SPATIAL EFFECT OF MARKET SENTIMENT ON HOUSING PRICE: EVIDENCE FROM SOCIAL MEDIA DATA IN CHINA

Junjie LI[1,*], Yu WANG[1], Chunlu LIU[2]

[1] *School of Management Engineering, Zhengzhou University, Zhengzhou, China*
[2] *School of Architecture and Built Environment, Deakin University, Geelong, Australia*

Abstract. Market sentiment has become more easily spread between cities through social media. This study investigates the spatial effect of market sentiment on housing price in a social media environment. In order to extract home-buyer sentiment from social media, we use text sentiment analysis techniques and build a novel housing market sentiment index. A spatial econometric model of housing price volatility is subsequently constructed and the housing market sentiment index is included as an independent variable in the model. Using panel data from 30 large and medium-sized cities in China for 20 quarters from 2016 to 2020, the spatial effect of market sentiment on housing price is empirically analyzed by calculating direct and indirect effects. The results show that market sentiment had a significant positive effect on housing prices in the local and neighboring cities over the research period. However, the impact of market sentiment on housing price was heterogeneous in terms of geographical region; the direct effect was stronger in the eastern region than in the central and western regions, and the indirect effect was significant only in the eastern region. These findings can provide references for government to formulate housing market regulation policies and measures based on market sentiment.

Keywords: housing price, market sentiment, sentiment analysis, social media, spatial Durbin model, spatial effect.

Introduction

Fluctuation in housing prices touches the nerves of policymakers, ordinary residents, and financial institutions (Dietzel, 2015) and has become an issue of general social concern. Any imbalance or contraction in the housing market can lead to financial instability, which in turn poses a macroeconomic threat (Lee & Park, 2018). Therefore, how to stabilize the housing market has become a top priority for the development of the real estate industry. So what are the forces causing high volatility in housing prices? Previous studies have shown that housing prices are influenced by economic fundamentals such as GDP, income level, population, and interest rates (Case et al., 2012). However, these standard economic explanations are difficult to reconcile with high volatility in housing prices over a corresponding period (Alkay et al., 2018). In a standard macroeconomic model with fully rational expectations, high volatility in housing prices is difficult to generate (Granziera & Kozicki, 2015).

In addition to economic fundamentals, a growing number of scholars have tried to explain the volatility of housing prices in terms of consumer psychology and irrational behavior. Case and Shiller (2003) argued that the "animal spirits" of investors play an important role in dramatic volatility in housing prices. These unexplained "irrational" factors, driven by "noise traders" or "irrational investors", are called market sentiment (Soo, 2013). The theoretical basis for sentiment analysis in the housing market is the noise trader approach, which divides investors into fully rational and imperfectly rational groups (De Long et al., 1990). The latter's investment behavior is thought to be more influenced by sentiment. The housing market is characterized by a high proportion of individual traders, market segmentation, information asymmetry, and lack of short-selling mechanisms, all of which make market participants vulnerable to sentiment-induced pricing errors (Clayton et al., 2009; Hui & Wang, 2014). Market sentiment has been widely accepted as a key driver of housing price volatility, particularly after several property market crises (Lin et al., 2009). Especially during periods of high sentiment, higher expectations of future housing returns lead to an influx of home buyers into the property market

and a rapid increase in transaction volumes (Fischer & Stamos, 2013). This overreaction to price changes may lead to significant upward momentum in housing prices (Case & Shiller, 2003).

With the help of social media (e.g., Twitter, Facebook, Sina Weibo), interactions among home buyers have become faster and easier, and market sentiment may spread and amplify rapidly through the Internet and so influence housing prices to some extent. Recent empirical studies have also suggested that interactions on social media are an important mechanism to explain housing price contagion (DeFusco et al., 2018). For example, using Facebook social media data, Bailey et al. (2018) found that social interactions played an important role in shaping individuals' housing market beliefs and investment behavior, confirming the theory that the spread of sentiment among investors can have an impact on expectations in housing markets (Gao et al., 2020). Bayer et al. (2021) further confirmed that individuals' experiences of housing prices in social networks are a significant source of heterogeneity in their housing market expectations and can influence home purchase decisions, with the resulting shift between optimism and pessimism about future housing price growth leading to housing price volatility. With the communication effect of social media, market sentiment "infection" occurs quickly and spreads continuously, generating "irrational expectations" and further increasing the possibility of rising housing prices (Li et al., 2020). Therefore, an in-depth exploration of the role of market sentiment in housing price volatility in the social media environment is crucial for stabilizing the property market.

Although explorations targeting the relationship between market sentiment and housing price volatility have gradually attracted the attention of scholars, research on market sentiment has still not received as much attention as research on the stock market (Balcilar et al., 2021). In particular, market sentiment in the social media environment is understudied and it is unclear how this plays a role in housing price volatility, despite its evident importance for irrational and volatile housing markets (Hui & Wang, 2014). More importantly, the housing market is not completely independent and, with the popularity of social media, market sentiment infection occurs rapidly and continues to spread across cities (Li et al., 2021), with a possible spatial effect on housing prices. However, this spatial effect has long been overlooked in previous studies of market sentiment as it affects housing prices.

Based on the above analysis, the objective of this research is to examine the spatial effect of market sentiment on housing prices in a social media environment. Three detailed questions need to be addressed: 1) How can we construct a housing market sentiment index from social media data which is a measure of housing market sentiment? 2) How does market sentiment affect housing prices through a spatial effect? and 3) Do differences in geographical location lead to differences in the spatial effect of market sentiment?

For the first question, it is first necessary to identify the subject of market sentiment and which social media platform to choose. The subject of market sentiment in our study is the home-buyer, who buy houses to live in themselves, not the investor, who sell their houses after its value has appreciated. This is due to the fact that our research focuses on representative large and medium-sized cities in China, which are economically developed and attract a large number of young people who leave their hometowns to work in these cities. After working, the next important thing for these young people is to get married, and in the traditional Chinese concept of "a house means a home", buying a house is often a prerequisite for young people to get married, which inevitably generates a huge demand for housing. In addition, there are currently very strict housing purchase restrictions in China, which will lead to a very limited speculative group in the housing market for sure. Thus, the sentiment of the home-buyers sets the tone of market sentiment. Furthermore, the majority of the homebuyer groups in China are young people who commonly use the largest Chinese social media platform – Sina Weibo – to express their views, which is the reason why Sina Weibo was chosen as the source of market sentiment data for the subsequent study in this paper. We then used a web crawler to obtain social media text data related to the housing market, which was then analysed using text sentiment analysis techniques to construct a housing market sentiment index for different cities at different times.

The second and third questions involve the spatial effect and required the use of spatial econometric models that decompose a spatial effect into direct and indirect effects (the indirect effect is also known as the spatial spillover effect). We incorporated the established housing market sentiment index as an independent variable in the spatial econometric model of house price volatility to empirically examine the direct and indirect effects of market sentiment on housing prices from the whole sample and different regions. The answers to these questions can help us further understand the laws of market sentiment affecting housing price and provide a reference for establishing a long-term mechanism for stable and healthy development of the housing market from the perspective of market sentiment management.

1. Literature review

While the role of sentiment in market conditions is easy to understand, the measurement of sentiment in a complex, asymmetrical information economy is not straightforward (Heinig & Nanda, 2018). Relying on direct surveys is a common measure (Case et al., 2012). Lambertini et al. (2013) used data from the University of Michigan Consumer Survey to provide evidence about the importance of consumer sentiment on the dynamics of the U.S. housing market. Ling et al. (2014) investigated survey-based sentiment measures of investor sentiment in private commercial real estate markets and returns. However, in the case of the highly geographical housing market, direct surveys are

difficult to use for scientific measurement of market sentiment in different regions and cities (Marcato & Nanda, 2016). Indirect characteristics of market sentiment with the help of macroeconomic indicators are other common measures of market sentiment (Baker & Wurgler, 2007). Balcilar et al. (2021) analyzed the application of sentiment indicators in the housing market using economic sentiment indicators from the E.U. and the U.S. Marcato and Nanda (2016) tested economic indicators as sentiment indicators to predict housing market returns. Wang and Hui (2017) used principal component analysis to select representative macroeconomic indicators that analyzed the impact of market sentiment on the real estate market in Hong Kong. However, as the macroeconomic indicators selected are usually ex-post data, the constructed sentiment index has a lag. In addition to these indirect proxies and direct surveys, text-based sentiment measurement analysis has recently attracted increasing attention since greatly increased computing power and wide availability of online data have gradually made accessible rich sources of information. The advantages of this approach over traditional sentiment measurement are the large number of sources of available data, the low cost and effort of data collection, and the relative straightforwardness of measurement. Heinig and Nanda (2018) applied textual analysis based on word lists to commercial real estate (CRE) data for London West End to capture the effect of sentiment. Hausler et al. (2018) examined the relationship between news-based sentiment, captured through a machine learning approach, and the US securitised and direct commercial real estate markets. Beracha et al. (2019) captured news-based market sentiment for real estate and its extent through textual analysis. Taking Guangzhou as an example, Gao and Zhao (2018) used text mining method to extract media influence from media texts, and constructed the buyer confidence index, and used it as a proxy variable for home-buyers' sentiment. However, mainstream media tend to carry a certain opinion bias (Ardèvol-Abreu & Gil de Zúñiga, 2017), which may also lead to mainstream media sentiment not necessarily being a true expression of market sentiment. For example, financial media may amplify market sentiment (Ren et al., 2021).

