Washington University in St. Louis

Washington University Open Scholarship

Honors in Management

Undergraduate Research

Spring 5-1-2020

Talk on the Block: Using public forum sentiment to predict housing prices

Luke Sammons

Josh An

Josh Hill

Karan Toshniwal

Olin Business School

Washington University in St. Louis

Talk on the Block:

Using public forum sentiment to predict housing prices

Josh An | Josh Hill | Luke Sammons | Karan Toshniwal

Advisor: Asaf Manela

Honors in Management

May 2020

Abstract

This paper develops a textual analysis methodology to quantify sentiment on public market forums to predict outcomes in the real estate market. This paper draws inspiration from Soo (2018) which quantified sentiment through real estate news media. We believe that analyzing public forums allows us to understand public sentiment in its most unedited, casual form; whereas real estate news media is limited to perspectives and interpretations of an editor. Antweiler and Frank (2004) showed that public forums are significant when predicting stock market outcomes, lending validity to our text source. Our methodology includes identifying a relevant dictionary of positive and negative words, scraping BiggerPockets real estate forums, running a textual sentiment analysis, and finally regressing against fundamental housing market indicators in 34 large metropolitan statistical areas (MSAs) to assess sentiment's predictability on home prices.

Our regression results suggest that sentiment significance varies more in the short-run with public forum text than it does with news media because news media is "marked-to-market daily." Marking-to-market is the practice of valuing securities, or portfolios of securities, at their current market value, as opposed to a book value. Because news media is updated every day, and sometimes more than once a day, we find that it is capturing current market home values much more quickly than forum sentiment. Additionally, we conclude that discussion on real estate public forums can predict housing prices in the long-run, suggesting that the users are engaging in conversation that is targeting long-term investments and trying to make sense of the potential future value of a home that they are considering buying and/or selling.

Introduction

With the recent outbreak of COVID-19, financial markets have been hit hard and lack the consumer confidence to recover from such unforeseen unemployment and production shocks. Despite all the efforts to recover from the "Great Recession" of 2008 by regulators, credit rating agencies, and investment banks, which have corrected for the malpractice and misuse of collateralized debt obligations (CDOs) and mortgage-backed securities (MBS), this new virus outbreak poses a serious threat to the current housing market on the uncertain path to recovery. Regardless of the outbreak, the underlying assets of these financial instruments are lagging in recovery. Home values have not regained their value in a majority of MSAs, which is evident in-home market value listings being well below their proper tax-appraised value ratio of 33.33% on popular real estate sites, such as Zillow. During the 20th Century, median home values appreciated at exponential rates, averaging a 6.2% increase YoY from 1968-2004. Home values will likely never appreciate as such again do to the lack of space for new construction in major cities and surrounding suburban areas. As supply continues to run dry, demand will rise to new heights and the real estate market will need to adjust. Across many industries, a popular form of gauging demand is through the use of public forums. Zillow operates its own public forum, on which homeowners and home-hunters alike can share opinions about certain markets, trends in their neighborhoods, home valuations, and desired locations.

Popular opinion would suggest that the most important aspect regarding real estate and home-buying is location. We believe that determining market temperature, that is whether certain markets are “hot” or “cold” for sellers and buyers, would help real estate companies, developers, and homebuyers find markets in which they can all benefit and equilibrate their respective housing markets. Additionally, in looking at text analysis to determine real estate market temperatures and trends, our group finds there are multitudes of data aggregation sites, such as newspaper articles and market indices, that track housing market sentiment. One such example is the monthly-reported University of Michigan Consumer Sentiment Index, which is regarded as a key temperature check on overall consumer confidence in the market. What the research does not capture is the chatter that occurs without a survey or the information that cannot be collected in a Wall Street Journal article. Thus, we believe that the best way to understand market temperature and bring home values back to the equilibrium of pre-2008 is to develop a predictive model based on public forums where raw emotions and sentiment are being shared every day.

In this vein, our marginal impact over Soo's 2018 paper "Quantifying Sentiment with New Media across Local Housing Markets," will be to demonstrate that beyond news media, a better predictive model for housing market temperature and home values is one that incorporates daily chatter on the individual consumer level; that is, unedited public discussion boards and forums may provide better housing market temperatures in local markets.

Literature & Dictionary Review

In "Quantifying Sentiment with New Media across Local Housing Markets," the author argues that direct impact on the housing prices is hard to ascertain since the degree of the price variation differs per area, thus requiring cross-sectional variables. After inspecting local news of 34 different cities in the U.S with textual analysis, the paper concludes that media sentiment is predominantly limited to local sources since housing prices are dictated by how each localized market is responding to home values and consumer confidence. On top of local media reflecting quantitative prediction for home prices, she also explains that these sentiment indexes can account for a wider range of variation, including post-housing bubble recovery with a time-lag.

