Local Housing Market Sentiments and Returns: Evidence from China

Shulin Shen[1] · Yiyi Zhao[1] · Jindong Pang[2]

Abstract

This paper examines the impacts of local housing sentiments on the housing price dynamics of China. With a massive second-hand transaction dataset, we construct monthly local housing sentiment indices for 18 major cities in China from January 2016 to October 2020. We create three sentiment proxies representing the local housing market liquidity and speculative behaviors from the transaction dataset and then use **partial least squares (PLS)** to extract a recursive **look-ahead-bias-free** local housing sentiment index for each city considered. The local housing sentiments are shown to have robust predictive powers for future housing returns with a salient short-run underreaction and long-run overreaction pattern. Further analysis shows that local housing sentiment impacts are asymmetric, and housing returns in cities with relatively inelastic housing supply are more sensitive to local housing sentiments. We also document a significant feedback effect between housing returns and market sentiments, indicating the existence of a pricing-sentiment spiral which could potentially enhance the ongoing market fever of Chinese housing markets. The main estimation results are robust to alternative sentiment extraction methods and alternative sentiment proxies, and consistent for the sample period before COVID-19.

✉ Jindong Pang
  jindong.pang@whu.edu.cn

  Shulin Shen
  shen_shulin@hust.edu.cn

  Yiyi Zhao
  yiyizhao@hust.edu.cn

[1] School of Economics, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, Hubei, China

[2] Economics and Management School, Wuhan University, Hubei 430072 Wuhan, China

Introduction

In the past several decades, housing prices have skyrocketed in major Chinese cities. The annual growth rate of real Chinese national housing prices could be as high as 16.7% from 2006 to 2010 according to Wu et al. (2014), and the annual growth rate of housing prices in first-tier cities of China was as high as 13% over 2003–2013 according to Fang et al. (2016). While some scholars (e.g., Chow & Niu, 2015; Tan et al., 2020; Wang & Zhang, 2014; Wu et al., 2016) have documented how these housing price appreciations are driven by fundamental factors of demand and supply, still a large part of these price appreciations remained unexplained.

Advances in behavioral finance have offered one potential explanation for this housing market fever in China, i.e., investor sentiment (or market sentiment). Originating from "animal spirits" by Pigou (1927) and Keynes (1936), Shiller (2000) proposes that the "irrational exuberance" of investors renders them to rely on many psychological factors for asset valuations. These early studies have laid the foundations for research on investor sentiment in the stock and many other financial markets (e.g., Baker & Wurgler, 2006, 2007). In behavioral finance models, investor sentiment, independent of market fundamentals, is believed to play a role in price determination and could induce some investors to possess systematic biases in their beliefs. Many studies have shown how market sentiments greatly affect stock returns (Baker & Wurgler, 2006, 2007; Da et al., 2015; Huang et al., 2015).

In viewing the important role of market sentiment in real estate markets, there is a growing literature examining the market sentiment on housing returns and dynamics, with the majority of the study focusing on commercial real estate markets (including REIT markets) in developed markets. Examples of these studies include Clayton and MacKinnon (2002, 2000), Gallimore and Gray (2002), Clayton et al. (2009), Lin et al. (2009), Ling et al. (2014), Jin et al. (2014), Das et al. (2015), Letdin et al. (2021) and so on. All these studies on commercial real estate markets and REIT markets highlight the significant role of market sentiments in the valuations of real estate properties.

However, empirical examinations of market sentiment on Chinese residential housing markets are limited. More formally, housing market sentiment could be defined as a misguided belief about housing price appreciations, which cannot be justified by the current economic information set available to housing market participants, as in Ling et al. (2015). As emphasized in Hui and Wang (2014) and Hui et al. (2017), the private (versus commercial) housing sector stands out as an ideal "victim" of investor sentiment for several reasons. First, market participants in private housing markets are mostly individuals and households with limited knowledge and information and hence more susceptible to sentiment. Second, the illiquidity of housing properties leads to high segmentation of the market and asymmetry of information. Among all housing sectors, the private housing market has the highest liquidity and makes it resemble the stock market. Lastly, short-selling restrictions render mispricing in the private housing market hard to

eliminate. As the real estate market is the mainstay of this emerging economy and housing composes a large part of household asset holdings, examining the impacts of investor sentiments on the private housing market is imperative, as emphasized in Case and Shiller (2003), Shiller (2010) and Case et al. (2012).

In this paper, we try to fill this gap by applying sentiment analysis techniques to housing markets in China, intending to examine how housing sentiments affect local housing markets in this emerging economy. To achieve this goal, we first construct city-specific housing sentiment indices for 18 major Chinese cities and then conduct a series of empirical analyses to examine the sentiment impacts on local housing returns. More specifically, we construct three sentiment proxies representing the local housing market liquidity and speculative behaviors from a massive second-hand transaction dataset and then use the recursive **look-ahead-bias-free** implementation of the **partial least squares** (**PLS**) method to extract a local housing sentiment index from these three proxies for each city considered. A panel regression of housing return forecasting is then examined with local housing sentiments as an explanatory variable. To investigate the persistence of sentiment impacts, we also estimate impulse responses of cumulative housing returns to sentiment shocks over different return horizons.

For sentiment construction, the first two proxies are the median time on the market of sold housing units and the turnover rate of listed housing units in a given month, serving as two housing market liquidity measures. These two housing market liquidity measures resemble the turnover rate of the stock market in Baker and Wurgler (2006). For the stock market, Baker and Stein (2004), Baker and Wurgler (2006), and Baker and Wurgler (2007) suggest that turnover, or more generally liquidity can serve as a sentiment index. For the real estate market, Clayton et al. (2008) and Ling et al. (2015) point out that increased liquidity is the amplifying channel of a pricing-sentiment spiral in the housing market. The third sentiment proxy is constructed as a small-house return premium over large houses. This small-house return premium measures the speculative investment in Chinese housing markets, as speculative investors in China tend to over-invest in smaller housing units because of less capital requirement and higher liquidity, and hence smaller housing units appreciate faster in cities with more speculative investors.

With these three local housing sentiment proxies, we first orthogonalize each of these proxies against a set of fundamental variables and then implement the recursive **look-ahead-bias-free** estimation of the **PLS** method by Kelly and Pruitt (2013, 2015) and Huang et al. (2015) to construct the local housing market sentiment for each city in our sample. Compared with traditional dimension reduction methods such as **principal component analysis** (**PCA**), **PLS** filters out irrelevant common components and summarizes the most useful information in these proxies to construct a single complex sentiment index with the largest covariance with the targeted series of predictions. To remove the potential **look-ahead bias** of the in-sample estimation of **PLS**, we recursively conduct the **look-ahead-bias-free** implementation of **PLS** with an expanding window scheme to obtain a **look-ahead-bias-free** estimate of the housing sentiment sequence.

