Market sentiment and housing bubbles

1. Introduction

Cycles are common in housing markets across the world. As pointed out in Glaeser et al. (2008), markets with inelastic supply tend to have more and longer cycles. The housing market falls squarely in this category. House prices go through boom and bust cycles that are typically decades long. The ups and downs in the housing market often take the whole economy with them due to the size of the real estate sector in national economy. The downturns in the housing market, or the burst of housing bubbles, is particularly painful, not only because it usually leads to serious economic contractions or even crisis, but also because of our natural tendency to resent losses (i.e., loss aversion). Not surprisingly, housing bubbles is one of the most debated topics in the study of cycles in the housing market.

Predicting housing bubbles is a very challenging undertaking. There is no lack of errors and embarrassment in the literature. For example, not long before the burst of the 2008 housing bubble in the USA, Smith and Smith (2006) analysed the rent and sale price data from ten US urban areas and concluded that "the bubble is not, in fact, a bubble in most of these areas: … buying a home at current market prices still appears to be an attractive long-term investment." Similarly, Himmelberg et al. (2005) pointed out four misconceptions about the assessment of bubbles, and concluded that up to 2004, there was no sign of bubbles in the US housing market. On the other hand, some early warnings of the bubbles, such as the evidence presented in Shiller's (2005) bestseller *Irrational Exuberance*, did not lead to any concrete actions to rein in the overheated housing market. It seems that we never know if it is a bubble until it bursts. What should be done to better understand bubbles?

According to Case and Shiller (2003), bubble "refers to a situation in which excessive public expectations of future price increases cause prices to be temporarily elevated". This makes the understanding of public expectations, or market sentiment, a crucial step in the study of bubbles. In Baker and Wurgler's (2007) definition, investor sentiment is a belief about future cash flows and investment risks that is not justified by the facts at hand. This definition implies not only that there are a good number of people who do not act rationally but also that their irrationality cannot be averaged out. Their biased view about the market situation is systematic and persistent so that a bubble can form from these irrational expectations. Therefore, bubbles by definition are an anomaly under standard economic theory.

Indeed, there is no lack of irrational decision makers in stock markets. Kyle (1985) named them 'noise traders': investors who irrationally act on noise as if it were information that would give them an edge. Delong et al. (1990) demonstrated that noise traders' activities actually discourage rational arbitrageurs from sufficiently betting against them to bring the price in line with fundamentals. As a result, prices can deviate significantly from fundamentals, so much so that bubbles can form. This makes sense when taking into account the limited arbitrage theory by Shleifer and Vishny (1997), by which betting against sentimental investors is costly and risky. There is also a significant social/mental cost to getting over market sentiment with conscious effort. Hassan and Mertens (2011) pointed out that getting over market sentiment is mentally costly in an information-laden, socially connected world. It is well established in the stock market that sentiment does move the market, and getting over market sentiment is not easy. Does this apply to the housing market as well?

It turns out that housing market is an even more fertile 'sentiment' land to feed bubbles. In a comprehensive study of the recent housing bubbles leading to the global financial crisis in 2008, Case et al. (2012) list many culprits that caused the housing market to go out of control: irresponsible lenders who generated mortgage loans recklessly, homebuyers under the influence of money illusion, credit rating agencies who suffered from agency problems, and government who failed to regulate the banking sector. But they argue that homebuyers' expectation and behaviour are the key driver; their long-term expectations were too optimistic. They are typically inexperienced and overwhelmed by the amount of information available. The large stake attached to the purchase of a home also put great psychological pressure on them. This is a recipe for noise traders. As a result, the housing market is far from efficient as assumed by the standard economic theory (Case and Shiller, 1989).

In this chapter we explore different ways of measuring market sentiment and their application in the housing market. We use the UK and the US housing market as cases to investigate whether market sentiment can help us to predict housing price turning points so that we can take actions before a bubble bursts. The case includes both traditional survey-based sentiment indices and an online search volume index as sentiment measurements. It gives us an opportunity to explore the potential of leveraging online information to measure sentiment more reliably, as discussed in the final section of the chapter.

2 The UK and USA housing markets

We have two countries included in this case study: the UK and the USA. Although the political and economic institutions in these two countries are similar, their housing markets are quite different, as can be seen in Figures 7.1 and 7.2. To facilitate comparison, we choose statistics published at OECD's official website for both countries. The indices included are real and nominal house price indices, rent index, and price-to-income ratio. The price-to-income ratio is constructed in a similar way as the one from *The Economist*, with values above 100 indicting over-valued housing market, and vice versa. The price-to-income ratio is on the secondary axis (right-hand side of the figure), and the other indices are on the left-hand-side axis.

In the UK housing market, the nominal house price index is well above its peak right before the global financial crisis. On the other hand, the real house price index has just recovered from the downturn. Rent index does not seem to be affected by housing cycles over the last six decades. It maintains a steady increasing trend throughout the time. The price-to-income ratio, an indicator of housing affordability, shows greater volatility than price indices during peak and trough stages of housing cycles. This is not surprising. When the market is booming credit is abundant. Lower income groups are able to purchase homes, all else being equal. The opposite is true during market downturns. As a result, the price-to-income ratio tends to swing much wider in both directions than house price indices. It is worth noting that the ratio raised above 100 since 2015, and only started to show a sign of dropping in the last quarter of 2018. Because the price-to-income ratio seems to resemble the pattern of house price trend closely, it looks like that the house price in the UK is going to drop.

