation in accumulated lagged sentiment predicts a future quarterly price appreciation of approximately 2.1 percentage points. This is still significant in magnitude relative to the average quarterly house price appreciation. Estimates for β_{High}, however, suggest that a the same increase in sentiment would lead to a 1.5 percentage point additional increase in future price growth. Thus, this means that sentiment in cities with a larger fraction of minority mortgage applicants has an impact nearly 2x greater than in those with fewer potential minority home buyers. In their analysis, Ferreira and Gyourko (2012) find that the share of minority home buyers are not able to explain the timing or magnitude of the house price booms across cities. These results suggest that while the fraction of minority purchasers may not be able to account for the changes in house prices alone, it may still have served as a factor in combination with changing sentiment during this period.

Speculators. Piazzesi and Schneider (2009) analyze housing answers from the Michigan Survey of Consumers and identify a set of homebuyers as "momentum" traders, or those that invested in housing because of an increased optimistic belief in rising house prices. Indeed, De Long et al. (1990*b*) show that traders may rationally respond to growing sentiment in the market and buy in anticipation of increasing prices. Chinco and Mayer (2014) document the presence of 2nd home buyers across local housing markets from 2000 to 2007, and find that non-local investors not only behave as speculators but also do not appear particularly well-informed. Bayer, Geissler and Roberts (2011) find that speculative activity, particularly among naive investors, increased during the housing boom. Regardless of whether these investors were rationally chasing or subject to sentiment themselves, these studies present an additional opportunity to test cross-sectional effects of sentiment across markets. Given a measure of speculators, evidence suggests that we should observe a larger impact of sentiment in markets where there is greater speculation.

Utilizing the transaction-level records in the DataQuick dataset, I define the number of speculators in a city by those who purchased one or more homes that are not owner-occupied.[9] I again divide the cities my sample into two equal groups, but now based on the presence of speculators across markets. *Low* thus refers to markets with a low number of speculators and *High* is an

[9]Similar strategies are used in to measure speculators in Ferreira and Gyourko (2012), **?**, and Bayer, Geissler and Roberts (2011).

indicator variable for cities with an above median number of speculators. I then run the same regression in Equation 5, but to test for cross-sectional differences instead based on speculation across markets. Thus, β now represents the baseline effect of sentiment for cities with low or no speculators, and β_{High} now represents the additional sentiment effect of having a high presence of speculators. As with the test across minority homebuyers, if there exist no cross-sectional differences across cities, β_{High} should not be significantly different than zero. If β_{High} is significantly greater than zero, then this would indicate that the impact of sentiment is greater in cities with greater speculation, consistent with a theory of sentiment.

Column 2 of Table 7 reports the results of estimating Equation 5 based on the cross-sectional differences in speculators. The results show that sentiment still has a positive and significant predictive effect on prices in cities with fewer speculators. The magnitude of β is slightly smaller than the original overall estimates of β from Equation 2, but is still significant in economic magnitude. Coefficient estimates for the *Low* group show that a one percent appreciation in accumulated lagged sentiment predicts a future quarterly price appreciation of approximately 2.1 percentage points, compared to a magnitude of 2.8 percentage points in Table 3. Coefficient estimates for β_{High}, however, suggest that a one percent increase in sentiment would lead to an additional 3.2 percentage point increase in future price growth. The magnitude of this estimate reveals a stark difference between the low and high groups, nearly twice the size of the estimates for the high minority group reported in column 1. Cities with a greater number of speculators observe a sentiment impact more than double than in those with fewer speculative investors. Thus as with the share of minorities, estimates reveal results consistent with a theory of sentiment.

Subprime Loans. Estimates in section 4.1 show that while the trend in subprime lending does rise and fall with prices, its pattern does not drive out the impact of sentiment on house prices. This result is consistent with studies such as Glaeser, Gottlieb and Gyourko (2010) that find while leverage expanded significantly, credit variables are unable to explain the substantial rise in house prices. Ferreira and Gyourko (2012) suggest that while their results also cannot explain the timing of the housing boom with patterns of subprime lending, it may still have played an important role in the later perpetuation of the housing boom. One possibility is that subprime lending served as a factor in combination with changing sentiment during this period. Mian and Sufi (2009), for example, find evidence that the increasing supply of credit from lenders led to the expansion of leverage taken out by subprime borrowers. If subprime lenders did indeed target the profile of certain borrowers, they likely targeted those buyers that were most subject to sentiment. We would simultaneously expect those most uninformed and vulnerable to sentiment to be more willing to take out risky, subprime loans. If this is the case, then this suggests an additional cross-sectional test: we should observe a larger impact of sentiment on prices in markets where more subprime loans were taken out.

Using the same measure of subprime loans defined in section 4.1, I divide markets into equal *Low* and *High* groups now based on the average number of subprime loans made each quarter. I again estimate the same Equation 5, but now to test for cross-sectional differences instead based on subprime lending across markets. As noted in the previous section, my sample of cities with full subprime lending coverage contains 23 cities, and the number of observations in this estimation should be smaller than previously. Similar to the two previous tests, this estimation tests the alternative hypothesis $\beta_{High} > 0$, indicating that sentiment effects on prices are greater in cities with greater subprime lending. Results in support of the null hypothesis, $\beta_{High} = 0$, would instead indicates that there is no discernible cross-sectional difference across markets based on the frequency of subprime lending.