The panel housing return forecasting regressions show that local housing sentiment indices have robust predictive powers for future housing market returns, while

the impulse response estimations indicate a salient pattern of short-run underreaction and long-run overreaction. Further analysis shows that local housing sentiment impacts are asymmetric in that local housing sentiments only have significant impacts on housing returns when the market sentiment is below the sample average. In addition, the sentiment effects are much stronger and more significant for cities with a relatively inelastic housing supply. Housing sentiments exhibit strong inertia and are positively correlated with past short-term cumulative housing returns, implying that housing sentiments have a backward-looking property and there exists a significant feedback effect between housing returns and sentiments.

The documented sentiment impacts are comparable to what has been documented for developed housing markets. As shown in the results section, a one-standard-deviation increase in local housing sentiments will cause the annual housing return to increase by about 1.68% in China, while this effect is approximately 5.58% for the U.S. housing market in Soo (2018) and approximately 3.73% in Bork et al. (2020).[1] These results indicate that the housing markets in China are also susceptible to irrational sentiments, with arbitrage and speculation prevalent in this emerging market.

To prove that the recursive **look-ahead-bias-free** implementation of the **PLS** method does not lead to "**false detection of predictability**", we conduct a placebo test in which three randomly-generated sentiment proxies are drawn from the standard normal distribution and then summarized into a "placebo" sentiment index by the recursive **look-ahead-bias-free** procedure. We find no significant predictive power of this "placebo" sentiment on future housing returns. The placebo test documents that the significant predictability of our extracted sentiment index is not generated mechanically by the recursive **look-ahead-bias-free** implementation of the **PLS** method.

Our main regression results are also robust to alternative sentiment construction methods, such as the **scaled PCA** method by Huang et al. (2022) and the traditional **PCA** method, and are consistent for the sample period before COVID-19. Also, we find consistently significant predictability power of an alternative sentiment index based on the number of newly listed properties (instead of the turnover rate as a proxy) on future housing returns.

Housing market sentiment is particularly difficult to measure compared with other financial markets for several reasons. First, as stated in Soo (2018), typical sentiment proxies for the stock market, such as closed-end fund discounts, mutual fund flows, and dividend premiums, do not have straightforward counterparts for the housing market. Second, housing markets are highly segregated geographically. The housing market conditions for different cities vary substantially even within a given region during the past housing cycle (Ferreira & Gyourko, 2012). Third, transactions in the housing market usually take several months or even years to complete. Hence, the impact of sentiments on housing market outcomes may take a long time to unfold.

In this paper, we choose to follow the studies on market-based (indirect) stock market sentiments to construct housing market sentiments (Baker & Wurgler, 2006; Da et al., 2015; Huang et al., 2015). The two housing market liquidity measures of

---

[1] For detailed analysis, see "Predicting Returns with Local Housing Sentiments" section.

our sentiment indices mimic the turnover rate of the stock market. The speculation measure of local housing markets could serve as an alternative to first-day returns on IPOs, which measures investor enthusiasm for the stock market. More importantly, by the richness of the transaction dataset, we can construct a sentiment index for each city considered, capable of reflecting heterogeneous local housing market conditions. And the impulse response analysis of cumulative housing returns to sentiment shocks over different return horizons enables us to examine how the sentiment impacts unfold with time.

Compared with the textual approach in Soo (2018), which develops housing sentiments for 34 cities across the United States by quantifying the qualitative tone of local housing newspaper articles, our market-based proxies may be a delayed reflection of market sentiments. However, these transaction-based proxies provide a good complementary to the textual sentiments when representative local housing newspapers are not available. Compared with the survey method to quantify market sentiments, such as those in Bork et al. (2020) and Ling et al. (2015) which use household survey responses to questions about buying conditions for houses in the U.S., our transaction-based sentiment index is an indirect measure of market sentiment.[2] But given the fact that surveys regarding Chinese customers are limited, our transaction-based sentiment index is a good alternative to those survey-based measures.

Moreover, even though internet usage is common nowadays, an aggregate search index, such as the mortgage default risk based on the Google search index of Chauvet et al. (2016) for the U.S. market, cannot provide localized sentiment measures given the heterogeneous properties of local housing markets. For Chinese real estate markets, Zheng et al. (2016) use Google search counts to construct city-specific confidence indices for 35 Chinese cities from 2005 to 2013. The confidence index constructed by Zheng et al. (2016) is calculated as the ratio of the count of positive entries to the total count of both positive and negative entries regarding the targeted real estate market. For the sample period considered in this paper (2016 to 2020) and the sample period of Zheng et al. (2016), Google search has been blocked and hence Google search entries may not work as a good indicator of confidence index or investor sentiments. The sentiment indices constructed in this paper are based on market transaction outcomes, which are the results of demand and supply factors,[3] and could serve as a complement to these methods listed above, especially the city-specific confidence index by Zheng et al. (2016).

The contribution of this paper lies in three parts. **First, we contribute to a growing literature evaluating the effect of market sentiments on housing**

---

[2] Other survey-based sentiment proxies, such as Anastasiou et al. (2021) which use European Commission's consumer surveys to construct sentiment indicators representing inflationary expectations and precautionary saving incentives, and Jin et al. (2014) which use the U.S. Conference Board Consumer Sentiment Index (a measurement of consumer perception on market conditions), are more of an overall confidence index towards the entire economy than a direct measure of housing sentiment.

[3] We do not consider proxies such as housing construction or housing investment in the construction of sentiment index, in that housing investment and construction, which are more of a supply side measure, usually takes several years to accomplish and hence may contain un-synchronous information with housing transaction outcomes.

markets by analyzing a representative sample of cities in China. Among existing studies on housing sentiments regarding developing markets (Hui & Wang, 2014; Hui et al., 2017, 2018; Lam & Hui, 2018; Soo, 2018; Zhou, 2018), most of them construct a sentiment index for a specific city and examine the relationship between sentiments and housing returns of the city such as Shanghai or Hong Kong, which lack a representative dataset. Aided by massive transaction data, we construct city-specific housing sentiment indices for 18 major cities in China, which are more representative of Chinese housing markets. This panel data structure of our sentiment indices provides more reliable and comprehensive empirical findings on the relationship between local housing sentiments and returns. This paper confirms that investor sentiment is an important factor in Chinese housing markets, suggesting arbitrage and speculation are prominent in this emerging market.