Social media is rapidly replacing non-digital media as a convenient and fast medium for communicating and sharing information, allowing users to interactively express their one-sided views on the housing market. The role of social media in sentiment detection has received increasing attention from researchers (Cerchiello & Nicola, 2018). Social media data have the characteristics of full samples, high frequency, and immediacy, and a sentiment index based on social media data is highly representative and forward-looking (Da et al., 2015). The measurement of market sentiment from social media has received increasing attention from researchers in recent years and has been widely used in financial markets such as stocks (Siikanen et al., 2018). For example, measuring sentiment from Twitter has been used for several purposes including analyzing the impact on share price volatility (Siikanen

et al., 2018), forecasting trends in the Dow Jones Industrial Average (Ranco et al., 2015), assessing the systemic risk of Italian banks (Cerchiello et al., 2017), and performing financial analysis and forecasting (Oliveira et al., 2017). Giannini et al. (2019) used Twitter data to examine how sentiment affected trading and how changes in investor disagreement were tracked. Yu and Yuan (2011) found that the relationship between company sentiment as reflected in social media and company stock returns and volatility was stronger than in traditional news media. Huang et al. (2015) used Sina Weibo as a data source for constructing a market sentiment index to study the relationship between market sentiment and financial markets. Topics and interactions on social media are more likely to stimulate investors' attitudes toward the market and thus influence the volatility of asset prices (Leitch & Sherif, 2017). All of these studies show that social media is a valuable resource and an important source of information to support decision-making. However, existing housing-related research on market sentiment is seriously lacking, most existing studies have relied on indirect indicators such as relatively lagging macroeconomic data, and the use of social media data, which can indicate the public's immediate attitudes toward the latest information, is still uncommon. The impact of this data source on home-buyer sentiment is evident in the modern information society (Cerchiello & Nicola, 2018). Sinyak et al. (2021) found that social media of the real estate market can be considered a valuable and innovative source of market sentiment and can provide real estate researchers and valuers with reliable leading market indicators. This is an interesting context from which to extract market sentiment in the housing market and further investigate its impact on housing prices, which is why this paper chooses to study market sentiment based on the social media environment.

To explore the characteristics of the impact of market sentiment on housing markets, several studies have used traditional measures such as multiple regression and vector autoregression, analyzing the predictive impact of media housing sentiment on housing prices (Soo, 2013), the effect of market sentiment on housing market returns (Ruscheinsky et al., 2018), and the impact of participants' sentiment on housing demand (Hui & Wang, 2014). Most of these methods use time-series data, but as the housing market has strong individual city characteristics, the impact of sentiment indices on housing price needs to be analyzed using panel data. For example, Lam and Hui (2018) used panel data and constructed panel regression models to study the impact of market sentiment on the future returns of Hong Kong residential properties. However, the housing market is spatially correlated, with housing prices spreading across cities (Luo et al., 2007; Wu et al., 2017), but traditional panel regression models ignore the spatial correlation of housing markets and do not take into account the spatial effect of sentiment on housing prices, leading to potentially biased estimates from the resulting models. Spatial econometric models, which incorporate spatial factors into the model by constructing a spatial

weight matrix, have been widely used in studies exploring housing price volatility (Vergos & Zhi, 2018). For example, Liang et al. (2020) used spatial econometric models to study the factors influencing housing prices and their spatial effect, which included direct and indirect effects. Therefore, this paper investigates the direct and indirect effects of market sentiment and each control variable on housing price based on the decomposition effect of the spatial econometric model, and provides insight into the impact of market sentiment on housing price from the perspective of the spatial effect, making up for the shortcomings in previous literature.

2. Methodology

2.1. Development of a housing market sentiment index based on social media data

With its high degree of openness and rich expression of ideas and opinions, Weibo provides a large amount of content-rich textual data for capturing market sentiment. This study uses Sina Weibo, a representative social media platform in China, as a data source to measure market sentiment and construct a Housing Market Sentiment Index. Sina Weibo is the world's largest Chinese-language social networking platform and, according to Sina Weibo's fourth quarter (Q4) 2020 financial results, Sina Weibo reached 521 million monthly active users in December 2020. The measurement of the housing market sentiment index based on Sina Weibo data is divided into the following four steps.

(1) Obtaining social media textual data. A web crawler was written in Python to automatically retrieve data from Sina Weibo related to the topic of housing price. This was done by using "housing price" in Chinese context as a seed keyword and combining it with Baidu's keyword search system to obtain words with high relevance to "housing price" and, after manual screening, the final keyword list (Table 1) was determined and entered into the crawler. Furthermore, in order to quantify the market sentiment in each city, the collected Sina Weibo textual data needed to be categorised by city. A new web crawler was written in Python to call Sina Weibo's user API interface, access the homepage of the user of each collected Sina Weibo tweet, obtain the city tag as a geo-

graphic tag for each tweet. We then sorted the collected Sina Weibo textual data by city so that we obtained a corpus of homebuyer comments on the housing market in each city.

(2) Data pre-processing. The Pandas tool in Python was used to read the Sina Weibo textual data and user data collected in the previous step. The Sina Weibo textual information included the tweet body, user nickname, user ID, and post time; the user information included the user ID, user nickname, and city. The "merge" function in Pandas was then used to stitch together the Sina Weibo textual and user information using the user ID as the connected column name to obtain a database of Sina Weibo texts with time and location information. Before sentiment annotation, the text needed to be pre-processed by word separation, deactivation, and cleaning of meaningless words. This study used the Jieba Chinese lexical analysis system for word separation and the Chinese stop words list in the Apache Lucene SmartChineseAnalyzer to remove stop words. In addition, the term frequency–inverse document frequency (TF-IDF) algorithm in the field of information retrieval was used to filter out words that occurred frequently but did not affect the sentiment polarity of the text, in preparation for the subsequent analysis of the sentiment polarity of the Sina Weibo text. The TF-IDF algorithm assumes that the importance of a word or phrase increases proportionally with the number of times it appears in the document, but at the same time decreases inversely with the frequency with which it appears in the corpus.

(3) Sentiment labeling. After the data pre-processing was completed, text sentiment analysis techniques were used to calculate the sentiment tendency value in each Sina Weibo text with the help of the Baidu AI open platform. The sentiment analysis module of the Baidu AI open platform is based on the bi-directional long short-term memory (Bi-LSTM) semantic model, which can better understand the semantics of text and accurately identify text sentiment. At the same time, the platform obtains a corpus across multiple domains through large-scale corpus-tagging propagation that makes up for the shortcomings of generic classifiers, ensuring excellent results for sentiment classification in various contexts. The sentiment analysis model of the Baidu AI open platform implemented sentiment labeling in the following three steps: a) transforming each word in each Sina Weibo text into a continuous semantic vector representation; b) transforming the word semantic sequence into a semantic representation of each Sina Weibo text through the Bi-LSTM network structure; and c) calculating the sentiment tendency value based on the semantics of each Sina Weibo text. For the accuracy of the sentiment analysis module of the Baidu AI open platform, we manually selected 5000 positive comments and 5000 negative comments from the processed tweets, and then called the sentiment analysis module of the Baidu AI open platform to perform sentiment annotation. The accuracy rates were found to be 86.56% and 87.02% respectively,

Table 1. Final keyword list in Chinese context

No.	Keyword	No.	Keyword
1	Housing price	9	Restricted purchase
2	Home loans	10	Loan restrictions
3	Provident fund	11	Used property
4	Down payment	12	New property
5	Purchasing a home	13	Real estate
6	Buying a home	14	Property market
7	Property tax	15	Commercial property
8	Property	16	Property market

proving that the Baidu algorithm is indeed capable of extracting sentiment from text.