Column 3 of Table 7 reports the cross-sectional results based on the differences across markets in subprime loans. Baseline estimates for sentiment in the *Low* group are still positive and significant in magnitude, but much smaller in comparison to overall estimates for the same sample in Table 6. A one percent appreciation in lagged sentiment predicts a future quarterly price appreciation of approximately 2.7 percentage points, compared to a magnitude of 4.8 percentage points in Table 6. Estimates for β_{High} in column 3, however, indicate that effect of sentiment is even higher in magnitude across markets with greater subprime lending. A one percent increase in sentiment predicts an additional 2.6 percentage points, or a total impact of 5.3 percentage point increase on future price growth. As in column 2, these estimates show a distinct difference between the *Low* and *High* groups based on subprime loans. Consistent with a sentiment hypothesis, cities with greater subprime lending experience a greater effect of sentiment on house prices.

Price of Homes. The tests in this section thus far have taken advantage of the cross-sectional differences in housing demand based across cities to tests predicted differences in sentiment. Nonetheless, since there exists different types of buyers within markets as well, we can also exploit cross-sectional differences in informed demand within a city. If the above results are evidence of an effect of sentiment, then varying levels of informed agents *within* local markets should imply varying impacts of sentiment as well. One way to identify different type of agents within a local city is through the separate markets they participate in. The most obvious and straightforward way to distinguish these markets is through price. Those with lower-income would only qualify for lower-priced homes, while higher-income buyers would be more active in higher-priced homes. Since minority households are often in the lower quartiles of the income distribution, Lusardi and Mitchell (2007) find that lower income profiles similarly report lower financial literacy levels. Thus as in the other tests in this section, a theory of sentiment would suggest that prices in markets dominated by buyers with less financial education should be more greatly affected by sentiment than those among a higher income profile.

FHFA house prices used in the estimations thus far only provide an index of the on average change in all house price levels for a metropolitan area. The Case-Shiller home price indices, however, also divide their indices into three price tiers: low, medium, and high. The tiers are calculated to be comparable across metro areas, so that a low-tier reflects the bottom third of sale prices while the high tier indicates sales in the top third of home prices. Thus in columns 4 and 5 of Table 7, I again estimate 5 but replace overall prices with first low-tier and then high-tier prices across metropolitan areas respectively. Note that since the Case-Shiller home price index only tracks house prices for 20 metropolitan areas, estimations in columns 4 and 5 are limited to this sample.

Column 4 first reports the total estimated effect of sentiment, β^{Low}, on the low-tier priced homes across cities. The results show that estimates are not only statistically significant but very large in magnitude, reporting a nearly 6 to 1 response in quarterly price growth to past changes in sentiment. Column 5 shows that estimates, β^{High}, for high priced homes are much smaller, reporting results of nearly half the magnitude of those for lower-priced homes. Baseline estimates based on the average price index for the sample of 20 Case-Shiller cities is approximately equal to 3.9, which is higher than estimates for the full 34 city sample. Nonetheless, column 4 shows that the total estimated effect for low-tier cities is still much higher with an estimated coefficient of 5.9 percentage points. Estimates for high-priced homes are equal to 3.3 percentage points, slightly smaller in magnitude than the baseline results though not necessarily statistically different.

Therefore, these estimates confirm that results are not only consistent with predictive effects of sentiment cross-sectionally across cities, but across market segments within cities as well.

5 Conclusion

While there has been much discussion and interest in the role of "animal spirits" in the most recent housing crisis, empirical tests of this hypothesis have been limited due to the lack of sentiment measures for the housing market. Any measure of expectations is naturally difficult to construct, and survey measures are expensive to implement and therefore limited in geographic scope and/or frequency. Housing markets are driven by local factors, however, and understanding why some markets experienced big price movements and others did not in the last housing cycle subsequently requires variables with cross-sectional variation. This paper contributes the first real-time measures of local housing sentiment across 34 major metropolitan markets by quantifying the tone of local housing news. Specifically, I capture the share of positive minus negative words across local housing news articles.

I find that patterns in my measure of media housing sentiment appear to lead movements in house prices by 2 years on average. In cities with big boom and busts of house prices, I find that media sentiment peaks in approximately 2004 while house prices peak in 2006. This leading pattern is also reflected in the Michigan Survey of Consumers (national) measure of housing confidence, which peaks slightly ahead of my composite-20 media housing sentiment index in 2004. I am also able to validate four of my local city sentiment measures against surveys of housing expectations from Case, Shiller and Thompson (2012). Though only available annually, Case, Shiller and Thompson (2012) measures exhibit similar leading patterns across cities and correlate highly with my media sentiment indices. In some ways, this pattern seems to contradict the perception that buyers were positive up through 2006 before prices began to fall. Articles in my media sentiment index are still positive through 2006, however, but reach its positive peak in 2004. Similarly in both the Michigan SOC and Case, Shiller and Thompson (2012), respondents are still very positive up to 2006, but their expectations are at their highest in 2004.

I find that changes in my measure of media sentiment have significant predictive power for future house price growth. The media housing sentiment index explains a significant amount of variation in house price changes above and beyond a set of observed economic factors that have been shown to predict house prices historically. These traditional factors appear to explain more variation in cities with more stable house price appreciation, but where media sentiment accounts for more of the variation in cities with large swings in house prices. This effect remains robust to the inclusion of additional controls for subprime lending and the availability of easy credit. The structure of the media sentiment index itself reflects a backward looking structure consistent with extrapolative expectations proposed in behavioral finance theories.

These results are consistent with two interpretations of the housing media sentiment index. The housing media index could proxy for investor sentiment in the housing market, or instead proxy for hard-to-measure fundamentals that are instead driving housing prices. Note that regardless of interpretation, the housing media sentiment index provides an useful methodology to measure unobservable factors in the housing market. Nonetheless, the effect of the media sentiment index on house prices does not appear to be driven by news stories that discuss housing fundamentals. The predictive power also remains robust when the identifying variation is narrowed to a smaller subset of sentiment-only news articles relative to the pattern of words across information-driven articles. Media sentiment also has a greater effect on house prices among those markets with a larger minority homebuyer presence, more speculative investors, and across lower priced homes. The predictive effect is also amplified in markets where more subprime loans were approved and taken out.