The US housing market is different from the UK one in several ways. First, the nominal and real house price indices are much more in agreement than is the case in the UK. Both indices are close to their historical peak before the global financial crisis. Second, price-to-income ratio had been predominately over 100, the cut-off point between over-valued and under-valued market, before the global financial crisis. However, it dropped steadily after 2006, and only went above the 100 marker in the last quarter of 2015. Unlike the UK housing market, the US market shows no sign of slowing down. All four indicators suggest an upward trend.

We choose these two countries to investigate whether the size and the structure of market affects the relationship between market sentiment and housing cycles. The USA is a much larger and complex country. To put things in perspective, the US population size is almost five times that of the UK, and the total land area of the USA is more than 15 times that of the UK. In other words, the UK is similar to one of the large states in the USA, such as California. Moreover, London basically dominates the rest of the UK in economic development and beyond. On the other hand, there are multiple similar sized cities in the US and consequently none of them can be as influential on the national housing market as London is in the UK. The USA is a multi-centric country, while the UK is monocentric. Will the power of sentiment be stronger in a smaller and more tightly knitted market like the UK? This is the question that we are going to answer with the data discussed in the next section.

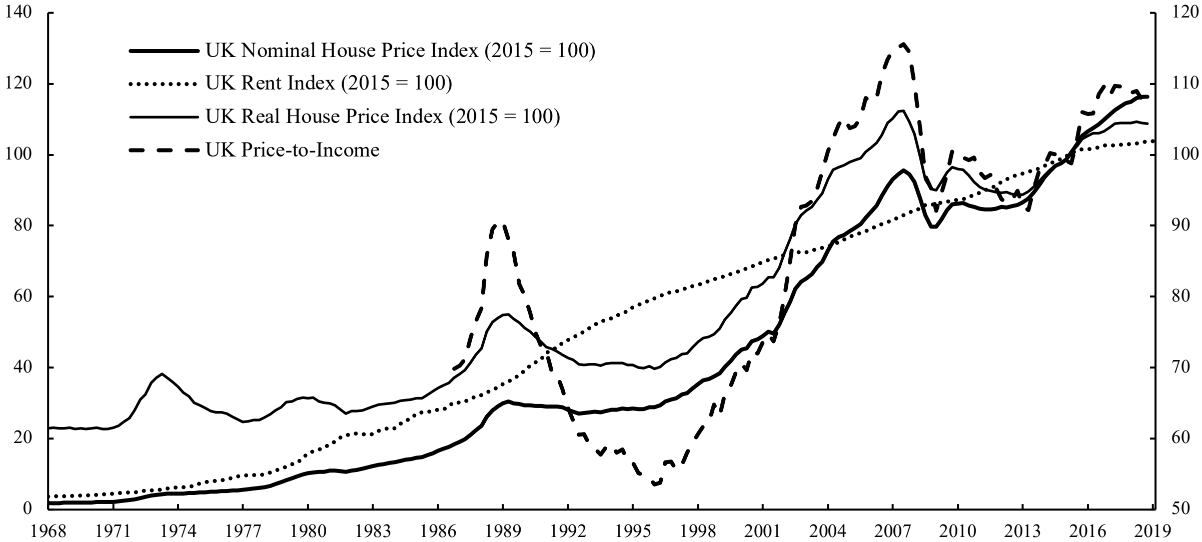

Figure 7.1: The UK Housing Market (1968 – 2019)

Source: OECD (https://data.oecd.org/).

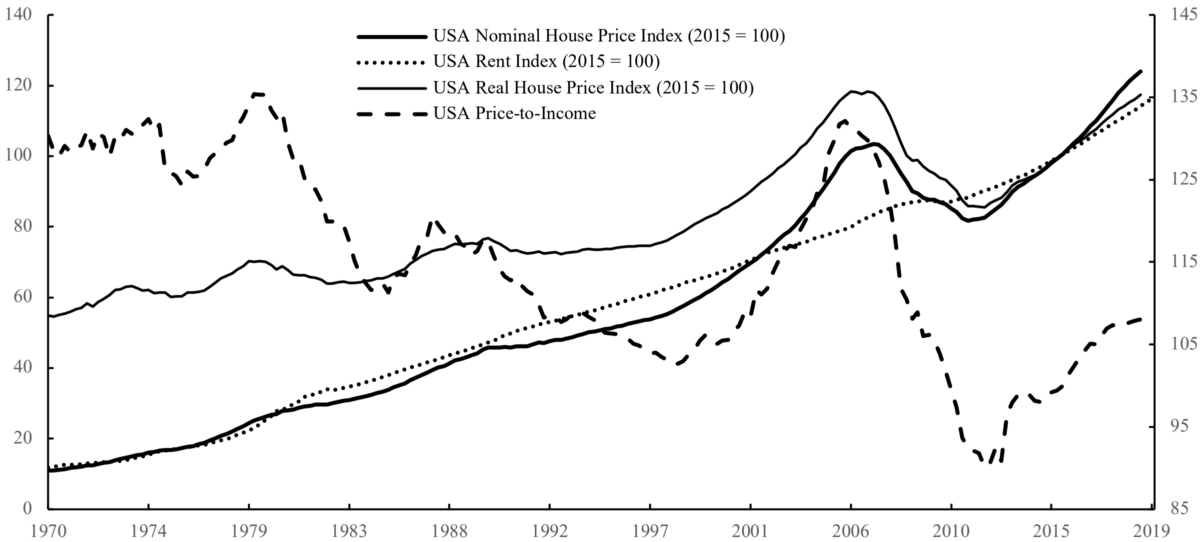

Figure 7.2: The USA Housing Market (1970 – 2019)

Source: OECD (https://data.oecd.org/).