Soo's dictionary is based on the Harvard IV-4 Psychological Dictionary, a widely used resource for this type of analysis, but is improved by the addition of varying tenses and inflections for each word, both positive and negative. For example, the dictionary includes "Blooming" and "Bloomed", as a variation of "Bloom." Soo modifies this dictionary to fit her textual input, financial news articles, and thus makes it a valuable resource in the undertaking of our research. Soo has generously provided us with this improved dictionary which serves as the foundation of our sentiment analysis.

Data Acquisition & Preparation

BiggerPockets Forum Data

To perform a sentiment analysis on the impact of forum discussion on housing prices in the largest metropolitan statistical areas (MSAs), our methodology will mimic that of Soo in her paper, "Quantifying Sentiment with News Media across Local Housing Markets" (2018). As stated in our introduction, our goal is to provide marginal impact by taking what we hypothesize to be a more "raw" form of sentiment from buyers and sellers directly, as opposed to an edited and filtered newspaper article(s), and demonstrate predictability from this raw sentiment. The forum for this text scrape is BiggerPockets, an online platform founded in 2004 which aims at providing a real estate investment community for local real estate markets. BiggerPockets boasts nearly 5 million forum posts and 800 new daily active users, there is a plethora of entries to be used for analysis. Our analysis will focus on the "Buying & Selling Real Estate Discussion" forum, which contains over 500 thousand unique posts and 17 thousand unique discussions. This community continues to grow and focuses on general questions about buying and selling real estate and homeownership. Some questions are general and have threads from people all over the country, as well as market-specific discussions. We believe that the frequency with which people interact on BiggerPockets, as well as its growing reputation amongst real estate investors, makes it one that will allow us to provide the marginal impact of unedited consumer sentiment.

Dictionary & Sentiment Score Development

We initially considered extending Soo's dictionary to include ideograms, emojis or emoticons, and casual vernacular, but we've decided to adopt her dictionary without modification. From cursory reviews of the forum, we observe only limited use of casual language likely due to the site's professional subject matter. As such, we do not believe that adding these elements will significantly affect our results. Also, by utilizing Soo's previously applied dictionary rather than a custom one, we allow our results to be more closely compared to the existing body of sentiment analysis research.

We begin our data preparation process by first filtering the raw data (courtesy of BiggerPockets) by location, removing any posts not included in our list of target MSAs. This step alone reduces our dataset from 500 thousand posts to approximately 280 thousand posts. Centering this more manageable dataset, we apply a series of textual treatments. The body text of each post is cleaned of HTML formatting before being tokenized or broken into a list of successive words. Next, we clean these tokenized lists of 'stop words' or words, like articles, that do not typically convey relevant information. The extensive stop word dictionary we apply is provided by the NLTK Project in its Natural Language Toolkit module. Finally, a negation rule is used to handle the reversing effect of negations on sentiment-carrying words. The final dataset contains more than 8.1 million individual words at an average of around 60 words per post. This preparation approximates Soo's process as described in her paper.

After applying these treatments, it is possible to generate the sentiment index. Using an extended form of Soo's dictionary to include negated terms, we generate counts of positive, negative, and total words on a per post basis. This per post data is summarized by location and month to generate a raw snapshot of popular sentiment. These aggregated count values are synthesized into sentiment scores using Soo's formulation by city$_i$ and time$_t$ period:

$$\mathbf{S_{it}} = \frac{(\mathbf{\#pos} - \mathbf{\#neg})}{\mathbf{\#\ of\ total\ words_{it}}}$$

That is, to provide a sentiment score on each thread, we take the difference in positive and negative words as a ratio of total words in the thread, for each city and time period. Alternatively, we test for sentiment as a function of the difference between positive and negative words, divided by the sum of positive and negative words, thus providing a more contained ratio.

Soo accounts for the negation of both positive and negative words by searching for one of six terms (no, not, none, neither, never, nobody) prior to the word in question. It is considered negated if one of those six words appears within five words preceding the word in question. We believe that a more robust form of testing for negation is to increase the search window size by reversing the sentiment value of every word between a negation and a punctuation. This ensures that we capture any negation even those that fall outside of the five-word window utilized by Soo.

Table 1.

Most Common Sentiment-Carrying Words			
Positive Words		**Negative Words**	
Word	**Frequency**	**Word**	**Frequency**
good	17.07%	contract	15.67%
great	8.24%	low	13.07%
well	7.40%	lower	8.58%
best	6.44%	foreclosure	3.85%
better	5.19%	flat	2.24%
high	3.84%	contracts	2.10%
higher	3.00%	discount	2.08%
big	2.57%	stop	2.02%
hope	2.04%	drop	2.00%
profit	1.80%	fall	1.94%

To validate our sentiment index, we compare our scores to the University of Michigan/Reuters Survey of Consumers (SOC), which asks, "Generally speaking, do you think now is a good or bad time to buy a house?" As is depicted in **Chart 1**, we validate by population-weighing our sentiment scores in our MSAs, and then plotting these against the "yes" respondents to the SOC question. Additionally, in **Chart 2**, we log-transform our sentiment index change over 2019 and plot these changes against the log-transformed changes of SOC affirmative respondents during the same time period.