I interpret these results as supportive of a sentiment interpretation of the housing market, and less consistent with an informational interpretation of the media index. A story of information would instead have to account for why certain unobserved fundamentals would have greater effects in markets with more homebuyers vulnerable to sentiment than others. The amplified effects of sentiment and subprime lending provide a potential explanation for why prior studies have been unable to account for the magnitude of house price changes with the share of minority buyers or expansion of easy credit alone. Without a strictly exogenous instrument for sentiment, this paper makes a careful effort to avoid making any conclusions about causality. Causes of the most recent housing cycle cannot likely be summarized by one single factor, however, and the cross-sectional analysis of this paper suggests that the driving factors behind the last boom are more complicated. Expectations and fundamentals likely have a more complex relationship; for example, where a subset of homebuyers may systematically overreact to a positive shock from lower interest rates or increases in credit supply. These results strongly suggest that sentiment should be taken seriously as a potential determinant of house prices and deserves greater attention in future research and policy concerns. In particular, the results of this paper suggest future work might address a greater understanding of what specific factors drive sentiment, whether the media plays a role in perpetuating financial mistakes, and if these factors can be improved with current financial education and literacy policies.

Appendix

A.1 Sentiment Index Robustness and Alternate Versions

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.

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, but find that term-weighted share of positive words has an almost identical impact on house price growth as the non-weighted baseline index.

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.

Agarwal, Sumit, Ran Duchin, Doug Evanoff, and Denis Sosyura. 2012. "In the Mood for a Loan: The Causal Effect of Sentiment on Credit Origination." *Working Paper.*

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.

Barber, Brad M., and Douglas Loeffler. 1993. "The "Dartboard" Column: Second-Hand Information and Price Pressure." *The Journal of Financial and Quantitative Analysis*, 28(2): pp. 273–284.

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

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics*, 49(3): 307 – 343.

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.

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., and Albert S. Kyle. 1993. "Smart Money, Noise Trading and Stock Price Behaviour." *The Review of Economic Studies*, 60(1): pp. 1–34.

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. 1988. "The Behavior of Home Buyers in Boom and Post-Boom Markets." *New England Economic Review*, 29–46.

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*.

Chinco, Alex, and Christopher Mayer. 2014. "Misinformed speculators and mispricing in the housing market." National Bureau of Economic Research.

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.

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.

Dougal, Casey, Joseph Engelberg, Diego Garcia, and Christopher A. Parsons. 2012. "Journalists and the Stock Market." *Review of Financial Studies*, 25(3): 639–679.

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.

Dyck, Alexander, and Luigi Zingales. 2003. "The Bubble and the Media." In *Corporate Governance and Capital Flows in a Global Economy.* , ed. P. K. Cornelius and B. Kogut, 83–102. New York, NY:Oxford University Press.

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. 2011. "Anatomy of the Beginning of the Housing Boom: U.S. Neighborhoods and Metropolitan Areas, 1993-2009." *NBER Working Paper 17374.*

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.

Fuster, Andreas, David Laibson, and Brock Mendel. 2010. "Natural Expectations and Macroeconomic Fluctuations." *The Journal of Economic Perspectives*, 24(4): pp. 67–84.

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

Garcia, D. 2012. "Sentiment during recessions." *Journal of Finance, Forthcoming.*

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, 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.

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.

Greenwood, Robin, and Andrei Shleifer. 2014. "Expectations of returns and expected returns." *Review of Financial Studies*, hht082.

Gurun, Umit G., and Alexander W. Butler. 2012. "Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value." *The Journal of Finance*, 67(2): 561–598.

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

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.

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

Keynes, J.M. 1936. *The General Theory of Employment, Interest and Money.* London:Macmillan.

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.

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.

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

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

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, A., and A. Sufi. 2011. "House Prices, Home Equity-based Borrowing, and the US Household Leverage Crisis." *The American Economic Review*, 101: 2132–2156.

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. 2005. "Rising Earnings Instability, Portfolio Choice, and Housing Prices."

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.

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.

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.

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.