(4) Construction of the housing market sentiment index. The range of sentiment tendency values from 0 to 1 was further subdivided into three sub-bands, where 0 to 0.4 represents negative, 0.4 to 0.6 represents neutral, and 0.6 to 1 represents positive (Tian et al., 2020). The housing market sentiment index (S) was then obtained for each city in each period according to Equation (1).

$$S_{it} = \frac{N\_pos_{it} - N\_neg_{it}}{N\_pos_{it} + N\_neg_{it}}. \tag{1}$$

In Equation (1), S_{it} denotes the housing market sentiment index for city i in period t; N_pos_{it} denotes the number of Sina Weibo texts marked as positive for city i in period t; and N_neg_{it} denotes the number of Sina Weibo texts marked as negative for city i in period t. The constructed housing market sentiment index (S) takes into account a combination of positive and negative comments and uses the difference between the number of both as a percentage of the total number of comments to measure the sentiment index, which measures the attitude of homebuyers towards the real estate market in each city.

2.2. Spatial econometric model

Popular spatial econometric models are the spatial lag model (SLM), spatial error model (SEM), and spatial Durbin model (SDM), which provide methods for analyzing spatial effect. Spatial effects, because of spatial correlation, can be characterized by the SLM and SEM. The SLM considers the spatial autocorrelation of the explanatory variables and introduces the spatial lag term of the explanatory variables on the basis of traditional panel regression models, while the SEM considers the spatial correlation of the error terms and introduces the spatial lag term of the random error terms on the basis of traditional panel regression models. According to the research of Elhorst (2014), the SDM is a formula which is an ordinary type of the SLM and SEM. Compared with the SLM and SEM, the SDM can visually analyze the spatial effect of each control variable and explanatory variable on the explanatory variables and can better describe the spatial effect through direct and indirect effects (Vega & Elhorst, 2015). Tests later in this paper will provide support by using SDM. The following SDM model of market sentiment affecting housing price was developed.

$$p_{it} = \rho \sum_{j=1}^{N} W_{ij} p_{it} + \sum_{k=1}^{n} (\beta_k X_{it,k} + \theta_k \sum_{j=1}^{N} W_{ij} X_{it,k}) + \mu_i + e_t + \varepsilon_{it}. \tag{2}$$

In Equation (2), W is the spatial weight matrix of $N \times N$; W_{ij} denotes the element in the i-th row and j-th column of the spatial weight matrix W; ρ, θ_k are the spatial regression coefficients; β_k are the corresponding parameters; N is the number of cities; n is the number of independent variables; p denotes the housing price and p_{it} denotes the housing price in city i in period t; X denotes

the control variables and $X_{it,k}$ denotes the value of the k-th independent variable for city i in period t.

Before constructing the model, a spatial autocorrelation test of the explanatory variable and the core explanatory variable was first required to determine whether or not a spatial econometric model was necessary. The most common method is to use the Moran index (*Moran's I*), which takes values in the range $[-1, 1]$ with values greater than 0 indicating a positive spatial correlation and less than 0 indicating a negative spatial correlation. *Moran's I* is calculated using Equation (3).

$$Moran's\ I = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}(X_i - \bar{X})(X_j - \bar{X})}{S^2 \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}}. \tag{3}$$

In Equation (3), $S^2 = \frac{1}{N}\sum_{i=1}^{1}(X_i - \bar{X})^2$; $\bar{X} = \frac{1}{N}\sum_{i=1}^{1}X_i$, X_i, X_j denotes the observed values of city i and city j respectively. In this paper, the values of housing price and housing market sentiment index are substituted into X in turn.

There is an individual effect in spatial econometric models; if the individual effect is related to the explanatory variable, then the individual effect is a fixed effect, and conversely if the individual effect is unrelated to the explanatory variable, then the individual effect is a random effect. A spatial econometric model is usually preceded by a Hausman test to determine whether a fixed effect or a random effect is chosen. The null hypothesis is that individual effects are not correlated with the explanatory variables. After the Hausman test, it is also necessary to select the appropriate spatial econometric model among the SLM, SEM and SDM. This is done by constructing the Lagrange multiplier (LM) and robust LM statistics based on model residuals based on a non-spatial panel model to test for spatial lags and spatial errors. If both the LM test and the robust LM test pass, the robustness of the model is further tested by a likelihood ratio (LR) test to verify whether the SDM simplified to the SLM or SEM.

If there is a spatial lag term, the regression coefficients of the econometric model cannot be used directly to capture the marginal effects of the explanatory variables. LeSage and Pace (2009) proposed a partial derivative approach to decompose a spatial effect into direct and indirect effects using spatial cross-sectional data. Elhorst (2014) extended this approach to spatial panel data. Referring to this approach, the SDM can be transferred to Equation (4).

$$p_{it} = (I - \rho W)^{-1} \sum_{k=1}^{n} (X_{it,k}\beta_k + WX_{it,k}\theta_k) + (I - \rho W)^{-1}(\mu_i + e_t + \varepsilon_{it}). \tag{4}$$

In Equation (4), I denotes a unit matrix of $N \times N$ with diagonal elements of 1 and other elements of 0. The direct and indirect effects of each explanatory variable on p are calculated by finding the matrix of partial derivatives of the expected value of p with respect to the corresponding

explanatory variable in unit 1 up to unit n in time. The direct effect measures the magnitude of the effect of each unit change in the dependent variable in the region; and the indirect effect measures the magnitude of the effect of each unit change in the independent variable on the dependent variable in other relevant regions.

2.3. Data sources and spatial weighting matrix

(1) Data sources

Considering the representativeness of the data, we selected 36 large and medium-sized cities in mainland China for the study, including 31 provincial capitals and 5 sub-provincial cities (Dalian, Qingdao, Ningbo, Xiamen, and Shenzhen), which have developed economies and large populations, and cover different regions of mainland China. However, due to the lack of relevant data or insufficient number of microblogs in 6 cities (Lhasa, Hohhot, Urumqi, Yinchuan, Xining, and Lanzhou), we excluded these cities in order to ensure data consistency, comparability between cities, and unbiased regression results. The 30 remaining large and medium-sized cities in mainland China were then selected as the study population and the specific city distribution is shown in Figure 1. In addition, China is a vast country with great imbalances in development. Taking the 30 large and medium-sized Chinese cities analyzed in this paper as an example, they are geographically divided into eastern, central and western cities, with cities located in the former region differing more from those in the latter two regions, while those in the latter two regions differing less. We have therefore further divided these 30 cities into the eastern region and the central and western regions according to geographical location and housing prices level. The 18 cities in the eastern region are Beijing, Changchun, Dalian, Fuzhou, Guanghzou, Hangzhou, Haikou, Herbin, Jinan, Nanjing, Ningbo, Qingdao, Shanghai, Shenyang, Shenzhen, Shijiazhuang, Tianjin, Xiamen; The 12 cities in the central and western regions are Changsha, Chengdu, Chongqing, Guiyang, Hefei, Kunming, Nanchang, Nanning, Taiyuan, Wuhan, Xian, Zhengzhou.
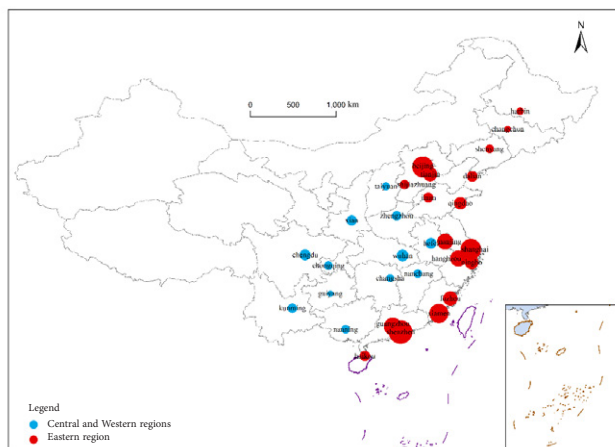

Figure 1. Map of 30 large and medium-sized cities in mainland China

Most studies of housing price trends examine annual changes in housing prices, but because housing prices often change significantly over the course of a year, examining housing price fluctuations on an annual cycle misses the true character of housing price changes. In addition, given the contagious and time-sensitive nature of market sentiment, the impact of market sentiment on the housing market is concentrated in the short term as homebuyers react to market sentiment in the short term, quickly forming expectations about future housing prices and thus acting directly on their own purchasing behaviour, so the impact of market sentiment on the housing market will be concentrated in the short term. However, too short a periodical examination (e.g., on a daily or monthly basis) is also not very meaningful. On the one hand, due to the large transaction characteristics of housing products, the volume of market transactions in the short term is very limited and hardly has an impact on price fluctuations in the overall housing market. On the other hand, the collection of social media data reveals that monthly data are insufficient for some cities, and quarterly data are even insufficient for the six excluded cities. In addition, the spatial Durbin model used in this paper to study spatial effects is applicable to short panel data with sample sizes larger than the number of periods. On balance, therefore, quarterly data as a research cycle for housing prices takes into account both the availability of data and the assurance that actual changes in housing prices can be captured.