3 Case data

The study of sentiment requires macro-level, or aggregated data. This is because sentiment, by definition, is a market level indicator. Individual's view about the market's position and direction is useful. However, what really matters is the collective view of all individuals involved. In an ideal world, one would have information from every resident in a country about their opinion about the housing market. The individual level data can then be aggregated for form a market-wide measurement of sentiment. Hence the name 'market sentiment'. In reality, such information is too costly, if not impossible, to collect. A large and representative enough sample is often what we can afford. To make the task even more challenging, housing cycles typically last for decades. Any changes in data collection or index calculation methods can potentially introduce noise into the analysis. Therefore, consistency is crucial for sentiment measurement. If collecting quality data in one year is difficult already, drawing such a sample on a regular basis to construct a market sentiment index is a very demanding task. The measurement of market sentiment went through three stages to address this challenge.

The first stage is the direct measurement of investor sentiment via questionnaire surveys. Case and Shiller's homebuyer survey since 1988 is one of the most well designed and implemented studies in this category. They have sent around 2,000 questionnaires to homebuyers in two hot (Los Angeles and San Francisco), one cold (Boston) and one stable market (Milwaukee) in USA since 1988, and annually from 2003 to 2014. Respondents were asked for their view about house prices: have they been 'rising rapidly' or 'falling rapidly' in the last 12 months? How much of a change did they expect there to be in the value of their home over the next 12 months? Answers to these questions are combined to gauge investor sentiment.

The response rate in the first year was impressive – 43.6%. However, this is because the questionnaires were sent with a letter hand signed by both Case and Shiller, who are well known and respected as the creator of the S&P/Case–Shiller US National Home Price Index. They also sent a postcard and a second email to nudge non-respondents to participate. Response rate dropped steadily in the following years, and the average response rate was 20% between 2002 and 2014. The limitation of this direct measurement method is obvious. First, sample size is small and sample representativeness is difficult to achieve. Case and Shiller had to choose four representative cities first, and then distribute questionnaires to randomly selected eligible residents in those cities (i.e., recent homebuyers). Yet, their sample size for each year is around 500, a fraction of homebuyers in the whole country. There is always a risk of misrepresenting the population. Second, even if the sentiment indicator leads housing prices, it can be used to make one-year-ahead forecasting at best. If the housing market responds to homebuyer sentiment within a year, researchers will not be able to forecast crucial turning points in time. This significantly limits the application of this approach in today's fast-moving investment environment.

In an effort to obtain more timely measurement of market sentiment, researchers turned to stock market data, which marks the second stage of sentiment measurement. Listed companies are legally bonded to collect and publish financial information on a regular basis; performance of these companies' stock is a good reflection of general economy and political environment. As a result, stock market information is available at a much higher frequency (e.g., daily vs. annually), with great consistency in terms of format and content. As summarized by Baker and Wurgler (2007), a wide range of stock market indicators, such as mutual fund flows, dividend premium, close-end fund discount, Initial public offering (IPO) first day returns and stock trading volume, have been used to measure stock market sentiment. They also created a sentiment index by combining multiple stock market indicators, and the index method has been well received in the literature.

These indirect measures of sentiment, be they individual indicators or indices that combine multiple indicators, are proxies. Investors' view about market trend is only inferred from these measurements, not directly quantified by using survey data. For example, when close-end fund discount is increasing, one can imply that retail investors are bearish. Although there is empirical evidence to support this theory, the link between the discount and investor sentiment has never been directly tested. It is still largely a 'black box', which sometimes leads to wrong policy suggestions. For example, historically researchers believe that it is retail investors who drive market sentiment, because they are less experienced and hence more irrational than institutional investors. This assumption is instrumental in sentiment literature, because it determines whom to ask in the survey methods, and whose transaction activities should be looked into. However, a recent study by DeVault et al. (2019) found evidence that sentiment index captured the demand shock of institutional investors, instead of individual investors. This demonstrates the limitations of indirect measurements of sentiment. We miss the real driver of sentiment.

Fortunately, with the advance in technology, we are able to directly measure investor sentiment in a much more efficient manner. We are now in the third stage of sentiment measurement: the internet-based sentiment indicators. Leveraging the vast amount of information from online users, researchers harvest data from newspapers websites, social media websites and online search engines to construct sentiment indices. Textual analysis is the commonly used method when comments and posts by online newspaper readers or social media users are analysed. For example, Chen et al. (2014) used this method to analyse the proportion of negative comments on articles posted on a professional website serving stock investors; Tetlock (2007) found that media pessimism based on online content in the *Wall Street Journal* columns predicts price drop followed by a reversal to fundamentals; Renault (2017) measured market sentiment by analysing content in a social media platform, StockTwits; Garcia (2013) also found that news content (not necessarily negative) predicted stock returns during recessions.

Instead of looking at what people said in these online platforms, another way of gauging sentiment is to examine what they searched. Seth Stephens-Davidowitz (2017) put it succinctly in his popular book: Everybody Lies! Human beings are social animals. We are programmed to be 'on the team'. This is why we are very good at giving the politically correct or socially appropriate answers. Or, in other words, we lie. Drawing on his work experience at Google, Seth Stephens-Davidowitz demonstrated how Google search data can reveal our real preferences on a wide range of interesting topics, such as elections, racial and gender discrimination, and sex. This is because when we are searching answers for our problems or desires behind closed doors, we are true to ourselves, and the search words capture more than what we are willing or able to say openly.

This makes online search information a powerful tool to measure sentiment. It is directly from individuals, and the information is more likely to be a true reflection of what people think. Online comments or social media website content analysis comes from individuals directly as well, but to a certain extent it still shares the same concern as with survey methods – what people say there is not necessarily what they really think.