Second, we provide more evidence on the role of local housing supply conditions in analyzing housing market sentiment impacts. Existing studies have shown that supply-side conditions greatly affect housing prices and constraints on housing supply can explain large differences in housing price dynamics in different regions (Glaeser & Ward, 2009; Glaeser et al., 2006, 2008; Green et al., 2005; Saiz, 2010). The role of housing supply may be more significant in Chinese housing markets as indicated by Liu (2014) and Wang et al. (2012). However, supply-side conditions have been overlooked in the empirical examinations of housing sentiment impacts on local housing returns (except Zheng et al., 2016). In this paper, we take advantage of supply elasticity estimates from Liu et al. (2019) to study whether local supply conditions alter market sentiment impacts. Our result shows that housing markets in cities with relatively inelastic housing supply are more sensitive to local housing sentiments than places with elastic housing supply. For cities with a much more elastic housing supply, the impacts of market sentiments on returns are much smaller or even negligible. The finding that local housing supply elasticities play an important role in the sentiment effects highlights the necessity of considering local market conditions in the analysis of sentiment analysis.

Third, we also contribute to the literature on China's high housing price puzzle. There exist many theoretical as well as empirical works documenting and explaining why the housing price level or growth rate is so high in China (Chen & Wen, 2017; Fang et al., 2016; Glaeser et al., 2017; Wei et al., 2012; Wu et al., 2016). Many of these works debate the existence of a bubble in Chinese housing markets by analyzing the fundamentals underlying this emerging market. The findings in this paper provide some evidence on this issue from the aspect of the nonfundamental market sentiments. Our local housing sentiment indices are shown to impact local housing returns significantly in addition to many fundamental controls, indicating nonfundamental sentiments also add up to the housing market fever of China. The short-run underreaction and long-run overreaction pattern of housing sentiments imply that there exist irrational investors in the Chinese housing market, exaggerating the turmoil of local housing markets in the short run. We also document that the expectations of Chinese real estate investors are backward-looking and there exists a significant feedback effect between housing returns and market sentiments. This pricing-sentiment spiral could enhance the ongoing market fever of Chinese housing

markets and potentially extend the length and magnitude of housing market cycles in China.

In brief, our empirical analysis enriches the literature on the role of sentiments in the housing market and provides better insights into the housing market dynamics, which may help policymakers to stabilize and improve the functioning of the housing markets. The rest of this paper is organized as follows. In "Data and Housing Price Index" section, we introduce our transaction data and construct second-hand housing price indices. Descriptions of sentiment proxies and the construction of city-level housing sentiment indices are discussed in "Housing Sentiment Index" section. "Local Housing Sentiments and Returns" section studies the relationship between housing market sentiments and returns. "Robustness Checks" section provides some robustness checks on pre-COVID-19 estimation, alternative sentiment construction methods, alternative sentiment proxies, and the placebo test. "Conclusions" section concludes.

Data and Housing Price Index

Transaction Data

In this paper, we use a massive second-hand housing transaction dataset to construct the housing price index and the sentiment index.[4] The private (residential) transaction records are retrieved from Lianjia.com, one of the most popular second-hand housing trading platforms in China. As a real estate brokerage firm, Lianjia helps housing sellers to list their properties online and provides brokerage services for potential sellers and buyers. Once a transaction (sale) has been concluded, Lianjia records the sale and posts its information in the *Transacted-Units* section on its website. Each transaction record includes detailed information on the characteristics of the sold housing unit: listing date, listing price, transaction date, transaction price, and project name (*xiao qu*), as well as unit attributes, such as floor area, layout, decoration, orientation, etc.

We collect all the transaction records from this *Transacted-Units* section on Lianjia.com for each city in our sample. In total, we obtain 1,799,272 transaction records from January 2015 to October 2020 covering 18 major cities in China. Table 1 lists all the cities and their corresponding sample periods. The earliest date of the transaction data is January 2015 for Shanghai and Tianjin. In the following, we first use the transaction data to construct the Hedonic Housing Price Index (HPI) and then derive local housing sentiment using the recursive **look-ahead-bias-free** implementation

---

[4] The reason for not using new house transaction data is that housing prices for new houses are highly regulated by local governments during our sample period. In contrast, second-hand housing transactions are not subject to price ceilings set by local governments and thus are more representative of local housing market conditions.

Table 1 Transaction data and sentiment sample periods

Num	City	Transaction Data	Look-ahead-bias-free Sentiment
1	Shanghai	2015/01 ~ 2020/10	2016/01 ~ 2020/10
2	Beijing	2016/07 ~ 2020/10	2017/07 ~ 2020/10
3	Nanjing	2015/10 ~ 2020/10	2016/10 ~ 2020/10
4	Xiamen	2015/11 ~ 2020/10	2016/11 ~ 2020/10
5	Hefei	2017/01 ~ 2020/10	2018/01 ~ 2020/10
6	Dalian	2015/08 ~ 2020/10	2016/08 ~ 2020/10
7	Tianjin	2015/01 ~ 2020/10	2016/01 ~ 2020/10
8	Guangzhou	2016/02 ~ 2020/09	2017/02 ~ 2020/09
9	Chengdu	2015/05 ~ 2020/10	2016/05 ~ 2020/10
10	Hangzhou	2016/03 ~ 2020/10	2017/03 ~ 2020/10
11	Wuhan	2016/02 ~ 2020/10	2017/02 ~ 2020/10
12	Shenyang	2017/01 ~ 2020/09	2018/01 ~ 2020/09
13	Jinan	2015/12 ~ 2020/10	2016/12 ~ 2020/10
14	Shenzhen	2015/09 ~ 2020/10	2016/09 ~ 2020/10
15	Suzhou	2015/08 ~ 2020/10	2016/08 ~ 2020/10
16	Chongqing	2017/01 ~ 2020/07	2018/01 ~ 2020/07
17	Changsha	2016/03 ~ 2020/10	2017/03 ~ 2020/10
18	Qingdao	2016/11 ~ 2020/10	2017/11 ~ 2020/10

Sample periods are chosen by their availability. As discussed below, we use the first twelve months' transaction data to get the first sentiment estimate. Hence, the sample periods for the look-ahead-bias-free sentiment start one year later than their transaction data

of the **PLS** method for each city. Since 2016, a large-scale purchase restriction policy and a series of tightening policies were introduced to fight against speculative housing investment and rapid housing price growth.[5] Hence, our analysis is under a general tightening policy environment.

Housing Price Index Based on a Hedonic Model

The first step of our empirical study is to construct a monthly quality-adjusted housing price index (HPI) for each city in our sample. We use the hedonic regression to build the HPI. The Hedonic model is introduced into the housing markets by Rosen (1974). According to the theory, housing prices contain a part that is driven by some attributes of the house itself (area, layout, orientation, and so on). Following this

---

[5] In March 2016, the Fourth Session of the 12th National People's Congress proposed to stabilize housing prices in first—and second-tier cities and destock in third—and fourth-tier cities. On September 30th, 2016, Beijing Municipal Commission of Housing and Urban–Rural Development and other departments issued the official document "*Several measures to promote the steady and healthy development of the real estate market in cities*", (http://zjw.beijing.gov.cn/bjjs/gcjs/tzgg36/391380/index.shtml) which raised the down payment ratio of the second house to 50% and started a new round of purchase and loan restriction policy nationwide.

idea, if these properties were decomposed from the price and then the remaining part is purely driven by supply and demand factors.