Based on the above considerations, this study uses a quarterly research cycle, using 30 large and medium-sized cities for 20 quarters from January 2016 to December 2020 as the research sample. Except for the housing market sentiment index (S), which needs to be constructed separately, housing price data and economic fundamentals data were mainly obtained from the website of the Chinese National Bureau of Statistics (n.d.). Housing price (p) is the explained variable, which is expressed as the average sales price of commodity houses in prefecture-level cities and obtained by dividing the total sales price of commodity houses by the area of commodity houses sold, in yuan/m^2. Housing price market sentiment (S) is the core explanatory variable and was constructed using text sentiment analysis techniques, as described in Section 2.1. The control variables are the amount of investment in property development, mortgage rates, and land prices, referring to the study of Huang et al. (2019), Hui and Ng (2016). The amount of investment in property development reflects the level of development of the housing market, with the symbol *inv*, expressed as the amount of completed investment in residential real estate development, in billions of yuan. Mortgage rates reflect the impact of national macro policies on housing price, with the symbol r in %. Nominal mortgage rates were taken from the People's Bank of China website. As nominal mortgage rates are the same across cities, it is not possible to explore the impact of mortgage rates on housing price, so the nominal mortgage rates were deflated by the Consumer Price Index (*CPI*)

for each city to obtain the real mortgage rates for each city. The land price reflects the cost of the property, with the symbol *l*, expressed in yuan/m² using the average floor area price of land sold in the residential category. In addition, to eliminate the effect of inflation *p*, *inv*, and *l* were transformed into real variables by deflating the *CPI*.

(2) Spatial weighting matrix

As can be seen from both the estimation of the model and the measurement of the spatial effect, the choice of the spatial weighting matrix is very important in relation to the empirical results. The spatial proximity weighting weights (assuming that things are connected only between regions with a common boundary, two regions with a common boundary take the value of 1, and two regions without a common boundary take the value of 0) and the spatial distance weighting weights distance (assuming that the spatial interaction between two regions is inversely proportional to the distance between the two regions) are the most intuitive and commonly used methods. The spatial proximity weighting matrix cannot truly reflect the spatial association due to the lack of data for cities such as Lhasa. Therefore, the spatial distance weighting matrix was chosen in this study and its matrix elements (W_{ij}) were obtained according to Equation (5).

$$W_{ij} = \begin{cases} 1/d_{ij}, & i \neq j \\ 0, & i = j \end{cases}. \tag{5}$$

In Equation (5), d_{ij} denotes the distance between city *i* and city *j*, and the specific values are obtained by calculating the spherical semi-positive vector distance from the latitude and longitude of city *i* and city *j*. The latitude and longitude data for cities were obtained from the National Geomatics Center of China (n.d.).

3. Empirical analysis

3.1. Descriptive analysis of data

Using the method described in Section 2.1, we ended up with a total of 3861862 Sina Weibo tweets. Our statistics show that the number of Sina Weibo tweets posted in China's top tier cities (Beijing, Shanghai, Shenzhen and Guangzhou) ranged from 150,000–200,000, and the number of tweets in other cities besides these four cities also remained between 100,000 and 150,000, and Figure 2 shows the temporal distribution of the number of Sina Weibo tweets.

From Figure 2, we can see that the average number of Sina Weibo tweets per city is greater than 5,000 per quarter in both the eastern cities and central and western cities, showing an upward trend over time. In addition, the average number of Sina Weibo tweets per city per quarter is greater in the eastern cities than in the central and western cities. This provides a sufficient data source for the next step of sentiment analysis. We also analysed the user groups posting the above collected Sina Weibo tweets and calculate the average age of Sina Weibo users of the collected tweets to be 33.4 years old. According to statistics, the average age of Chinese homebuyers is 33.2 years old, and the average age of our selected group of Sina Weibo users is roughly the same as the average age of homebuyers, indicating that our selected users are likely to be potential homebuyers. This further suggests that it is appropriate to characterize the sentiment of homebuyers by measuring the sentiment of these Sina Weibo groups. In addition, we plotted a line graph of the average housing price in the eastern cities and the central and western cities to reflect the characteristics of housing price fluctuations in the two regions, and the results are shown in Figure 3.
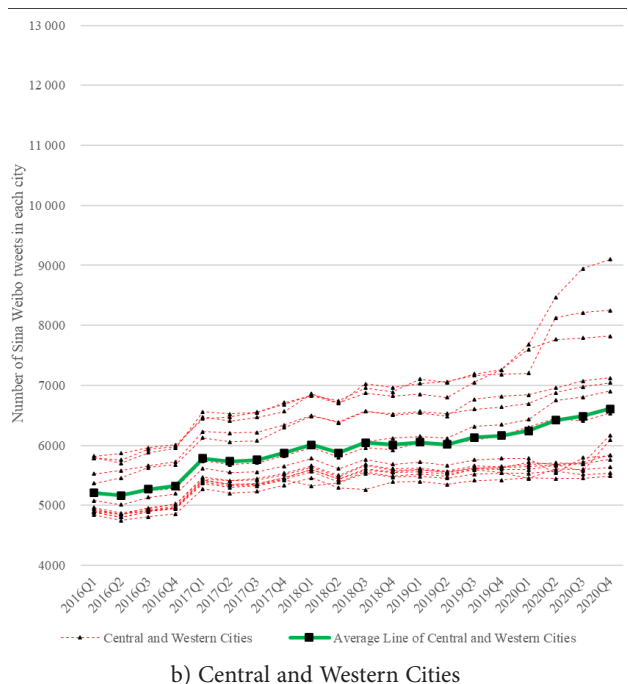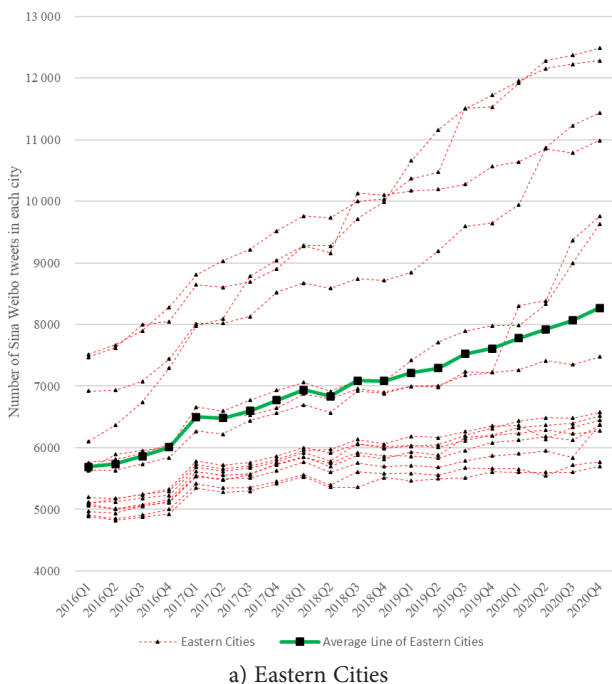
a) Eastern Cities

b) Central and Western Cities

Figure 2. The number of quarterly Sina Weibo tweets in 30 cities in mainland China