In 2006, Google launched Google Trends, a website that provides and analyses information on search queries in Google searches. Google Trends generate search volume index (SVI) for a keyword or a combination of several search words. The index is scaled such that the numbers are relative measurements of search volume within the specified period. For example, a value of 100 is the peak popularity for the search word; a value of 50 means that the search word is half as popular. As early as in 2009, two analysts from Google demonstrated how to use Google Trends to predict retail, automotive and home sales in the USA, as well as tourist arrivals in Hong Kong (Choi and Varian, 2009, 2012). Applications in stock markets are quick to follow (Da et al., 2011, 2015). Researchers found that SVI – based sentiment indicators are good predictors in housing markets in the USA (Choi and Varian, 2009; Dietzel, 2016; Hohenstatt et al., 2011), the UK (Hohenstatt and Kaesbauer, 2014), Netherlands (van Veldhuizen et al., 2016), and India (Venkataraman et al., 2018) as well.

The discussions above clearly demonstrate the benefit of using direct measurement of sentiment. Correspondingly, we use direct measurement of market sentiment based on survey and Google searches. We will check whether these sentiment indices are helpful in predicting house prices and rents. For survey-based sentiment indices, we obtained two sentiment measurements from the OECD's database: Business Confidence Index (BCI) and Consumer Confidence Index (CCI).

BCI is based on business tendency surveys that seek enterprises' assessment of production, orders and stocks, as well as their current position and expectations for the immediate future. For example, a manager from a construction company will be asked about how the building activity of her firm has changed over the past three months, and how she expects the building activity to change over the next three months; financial firms will answer similar questions but with regard to their profitability and capital expenditure. BCI is then calculated by aggregating answers from individual firms across several main sectors such as manufacturing, construction, retail trades, and financial services. Values above 100 indicate that economic conditions are better than normal (or the long-term average), and vice versa.

CCI is calculated in a similar way, except that it is based on responses in the consumer tendency surveys where households report their plans for major purchases and their economic situation, both currently and their expectations for the immediate future. The survey questionnaire contains questions regarding respondents' past-12 months' assessment and 12-months' ahead forecast of their own financial situation, general economic situation in the country, consumer prices, and unemployment. The survey also has questions about respondents' plan to make long-term financial commitments, such as buying a house or a car in the next 12 months, and whether they plan to save more or less. These questions do not ask for respondents' assessment of the current economic condition, but are obviously good indicators of consumer sentiment. A list of all questions included in the business and consumer tendency surveys can be found at OECD's website.

We also generated a Google search volume index from the Google Trends website for the search phrase 'mortgage loan' for both the UK and the USA. The index is available from 2004 onwards. Therefore, to include this sentiment measurement in this case study, we limit the sample period to 2004 to 2019. Note that there are many other candidates besides 'mortgage loan' that can be used to generate the SVI. For example, one can use 'buy home', 'buy house', or a combination of 'buy home' and 'mortgage loan'. The choice of search words for SVI generation is a very important topic, but beyond the scope of this case study. We use 'mortgage loan' because it stands a good chance of capturing the market sentiment related to both the demand of housing and the availability of credit. The focus of this case study is to demonstrate the differences between sentiment indices generated from traditional survey and online search platforms, instead of the difference of SVIs generated by using different search terms. We use 'mortgage loan' in this study mainly for demonstration purposes. Readers who are interested in exploring alternative search words may feel free to experiment at Google Trends website on their own.

Table 7.1: Variable definition and descriptive statistics

Variable	Description	Remark
PI	Index of real house price	2010 = 100
RI	Index of rental price	2010 = 100
GDPR	Growth rate of quarterly GDP	Unit: %
INC	Per Capita Real Disposable Personal Income	Seasonally Adjusted Annual Rate
DST	USA: Housing Starts (New Privately Owned Housing) UK: Permanent dwellings started	Thousands of Units, Seasonally Adjusted
IR	Long-term interest rates	Unit: %
BCI	The business confidence index (BCI)	Amplitude adjusted, Long-term average = 100
CCI	The consumer confidence index (CCI)	Amplitude adjusted, Long-term average = 100
DESVI	Google search volume index (SVI) by using 'mortgage loan' as the search word	Seasonally adjusted, Historical peak = 100

Note: All variables are obtained from OECD*, except for INC, DST, and DESVI. The data source for INC and DST is the National Statistics Office for the UK market, and the Federal Reserve Economic Database for the US market. DESVI is obtained from Google Trends website.*

4 Case questions and discussions

The objective of this case study is to investigate whether market sentiment can help predicting market turning points, and if yes, which type of sentiment index (survey- or search-volume-based) is more reliable. We use real house price index and rent index as the dependent variables in this analysis. We also include some important macroeconomic indicators such as GDP growth rate, per capita income, interest rates, and new housing stock in the model in order to separate the net effect of market sentiment. Apparently most, if not all, of these variables are highly correlated among each other, and are not stationary over time. Consequently, we use time series analysis methods such as the Vector Autoregressive model to study the dynamic relationship among them.