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

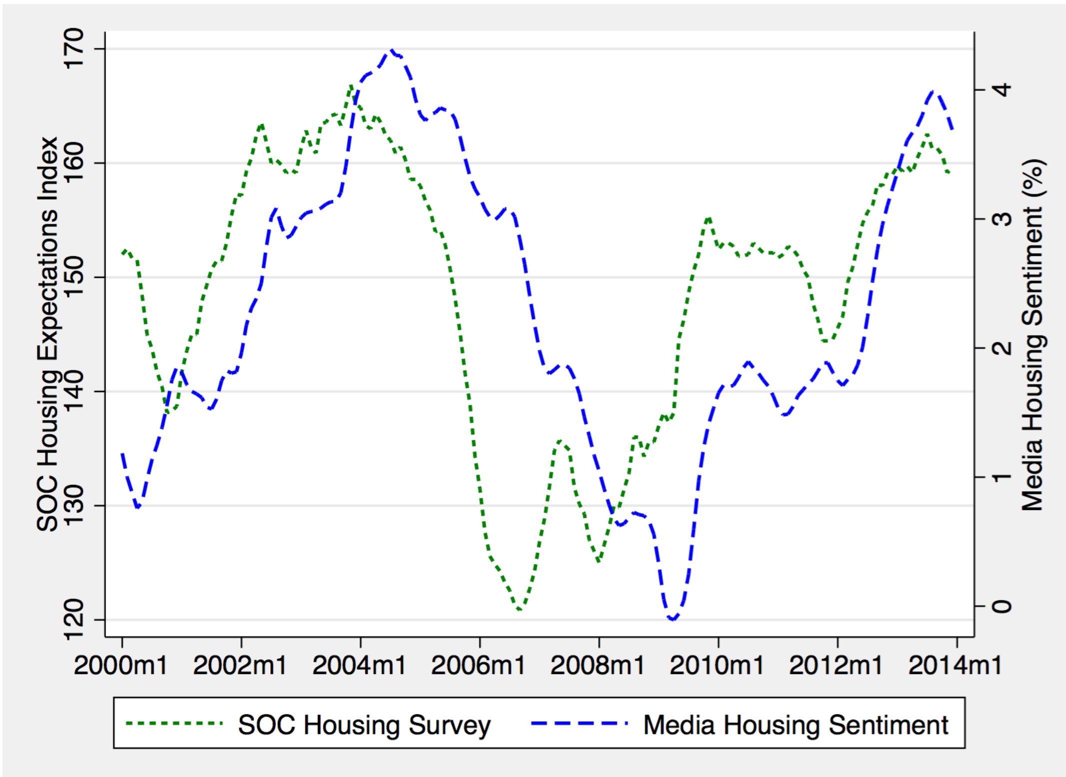
Tetlock, Paul C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *The Journal of Finance*, 62(3): pp. 1139–1168.

Tetlock, Paul C. 2011. "All the News That's Fit to Reprint: Do Investors React to Stale Information?" *The Review of Financial Studies*, 24(5): pp. 1481–1512.

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.

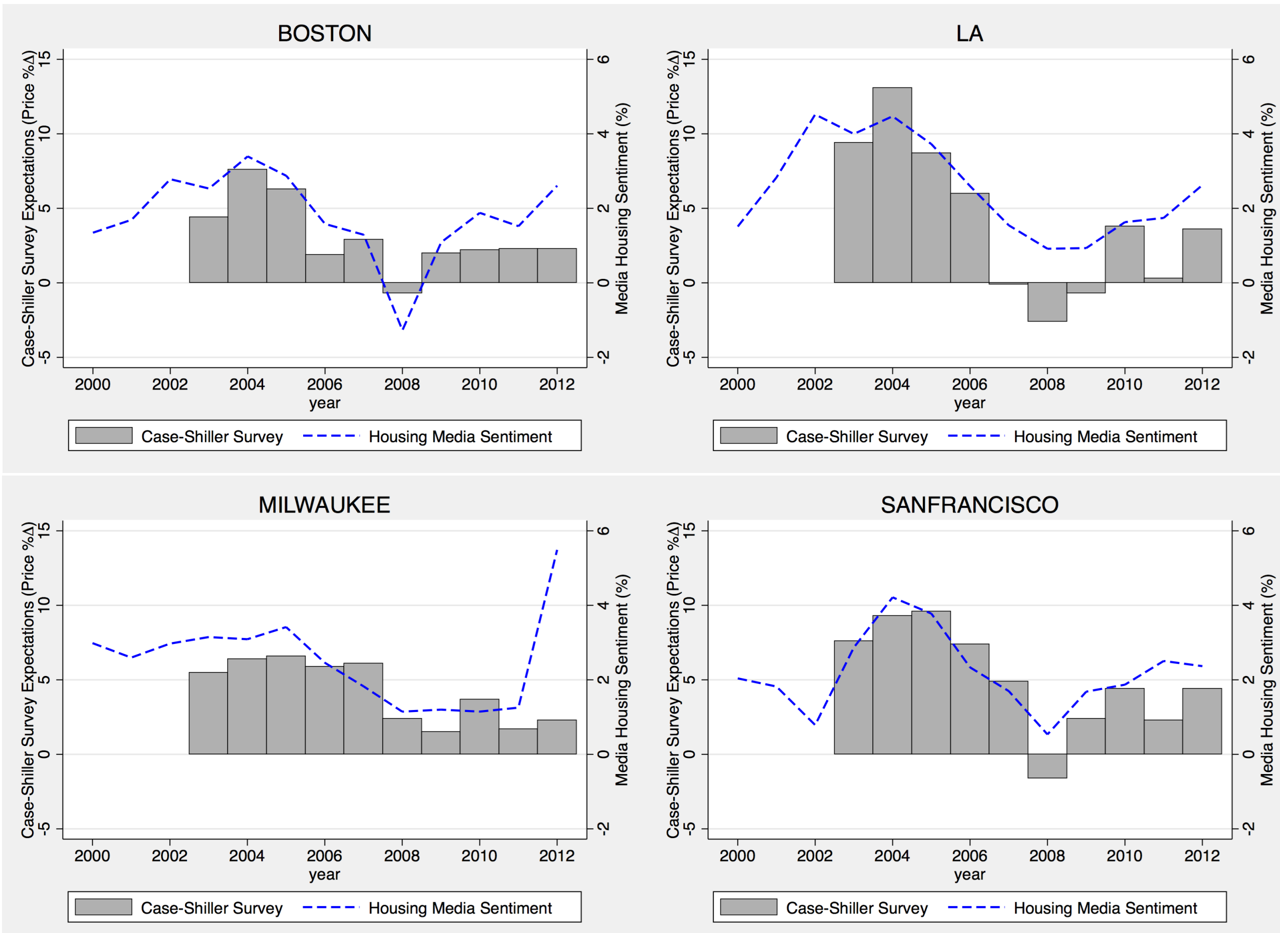
Tracy, Joseph, Henry Schneider, and Sewin Chan. 1999. "Are Stocks Overtaking Real Estate in Household Portfolios?" *Current Issues in Economics and Finance.*