Accordingly, our HPI construction is based on the following equation:

$$\ln\left(P_{ilt}\right) = \alpha + \sum_{t=2}^{T} \beta_t D_t + \gamma X_{ilt} + \pi_l + \varepsilon_{ilt}, \tag{1}$$

where P_{ilt} represents the trading price per square meter for housing unit i in the project (*xiao qu*) l on transaction date t. X_{ilt} denotes a vector of unit attributes, including floor area, elevator, number of bedrooms, floor level, orientation, and decoration type (fancy, simple, or other). We also control for the project-fixed effect π_l. Note that a project in Chinese cities is a very small geographic unit, similar to a Census block in the U.S. The project-fixed effects control for neighborhood amenities, including the school district, crime rate, traffic service, distance to CBD, and so on. The error term is denoted by ε_{ilt}.

A series of month dummies, D_t, are included in the regression to capture the time variation in the housing prices clear of quality changes. The value of HPI for month t is then calculated as $exp(\beta_t) \times 100$ and the value is set as 100 for the origin month. We apply the hedonic regression for each city in our sample and obtain 18 city-specific monthly HPI series from January 2015 to October 2020. It is worth noting that the final HPI series is an unbalanced panel with the earliest date in January 2015 for Shanghai and Tianjin.

Housing Sentiment Index

In this section, we use second-hand transaction data to construct our housing sentiment indices. We first introduce our sentiment proxies and then illustrate the **recursive look-ahead-bias-free** procedure of the **PLS** method from which the housing sentiment index is constructed. Finally, we display some correlation analysis between our sentiment indices and several confidence indices from official sources.

Sentiment Proxies

In the construction of sentiment proxies, we majorly follow the approach of Zhou (2018). We construct three sentiment proxies based on our trade-by-trade data. The first proxy is *MedianIntv*, the natural logarithm of the median holding period of sellers, i.e., the median number of days on the market of sold housing units in a given month. Intuitively, a lower level of *MedianIntv* would indicate a relatively higher market sentiment.

The second proxy is *Turnover*, the ratio of housing areas being sold to the total housing areas for sale in a given month. To construct this proxy, for each month, we divide the total floor areas transacted in this month by the floor areas of listed properties including transacted houses and available-for-sale houses (which have been listed but not yet sold until the current month). Note that this *Turnover* ratio is a quasi-turnover ratio of the housing market. The real turnover rate of the housing

market is the ratio of trading volume to the total housing stock as in Ling et al. (2014). Usually, the real turnover rate of the housing market is small, since most housing units are not for sale. By definition, this **Turnover** proxy measures the liquidity of listed properties in the housing market and usually, it is much larger than the real turnover ratio of the housing market.

Essentially, constructed in the same spirit as measures in Clayton et al. (2008), Ling et al. (2014), and Ling et al. (2015), **MedianIntv** and **Turnover** measure the liquidity of real estate markets. Unlike **MedianIntv**, **Turnover** should be positively correlated with the housing market liquidity since a liquid housing market should be associated with a shorter time on the market and a high turnover rate. As the correlation matrix in Table 9 of Appendix A shows, **Turnover** is significantly negatively correlated with **MedianIntv**. Meanwhile, **Turnover** exhibits a positive correlation with the contemporaneous housing return R_t while **MedianIntv** exhibits a negative correlation with R_t. These correlation coefficients provide some support for the positive (negative) correlation between **Turnover** (**MedianIntv**) and market liquidity.

The third proxy is **SMB**, the small-house return premium over large houses. Each year we calculate the quintile breakpoints of the transacted houses' floor areas and then divide houses into five groups according to the latest breakpoints. Small (big) houses are those in the first (fifth) quintile. Then we use the Hedonic method in "Housing Price Index Based on a Hedonic Model" section to construct a housing price index for small houses and big houses, respectively. In the end, **SMB** is equal to the small-house return minus the big-house return (in percentage).

Zhou (2018) states that over-optimistic investors tend to buy big houses and push up prices for these large houses. Hence, the proxy **SMB**, which measures the return premium of smaller houses, is expected to be negatively correlated with sentiments according to Zhou (2018). However, in contrast to this argument, speculative investors in the Chinese housing market tend to over-invest in smaller houses for several reasons. First of all, the financial pressure of speculating in small houses is relatively low because of lower total costs and hence less capital requirement. Secondly, the risk in investing in small houses is much lower as a flipper of small houses could more easily find a buyer in the market because of the higher liquidity and the dominant demand for small houses. Thirdly, the potential return on investing in small houses is higher since the price appreciation tends to be higher for smaller houses.

For the Singapore housing market, Fu and Qian (2014) also find that short-term speculators typically target smaller units as they require less capital and are easier to sell. Furthermore, some online news reports also reveal that small housing units, the primary target of speculators, are always sold out even during periods of a market downturn.[6] Some speculators were personally interviewed and admitted that "*they invest in small houses because of the high return-on-investment ratio*".[7] In short, from the point of view of speculating behavior, the prices for smaller housing units

---

[6] An agent of a real estate firm in Yantai (a Chinese city) suggests that "*while the big houses stay lag in sale, the smaller houses are sold well, even in the periods of a housing market downturn*" according to the report available at http://sd.dzwww.com/sdgd/sdgdxw/201112/t20111214_6817566.htm.

[7] Please find the news report at http://news.sohu.com/20170405/n486498292.shtml.

appreciate faster in cities with more speculative investors, and hence **SMB** should be positively correlated with housing sentiments. We will examine the correlation between **SMB** and housing sentiment in the next section.

There are two more proxies used by Zhou (2018), namely, the newly-opened housing construction area, and the housing sector investment to construct the housing sentiment for Shanghai. In this paper, we choose to construct the housing sentiment using **MedianIntv**, **Turnover**, and **SMB** only for three reasons. Firstly, since we are constructing sentiments for 18 Chinese cities, data for the other two proxies are not available for some cities in our sample. Secondly, this paper tries to construct a market outcome-based sentiment. Newly-opened housing construction and housing investment reflect sentiments of the supply side, which has been reflected in market outcomes to some extent. So, omitting these two proxies in the construction of the sentiment will not result in big information loss. Finally, housing investment and construction usually take several years to accomplish. So, sentiments from the supply side may not be synchronous with these based on market outcomes. Hence, omitting the housing investment and construction variables and relying on the other three transaction-based proxies will leave us with a more synchronous sentiment measure.