Chart 1.

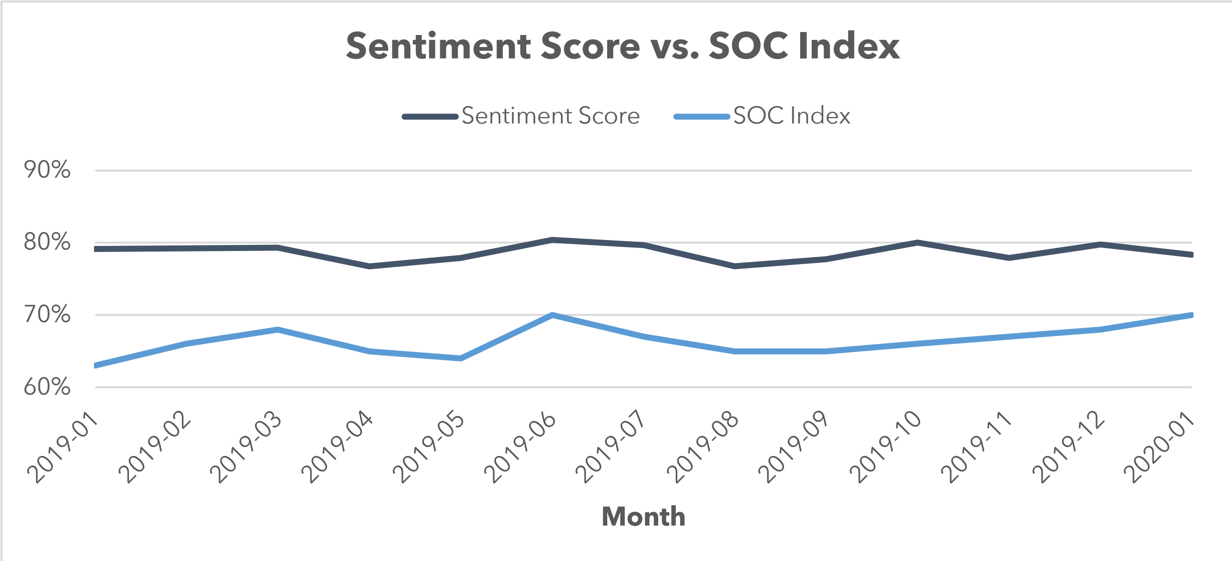
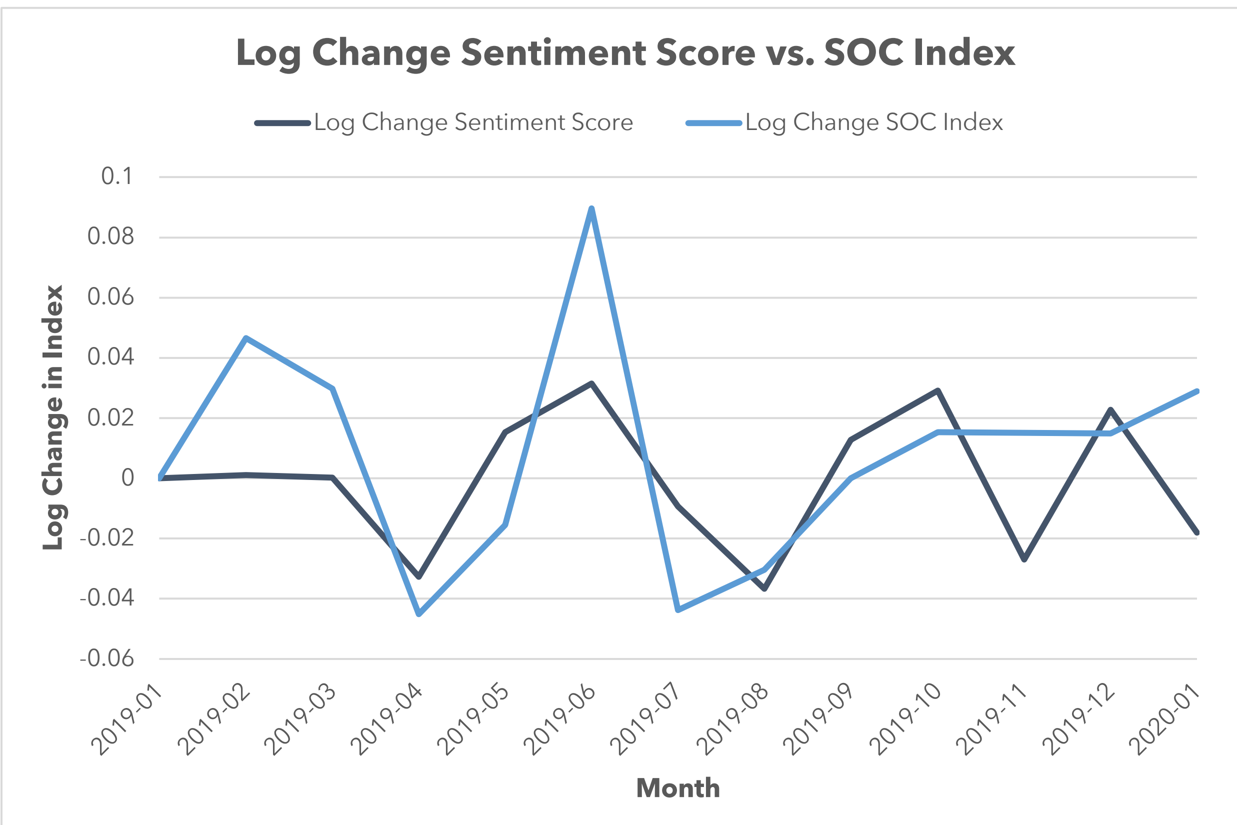