Figure 7.3 and Figure 7.4 give time series charts of house price and rent indices, and sentiment indices for the UK and the USA respectively. In both countries, the original SVIs show significant seasonal patterns. Search volume drops notably in the last quarter of the year, and then gradually reach its peak in the fall. There are two possible reasons behind this pattern. December is the holiday season. Homebuyers may deliberately avoid this season, so their Christmas and New Year plans will not be disrupted by home moving. Secondly, and more interestingly, a psychological factor called Seasonal Affective Disorder (SAD) might be at play. SAD is defined as a condition characterized by recurrent depressive episodes that occur annually, usually during fall and winter when daytime is short (Rosenthal et al., 1984). The proportion of population suffering from SAD is not small – as much as 50% of the global population is affected by SAD at certain stage of their life. As a result, SAD affects our behaviours in many areas, including in the decisions of buying or selling a house. Kaplanski and Levy (2012) find that latitude and changes in the number of daylight hours affect real estate prices in the USA, the UK, and Australia. Winter Blues is rather persistent and consistent in the three large real estate markets examined. The SVI indices that we generated in this case study obviously support their conclusion. The seasonal pattern is consistent with what was found in the literature. Moreover, the seasonal pattern is stronger in the UK than in the USA, which means latitude does play a significant role in affecting people's mood.

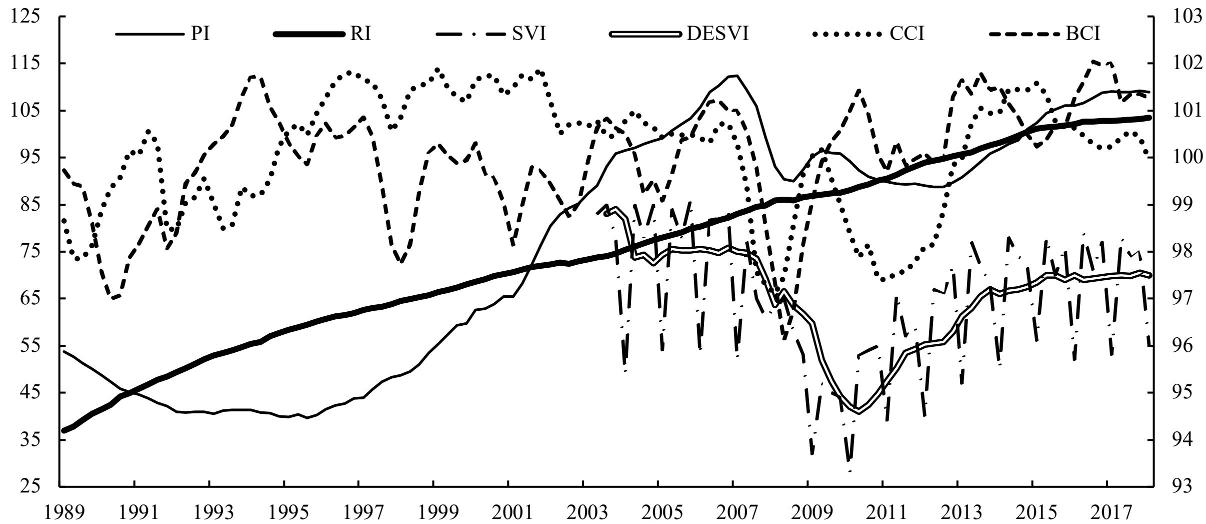

Figure 7.3: UK house pricing market and sentiment indices

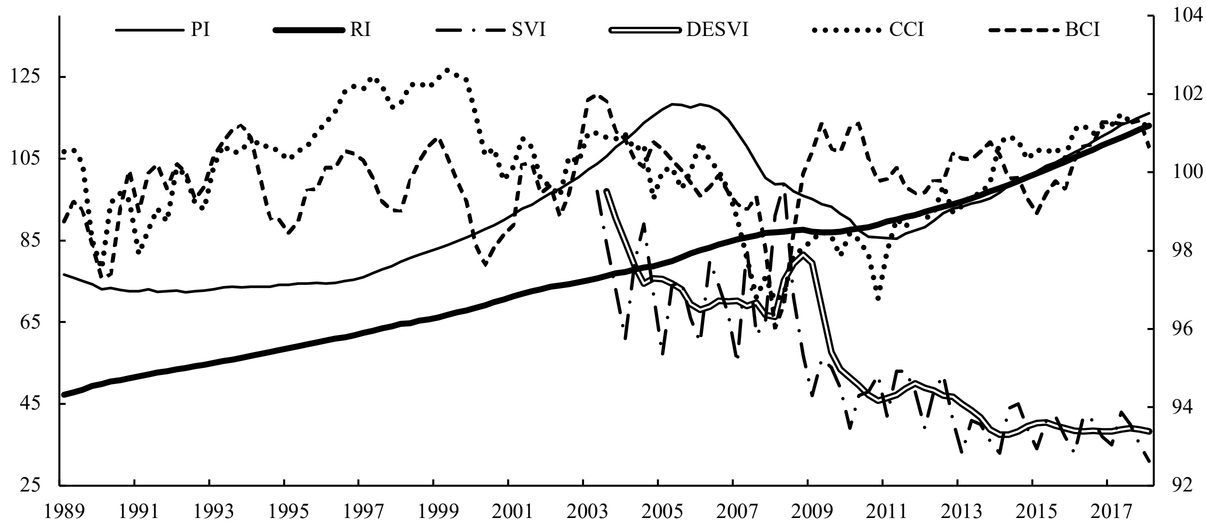

Figure 7.4: US house pricing market and sentiment indices

SVI is a good candidate for sentiment measurement, because it does capture people's emotion objectively. When we are not happy, we simply do not search for houses, and the search volume shows even if we do not realise the onset of the Winter Blues. However, the seasonal variation in SVI within a year does not help our analysis, because we are studying long-term relationships. This is why we used seasonally adjusted versions of all variables when applicable. In Figure 7.3 and Figure 7.4, we presented both the original SVI series and the seasonally adjusted (i.e., keeping the long-term trend only) to illustrate the reason to de-seasonalise time series when studying long-term relationships. In this case study we used a straightforward four-quarter moving average to remove the seasonal variation. From Table 7.2 where descriptive statistics of variables are presented, the de-seasonalisation process does not change the means and standard deviation of the series significantly. The procedure removes consistent within-year variations while keeping long-term trend. The patterns show in Figures 7.2 and 7.3 clearly show such an effect.