Figure 1: Validating Media Housing Sentiment with Survey of Consumers



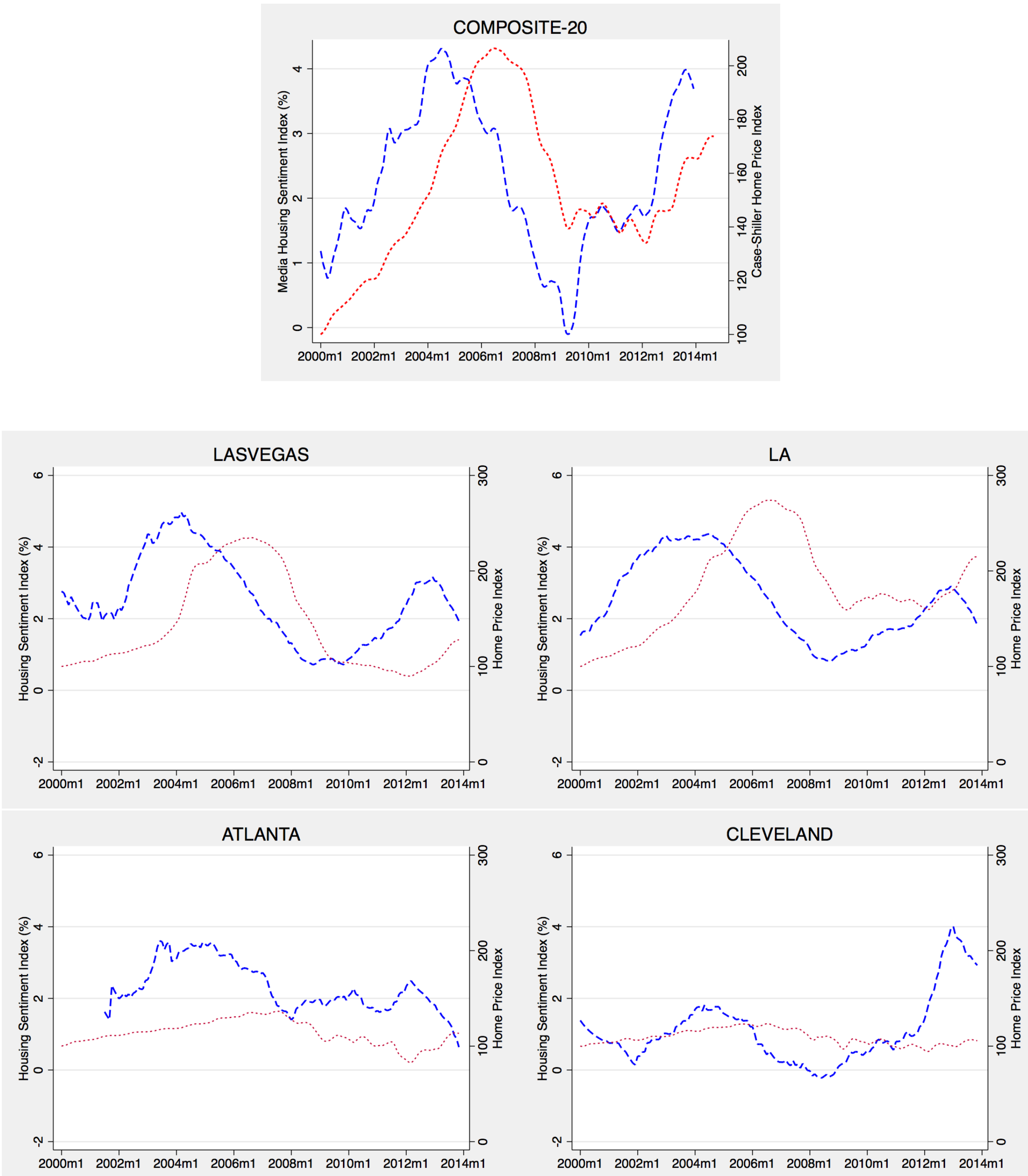
Note: Figure 1 compares the patterns of the composite-20 housing media sentiment index with the Michigan/Reuters Survey of Consumers (SOC) survey of home buyers. The SOC surveys a nationally representative sample of 500 consumers and asks whether they think it is a good or bad time to buy a home. The SOC cannot be broken down by city, but provides a validation of the media sentiment index on a national level. The green dashed line plots the SOC relative index, which equals the percentage who answered "Good" - "Bad" + 100. The blue dashed line plots the housing media sentiment index, which equals the fraction of positive - negative words across all housing news articles per month. The media sentiment index lags the SOC survey index by several months but mirrors a very similar pattern. The correlation between the SOC survey and (lagged) media sentiment is approximately 0.84.

Figure 2: Validating Media Housing Sentiment with Case-Shiller Surveys



Note: Figure 2 compares the patterns of the housing media sentiment index with surveys by Case, Shiller and Thompson (2012). Case, Shiller and Thompson (2012) survey home buyer expectations in four cities annually from 2003 to 2012. The grey bars plot the percentage respondents think home prices will increase or decrease over the next year. The Case-Shiller survey is limited to an annual frequency, but provides a validation of the media sentiment index at the city level. The blue dashed line plots the housing media sentiment index, which equals the fraction of positive - negative words across all housing news articles per month. The media sentiment generally follows a very similar trending pattern city by city. The correlation between the Case-Shiller Survey and media index is equal to approximately 0.74.

Figure 3: Media Housing Sentiment and Case-Shiller Home Price Indexes



Note: Figure 2 plots the housing sentiment index and housing price indexes for the composite-20 index and a sample of four individual cities. Lines are smoothed for seasonal variation and noise with a one-year backward and forward moving average. The full set of city plots and summary statistics are located in the Appendix.