At the end of this section, we display some summary statistics for these three proxies. To save space, Table 2 only shows the average values of these three proxies for Beijing, Shanghai, Guangzhou, and Shenzhen, respectively. For example, the average median days on the market of second-hand properties in Beijing is 23.54 days (exp(3.159)) in 2016, while this value increases to 70.39 days (exp(4.25)) in 2020. Also, the turnover rate of listed second-hand properties has decreased from 46 to 29% during the same period for Beijing. The average small-house return premium is about 0.24% for Beijing in 2016, decreasing to a negative value of -0.17% in 2020.

Housing Sentiment Index

With these above three proxies, we then construct sentiment indices for the 18 cities in our sample. First, all these raw proxies are standardized with zero means and unit standard deviations. Second, to eliminate the impact of the business cycle, we regress the standardized proxies on some macroeconomic variables and obtain the residuals. Following Zhou (2018), we choose four macroeconomic variables including Purchasing Managers' Index (*PMI*), the growth rate of consumer price index (*CPI*), the growth rate of M2 (*M2G*), and the difference between the yield of AA-grade corporate bonds and AAA-grade corporate bonds (*Default*). The residuals from these regressions contain sentiment information that is orthogonal to business cycle factors.

Third, to smooth out jumps, we impose a three-month moving average to the residuals, as in Huang et al. (2015). The smoothed residuals are called "clean" proxies. Fourth, considering that the clean proxies may have a lag in reflecting the sentiment of market entities, we need to choose between the current value and the lagged value of each proxy, as in Baker and Wurgler (2006). In detail, we first perform principal

Table 2 Housing sentiment proxies from 2016 to 2020

City	Year	MedianIntv	Turnover	SMB
Beijing	2016	3.159	0.462	0.235
	2017	3.511	0.306	-0.283
	2018	3.809	0.252	0.239
	2019	4.157	0.212	0.211
	2020	4.254	0.296	-0.172
Shanghai	2016	6.108	0.102	1.320
	2017	6.679	0.117	-0.178
	2018	4.451	0.171	-0.501
	2019	4.136	0.220	0.074
	2020	4.055	0.313	0.055
Guangzhou	2016	3.836	0.195	0.698
	2017	3.776	0.317	0.402
	2018	3.867	0.193	0.561
	2019	4.332	0.142	-0.323
	2020	4.530	0.273	-0.941
Shenzhen	2016	3.591	0.270	0.460
	2017	4.006	0.235	0.072
	2018	4.077	0.190	0.254
	2019	4.274	0.195	0.302
	2020	3.940	0.353	0.146

Values are averaged by year for each city presented. *MedianIntv* is the natural logarithm of the median number of days on the market. *Turnover* equals the total area of transacted houses divided by the total area of houses available for sale in a given month, and *SMB* equals the small-house return minus the big-house return (in percentage)

component analysis on ***MedianIntv$_t$***, ***MedianIntv$_{t-1}$***, ***Turnover$_t$***, ***Turnover$_{t-1}$***, ***SMB$_t$***, and ***SMB$_{t-1}$***, to extract its first principal component. Then the correlation between the first component and each of the six variables is calculated to determine whether to choose the current value or the lagged value. Finally, the value which is more correlated with the first component is chosen as the final sentiment proxy.

Fifth, we conduct the **recursive look-ahead-bias-free** implementation of the **PLS** approach of Kelly and Pruitt (2013, 2015) and Huang et al. (2015) to construct a **look-ahead-bias-free** housing sentiment index for each city in our sample.

Specifically, at month *t*, for each sentiment proxy *proxy$_i$*(***MedianIntv***, ***Turnover***, or ***SMB***), we run the following **first-stage regression:**

$$proxy_{i,s-1} = \pi_{i,0} + \pi_{i,1} Return_s + u_{i,s-1}, s \leq t, \tag{2}$$

in which *proxy$_{i,s-1}$* denotes the lagged sentiment proxy (proxy *i* at time *s* − 1) and *Return$_s$* denotes housing return (in percentage) at time *s*. To obtain the loadings for month *t*, we only use data up to month *t* in the above regression. The latest return (*Return$_t$*) used on the right-hand side is from time *t*-1 to *t*. The latest
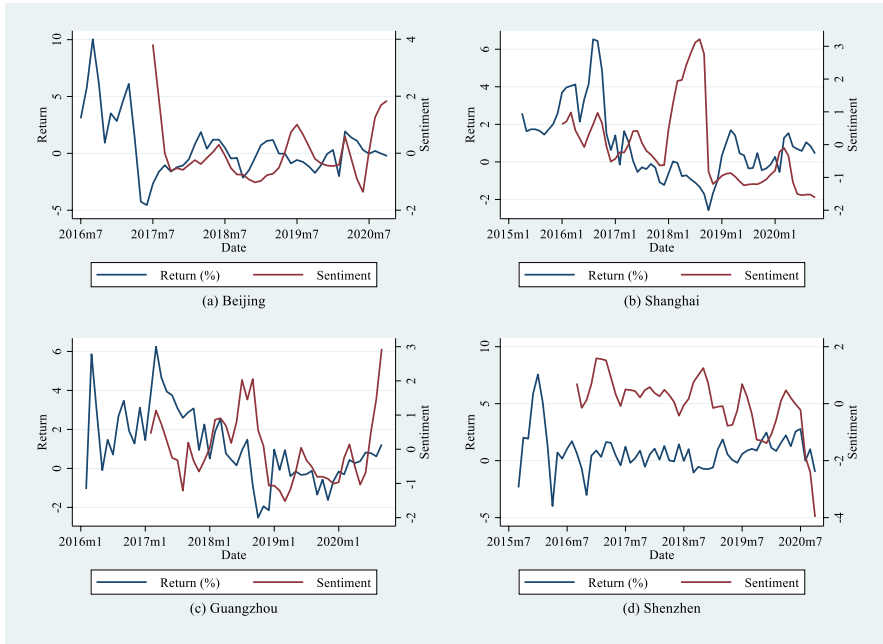
Fig. 1 Housing sentiments and returns. *Notes*. The blue line represents hedonic HPI returns (in percentage) and the red line represents housing sentiments

sentiment proxies used on the left-hand side is $proxy_{i,t-1}$. Thus, the first-stage coefficient estimates $\hat{\pi}_{i1}$ are in the time t information set since they use monthly returns $\{Return_2, \ldots, Return_t\}$ and monthly proxies $\{proxy_{i,1}, \ldots, proxy_{i,t-1}\}$.