a) Eastern Cities

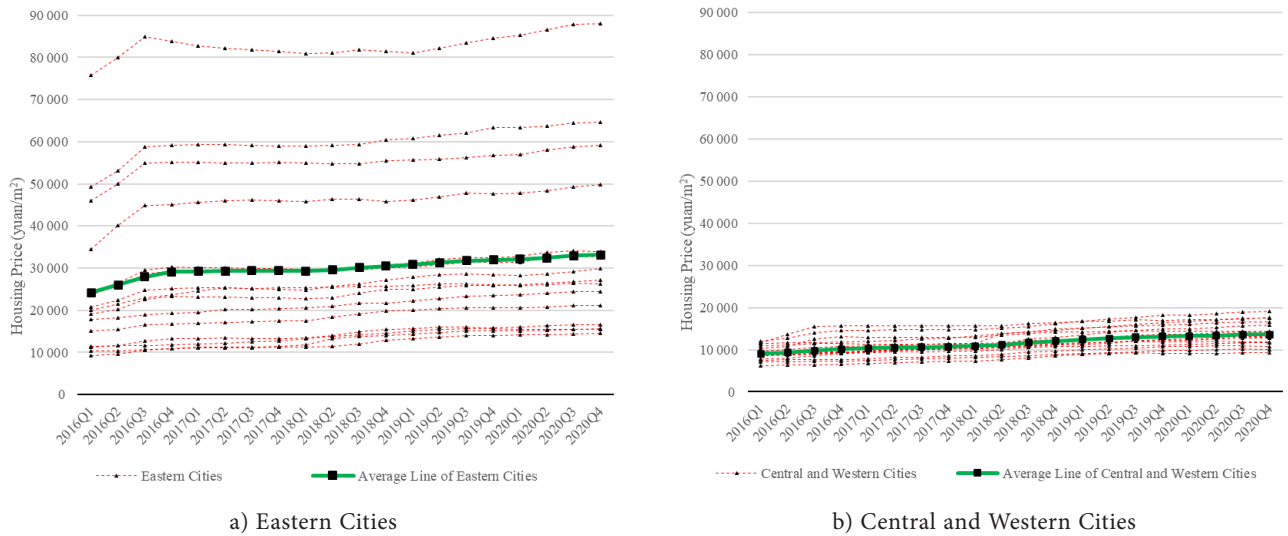b) Central and Western Cities

Figure 3. The quarterly housing price in 30 cities in mainland China

Table 2. Descriptive statistics for each variable

Variable	Mean	Std. Dev.	Min	Max
$log(p)$	9.748	0.618	8.840	11.336
S	−0.349	0.198	−0.681	0.396
r	4.574	0.682	2.797	6.345
$log(inv)$	4.470	0.898	1.121	5.713
$log(l)$	9.252	0.941	7.858	11.403

Table 3. *Moran's I* values for *p* and *S* from 2016 to 2020

Year	p	S
2016	0.316* (0.064)	0.106* (0.076)
2017	0.312* (0.067)	0.108* (0.099)
2018	0.337* (0.050)	0.119* (0.097)
2019	0.329* (0.061)	0.237** (0.022)
2020	0.326* (0.058)	0.151** (0.041)

Note: ** and * indicate coefficients are statistically significant at 5% and 10%, respectively.

From Figure 3, we can see that housing prices in the eastern cities are generally higher than those in the central and western cities. During the period from 2016 to 2020, the housing prices of 30 large and medium-sized cities all show an upward trend. The average value of housing prices in eastern cities is above 20,000 yuan per sqm, while the housing price level in central and western cities is less than 15,000 yuan per sqm, thus also illustrating the huge difference in housing prices between the eastern cities and the central and western cities. Therefore, it is not only necessary to analyse the spatial effects of the whole sample, but also to subdivide the study sample and analyse the spatial effects of different regions.

In the subsequent regression analysis, the values of *p*, *inv*, and *l* were taken in logarithmic form to eliminate the effect of heteroskedasticity in the empirical process. Descriptive statistics for each variable are shown in Table 2. The results of the sentiment index are in line with expectations. China's housing prices as a whole are on an upward trend, while the group we are looking at does not want to see this, so the average level of overall sentiment is negative, expressed as the mean sentiment index of less than zero.

3.2. Statistical tests for model form selection

Table 3 shows the test values of the *Moran's I* index for *p* and *S* for the 30 large and medium-sized cities. As can be seen from Table 3, the *Moran's I* values are significant and

greater than 0 for each year for both *p* and *S*, indicating a significant positive spatial correlation between housing price and market sentiment in these cities. The *Moran's I* value for housing price was relatively stable from 2016 to 2020, with values around 0.3, indicating a strong positive spatial correlation with housing price. The *Moran's I* value of *S* had a small but year-on-year increase in value, reaching a maximum in 2019 and falling back in 2020. Both *p* and *S* are spatially autocorrelated, so it was necessary to consider the spatial correlation between housing price and market sentiment in the empirical process for each city. A spatial econometric model was needed to further analyze the spatial effect of market sentiment on housing price, otherwise the regression results would be biased.

Hausman test results indicate that the analysis rejects the null hypothesis and passes the significance test at a 1% level of confidence, so a spatial econometric model that includes the fixed effect needed to be developed. Balta (2009) also pointed out that the fixed effect is generally the best choice when the study sample is restricted to certain specific individuals.

In order to determine which model was optimal among the SLM, SEM, and SDM, the LM statistic needed to be calculated and the results are shown in Table 4. The null hypothesis for the LM_lag and robust LM_lag tests is that there are no spatial lag effects in the model. The null hypothesis for the LM_error and robust LM_error tests is

that there are no spatial error effects in the model. If there are only spatial lag effects, the SLM is optimal; if there are only spatial error effects, the SEM is optimal; if there are both spatial lag effects and spatial error effects, the SDM is optimal. The test results show that the non-spatial panel model passes the LM lag, robust LM lag, LM error, and robust LM error tests at a 1% level of significance. This suggests both null hypotheses are rejected and there are both spatial lag effects and spatial error effects in the model. To sum up, this study builds an SDM.

Table 4. Statistical tests results for form selection

	Statistic
LM_lag	550.714***
Robust LM_lag	66.744***
LM_error	486.885***
Robust LM_error	2.916*

Note: *** and * indicate coefficients are statistically significant at 1% and 10%, respectively.

Before spatial econometric model regression, an LR test was used to verify whether the SDM would be converted into the SLM or SEM. The two null hypotheses of the LR test were: 1) that the SDM was simplified to the SEM; and 2) that the SDM was simplified to the SLM. If the LR test reaches significance, the null hypothesis is rejected. Table 5 presents the Z and P values for the LR test. According to the test results, both null hypotheses were rejected, thus indicating that the SDM under a fixed effect is robust and more suitable for the data characteristics studied here.

Table 5. LR test results

Null hypothesis	Z_value	P_value
SDM simplified to SLM	21.84***	0.0002
SDM simplified to SEM	302.40***	0.0000

Note: *** indicates coefficients are statistically significant at 1%.

3.3. Overall effect of market sentiment on housing price

Based on the established SDM for regional housing price (Equation 2), regressions were conducted on panel data from 30 large and medium-sized cities in China using STATA version 16 (Stata Corporation, College Station, Texas, USA) and the estimated results are shown in the 4th column of Table 6. The estimation results for the fixed effects non-spatial panel models without and with the inclusion of the sentiment index are given in the 2nd and 3rd columns of Table 6. From the 2nd and 3rd columns, the R^2 of the model increases from 0.4526 to 0.4935 supporting the idea that that adding a sentiment index to the underlying market model enhances the model's ability to explain housing price volatility (Ling et al., 2014). The results comparing between the 3rd and 4th columns show that spatial correlation should be fully considered in the analysis of housing price, meaning that it is reasonable to use the SDM. Firstly, the spatial autoregressive coefficient ($\rho = 0.5783$) is significant, which suggests that housing price is spatially correlated. Secondly, compared to the fixed-effects panel model, the R^2 of the SDM has improved, with the regression coefficient of *inv* decreasing from 0.327 to 0.076, the regression coefficient of *l* decreasing from 0.213 to 0.077, and the regression coefficient of *S* decreasing from 0.179 to 0.080, indicating that the fixed-effects panel model does not take into account the spatial correlation of housing markets between cities and overestimates the impact of these factors on housing price.