Table 1: Major Newspaper Publications by City

City	Newspaper Publication	City	Newspaper Publication
Atlanta	The Atlanta Journal-Constitution	Milwaukee	The Milwaukee Journal Sentinel
Austin	Austin American-Statesman	Minneapolis	Star Tribune
Baltimore	The Baltimore Sun	NYC	New York Times
Boston	Boston Herald/Boston Globe	Orlando	Orlando Sentinel
Charlotte	The Observer	Philadelphia	The Philadelphia Inquirer
Chicago	Chicago Tribune	Phoenix	The Arizona Republic
Cincinnati	The Cincinnati Enquirer	Pittsburgh	Pittsburgh Post-Gazette
Cleveland	The Plain Dealer	Portland	The Oregonian
Columbus (OH)	The Columbus Dispatch	Sacramento	Sacramento Bee
Dallas	The Dallas Morning News	San Antonio	San Antonio Express-News
Denver	The Denver Post	San Diego	The San Diego Union-Tribune
Detroit	Detroit News/Detroit Free Press	San Francisco	The San Francisco Chronicle
Indianapolis	The Indianapolis Star	San Jose	San Jose Mercury News
Kansas City	The Kansas City Star	Seattle	The Seattle Times
LA	LA Times/LA Daily News	St. Louis	St. Louis Post-Dispatch
Las Vegas	Las Vegas Review-Journal	Tampa	Tampa Tribune
Miami	The Miami Herald	Washington, D.C.	The Washington Post

Note: Table 1 lists each city and its corresponding newspaper my sample of housing news articles (N=33998). I draw from one major newspaper publication for most cities, with the exception of Boston, Detroit, and Los Angeles, in which I draw form the two major newspapers in the area. My sample covers articles from January 2000 to December 2013.

Table 2: Sample Positive Words and Word Counts

word	% of Total PosWord Count	Freq.
BIGGEST	3.33	18809
BOOM	1.83	10394
BOOST	0.56	3176
EXTRAORDINARY	0.94	5305
EXPLODE	0.57	3197
FASTEST	2.76	15633
FRENZY	2.91	16513
FUEL	0.69	3884
GREAT	4.33	2458
HOT	0.85	917
HUGE	0.72	4077
JUMP	0.96	5411
RECORD	2.34	13251
ROCKET	0.13	753
POSITIVE	0.29	1672
PROPEL	0.18	991
PROLIFIC	0.30	1688
SIZZLE	0.70	3956
SKYROCKET	0.34	101
SOAR	0.58	3298
SPIKE	0.13	745
SPEEDING	0.24	1332
STRONG	2.02	11413
SURGE	0.44	2481

Note: The word counts for each listed word includes different tenses and inflections. So for example "boom" includes counts for "booms", "boomed", and "booming."

Table 3: Predicting House Price Appreciation with Media Sentiment

	Dep Var: House Price Growth		
	(1)	(2)	(3)
Lagged Media Sentiment	2.815	2.815	2.815
	(6.20)	(4.50)	(3.89)
Year 1 ($L^1 + ... + L^4$)	1.012	1.012	1.012
	(5.44)	(4.03)	(3.96)
Year 2 ($L^5 + ... + L^8$)	1.112	1.112	1.112
	(5.33)	(5.03)	(3.55)
Year 3 ($L^9 + ... + L^{12}$)	0.690	0.690	0.690
	(4.32)	(2.91)	(2.99)
Std Errors: Newey-West	✓	.	.
Std Errors: Driscoll-Kray	.	✓	.
Std Errors: Clustered by (i, t)	.	.	✓
Observations	1450	1450	1450
Adjusted R^2	0.48	0.48	0.48

Note: Table 3 reports the estimates for β in equation (2). Each coefficient estimate measures the impact of a one percent increase in the quarterly growth of sentiment on the future quarterly growth in housing prices. All columns control for past price changes and a vector of housing fundamentals, $\triangle x_t$, described in the text. Lagged Media Sentiment reports the cumulative effect of all lags of sentiment from $t-12$ to $t-1$. The rows below break down total sum of the quarterly lags of sentiment by lagged years. "Year 1" coefficient equals the sum of lagged expectations from L^1 to L^4, "Year 2" is the sum of lags L^5 to L^8, "Year 3" is the sum from lags L^9 to L^{12}. The corresponding t-statistics for the linear combination of estimates are reported in parentheses below based on calculated standard errors. Column 1 calculates Newey and West (1987) standard errors, robust to heteroskedasticity and auto-correlation up to $T-1$ lags, Column 2 calculates Driscoll and Kraay (1998) standard errors to account for cross-section spatial dependence, and Column 3 calculates standard errors clustered by quarter and city.

Table 4: Predicting Housing Media Sentiment with Past Returns

	Dep Var: Media Sentiment S_{it}			
	6 months	1 year	3 years	5 years
	(1)	(2)	(3)	(4)
Cumulative Price Appreciation	10.578	7.127	2.614	1.020
	(6.23)	(6.62)	(3.61)	(1.90)
$log(P)$	0.404	0.245	-0.325	-0.331
	(1.03)	(0.63)	(-0.63)	(-0.54)
30-year Mortgage Rate	-0.597	-0.597	-0.700	-0.700
	(-5.35)	(-6.25)	(-6.86)	(-7.33)
Observations	1850	1850	1850	1850
Adjusted R^2	0.207	0.222	0.198	0.180

Note: Table 4 explores the determinants of media sentiment, and reports the estimates for λ and δ in equation (3). Standard errors are clustered by quarter and city, corresponding t-statistics are reported in parentheses below. Estimates show that when past returns are high, media sentiment is also high. This effect is robust up to a window of 3 years, but the magnitude of the effect appears to decline with a longer horizon.