In the **second-stage regression**, for month t, we run the cross-sectional regression as follows:

$$proxy_{i,t} = \alpha + S_t^{PLS}\hat{\pi}_{i1} + v_{i,t}, i = 1, 2, \ldots, N. \tag{3}$$

In the above regression, the independent variable is the loadings that we obtain in Eq. (2). Then the slope estimate S_t^{PLS} obtained in the above regression is the **look-ahead-bias-free** sentiment estimate for month t.

For sentiment estimate S_{t+1}^{PLS} at month $t+1$, we re-estimate the above first-stage regressions by expanding the first-stage sample to $\{s: s \leq t+1\}$ and re-estimate the second-stage regression for time-period $t+1$. By recursively implementing the above procedure with an expanding estimation window, we obtain a recursive **look-ahead-bias-free** estimate of the housing sentiment sequence $\{S_t^{PLS}, t = t_0, \ldots, T\}$.

Note that the first available sentiment estimate comes on the date t_0, which is usually later than the starting period of the proxy and housing return sample. This sample loss is because we need to use an initial training data set $\{s: s \leq t_0\}$ to estimate the first-stage regression for the month t_0. In the empirical implementation, we use the first twelve months as the initial training sample and hence our final sentiment sample starts one year later than the transaction data as exhibited in Table 1.

Table 3 Correlation of local housing sentiments with three sentiment proxies

Num	City	MedianIntv	Turnover	SMB
1	Shanghai	-0.366	0.035	0.023
2	Beijing	-0.554	0.206	-0.159
3	Nanjing	-0.400	0.270	0.003
4	Xiamen	-0.379	-0.030	-0.080
5	Hefei	0.197	0.488	0.002
6	Dalian	-0.284	0.219	0.103
7	Tianjin	-0.558	0.113	0.117
8	Guangzhou	0.433	-0.471	0.010
9	Chengdu	-0.684	0.481	-0.058
10	Hangzhou	-0.790	0.486	0.075
11	Wuhan	-0.651	0.187	0.003
12	Shenyang	-0.438	-0.549	-0.071
13	Jinan	-0.776	0.321	0.129
14	Shenzhen	-0.287	0.412	0.157
15	Suzhou	-0.795	0.454	0.098
16	Chongqing	0.475	-0.230	-0.035
17	Changsha	-0.635	0.280	0.098
18	Qingdao	-0.590	0.246	-0.061

Values are the correlation coefficients between the local housing sentiment index and three sentiment proxies for each city. *MedianIntv* is the logarithm of the median days on the market. *Turnover* equals the total area of transacted houses divided by the total area of houses available for sale in a given month, and *SMB* equals the small-house return minus the big-house return (in percentage)

We display the housing sentiment index (in red) and the hedonic HPI returns (in blue) for selected cities in Fig. 1. As Fig. 1 shows, the trend of housing sentiments is similar to that of housing returns, especially in Beijing, Guangzhou, and Shenzhen. This co-movement pattern indicates a positive sentiment-return relationship. We also calculate the contemporaneous correlation between the sentiment index (S_t^{PLS}) and the housing return R_t as shown in Table 9 in Appendix A. The correlation coefficient (0.327) is similar to that of Zhou (2018) with a 0.26 correlation between her sentiment indices and returns for the Shanghai housing market.

Correlation with Proxies and Confidence Indices

To confirm the reliability of our sentiment indices, we first calculate the correlations between S_t^{PLS} and these three sentiment proxies. As Table 9 in Appendix A shows, the housing sentiment S_t^{PLS} is positively correlated with **Turnover** and negatively correlated with **MedianIntv**. These results are consistent with the findings on the positive correlation between market liquidity and sentiments in the literature. To further investigate the correlation between local sentiment and proxies, we calculate their correlation for each city and display the results in Table 3. Results in Table 3 verify the positive correlation

between sentiment and ***Turnover*** and the negative correlation between sentiment and ***MedianIntv***.

Interestingly, distinct from Zhou (2018), we find that 12 out of 18 cities show a positive correlation between sentiment and ***SMB***. As argued by Zhou (2018), ***SMB*** should be negatively correlated with housing sentiment. However, this is not the case for Chinese housing markets. As mentioned above, speculative investors in Chinese housing markets tend to over-invest in smaller houses, and hence prices for smaller housing units appreciate faster in cities with more speculative investors. So, the proxy ***SMB*** reflecting housing investors' speculative behavior should be positively correlated with housing sentiments. Our empirical results in Table 3 verify the prevalence of speculative investors in Chinese housing markets. Cities such as Shanghai, Nanjing, Hefei, Dalian, Tianjin, Guangzhou, Hangzhou, Wuhan, Jinan, Shenzhen, Suzhou, and Changsha, all show significantly positive correlations between local housing sentiments and the small-house return premium. These findings indicate that speculative behaviors are prevalent in Chinese housing markets, and over-optimistic investors are not buying bigger houses but instead over-investing in smaller housing units.

We also compare our sentiment index with some official indices for Beijing and Shanghai, the two largest cities in China. Since housing buyers include both consumers and investors (Han, 2013; Miller & Pandher, 2008), we consider consumer confidence indices and investor confidence indices. As Table 10 of Appendix A shows, our housing sentiment index of Shanghai is significantly positively correlated with the *Housing Boom Index of China* (HB_{CN}) but negatively correlated with the *Investor Confidence Index for Economic Policy* (IC_{CN_P}) and the *Investor Confidence Index of China* (IC_{CN}). The sentiment index of Beijing exhibits positive but insignificant correlations with the *Consumer Confidence Index of Beijing* (CC_{BJ}) and the *Consumer Expectation Index* of Beijing (CE_{BJ}). While these results provide some support to the reliability of our second-hand transaction-based housing market sentiment index, it also reveals that housing sentiment may exhibit distinct characteristics relating to macroeconomic indices.

Local Housing Sentiments and Returns

Predicting Returns with Local Housing Sentiments

In this section, we examine the relationship between local housing sentiments and housing market returns. The basic regression equation is set as follows:

$$R_{i,t+1} = \alpha + \beta S_{it}^{PLS} + \gamma Spring_t + \eta Autumn_t + \theta Macros_t + \delta Fundamentals_{it} + \pi_i + \lambda_t + \varepsilon_{it},$$
(4)

where $R_{i,t+1}$ denotes the housing market return of city i at time $t + 1$, calculated by the log return (in percentage) of HPI constructed in "Housing Price Index Based on a Hedonic Model" section. S_{it}^{PLS} is the housing sentiment of city i at time t. $Spring_t$ and $Autumn_t$ are seasonality dummy variables, standing for the cold and hot seasons

for Chinese housing markets, respectively.[8] $Macros_t$ refers to some macroeconomic control variables which only vary across time but not across cities, including Purchasing Managers Index (PMI_t), the growth rate of M2 ($M2G_t$), and the yield spread between AA-grade corporate bonds and AAA-grade corporate bonds ($Default_t$).