Chart 2.



Zillow Housing Data

We gather the home sale prices and housing inventory levels in our MSAs of question through Zillow Group, an online platform for buying, renting, and selling real estate. Zillow has extensive amounts of research due to the number of listings across the country that are shared on its site. In fact, this research is so robust that the company has started providing its own home loans to people searching for real estate on their website, essentially serving as a one-stop-shop for searching and financing. Our analysis incorporates Zillow data for both historic home prices ("Median Sale Price" (seasonally-adjusted)), rent levels, and monthly inventory levels ("Monthly, For-Sale Inventory" (seasonally-adjusted)) of different MSAs, dating back to 2010.

Time-Series Regression Model

To test whether or not forum sentiment analysis can be used to predict home prices, our group replicates the regression of Soo's 2018 study. Considering the monthly-nature of our data, our group finds it best to replicate Soo's regression without log-transformation or differencing each variable down to the month. Thus, our time-series regression model for the 34 MSAs between 2013-2019 is as follows:

$$\begin{aligned}\mathbf{Price_{it+k}} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{Sentiment_{it}} + \boldsymbol{\gamma}\mathbf{Price_{it}} + \boldsymbol{\delta}\mathbf{Rent_{it}} + \boldsymbol{\pi}\mathbf{Inventory_{it}} + \boldsymbol{\theta}\mathbf{Population_{it}} \\ + \boldsymbol{\vartheta}\mathbf{Unemployment_{it}} + \boldsymbol{\rho}\mathbf{Rate_{it}} + \boldsymbol{\varphi}\mathbf{Income_{it}} + \boldsymbol{\varepsilon}\end{aligned}$$

Similar to our sentiment index, let *i* denote the city and *t* denote the time period, which in our case is monthly, spanning as far back as 2013. Our dependent variable is led at a lead by *k*, where *k* = *number of* months. The following table contains a descriptive list of our variables, including fundamental economic variables such as rent, population, income per capita, employment statistics, and mortgage rates of a given city. In addition to these fundamentals, we are also adding another variable to this equation, which assesses "Monthly, For-Sale Inventory (seasonally-adjusted)," according to the data provided by Zillow. This will serve as a proxy for market temperature and control for changing house prices as a result of a market cooling off, rather than sentiment being the main contributing factor.

Table 2.

Variable	**Description**	**Source**
$Price_{it}$	Median Home Sale Price in $city_i$ and $time_t$	Zillow Research
$Sentiment_{it}$	Sentiment Index in $city_i$ and $time_t$	Bigger Pockets Forums
$Rent_{it}$	Average Rent in $city_i$ and $time_t$	Zillow Research
$Inventory_{it}$	Monthly For-Sale Inventory in $city_i$ and $time_t$	Zillow Research
$Population_{it}^{*}$	Annual Population in $city_i$	FRED (St. Louis)
$Unemployment_{it}$	Monthly Unemployment in $city_i$ and $time_t$	FRED (St. Louis)
$Rate_{it}$	Monthly 30-yr. Fixed Mortgage Rate in $city_i$ and $time_t$	FRED (St. Louis)
$Income_{it}^{*}$	Annual Per Capita Income in $city_i$	Bureau of Economic Analysis

*Note: Population and Per Capita Income by MSA are calculated on an annual rate, requiring that the variables by linearly interpolated to attain monthly measurements.

The goal with these variables is to pull as much out of the error term as possible, in the hopes that our beta will be statistically significant and different from zero. Additionally, we seek to acknowledge and understand that prior to 2000, most of the predictive power (70%) on home prices was nested in the fundamentals; yet, today, those same variables only account for 10% of the variation in home prices, according to Soo. In that vein, our hypothesis test is as follows:

> **H_o: $\beta = 0$.** If sentiment shared through BiggerPockets forums do not accurately predict price movements from month-to-month, or are simply captured in the fundamentals already, then our beta coefficient should not be different from zero.

> **H_a: $\beta \neq 0$.** If sentiment shared through BiggerPockets forums are in some way impacting the future median home prices of given MSAs as reported by Zillow, then the beta coefficient will be different from zero and statistically significant at $\alpha = 0.10$.