Table 7.2: Descriptive Statistics

	UK				USA			
	Mean	**Std Dev**	**Min**	**Max**	**Mean**	**Std Dev**	**Min**	**Max**
PI	76.02	25.84	39.65	112.39	91.42	14.83	72.32	118.29
RI	74.93	18.62	36.94	103.47	76.71	17.94	47.25	113.07
GDPR	0.49	0.59	-2.17	1.93	0.61	0.59	-2.16	1.83
IR	5.10	2.74	0.84	12.32	4.62	1.92	1.56	8.70
INC	4,168	602	3,015	4,972	35,132	5,161	26,962	44,831
DST	45,590	8,721	22,270	64,710	1,297	406	505	2,151
CCI	100.08	1.29	97.11	101.90	100.06	1.40	96.73	102.61
BCI	99.99	1.22	96.11	102.04	99.84	1.04	96.02	101.99
SVI	64.92	14.74	28.00	86.00	54.37	17.93	31.00	99.00
DESVI	65.29	10.74	41.00	84.00	56.16	16.88	37.50	97.00

We estimate vector autoregressive (VAR) models by using house price index and rent index as the dependent variables. Because house price and rent are likely to affect each other, and both of them are likely to be affected by the macroeconomic and sentiment variables considered in this study, it is necessary to estimate the house price and rent equations jointly. An important requirement in VAR estimation is that all variables should be stationary, which means these variables should at least have constant means over the sampling period studied. Unfortunately, none of our variables meets this requirement. A standard way of checking stationarity is the Dicky–Fuller test that is widely available in most statistical software packages. However, after taking the first difference (e.g., the first difference of PI_t is $\Delta PI_t = PI_t - PI_{t-1}$), all series passed the stationary test. When first differenced time series are used in a regression model, the coefficient should be interpreted as the average one-period change in Y according to one-period change in X. This transformation serves our purposes well in this study. Because we are interested in predicting market turning points, quarterly changes in house prices and rent are the focus. Therefore, we use the first difference of all variables in the VAR models estimated below. The results are presented in Table 7.3 and Table 7.4 for the UK and the US markets respectively.

In each table statistical significance is denoted by asterisks. Only variables with coefficients that are significant at 10% or less are considered in the discussions below, because the statistically insignificant ones are essentially zero at the 10% level. For each country we estimated six models. The first three models are baseline models with one sentiment index as the independent or exogenous variable only; the last three models are full models that have macroeconomic indicators as additional independent variables. The difference between baseline and full models can help us to understand how much these sentiment indices share information in common with the macroeconomic factors considered. For example, if one sentiment index is correlated with one or several macroeconomic factors, the sentiment index may be significant in the baseline model and insignificant in the corresponding full model. Such a sentiment index might not be a good measurement of market sentiment, as it does not contain much unique information that has not been captured by those macroeconomic factors.

Each model has separate coefficient estimates for house price and rent model, indicated by the name of the dependent variable in each model in the top row of the tables. Although the outputs for the rent model are insignificant across the board, rent and house price do affect each other, as indicated by the significant coefficient estimates of the lagged term of PI and RI in most of the models. We keep the rent index and the VAR framework in this study for the purpose of demonstrating how the two market segments react differently to the same set of predictors.

4.1 Can market sentiment help predict market turning points?

We include a one-period lag term of the sentiment indices in all models. These lagged terms are used to check if any of these sentiment indices can give us an early warning. Specifically, if the lagged term of a sentiment index is statistically significant, it means that the index is able to predict the quarterly change of house price or rent index one quarter ahead of time. We also include the contemporary term of these indicators because it is entirely possible that the market can react to these indicators within a quarter. However, such relationship is not of much practical value to us, because our objective is to find an indicator that can help us to predict market turning points or give us early warnings for such events.

In the UK market, all three sentiment indices have significant contemporary terms with expected signs in the baseline models (i.e., Models 1 through 3). CCI, the consumer confidence index by OECD, has its one-period-lagged term significant as well. After macroeconomic indicators are added to these models, CCI's contemporary term is not significant anymore. However, its lagged term is still an important determinant of house price index. The results for Google Trend search index (DESVI) and BCI the business confidence index are qualitatively identical between the baseline and full models. All significant coefficients have the positive sign as expected. This indicates that the house price index moves in the same direction as market sentiment. Market sentiment is helpful in prediction of house price turning points.

The results from the US market tell a different story. First, Google search volume index does not help in predicting quarterly changes in house prices at all. Second, although CCI's lagged term is positive and significant in its full model (Model 5), the effect size is much smaller than that in the UK, and is significant at 10% level only. Finally, BCI's coefficient estimate in Model 6 is counterintuitive because it is negative and significant at the 1% level. It is even marginally significant, with a negative sign too, in the corresponding rent index model. We can only interpret the unexpected results of BIC in Model 6 as an indication of the unsuitability of BIC as a sentiment indicator for housing studies. The information collected in the Business Tendency Survey may not be relevant enough to the housing market that we are studying. Overall, the ability of sentiment indices to predict house price turning points appears to be weaker in the US market.