From column 3 of Table 6, the coefficient for *S* is 0.080, which is significant at the 1% level. This indicates that market sentiment has a significant positive effect on housing price. It indicates that positive market sentiment tends to boost public expectations of future housing prices, stimulating enthusiasm to invest in the property market, which in turn drives up housing prices. The coefficient of $W \times S$ is 0.011, which implies that market sentiment in the local city has a spatial effect of promoting higher housing prices in neighboring cities. The coefficients of *inv* and *l*

Table 6. Estimated results for whole sample

Variable	Non-spatial panel model without S	Non-spatial panel model	SDM
S		0.179***(5.64)	0.080***(3.52)
inv	0.355***(5.18)	0.327***(5.12)	0.076**(2.38)
r	−0.009(−0.54)	−0.014(−0.80)	−0.036(−1.28)
l	0.243***(6.73)	0.213***(6.52)	0.077***(14.00)
W×S			0.011(0.63)
W×inv			0.039(1.01)
W×r			0.031(1.26)
W×l			0.065***(7.55)
ρ			0.5783***(31.21)
R²	0.4526	0.4935	0.6451
Log_L			2492.3507

Note: *** and ** indicate coefficients are statistically significant at 1% and 5%, respectively. Z-values are given in parentheses.

are positive and significant, indicating that increased investment by property developers and higher land prices have positive driving effects on housing prices. The coefficient of r is negative, indicating that mortgage rates have a negative effect on housing prices.

Due to the inclusion of a spatial lag term in the model, the coefficients of the estimation results do not directly reflect the marginal effects of the variables. The parameters of the coefficients do not represent the exact degree of influence and only indicate whether independent variable has a significant effect on housing prices. Therefore, a partial differential method was used to calculate the direct effects, indirect effects, and total effects to measure the exact degree of influence of the variables, with the results shown in Table 7. As can be seen from column 4 of Table 7, the total effects of S, inv, and l are 0.216, 0.278, and 0.339, respectively, and all pass the significance test, indicating that market sentiment, investment in property development, and land price are all important factors in driving up housing price.

As shown in Table 7, market sentiment has a significant positive direct effect on housing prices. From row 1 of Table 7, the direct effect coefficient of S on housing price is 0.185, which is significant at the 1% level. This is mainly due to the particularities of China's housing market, which has developed over a relatively short period of time, does not have a robust information-disclosure mechanism, is less transparent, and is typically a non-effective market. Unlike the stock market where there are more institutional investors, the majority of participants in China's housing market are individuals or households, who have less information, and this asymmetry of information may lead to a tendency to follow the actions of others who are perceived to be better informed. This "herding behavior" allows emotions to play a greater role. In addition, the peculiarities of the illiquid housing market and the restrictions on short selling, combined with a series of regulatory policies in China such as purchase and sale restrictions, further reduce the likelihood of rational and sophisticated participants entering the market to offset mispricing. As a result, market sentiment is more likely to have a strong and lasting impact on China's housing market.

However, it is important to note that a positive coefficient on the housing market sentiment indicator does not indicate that sentiment is positive, but only those changes in market sentiment are in the same direction as fluctuations in housing prices. As sentiment in the Chinese housing market is generally negative, a positive change in market sentiment implies a weakening of the negative course of market sentiment. This is mainly due to the fact that the direction of change in market sentiment is determined by the size of the increase in housing prices rather than the growth or fall in housing prices. On the one hand, when housing prices rise sharply, the negative voice in the market increases and the negative change in market sentiment is reflected in a fall in the sentiment index. As a result of the severe negativity in market sentiment, home buyers are full of complaints about the sharp rise in housing prices. In order to eliminate the negative impact of rising negative sentiment on society, the government will enact strict property market regulation measures to curb the rise of property prices. On the other hand, when housing price rises become smaller, although market sentiment is still negative overall, the degree of negativity decreases and is reflected in an increase in the market sentiment index, which is an indication of a positive change in market sentiment. With the positive change in market sentiment, the negative voices in market sentiment diminish, the market becomes relatively calm and the government adopts mild real estate market controls, which in turn promote higher housing prices. We thus find that positive changes (the negative voice in housing market decreases) in market sentiment boost housing prices and negative changes (the negative voice in housing market increases) dampen them, suggesting that changes in market sentiment and housing price fluctuations are consistent, indicating that changes in market sentiment have a positive effect on housing price fluctuations, resulting in a positive coefficient for the sentiment index. As the rate of increase in housing prices in China's 30 large and medium-sized cities is currently levelling off (as shown in Figure 3), the negative voices in the market have diminished and market sentiment has changed positively, which in turn has contributed to the rise in housing prices. This explains the phenomenon that housing prices are rising despite the fact that market sentiment is the overall negative.

Row 1 of Table 7 also shows that the indirect effect coefficient of S is 0.031, which is significant at the 10% level, indicating that market sentiment has a significant positive spatial spillover effect on housing price, with a positive change in market sentiment not only boosting housing prices in the city but also contributing to a rise in housing prices in other cities. We argue that the reason for the spatial effect of market sentiment on housing prices comes from the rigid demand of the homebuyer group. The Chinese believe that "a house means a home", plus in China the youth group is the main home buyer and owning a home is a pre-requisite for them to get married, so home buyers are more sensitive to information on housing prices. When the market sentiment index of a city rises, it indicates that the previous housing price rises in that city have previously become smaller. Home buyers in the city will share their views online, and home buyers in neighbouring cities will receive these messages

Table 7. The direct, indirect, and total effects of the whole sample

Variable	Direct effect	Indirect effect	Total effect
S	0.185***(4.57)	0.031*(1.73)	0.216***(4.53)
inv	0.184***(2.96)	0.094***(3.20)	0.278***(4.21)
r	−0.040(−1.03)	0.030(1.32)	−0.010(−0.26)
l	0.237***(7.58)	0.102***(9.00)	0.339***(13.11)

Note: *** and * indicate coefficients are statistically significant at 1% and 10%, respectively. Z-values are given in parentheses.

through social media. Due to rigid demand, they will have expectations that housing price rises in their city will also become smaller, at which point the market sentiment changes positively and the market sentiment index will rise. As changes in market sentiment have a positive effect on house price volatility, it can lead to a small increase in housing prices of neighboring cities. In addition, the indirect effect is smaller than the direct effect due to the time-sensitive nature of the spread of market sentiment and the fact that local home buyers are more likely to be affected by local market sentiment and thus engage in home-buying behavior causing housing price volatility. Although the spatial spillover effect is smaller than the direct effect, it is still not negligible.

Among the control variables, investment in property development and land price have a significant positive effect on housing price in the city, while mortgage interest rates have a negative but non-significant effect on housing price. The direct and indirect effects of *inv* are 0.184 and 0.094, respectively, both significant at the 1% level, as shown in row 2 of Table 7. Row 3 of Table 7 shows that the direct and indirect effects of *l* are 0.237 and 0.102, both significant at the 1% level. The direct and indirect effects of *r* do not pass the significance test, as shown in row 4 of Table 7. This may be related to the excessive rise in housing prices in China, where a reduction in mortgage rates is not effective in increasing market demand due to high housing prices, and therefore does not have a significant effect on housing price volatility. The above analysis suggests that positive changes in housing market sentiment, increased investment in property development, and rising land prices all contribute to higher housing prices in local and neighboring cities.

3.4. Comparing results for different regions

When analyzing the spatial effect, it is necessary to examine the geographical regional differences in market sentiment affecting housing price. This study therefore divides the 30 large and medium-sized cities into the eastern region and the central and western regions by geographical location. In China, the eastern region of the country was the first to open up and has a high level of economic development, while the central and western regions are less economically developed. Table 8 gives the estimation results from the SDM for the eastern region and the central and western regions. The results of two sub-samples are consistent with the whole sample, thus demonstrating the robustness of the model.