<u>Results</u>

To properly assess the perceived value of our sentiment score, we lag our dependent variable with numerous different time-horizons, in time intervals of 3 months (i.e., 1 quarter). Soo notes that sentiment in her study has predictive power up to 8 quarters. Thus, we assess our regression equation at k = 3, 6, 9, 12, 15, 18, 21, 24 months to validate this claim. The results for all regressions and their heteroskedastic-adjusted "robust" transformations can be found in our **Appendix**.

Without testing for heteroskedasticity, all of our regressions come back with strong R-squared values and most independent variables are significant at least under $\alpha = 0.10$ significance, except for our variable of interest. Understanding that non-constant variance of the error terms due to clustering across time and city is a common flaw of panel data, we correct for heteroskedasticity by using the "robust" option in STATA. Adjusting for heteroskedasticity, the regressions that report a significant sentiment index coefficient are those with 15, 18, and 21-month leads on the MSA median home sale price. The results of these regressions are in **Table 3** below:

Table 3.

Variable	**Lead, k = 15 months**	**Lead, k = 18 months**	**Lead, k = 21 months**
	Forward Price*	Forward Price	Forward Price
Sentiment	35694.9* (19694.6)	50769.7** (19695.8)	40481.6* (23723.8)
Price	0.513*** (0.145)	0.468*** (0.156)	0.485*** (0.146)
Rent	81.27** (30.50)	105.2*** (38.24)	104.3** (41.34)
Inventory	0.143 (0.538)	0.180 (0.546)	0.241 (0.506)
Population	0.162 (0.153)	0.113 (0.166)	0.0690 (0.169)
Unemployment	52086.7 (136221.3)	72086.6 (153053.4)	116708.2 (150439.1)

Rate	-104764.3 (145319.2)	-373244.1** (162990.2)	-477934.8*** (150987.1)
Income	24.58 (14.92)	22.80 (14.15)	25.39** (10.92)
Constant	-147025.6* (72887.8)	-132907.5 (79091.8)	-126400.3 (75682.1)
N	*2290*	*2189*	*2088*
R^2	*0.859*	*0.847*	*0.855*
Prob > F	*0.0000*	*0.0000*	*0.0000*
Rho	*.95705311*	*.94475744*	*.93872749*

*Note: "Forward Price" = $Price_{it+k}$
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

The addition of monthly inventory statistics from Zillow is not statistically significant in these regressions. Average monthly rent from Zillow is statistically significant at least at the $\alpha = 0.05$ level in each of the regressions above. Sentiment is positively correlated with median home sale prices. All else equal and averaging across these three models, for each 100% increase in sentiment score, we can anticipate an approximately $42,000 increase in median home sale price. As we would expect, all else equal, there is a negative correlation between mortgage rates and the median sale price of homes in a given MSA. In other words, as debt becomes "cheaper" for a homeowner, we expect to see people purchase more expensive homes, likely inflating the median home sale price in certain metropolitan areas. Additionally, as per capita income increases, we will observe an increase in median home sales prices. Finally, this regression proves that past home prices are significantly indicative of the forward home prices (the lead prices).

On the surface, these results are not as consistent as Soo's results, through which she saw significance in her sentiment index through 8 quarters of data. Given this discrepancy, we attempt an adjustment to the regression model by testing sentiment index significance beyond two years (i.e., k = 24 months). We want to assess whether or not our sentiment score could be significant beyond the two years of significance that Soo found because it may hint at the significance that time horizons can play in text analysis.

To test this change, we lead our dependent variable by 4 more quarters, k = 27, 30, 33, and 36 months. Under our standard regression equation, sentiment is significant at k = 27, 30, and 36 months. In this case, our variable of interest is not significant under the heteroskedastic-adjusted form of the regression for any of these dates, potentially due to the seasonality and cyclical nature of our dataset. These results are below:

Table 4.

Variable	**Lead, k = 27 months**	**Lead, k = 30 months**	**Lead, k = 36 months**
	Forward Price*	Forward Price	Forward Price
Sentiment	54286.3^{*} (29323.6)	52153.1^{*} (27943.1)	58594.3^{*} (30815.7)
Price	0.579^{***} (0.0256)	0.551^{***} (0.0263)	0.466^{***} (0.0332)
Rent	118.4^{***} (6.254)	135.5^{***} (6.156)	95.97^{***} (8.085)
Inventory	0.257^{**} (0.118)	0.405^{***} (0.121)	0.601^{***} (0.152)
Population	-0.181^{***} (0.0667)	-0.228^{***} (0.0672)	-0.126 (0.0912)

Unemployment	76154.8** (38337.4)	-2487.0 (37069.3)	-45079.8 (44102.1)
Rate	-539744.9*** (98261.6)	-477008.5*** (92803.3)	-315227.9*** (107683.9)
Income	15.40*** (2.324)	9.437*** (2.277)	29.03*** (3.300)
Constant	-24652.0 (25624.4)	2961.5 (25589.3)	-39452.1 (34486.8)
N	*1885*	*1783*	*1579*
R^2	*0.877*	*0.884*	*0.840*
Prob > F	*0.0000*	*0.0000*	*0.0000*
Rho	*.94972542*	*.96856025*	*.96695985*

*Note: "Forward Price" = $Price_{it+k}$
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

The main difference to note from these results is that in at least one of the above regressions, all of the independent variables are significant. Variables, such as Inventory, Population, and Unemployment, that were not significant in the short-run (i.e., k < 24 months) are now statistically significant at points beyond two years (i.e., k > 24 months).