4.2 Which type of sentiment index (survey- or search-volume-based) is more reliable?

It depends. Our results indicate that the answer to this question depends on the suitability of a sensitive index for the study area. Google search volume index does a reasonably good job in the UK market, which is small and more centrally controlled. For larger geographic regions such as the USA, it will be much more challenging to identify the search terms to generate the SVIs. Web users' habit of searching for information may vary across the study region, and there is no evidence that the variations can be averaged out. Instead of discrediting Google SVIs as reliable sentiment measurements in housing studies, we would rather argue that the UK results show the potential of this helpful tool, and the US results highlight the potential pitfalls of using it in certain types of housing studies. Further empirical studies are needed to help us to master the use of SVIs or similar online tools in our field.

The same observation can be made for survey-based sentiment measurement as well. Our results show that CCI consistently outperformed DESVI and BCI in both countries. It has expected sign in all models. In addition, CCI is able to give earlier warnings for market turning points (i.e., its lagged term is significant). This is likely due to the way in which CCI is constructed. It is based on the Consumer Tendency Survey, where the consumer's view about the economic condition is directly measured based on their answers to survey questions. The sample is representative for the housing market that we are studying in this case. The BCI, on the other hand, is based on a Business Tendency Survey where questionnaires were completed by managers of randomly selected firms listed as members of chambers of commerce or employers' associations. It might not be a good indictor to capture the sentiment among home buyers and sellers, at least not in the USA market.

Table 7.3: VAR estimations (UK)

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	PI	RI	PI	RI	PI	RI	PI	RI	PI	RI	PI	RI
PI(-1)	0.798***	0.053	0.760***	0.085	0.745***	0.077	0.764***	0.076	0.741***	0.090	0.724***	0.087
RI(-1)	-0.148**	0.028	-0.162**	-0.047	-0.115*	-0.014	-0.109*	0.064	-0.131**	-0.036	-0.085	-0.014
GDPR							0.665***	0.465	0.248	0.133	0.312*	0.157
INC							0.009***	0.002	0.006***	0.0004	0.008***	0.001
IR							0.294	-0.223	0.201	-0.113	-0.091	-0.018
DST							0.00003*	-0.00009**	0.00002	-0.00004**	0.00002	-0.00004*
DESVI	0.201***	0.092					0.170***	0.115				
DESVI(-1)	-0.045	-0.098					-0.041	-0.127				
CCI			0.415*	0.120					0.284	0.169		
CCI(-1)			0.752***	0.042					0.637***	0.069		
BCI					0.653***	-0.064					0.58***	-0.047
BCI(-1)					-0.025	-0.247					0.064	-0.199
CONSTANT	0.111	0.293	0.204**	0.508***	0.195*	0.503***	0.049	0.227	0.119	0.481***	0.058	0.485***

*Note: ***: p-value < 0.01. **: p-value < 0.05. *: p-value < 0.10.*

Table 7.4: VAR estimations (USA)

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	PI	RI	PI	RI	PI	RI	PI	RI	PI	RI	PI	RI
PI(-1)	0.874***	-0.063	0.887***	-0.068	0.918***	-0.066	1.027***	-0.119	0.976***	-0.105	0.985***	-0.103
RI(-1)	-0.629***	0.026	-0.629***	0.046	-0.636***	0.050	-0.684***	0.18**	-0.663***	0.144**	-0.667***	0.142**
GDPR							-0.050	-0.192	0.007	-0.074	0.073	-0.069
INC							-0.001***	-0.002***	-0.001***	-0.002***	-0.001***	-0.002***
IR							-1.204***	-0.522	-0.65***	-0.101	-0.231	0.169
DST							0.00078	-0.00069	0.00029	-0.00009	0.00022	-0.00047
DESVI	0.079	0.021					0.063	0.038				
DESVI(-1)	-0.079	-0.002					-0.033	0.001				
CCI			0.205	0.017					0.066	-0.148		
CCI(-1)			-0.063	-0.221					0.222*	0.100		
BCI					0.181	0.067					0.107	0.071
BCI(-1)					-0.494***	-0.290					-0.371***	-0.246*
CONSTANT	0.260	0.473**	0.373***	0.525***	0.373***	0.525***	0.504***	0.860***	0.519***	0.817***	0.536***	0.825***

*Note: ***: p-value < 0.01. **: p-value < 0.05. *: p-value < 0.10.*

5 Summary

In this chapter we first introduce different theories regarding the role of sentiment in the formation of housing cycles. A case study of the UK and the USA housing markets is used to demonstrate how to use survey and online search volume based sentiment indices to predict housing price turning points. The result of our case analysis suggests that market sentiment is helpful in predicting house price turning points, and the choice of sentiment indicator is an empirical issue. All three sentiment indices considered are useful at certain circumstances, yet none of them can produce robust results across the board. This is particularly true for the Google search volume index, which is completely irrelevant in the US housing market but very helpful in the UK study. This begs further empirical investigations on the best practice of using this efficient tool in housing studies.

To move forward in this direction, let us take a moment to review an important theory underlying the study of sentiment, the attention theory by Barber and Odean (2008). In the stock market, individual investors are net buyers of attention-grabbing stocks. Applying this theory in the housing market, it implies that buyers' attention can be divided and manipulated to a much greater extent than sellers. They are the ones who can be led by advertisement and news reports, and can bid house prices up to a level that is well above fundamental values. There are many houses to choose from, and the information can be overwhelming, enough to trigger heuristic thinking and behavioural biases. Sellers, on the other hand, have only one house to offer and their attention is much less divided when compared to the buyer. This theory has been confirmed in the stock market by Da et al. (2011). This indicates that we should focus more on homebuyers when measuring market sentiment because their views are more likely to be influenced by noise.