Table 5: Is Media Sentiment Driven By News Stories on Fundamentals?

	Dep Var: Housing Price Growth, t=quarterly						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Lagged Media Sentiment	2.971	3.494	3.386	3.339	3.904	3.805	3.494
	(3.95)	(4.52)	(4.59)	(4.50)	(4.66)	(4.63)	(4.16)
Media Rents	✓	✓	✓	✓	✓	✓	✓
Media Labor Market Conditions	.	✓	✓	✓	✓	✓	✓
Media Housing Supply	.	.	✓	✓	✓	✓	✓
Media User Costs	.	.	.	✓	✓	✓	✓
Media Demographics	.	.	.	.	✓	✓	✓
Media Local GDP & Inflation	.	.	.	.	.	✓	✓
Media Credit Conditions	.	.	.	.	.	.	✓
Observations	1432	1432	1432	1432	1432	1432	1432
Adjusted R^2	0.472	0.473	0.477	0.475	0.476	0.474	0.474

Note: Table 5 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. Standard errors are clustered by quarter and city, corresponding t-statistics are reported in parentheses. Estimates of lagged media represent 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. Estimates remain robust to the inclusion of "media fundamentals," suggesting that the estimated effect is not driven by a particular set of stories on fundamentals.

Table 6: Controlling for Easy Credit and Subprime Lending

	(1)	(2)	(3)	(4)
Lagged Media Sentiment	4.801	4.812	4.808	3.388
	(3.58)	(3.45)	(3.42)	(2.50)
Year 1 ($L^1 + ... + L^4$)	1.649	1.650	1.659	1.230
	(3.67)	(3.67)	(3.68)	(2.72)
Year 2 ($L^5 + ... + L^8$)	2.023	2.029	2.026	1.350
	(3.53)	(3.34)	(3.30)	(2.27)
Year 3 ($L^9 + ... + L^{12}$)	1.129	1.134	1.123	0.809
	(2.69)	(2.53)	(2.46)	(1.81)
Libor Rate	.	✓	✓	✓
Loan-To-Value Ratios	.	.	✓	✓
% of Subprime Lending	.	.	.	✓
Observations	762	762	762	762
Adjusted R^2	0.481	0.481	0.477	0.506

Note: Table 6 reports the estimates for β in equation (2) with additional controls for subprime lending and ease of credit. As in Table 3, each coefficient estimate measures the impact of a one percent increase in the quarterly growth of sentiment on the future quarterly growth in housing prices. All columns control for past price changes and a vector of housing fundamentals. Lagged Media Sentiment reports the cumulative effect of all lags of sentiment from $t-12$ to $t-1$. The rows below break down total sum of the quarterly lags of sentiment by lagged years. "Year 1" coefficient equals the sum of lagged expectations from L^1 to L^4, "Year 2" is the sum of lags L^5 to L^8, "Year 3" is the sum from lags L^9 to L^{12}. The corresponding t-statistics for the linear combination of estimates are reported in parentheses below based on calculated standard errors clustered by quarter and city.

Table 7: Predicting Price Growth Using Weekend and Narrative Sentiment

	Baseline	Weekend	Narrative
	(1)	(2)	(3)
Sum of Lagged Sentiment	2.815	2.545	2.621
	(3.88)	(3.99)	(3.78)
Year 1 ($L^1 + ... + L^4$)	1.012	0.541	0.907
	(3.96)	(3.06)	(3.88)
Year 2 ($L^5 + ... + L^8$)	1.112	1.195	1.047
	(3.55)	(3.95)	(3.80)
Year 3 ($L^9 + ... + L^{12}$)	0.690	0.809	0.667
	(2.99)	(3.38)	(2.64)
Observations	1450	1450	1450
Adjusted R^2	0.48	0.47	0.48

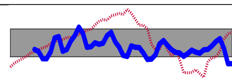
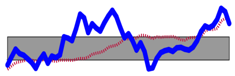
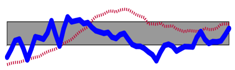
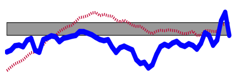
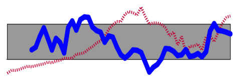
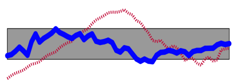
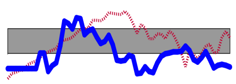
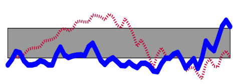
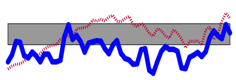
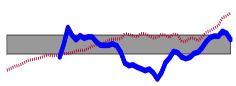
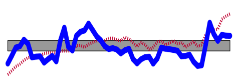
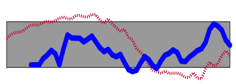
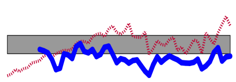
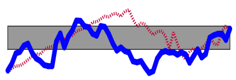
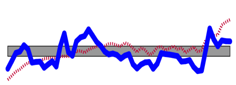
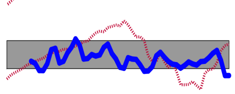
Note: Table 7 reports the estimates for β in equation (2) using alternate sentiment indices $S^{Weekend}$and $S^{Narrative}$. Each coefficient estimate measures the impact of a one percent increase in the quarterly growth of sentiment on the future quarterly growth in housing prices. All columns control for past price changes and a vector of housing fundamentals. Lagged Media Sentiment reports the cumulative effect of all lags of sentiment from $t - 12$ to $t - 1$. The rows below break down total sum of the quarterly lags of sentiment by lagged years. "Year 1" coefficient equals the sum of lagged expectations from L^1 to L^4, "Year 2" is the sum of lags L^5 to L^8, "Year 3" is the sum from lags L^9 to L^{12}. The corresponding t-statistics for the linear combination of estimates are reported in parentheses below based on calculated standard errors clustered by quarter and city.