The results in Table 8 show that the total effects of *S*, *inv*, and *l* are positive and significant in both the eastern region and the central and western regions. The total effects of *S* are 0.240 and 0.149 in the eastern region and the central and western regions, respectively; the total effects of *inv* are 0.287 and 0.167 in the eastern region and the central and western regions, respectively; and the total effects of *l* are 0.333 and 0.746 in the eastern region and the central and western regions, respectively. This indicates

Table 8. Results of SDM for Eastern region and Central and Western regions

	Variable	Eastern	Central and Western
	S	0.089***(3.86)	0.066**(2.44)
	inv	0.076*(2.39)	0.1269***(2.84)
	r	−0.037(−1.32)	−0.020(−0.57)
	l	0.077***(3.97)	0.274***(11.17)
	$W \times S$	0.012(0.70)	0.019(0.94)
	$W \times inv$	0.044(1.14)	−0.032(−0.66)
	$W \times r$	0.031(1.26)	0.013(0.35)
	$W \times l$	0.065***(3.48)	0.1553***(9.18)
Direct effect	S	0.205***(5.19)	0.117***(3.28)
	inv	0.192***(3.12)	0.129***(3.16)
	R	−0.042(−1.09)	−0.016(−0.51)
	l	0.233***(7.23)	0.536***(9.89)
Indirect effect	S	0.035*(1.90)	0.032(1.53)
	inv	0.095***(3.23)	0.038(0.67)
	r	0.030(1.32)	0.004(0.10)
	l	0.102***(9.45)	0.210***(14.05)
Total effect	S	0.240***(5.21)	0.149***(3.54)
	inv	0.287***(4.39)	0.167***(2.98)
	r	−0.012(−0.32)	−0.012(−0.31)
	l	0.333***(12.74)	0.746***(13.17)
	R^2	0.6485	0.8999
	Log_L	2495.3673	700.7661

Note: ***, **, and * indicate coefficients are statistically significant at 1%, 5%, and 10%, respectively. Z-values are given in parentheses.

that housing prices in eastern cities are more influenced by market sentiment. Among the economic fundamentals factors, property development investment has a stronger effect in the eastern region than in the central and western regions, while land price has a greater degree of influence in the central and western cities. Unlike the other variables, the total effect of *r* is negative and insignificant for housing prices in both the eastern region and the central and western regions, consistent with the whole sample, indicating that an increase in mortgage rates does not have a significant negative impact on housing price volatility.

The direct effects of *S* are 0.205 and 0.117 in the eastern region and the central and western regions, respectively, both significant at the 1% level. This indicates that housing price is significantly influenced by the local market sentiment and that this effect is stronger in the eastern region than in the central and western regions. The eastern housing market is the core component of China's housing market, with a higher level of development, a larger resident population, high resident income, strong purchasing power, higher public market participation, more active transactions, and information asymmetry. However, at the same time, due to high housing prices, the rate of home ownership is rather lower, resulting in a greater demand for rigid home ownership. These conditions make homebuyers also

more susceptible to market sentiment, with stronger levels of disposition effects, noisy transactions, and herd utility in the housing market, leading to irrational home-buying behavior, which further exacerbates irrational housing price increases. So the direct effect of market sentiment on housing price is stronger in the eastern region.

In the eastern region, the indirect effect of *S* is 0.035, indicating that market sentiment has a positive spatial spillover effect on housing price in the eastern region; that is, higher market sentiment has a positive impact on neighboring cities' housing prices. The indirect effect of *S* in the central and western regions is not significant, indicating that no significant spatial spillover effect of market sentiment occurs in the central and western cities. This shows that the spatial spillover effect of market sentiment in China's housing market mainly comes from the eastern region. This phenomenon may be due to the differences in the levels of city development and population movement between the eastern region and the central and western regions. Cities in the eastern region are mostly municipalities directly under the central government, regional central cities, economically developed eastern regions, and open coastal cities, which not only have strong economic bases and large middle class populations, but also have convenient transportation and the ability to radiate to many surrounding cities. Cities in the eastern region therefore have strong demographic competitiveness, making it easier for them to attract an inflow of people, especially young people, which in turn spreads market sentiment in the cities in the eastern region and has a strong spatial spillover effect on housing price.

Conclusions

This study has tested the impact of market sentiment on housing price in the social media environment from the perspective of behavioral finance theory. We used text sentiment analysis techniques to capture housing market sentiment in social media and constructed a housing market sentiment index for 30 large and medium-sized cities in China, which uses the advantages of big data such as timeliness and foresight to make up for the shortcomings of using indirect indicators such as macroeconomic indicators to characterize sentiment. Considering the existence of correlation in the housing market and the phenomenon of propagation of market sentiment, an SDM has been constructed and empirically analyzed to verify the spatial effect of market sentiment on housing price volatility, while avoiding the problem of omission of spatial utility found in most studies and improving the precision of parameter estimation, which is novel in this research area in relation to studying the spatial spillover effect of market sentiment. In addition, this study has explored the heterogeneity of the spatial effect of market sentiment on housing price from the perspective of geographical region.

The study concludes that market sentiment, investment in property development, and land price all affect housing price. Ignoring the spatial correlation of housing markets would overestimate the impact of each factor on housing prices. For the total sample, a positive change in housing market sentiment leads not only to an increase in local housing price (a direct effect), but also to increases in neighboring cities (an indirect effect). In other words, market sentiment affects housing price inflation in two ways: the first is by directly influencing the expectations of local home buyers, which in turn increases price changes in the local housing market, as reflected by the direct effect coefficient; and the second is by influencing the expectations of home buyers in neighboring cities through spatial spillover effects, leading to increased price changes in the neighbouring cities' housing markets, as reflected by the indirect effect coefficient. The indirect effect is smaller than the direct effect. This process provides us with the possibility to explore the fluctuation of house prices and the future trend of the real estate market. In addition, the spatial effect of market sentiment on housing price is heterogeneous in terms of geographical region. In the eastern region, the direct effects of market sentiment are stronger than in the central and western regions, while spatial spillover effects are observed only in the eastern region.

The study also has policy implications, as the findings can help governments and policymakers understand future housing price volatility based on market sentiment and become better able to formulate strategies to achieve a healthy housing market. Governments and policymakers can use this housing market sentiment index as a monitoring indicator of online market sentiment and incorporate it into an early-warning system for the housing market. When the housing market sentiment index changes too much, timely measures can be taken to stabilize market sentiment and thus stabilize local and neighboring cities' housing prices. The government should also consider the housing market sentiment index to reflect public expectations and pay attention to changes in public opinion when formulating real estate regulation policies, strengthen the management of housing market sentiment in social media, and consider the strength and direction of policies, so as to effectively achieve its macroeconomic regulation objectives for the housing market.

Funding

References

Alkay, E., Watkins, C., & Keskin, B. (2018). Explaining spatial variation in housing construction activity in Turkey. *International Journal of Strategic Property Management, 22*(2), 119–130. https://doi.org/10.3846/ijspm.2018.443

Ardèvol-Abreu, A., & Gil de Zúñiga, H. (2017). Effects of editorial media bias perception and media trust on the use of traditional, citizen, and social media news. *Journalism & Mass Communication Quarterly*, *94*(3), 703–724. https://doi.org/10.1177/1077699016654684

Bailey, M., Cao, R., Kuchler, T., & Stroebel, J. (2018). The economic effects of social networks: evidence from the housing market. *Journal of Political Economy*, *126*(6), 2224–2276. https://doi.org/10.1086/700073

Baker, M., & Wurgler, J. (2007). *Investor sentiment in the stock market* (Working Paper No. 13189). National Bureau of Economic Research. https://doi.org/10.3386/w13189

Balcilar, M., Bouri, E., Gupta, R., & Kyei, C. K. (2021). High-frequency predictability of housing market movements of the United States: the role of economic sentiment. *Journal of Behavioral Finance*, *22*(4), 490–498. https://doi.org/10.1080/15427560.2020.1822359

Bayer, P., Mangum, K., & Roberts, J. W. (2021). Speculative fever: investor contagion in the housing bubble. *American Economic Review*, *111*(2), 609–651. https://doi.org/10.1257/aer.20171611

Beracha, E., Lang, M., & Hausler, J. (2019). On the relationship between market sentiment and commercial real estate performance–a textual analysis examination. *Journal of Real Estate Research*, *41*(4), 605–638. https://doi.org/10.22300/0896-5803.41.4.605

Case, K. E., & Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, *2003*(2), 299–362. https://doi.org/10.1353/eca.2004.0004

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*, *2012*(1), 265–315. https://doi.org/10.1353/eca.2012.0014

Cerchiello, P., Giudici, P., & Nicola, G. (2017). Twitter data models for bank risk contagion. *Neurocomputing*, *264*, 50–56. https://doi.org/10.1016/j.neucom.2016.10.101

Cerchiello, P., & Nicola, G. (2018). Assessing news contagion in finance. *Econometrics*, *6*(1), 5. https://doi.org/10.3390/econometrics6010005