Conclusion

After reviewing our data, we analyze the fact that under our panel regression model our sentiment score becomes more significant as we lag our variables more. The lack of consistency in significance may suggest that news media sentiment better captures market temperatures in the short-run, as posited by Soo. However, we believe that our analysis of public forum conversation hints at something larger. If you stop to consider who may be interacting on these real estate investment forums, you will likely uncover that these people are homeowners and brokers that are trying to buy and sell homes at a very localized level. These are individuals that are seeking to invest in a home for more than just a few months. In fact, as of the second quarter in 2019, the average duration of homeownership in the United States achieved a record high of 8.09 years (Attom Data Solutions, 2019). This tenure trend has been rising since 2010. Understanding this homeowner trend, the results of our regression confirm that home values may be predicted by market sentiment in the long-run, and in some cases, longer than 2 years. The conversations that people are having in a public forum are perhaps more indicative of long-term housing market trends and include rhetoric meant to be suggestive of performance of a long-term investment, such as a family home.

Additionally, we believe that the reason for our significance not being consistent across different time horizons is in large part due to the fact that public forums capture raw emotions, questions, or opinions of the current, and more importantly, the future state of local housing markets than can be interpreted from an edited news media source or article. Similarly, with the pressure of a twenty-four news cycle, people are looking for information that is "marked-to-market" daily; that is, people are searching for the most up-to-date information on home values. This may be captured in newspapers that reflect the current market value right now (i.e., they are "marked-to-market"), as opposed to public forum sentiment, which may take time to impact market values.

In this vein, our marginal contribution is two-fold:

(i) Our regression results suggest that sentiment significance varies more in the short-run with public forum text than it does with news media, because news media is marked-to-market daily to reflect even the most volatile of sentiments. That is, sentiment is adjusted daily (and sometimes faster) with news media; whereas, public forum sentiment adjusts slowly in the short-run.

(ii) Discussion on real estate public forums can predict housing prices in the long-run, suggesting that the users are engaging in conversation that is targeting long-term investments and trying to make sense of the potential future value of a home that they are considering buying and/or selling.

Future work on this subject may seek to answer some tangential questions to this topic of public forum sentiment. Specifically, real estate investors might want to validate Soo's claim that socio-economic fundamental data only accounts for 10% of the variation in home prices today. Our rho statistics in **Tables 3 and 4** suggest that consistently more than 90 percent of the variation in the model is due to differences across panels. This is evidence that there still exists a lot of variation in the fundamental home price indicators that may be affecting the median home sale price in these 34 MSAs. However, the lack of significance of key fundamental metrics (as in **Table 3**), suggests that in the short-term, some of these fundamentals (i.e., Population, and Unemployment) have lost significance relative to the power of sentiment. Furthermore, the results of this paper suggest that more research should be done to understand what drives sentiment in public forums, and whether or not public forum sentiment is driven by news media sentiment, or vice-versa. Determining the direction of a potential causal relationship between these two forms of text sentiment may help to understand the true driver(s) of home values in the future.

References

Antweiler, Werner and Frank, Murray. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. Journal of Finance. 59.1259-1294.10.2139/ssrn.282320.

Attom Data Solutions. (Posted July 16, 2019). U.S. Median Home Prices Reach A New Peak In Q2 2019. https://www.attomdata.com/news/market-trends/q2-2019-u-s-home-sales-report/.

Kelly, Bryan and Manela, Asaf and Moreira, Alan. (November 22, 2019). Text Selection. Available at SSRN: https://ssrn.com/abstract=3491942.

Carlson, Ben. (Posted March 8th, 2019). The Real Estate Market in Charts. A Wealth of Common Sense. https://awealthofcommonsense.com/2019/03/the-real-estate-market-in-charts/.

Soo, Cindy. (Updated 2018). Quantifying Sentiment with News Media across Local Housing Markets. Review of Financial Studies. 31. 3689-3719. 10.1093/rfs/hhy036.

Trimbur, Thomas and Bell, William. (November 13, 2008). Seasonal Heteroskedasticity in Time Series Data: Modeling, Estimation, and Testing. Available at U.S. Census Bureau: https://www.census.gov/ts/papers/rrs2008-11.pdf.