Table 8: Cross-Sectional Effects of Media Sentiment

	(1)	(2)	(3)	Low Price Homes (4)	High Price Homes (5)
Lagged Media Sentiment	2.055	2.020	2.768	5.921	3.335
	(3.51)	(3.40)	(2.42)	(4.15)	(4.63)
× *High* Minority Buyers	1.568	.	.	.	.
	(2.64)	.	.	.	.
× *High* 2nd Home Buyers	.	3.202	.	.	.
	.	(3.63)	.	.	.
× *High* Subprime Loans	.	.	2.576	.	.
	.	.	(2.70)	.	.
Observations	1450	1450	773	712	712
Adjusted R^2	0.483	0.490	0.483	0.683	0.753

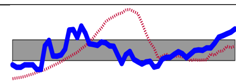
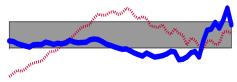
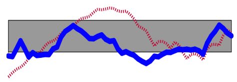
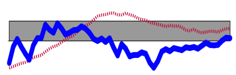
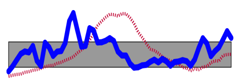
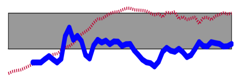
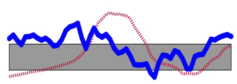
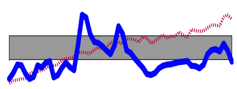
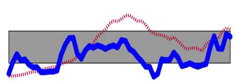
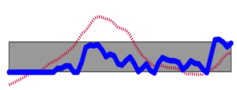
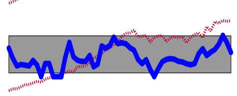
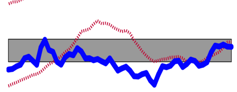
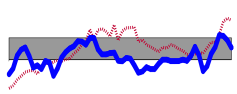
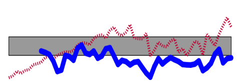
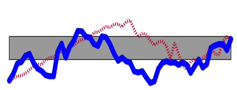
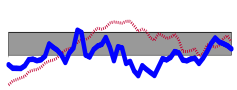
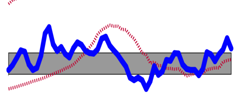
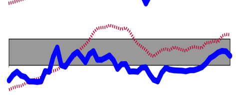
Note: Table 68explores the cross-sectional effects of sentiment on house price changes. reports the estimates for β and β^{High}in equation 5. Each coefficient estimate measures the impact of a one percent increase in the quarterly growth of sentiment on the future quarterly growth in housing prices. The baseline estimate for β represents the effect of the *Low* group described in the text, while estimates β^{High} represent the additional effect of sentiment on prices for the corresponding *High* group. All columns still control for past price changes and the same vector of housing fundamentals in all previous tables. Lagged Media Sentiment reports the cumulative effect of all lags of sentiment from $t-12$ to $t-1$. The corresponding t-statistics for the linear combination of estimates are reported in parentheses below based on calculated standard errors clustered by quarter and city.

Table A.1 Housing Sentiment and Prices by City

		Media Sentiment		House Prices	
		Mean	Sd	Mean	Sd
Atlanta		2.30	1.45	170.48	17.09
Austin		3.39	2.55	242.36	37.39
Baltimore		2.55	1.64	209.91	46.43
Boston		1.97	1.61	233.21	31.09
Charlotte		2.68	2.28	169.90	17.08
Chicago		2.70	1.31	190.52	28.31
Cincinnati		1.95	2.31	164.29	10.25
Cleveland		1.19	1.70	157.75	11.51
Columbus (OH)		1.66	1.94	168.99	10.42
Dallas		2.56	1.98	164.07	14.64
Denver		2.03	1.72	266.80	22.88
Detroit		1.95	1.85	168.92	35.11
Indianapolis		1.66	1.31	152.14	7.77
Kansas City		3.01	2.10	180.92	13.81
Las Vegas		2.59	1.76	164.39	57.22
Los Angeles		2.60	1.33	187.34	51.25

Note: This figure plots the housing sentiment and price indexes and lists the respective mean and standard deviation by city from 2000-Q1 to 2013-Q4. The solid line represents the media sentiment index and dashed line plots the price index. The shaded region plots the interquartile range of the house price index. The media sentiment index equals the share of positive minus negative words across housing newspaper articles in each city-month.

(Table A.1 Continued)

		Media Sentiment		House Prices	
		Mean	Sd	Mean	Sd
Miami		2.01	1.90	260.85	78.13
Milwaukee		2.77	1.65	206.74	23.08
Minneapolis		2.52	1.59	220.55	28.45
NYC		2.57	1.56	221.92	42.53
Orlando		2.68	1.87	188.80	50.38
Philadelphia		1.75	1.37	196.68	44.04
Phoenix		2.73	2.05	221.07	60.06
Pittsburgh		3.00	2.36	171.44	20.21
Portland		2.33	1.70	264.37	50.01
Sacramento		1.45	1.60	168.89	45.17
San Antonio		2.66	1.74	185.99	28.61
San Diego		2.50	1.52	215.39	48.04
San Francisco		2.35	1.58	231.34	33.92
San Jose		2.36	1.73	229.34	32.93
Seattle		2.68	2.10	229.04	42.87
St. Louis		2.21	1.82	185.09	19.57
Tampa		1.62	2.06	211.07	50.65
Washington, D.C.		1.93	1.41	216.78	49.73