We also control for some city-specific economic fundamentals ($Fundamentals_{it}$), including the growth rate of the consumer price index (CPI_{it}), the growth rate of the population ($population_{it}$), the unemployment rate ($unemployment_{it}$), the growth rate of the average income of on-the-job employees ($income_{it}$), the growth rate of the average monthly per-square-meter rents ($rents_{it}$), the house price to rent ratio (PRR_{it}).[9] City-fixed effects (π_i) are included in the regression, and time-fixed effects, λ_t, are also controlled for in the last specification in the result table. All these control variables are obtained from the Wind dataset. Finally, ε_{it} denotes the regression residual.

Table 4 presents the results from Eq. (4) with standard errors clustered at the city level while Table 11 in Appendix A provides the detailed results of this regression. From Column (1) to Column (6), control variables are added step by step. In the first column, we only control for seasonality dummies ($Autumn_t$ and $Spring_t$) besides the sentiment variable and the city-fixed effects. In the second column, we further control for time-varying macroeconomic variables. City-specific fundamental controls are added in Column (3) and lagged macroeconomic and fundamental variables are further included in Column (4) of Table 4. In Column (5), we also include the first lag of housing returns in consideration of the persistence of housing returns. In the last column of Table 4, time-fixed effects are controlled for and hence these time-varying macroeconomic variables and seasonality dummies are omitted.

As shown in Table 11 in Appendix A, we can see that housing returns exhibit a salient seasonality pattern (with lower returns in Autumn) and decrease with the default risk ($Default_t$). Other macroeconomic variables' impacts are not persistent across different specifications. For fundamental controls, the unemployment rate and the inflation rate display significant negative impacts on housing returns according to the last three columns. Other fundamental controls' impacts are not significant or persistent across different specifications.

The most important finding from Tables 4 and Table 11 is that the local housing sentiment has a significant (at the 5% significance level) positive impact on future housing returns. This significant impact is consistent across different specifications, although with different magnitudes. Our conclusions are mainly based on Columns (6) with time-fixed effects of Tables 4 and Table 11. Column (6) shows that housing sentiments can strongly predict future housing returns even when we control for time-fixed effects. In detail, a one-standard-deviation increase in housing sentiments is positively associated with a future monthly return appreciation of approximately 0.14%. This result indicates that the annual housing return will increase by about

---

[8] $Spring_t$ equals 1 if month t is in March, April, or May, while $Autumn_t$ equals 1 if month t is in September, October, or November.

[9] PRR_{it} is calculated by dividing the hedonic HPI by $rents_{it}$. Note that $population_{it}$, $unemployment_{it}$, $income_{it}$ are yearly data, and we convert them into monthly series by linear interpolation.

Table 4 Predicting future returns by local housing sentiments

	(1)	(2)	(3)	(4)	(5)	(6)
	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
S_{it}^{PLS}	0.451**	0.447**	0.522***	0.296**	0.113*	0.137**
	(0.176)	(0.177)	(0.174)	(0.120)	(0.060)	(0.062)
$Autumn_t$	-0.991***	-1.025***	-0.922***	-0.638***	-0.619***	
	(0.197)	(0.202)	(0.186)	(0.176)	(0.152)	
$Spring_t$	0.240	0.212	0.280	0.289**	-0.191	
	(0.155)	(0.143)	(0.185)	(0.136)	(0.153)	
$R_{i,t}$					0.606***	0.598***
					(0.067)	(0.068)
$Macros_t$	No	Yes	Yes	Yes	Yes	No
$Fundamentals_{it}$	No	No	Yes	Yes	Yes	Yes
$laggedControl_{it}$	No	No	No	Yes	Yes	Yes
$CityFE$	Yes	Yes	Yes	Yes	Yes	Yes
$TimeFE$	No	No	No	No	No	Yes
$Num.ofcities$	18	18	18	18	18	18
R^2	0.114	0.129	0.174	0.352	0.494	0.519
$Obs.$	799	799	786	780	780	780

$Control_{it}$ contains both of $Macros_t$ and $Fundamentals_{it}$. $Macros_t$ refers to macroeconomic variables including PMI_t (Purchasing Managers Index), $M2G_t$ (the growth rate of M2), and $Default_t$ (the yield spread between AA-grade corporate bonds and AAA-grade corporate bonds). $Fundamentals_{it}$ contains CPI_{it} (the growth rate of the consumer price index), $population_{it}$ (the growth of population), $unemployment_{it}$ (the unemployment rate), $income_{it}$ (the growth rate of the average income of on-the-job employees), $rents_{it}$ (the growth rate of the average rents per square meter per month), and PRR_{it} (the house price to rent ratio). City-fixed effects are controlled for in each column. Time-fixed effects are controlled for in Column (6). In the last three columns, the first lagged values of $Macros_t$ and $Fundamentals_{it}$ are also included. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively. City-clustered standard errors are reported in parentheses

1.68% (0.14% × 12) in China given a one-standard-deviation positive shock to housing sentiments.

Soo (2018) also finds a positive predicting power of news media sentiments on quarterly housing returns for the U.S. housing markets. It is shown that for every one percent increase in four quarters of accumulated lagged sentiments, the future quarterly price appreciation is approximately 0.93%. With standard deviation of the news sentiments of value 1.5, annual housing appreciation will be 5.58% (1.5 × 0.93% × 4) to a positive one standard deviation of the news sentiment shock. The housing sentiment impacts on Chinese markets estimated in this paper are comparable (but smaller) to that inferred from Soo (2018). More recently, Bork et al. (2020) use household survey responses from the University of Michigan consumer surveys to construct a housing sentiment index for the U.S. housing market. The housing sentiment constructed by Bork et al. (2020) is associated with a standard deviation of 0.08 and is estimated to impact the quarterly housing price growth rate with a coefficient of 11.67. So, one standard deviation increase in housing sentiments will lead to a 3.73 (11.67 × 0.08 × 4)

percentage points increase in the annual housing appreciation. Our estimate of housing sentiment impacts on the Chinese housing markets is more comparable to the estimate from Bork et al. (2020).

We offer one explanation for this disparity as well as the similarity among these sentiment impacts. Soo (2018) draws the conclusion based on city-specific sentiment indices from media tone in local newspaper articles for 34 U.S. cities from 2000 to 2013, while Bork et al. (2020) construct an aggregate housing sentiment index for the U.S. housing market based on the University of Michigan consumer surveys for a much longer period of 1975 to 2017. In contrast, the housing sentiment indices constructed in this paper are based on market transaction outcomes for 18 major Chinese cities covering the period of 2016 to 2020. Within this sample period, Chinese housing markets are under strict regulation policies, such as purchasing restrictions on multi-house owners and so on. However, even given these strict regulations, the estimated annual housing sentiment impact is still as high as 1.68%, indicating that Chinese housing markets may be also highly susceptible to irrational sentiment compared with the U.S. market. As Baker and Wurgler (2006) suggested, assets with limited arbitrage were prone to be influenced by sentiment more easily. This comparable sentiment impact on housing returns in China under a stringent regulatory environment also indicates that there may be many arbitrage and speculation behaviors in the Chinese housing market.