Appendix

Appendix A
(Summary Statistics)

Variable	Mean	Std. Dev.	Min.	Max.
Sentiment	.0183083	.0092935	-.0454545	.1538462
Price	263328.2	157952.9	102900	1130700
Rent	1523.795	538.4578	872	3310
Inventory	19913.61	18843.85	1544	120656
Unemployment	.0497052	.0145858	.021	.108
Population*	359243.9	299600.7	130977.7	1611232
Rate	.0403423	.0035986	.0341	.0487
Income*	4555.832	971.5054	2966.083	8851.083
Observations	*2395*			

*Note: Population and Per Capita Income by MSA are calculated on an annual rate, requiring that the variables by linearly interpolated to attain monthly measurements.

Appendix B
(Data tables for k = 3-12 month lead, Fixed Effects Model)

Variable	**k = 3**	**k = 6**	**k = 9**	**k = 12**
	Forward Price*	Forward Price	Forward Price	Forward Price
Sentiment	8499.9 (16693.7)	21613.9 (24192.2)	36981.0 (28553.3)	28391.8 (31177.7)
Price	0.943*** (0.00847)	0.836*** (0.0124)	0.702*** (0.0146)	0.585*** (0.0161)
Rent	-3.467 (2.941)	12.31*** (4.286)	33.63*** (5.088)	54.18*** (5.567)
Inventory	-0.204*** (0.0602)	-0.213** (0.0878)	-0.123 (0.104)	0.0328 (0.114)
Population	0.0622** (0.0299)	0.112** (0.0437)	0.160*** (0.0517)	0.204*** (0.0567)
Unemployment	18321.5 (20931.1)	51776.5* (30513.3)	73631.0** (36094.4)	43021.7 (39427.8)
Rate	-356644.4*** (46387.3)	-487860.2*** (67585.9)	-361522.6*** (80075.0)	-96834.7 (87679.2)
Income	6.222*** (1.093)	12.69*** (1.597)	20.15*** (1.889)	24.53*** (2.070)
Constant	-8593.5 (11168.8)	-43826.4*** (16269.7)	-96714.0*** (19262.9)	-141060.7*** (21072.9)
N (Obs)	*2378*	*2387*	*2389*	*2392*
R^2	*0.969*	*0.932*	*0.903*	*0.883*

*Note: "Forward Price" = $Price_{it+k}$
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Appendix C
(Data tables for k = 15-24 month lead, Fixed Effects Model)

Variable	k = 15	k = 18	k = 21	k = 24
	Forward Price*	Forward Price	Forward Price	Forward Price
Sentiment	35694.9 (33430.7)	50769.7 (33605.7)	40481.6 (31806.1)	24810.3 (31256.1)
Price	0.513*** (0.0180)	0.468*** (0.0194)	0.485*** (0.0209)	0.565*** (0.0230)
Rent	81.27*** (6.009)	105.2*** (6.146)	104.3*** (6.051)	92.62*** (6.220)
Inventory	0.143 (0.124)	0.180 (0.127)	0.241* (0.123)	0.220* (0.118)
Population	0.162*** (0.0628)	0.113* (0.0667)	0.0690 (0.0670)	-0.0550 (0.0678)
Unemployment	52086.7 (42994.2)	72086.6 (44029.0)	116708.2*** (42330.3)	111296.3*** (40027.4)
Rate	-104764.3 (101540.1)	-373244.1*** (108547.7)	-477934.8*** (109632.0)	-494019.2*** (105351.5)
Income	24.58*** (2.261)	22.80*** (2.374)	25.39*** (2.351)	26.29*** (2.438)
Constant	-147025.6*** (23511.0)	-132907.5*** (25123.0)	-126400.3*** (25529.2)	-82765.1*** (26030.8)
N (Obs)	*2290*	*2189*	*2088*	*1987*
R^2	*0.859*	*0.847*	*0.855*	*0.868*

*Note: "Forward Price" = Price_{it+k}
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Appendix D

(Data tables for k = 27-36 month lead, Fixed Effects Model)

Variable	k = 27	k = 30	k = 33	k = 36
	Forward Price*	Forward Price	Forward Price	Forward Price
Sentiment	54286.3* (29323.6)	52153.1* (27943.1)	39836.1 (28782.4)	58594.3* (30815.7)
Price	0.579*** (0.0256)	0.551*** (0.0263)	0.482*** (0.0287)	0.466*** (0.0332)
Rent	118.4*** (6.254)	135.5*** (6.156)	120.9*** (6.746)	95.97*** (8.085)
Inventory	0.257** (0.118)	0.405*** (0.121)	0.630*** (0.135)	0.601*** (0.152)
Population	-0.181*** (0.0667)	-0.228*** (0.0672)	-0.137* (0.0750)	-0.126 (0.0912)
Unemployment	76154.8** (38337.4)	-2487.0 (37069.3)	-623.6 (39267.5)	-45079.8 (44102.1)
Rate	-539744.9*** (98261.6)	-477008.5*** (92803.3)	-357996.7*** (95027.5)	-315227.9*** (107683.9)
Income	15.40*** (2.324)	9.437*** (2.277)	20.39*** (2.498)	29.03*** (3.300)
Constant	-24652.0 (25624.4)	2961.5 (25589.3)	-43918.3 (28147.8)	-39452.1 (34486.8)
N (Obs)	*1885*	*1783*	*1681*	*1579*
R^2	*0.877*	*0.884*	*0.870*	*0.840*