Therefore, to find a good measurement of sentiment, one should start with well-tested attention measurements first, and then choose the one that can reliably and effectively identify buyer and seller attention. In this sense, Google SVI is a starting point. Da et al. (2011) used Google SVI as a direct measurement of investor attention, or sentiment. Google SVI captures attentions from individual investors primarily. News and media reports measure information available, not how much attention they grabbed from investors. They could be poor predictors of attention. Google SVI measures revealed attention. Askitas (2016) and Askitas and Zimmermann (2015) constructed a buyer vs. seller ratio index based on Google SVI. This is a good way of measuring sentiment. Basically, if there is more 'buy' search than 'sell' search, the market is booming, and vice versa. In a similar vein, as long as the role of buyer and seller can be clearly and reliably identified, social media information can be used to compose a buyer vs. seller ratio index as well. There is great potential of exploring online big data along this direction.

The next step is to do short-term nowcast, not a one-month forecast. Housing market data typically lag behind the market for months. For example, Standard & Poor's/Case-Shiller® US National Home Price Index has a two-month lag; The Hong Kong Centa City Property Leading Index lags the market by a week. Google SVI and social media big data, on the other hand, are almost 'real time' – they have a time lag of zero. This is ideal for short-term forecasting, or nowcast. This will significantly enhance our ability to detect a bubble in a timely way, and subsequently take necessary actions promptly. With advances in technology, this is possible.

References

1 Askitas, N., 2016. Trend-spotting in the housing market. *Cityscape 18*, 165–177.

2 Askitas, N., Zimmermann, K.F., 2015. The internet as a data source for advancement in social sciences. *International Journal of Manpower 36*, 2–12.

3 Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. *Journal of Economic Perspectives 21*, 129–151.

4 Barber, B.M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*, 785–818.

5 Case, K.E., Shiller, R.J., 1989. The efficiency of the market for single-family homes. *American Economic Review 79*, 125–137.

6 Case, K.E., Shiller, R.J., 2003. Is there a bubble in the housing market? *Brookings Papers on Economic Activit*y, 299–362.

7 Case, K.E., Shiller, R.J., Thompson, A.K., 2012. What have they been thinking? Homebuyer behavior in hot and cold markets. *Brookings Papers on Economic Activity*, 265–315.

8 Chen, H.L., De, P., Hu, Y., Hwang, B.H., 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies 27*, 1367–1403.

9 Choi, H.Y., Varian, H., 2009. Predicting the present with Google Trends. Working Paper, Google Inc.

10 Choi, H.Y., Varian, H., 2012. Predicting the present with Google trends. *Economic Record 88*, 2–9.

11 Da, Z., Engelberg, J., Gao, P.J., 2011. In search of attention. *Journal of Finance 66*, 1461–1499.

12 Da, Z., Engelberg, J., Gao, P.J., 2015. The sum of all fears investor sentiment and asset prices. *Review of Financial Studies 28*, 1–32.

13 Delong, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial-markets. *Journal of Political Economy 98*, 703–738.

14 DeVault, L., Sias, R., Starks, L., 2019. Sentiment metrics and investor demand. *Journal of Finance 74*, 985–1024.

15 Dietzel, M.A., 2016. Sentiment-based predictions of housing market turning points with Google trends. *International Journal of Housing Markets and Analysis 9*, 108–136.

16 Garcia, D., 2013. Sentiment during recessions. *Journal of Finance 68*, 1267–1300.

17 Glaeser, E.L., Gyourko, J., Saiz, A., 2008. Housing supply and housing bubbles. *Journal of Urban Economics 64*, 198–217.

18 Hassan, T.A., Mertens, T.M., 2011. Market sentiment: A tragedy of the commons. *American Economic Review 101*, 402–405.

19 Himmelberg, C., Mayer, C., Sinai, T., 2005. Assessing high house prices: Bubbles, fundamentals and misperceptions. *Journal of Economic Perspectives 19*, 67–92.

20 Hohenstatt, R., Kaesbauer, M., 2014. GECO's weather forecast for the UK housing market: To what extent can we rely on Google econometrics? *Journal of Real Estate Research 36*, 253–281.

21 Hohenstatt, R., Käsbauer, M., Schäfers, W., 2011. "Geco" and its potential for real estate research: Evidence from the US housing market. *Journal of Real Estate Research 33*, 471–506.

22 Kaplanski, G., Levy, H., 2012. Real estate prices: An international study of seasonality's sentiment effect. *Journal of Empirical Finance 19*, 123–146.

23 Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica 53*, 1315–1335.

24 Renault, T., 2017. Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking & Finance 84*, 25–40.

25 Rosenthal, N.E., Sack, D.A., Gillin, J.C., Lewy, A.J., Goodwin, F.K., Davenport, Y., Mueller, P.S., Newsome, D.A., Wehr, T.A., 1984. Seasonal Affective Disorder: A description of the syndrome and preliminary findings with light therapy. *Archives of General Psychiatry 41*, 72–80.

Stephens-Davidowitz, S., and S. Pinker (2017). Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are. New York, Dey Street Books.

26 Shiller, R.J., 2005. *Irrational Exuberance*, 2nd ed. Currency/Doubleday, New York.

27 Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. *Journal of Finance 52*, 35–55.

28 Smith, M.H., Smith, G., 2006. Bubble, bubble, where's the housing bubble? *Brookings Papers on Economic Activity*, 1–67.

29 Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance 62*, 1139–1168.

30 van Veldhuizen, S., Vogt, B., Voogt, B., 2016. Internet searches and transactions on the Dutch housing market. *Applied Economics Letters 23*, 1321–1324.

31 Venkataraman, M., Panchapagesan, V., Jalan, E., 2018. Does internet search intensity predict house prices in emerging markets? A case of India. *Property Management 36*, 103–118.