Clayton, J., Ling, D. C., & Naranjo, A. (2009). Commercial real estate valuation: Fundamentals versus investor sentiment. *The Journal of Real Estate Finance and Economics*, *38*(1), 5–37. https://doi.org/10.1007/s11146-008-9130-6

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1–32. https://doi.org/10.1093/rfs/hhu072

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738. https://doi.org/10.1086/261703

DeFusco, A., Ding, W., Ferreira, F., & Gyourko, J. (2018). The role of price spillovers in the American housing boom. *Journal of Urban Economics*, *108*, 72–84. https://doi.org/10.1016/j.jue.2018.10.001

Dietzel, M. (2015). Sentiment-based predictions of housing market turning points with Google trends. *International Journal of Housing Markets and Analysis*, *9*, 108–136. https://doi.org/10.15396/eres2015_3

Elhorst, J. P. (2014). *Spatial econometrics*. Springer. https://doi.org/10.1007/978-3-642-40340-8

Fischer, M., & Stamos, M. Z. (2013). Optimal life cycle portfolio choice with housing market cycles. *The Review of Financial Studies*, *26*(9), 2311–2352. https://doi.org/10.1093/rfs/hht010

Gao, Q., & Zhao, T. (2018). The influence of home buyer sentiment on Chinese housing prices–based on media text mining. *International Journal of Economics and Finance*, *10*, 145–145. https://doi.org/10.5539/ijef.v10n9p145

Gao, Y., Zhao, K., Wang, C., & Liu, C. (2020). The dynamic relationship between internet attention and stock market liquidity: a thermal optimal path method. *Physica A: Statistical Mechanics and its Applications*, *550*, 124180. https://doi.org/10.1016/j.physa.2020.124180

Giannini, R., Irvine, P., & Shu, T. (2019). The convergence and divergence of investors' opinions around earnings news: evidence from a social network. *Journal of Financial Markets*, *42*, 94–120. https://doi.org/10.1016/j.finmar.2018.12.003

Granziera, E., & Kozicki, S. (2015). House price dynamics: fundamentals and expectations. *Journal of Economic Dynamics and Control*, *60*, 152–165. https://doi.org/10.1016/j.jedc.2015.09.003

Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: a machine learning approach. *Journal of Property Research*, *35*, 344–371. https://doi.org/10.1080/09599916.2018.1551923

Heinig, S., & Nanda, A. (2018). Measuring sentiment in real estate – a comparison study. *Journal of Property Investment & Finance*, *36*(3), 248–258. https://doi.org/10.1108/JPIF-05-2017-0034

Huang, R., Zuo, W., & Bi, L. (2015). Predicting the stock market based on microblog mood. *Journal of Industrial Engineering/ Engineering Management*, *29*(01), 47-52+215. https://doi.org/10.13587/j.cnki.jieem.2015.01.006

Huang, Y., Hong, W., & Yu, H. (2019). How does the market sentiment affect urban housing price. *Economic Theory and Business Management*, *07*, 75–88 (in Chinese).

Hui, E. C. M., & Ng, I. M. H. (2016). Access to mortgage credit and housing price dynamics. *International Journal of Strategic Property Management*, *20*(1), 64–76. https://doi.org/10.3846/1648715X.2015.1103802

Hui, E. C., & Wang, Z. (2014). Market sentiment in private housing market. *Habitat International*, *44*, 375–385. https://doi.org/10.1016/j.habitatint.2014.08.001

Lam, C. H. L., & Hui, E. C. M. (2018). How does investor sentiment predict the future real estate returns of residential property in Hong Kong? *Habitat International*, *75*, 1–11. https://doi.org/10.1016/j.habitatint.2018.02.009

Lambertini, L., Mendicino, C., & Punzi, M. T. (2013). Expectation-driven cycles in the housing market: evidence from survey data. *Journal of Financial Stability*, *9*(4), 518–529. https://doi.org/10.1016/j.jfs.2013.07.006

Lee, C., & Park, K. (2018). Analyzing the rent-to-price ratio for the housing market at the micro-spatial scale. *International Journal of Strategic Property Management*, *22*(3), 223–233. https://doi.org/10.3846/ijspm.2018.1416

Leitch, D., & Sherif, M. (2017). Twitter mood, CEO succession announcements and stock returns. *Journal of Computational Science*, *21*, 1–10. https://doi.org/10.1016/j.jocs.2017.04.002

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420064254

Li, J., Hu, Y., & Liu, C. (2020). Exploring the influence of an urban water system on housing prices: case study of Zhengzhou. *Buildings*, *10*(3), 44. https://doi.org/10.3390/buildings10030044

Li, J., Zheng, L., Liu, C., & Shen, Z. (2021). Information spillover effects of real estate markets: evidence from ten metropolitan cities in China. *Journal of Risk and Financial Management*, *14*(6), 244. https://doi.org/10.3390/jrfm14060244

Liang, C.-M., Lee, C.-C., Lin, Y.-H., Yu, Z., & Yeh, W.-C. (2020). The impact of luxury housing on neighborhood housing prices: an application of the spatial difference-in-differences method. *International Journal of Strategic Property Management*, *24*(6), 456–473. https://doi.org/10.3846/ijspm.2020.13649

Ling, D. C., Naranjo, A., & Scheick, B. (2014). Investor sentiment, limits to arbitrage and private market returns: investor sentiment, limits to arbitrage and private market returns. *Real Estate Economics*, *42*(3), 531–577. https://doi.org/10.1111/1540-6229.12037

Luo, Z. Q., Liu, C., & Picken, D. (2007). Housing price diffusion pattern of Australia's state capital cities. *International Journal of Strategic Property Management*, *11*(4), 227–242. https://doi.org/10.3846/1648715X.2007.9637571

Marcato, G., & Nanda, A. (2016). Information content and forecasting ability of sentiment indicators: case of real estate market. *Journal of Real Estate Research*, *38*, 165–203. https://doi.org/10.1080/10835547.2016.12091442

National Bureau of Statistics of China. (n.d.). *Statistical data*. www.stats.gov.cn/tjsj/

National Geomatics Center of China. (n.d.). *National catalogue service for geographic information*. https://www.webmap.cn

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, *73*, 125–144. https://doi.org/10.1016/j.eswa.2016.12.036

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS One*, *10*(9), 1–21. https://doi.org/10.1371/journal.pone.0138441

Ren, J., Dong, H., Padmanabhan, B., & Nickerson, J. V. (2021). How does social media sentiment impact mass media sentiment? A study of news in the financial markets. *Journal of the Association for Information Science and Technology*, *72*(9), 1183–1197. https://doi.org/10.1002/asi.24477

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410–428. https://doi.org/10.1108/JPIF-07-2017-0050

Siikanen, M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., & Hussain, A. (2018). Facebook drives behavior of passive households in stock markets. *Finance Research Letters*, *27*, 208–213. https://doi.org/10.1016/j.frl.2018.03.020

Sinyak, N., Tajinder, S., Madhu Kumari, J., & Kozlovskiy, V. (2021). Predicting real estate market trends and value using pre-processing and sentiment text mining analysis. *Real Estate: Economics, Management*, (1), 35–43. https://doi.org/10.22337/2073-8412-2021-1-35-43

Soo, C. K. (2013). Quantifying animal spirits: news media and sentiment in the housing market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2330392

Tian, H., Gao, C., Xiao, X., Liu, H., He, B., Wu, H., Wang, H., & Wu, F. (2020). SKEP: Sentiment knowledge enhanced pretraining for sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 4067–4076). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.374

Vega, S. H., & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*, *55*(3), 339–363. https://doi.org/10.1111/jors.12188

Vergos, K., & Zhi, H. (2018). Is it a curse or a blessing to live near rich neighbors? Spatial analysis and spillover effects of house prices in Beijing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3289359

Wang, Z., & Hui, E. C. (2017). Fundamentals and market sentiment in housing market. *Housing, Theory and Society*, *34*(1), 57–78. https://doi.org/10.1080/14036096.2016.1196240

Wu, Y.-L., Lu, C.-L., Chen, M.-C., & Chu, F.-N. (2017). What forces drive the dynamic interaction between regional housing prices? *International Journal of Strategic Property Management*, *21*(3), 225–239. https://doi.org/10.3846/1648715X.2016.1254120

Yu, J., & Yuan, Y. (2011). Investor sentiment and the mean–variance relation. *Journal of Financial Economics*, *100*(2), 367–381. https://doi.org/10.1016/j.jfineco.2010.10.011