*Note: "Forward Price" = $Price_{it+k}$
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Appendix E

(Data tables for k = 3-12 month lead, Fixed Effects Model, Robust Se)

Variable	**k = 3**	**k = 6**	**k = 9**	**k = 12**
	Forward Price*	Forward Price	Forward Price	Forward Price
Sentiment	8499.9 (10653.3)	21613.9 (22838.4)	36981.0 (21990.4)	28391.8 (20736.4)
Price	0.943*** (0.0150)	0.836*** (0.0501)	0.702*** (0.0938)	0.585*** (0.129)
Rent	-3.467 (7.084)	12.31 (15.38)	33.63* (19.59)	54.18** (24.10)
Inventory	-0.204 (0.133)	-0.213 (0.270)	-0.123 (0.400)	0.0328 (0.491)
Population	0.0622 (0.0411)	0.112 (0.0781)	0.160 (0.111)	0.204 (0.136)
Unemployment	18321.5 (29950.2)	51776.5 (61424.8)	73631.0 (96771.9)	43021.7 (121116.1)
Rate	-356644.4*** (95892.7)	-487860.2*** (147664.1)	-361522.6** (155465.5)	-96834.7 (139253.0)
Income	6.222** (2.467)	12.69** (6.123)	20.15* (11.69)	24.53 (14.88)
Constant	-8593.5 (11953.5)	-43826.4 (27157.7)	-96714.0** (46849.2)	-141060.7** (64797.4)
N (Obs)	*2378*	*2387*	*2389*	*2392*
R^2	*0.969*	*0.932*	*0.903*	*0.883*

*Note: "Forward Price" = $Price_{it+k}$
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Appendix F

(Data tables for k = 15-24 month lead, Fixed Effects Model, Robust Se)

Variable	k = 15 months	k = 18 months	k = 21 months	k = 24 months
	Forward Price*	Forward Price	Forward Price	Forward Price
Sentiment	35694.9* (19694.6)	50769.7** (19695.8)	40481.6* (23723.8)	24810.3 (23364.6)
Price	0.513*** (0.145)	0.468*** (0.156)	0.485*** (0.146)	0.565*** (0.110)
Rent	81.27** (30.50)	105.2*** (38.24)	104.3** (41.34)	92.62** (40.33)
Inventory	0.143 (0.538)	0.180 (0.546)	0.241 (0.506)	0.220 (0.430)
Population	0.162 (0.153)	0.113 (0.166)	0.0690 (0.169)	-0.0550 (0.183)
Unemployment	52086.7 (136221.3)	72086.6 (153053.4)	116708.2 (150439.1)	111296.3 (144568.3)
Rate	-104764.3 (145319.2)	-373244.1** (162990.2)	-477934.8*** (150987.1)	-494019.2*** (146039.9)
Income	24.58 (14.92)	22.80 (14.15)	25.39** (10.92)	26.29*** (6.181)
Constant	-147025.6* (72887.8)	-132907.5 (79091.8)	-126400.3 (75682.1)	-82765.1 (63224.8)
N	*2290*	*2189*	*2088*	*1987*
R^2	*0.859*	*0.847*	*0.855*	*0.868*

*Note: "Forward Price" = $Price_{it+k}$
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Appendix G

(Data tables for k = 27-36 month lead, Fixed Effects Model, Robust Se)

Variable	k = 27 months	k = 30 months	k = 33 months	k = 36 months
	Forward Price*	Forward Price	Forward Price	Forward Price
Sentiment	54286.3 (32196.9)	52153.1 (33199.8)	39836.1 (33448.5)	58594.3 (34818.2)
Price	0.579*** (0.0893)	0.551*** (0.0969)	0.482*** (0.0989)	0.466*** (0.100)
Rent	118.4** (43.33)	135.5*** (48.26)	120.9** (49.47)	95.97** (45.65)
Inventory	0.257 (0.369)	0.405 (0.379)	0.630 (0.440)	0.601 (0.418)
Population	-0.181 (0.214)	-0.228 (0.224)	-0.137 (0.219)	-0.126 (0.236)
Unemployment	76154.8 (136402.6)	-2487.0 (126977.3)	-623.6 (120867.0)	-45079.8 (104964.1)
Rate	-539744.9*** (178823.3)	-477008.5** (185531.4)	-357996.7** (170031.9)	-315227.9** (152707.4)
Income	15.40** (5.964)	9.437 (9.507)	20.39** (8.947)	29.03*** (8.016)
Constant	-24652.0 (64805.4)	2961.5 (69649.7)	-43918.3 (67964.9)	-39452.1 (75515.7)
N	*1885*	*1783*	*1681*	*1579*
R^2	*0.877*	*0.884*	*0.870*	*0.840*

*Note: "Forward Price" = $Price_{it+k}$
Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$