Underreaction and Overreaction

Further, it is meaningful to explore whether the transaction-based sentiment captures fundamental information or just market sentiments. According to behavioral asset-pricing theories, if the sentiment captures investors' behavior bias, the cumulative response curve of asset price should exhibit short-run underreaction and long-run reversal. Liu et al. (2019) show that the effect of sentiments on stock returns behaves the pattern of short-run underreaction and long-run reversal in China's stock market. In this section, we want to explore whether the same pattern will appear in the housing market.

Following Liu et al. (2019), we adopt the local projection method of Jordà (2005) to estimate cumulative impulse responses of housing returns to local housing sentiment indices. As shown by Jordà (2005), this local projection method is more robust to misspecifications than conventional vector-autoregression (VAR) for estimating impulse responses. Specifically, we consider the following multiple-horizon predictive regressions:

$$\log\left(P_{ib}\right) - \log\left(P_{it}\right) = \alpha_b + \beta_b S_{it}^{PLS} + \delta_b Control_{it} + \pi_i + \lambda_t + \varepsilon_{it}, \qquad (5)$$

where b ranges from month $t+1$ to month $t+12$, extending the return horizon from 1 month to 1 year. P_{it} represents the second-hand HPI constructed in "Housing Price Index Based on a Hedonic Model" section. $\log\left(P_{ib}\right) - \log\left(P_{it}\right)$ measures the cumulative logarithmic return (in percentage) for city i from t to $t+b$, denoted by $R_{i,[t,t+b]}$, for $b = 1, \ldots, 12$. The cumulative impulse response of housing prices to S_{it}^{PLS} is captured by β_b at different horizons, which can be separately estimated by ordinary least squares. As in Column (6) of Table 4, we control for city-fixed effects (π_i), time-fixed effects (λ_t), and some fundamental variables including CPI_{it},

$population_{it}$, $unemployment_{it}$, $income_{it}$, $rents_{it}$, PRR_{it}, and their lagged terms in the estimation (variable definitions see "Predicting Returns with Local Housing Sentiments" section).[10]

Table 5 reports the estimated housing sentiment coefficients at different return horizons. We can find that the cumulative impulse response gradually increases, peaks at around 5 months, and reverses to a lower and statistically insignificant level at longer horizons. Figure 2 depicts the coefficients of housing sentiment β_b from Table 5 with the time horizon. The salient pattern of short-run underreaction and long-run overreaction is similar to what Liu et al. (2019) document for the Chinese stock market. The significant short-run impacts and insignificant long-run impacts indicate that the sentiment indeed leads to mispricing in China's housing markets in the short run. And the lack of permanent effects of housing sentiments on cumulative returns in the long run indicates that there is no fundamental information contained in the housing sentiments. This result verifies the existence of irrational traders in the Chinese housing market by standard behavioral asset-pricing models as in Barberis et al. (1998); De Long et al. (1990); Daniel et al. (1998); Hong and Stein (1999).

Asymmetric Effects of Sentiments

In this section, we decompose our sentiment S_{it}^{PLS} into a positive part and a negative part to investigate potential asymmetric housing sentiment impacts. We define $S_{it}^{pos} = max\{S_{it}^{PLS}, 0\}$ as the positive sentiment, and $S_{it}^{neg} = min\{S_{it}^{PLS}, 0\}$ as the negative sentiment.[11] Then the S_{it} in Eq. (5) is replaced by S_{it}^{pos} and S_{it}^{neg} while other control variables remain the same. The empirical results are displayed in Table 6. Table 6 shows that S_{it}^{neg} has significantly positive impacts on cumulative returns at short horizons (b = 1, 3, 6) while S_{it}^{pos} has no significant coefficients for all horizons. These results indicate that local housing sentiments have an asymmetric impact on housing returns, with only negative sentiments having significant positive effects on future housing returns.

Our findings are distinct from Zhou (2018) who shows that high positive sentiments are followed by low housing returns. On one hand, our conclusions are based on panel data analysis for 18 cities other than Shanghai only. The broad set of cities makes our findings more representative of the Chinese housing markets. On the other hand, this different pattern may be related to our sample period. Zhou (2018) constructed a trade-based sentiment for Shanghai from January 2010 to May 2015, in which the housing price experienced several rounds of upswing. In contrast, our sample covers the period of January 2016 to October 2020, during which a series of tightening policies were introduced to curb the rapid growth of housing prices. The tightening policy trend and the global COVID-19 epidemic may result in a generally low level of housing sentiment relative to the period before 2016.

---

[10] As in Column (6) of Table 4, seasonality dummies and macroeconomic variables are omitted when we control for time-fixed effects.

[11] Note the sample mean of housing sentiment series are set to zero. So, the negative (positive) sentiment stands for the part of sentiment below (above) the sample average in the sample period considered.

Table 5 Multi-Horizon Return Regressions

Return Horizon $R_{i,[t,t+b]}$												
	1 Month	2 Months	3 Months	4 Months	5 Months	6 Months	7 Months	8 Months	9 Months	10 Months	11 Months	12 Months
S_{it}^{PLS}	0.137**	0.311*	0.581*	0.766*	0.844*	0.845	0.717	0.519	0.350	0.320	0.353	0.280
	(0.062)	(0.169)	(0.316)	(0.409)	(0.484)	(0.525)	(0.613)	(0.654)	(0.667)	(0.656)	(0.635)	(0.643)
R^2	0.519	0.542	0.528	0.527	0.506	0.516	0.516	0.500	0.480	0.490	0.481	0.449
Obs.	780	762	744	724	703	681	656	631	608	589	570	550

This table reports the estimated sentiment coefficient β_b in $\log(P_{ib}) - \log(P_{it}) = \alpha_b + \beta_b S_{it}^{PLS} + \delta_b Control_{it} + \pi_i + \lambda_t + \epsilon_{i,t}$, with b varying from 1 to 12 months. P_{it} denotes second-hand housing price index. Control variables include lagged monthly returns, CPI_{it}, $population_{it}$, $unemployment_{it}$, $income_{it}$, $rents_{it}$, PRR_{it}, and their lagged values. City-fixed effect π_i, and time-fixed effect λ_t, are controlled for in regression. Finally, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. City-clustered standard errors are reported in parentheses
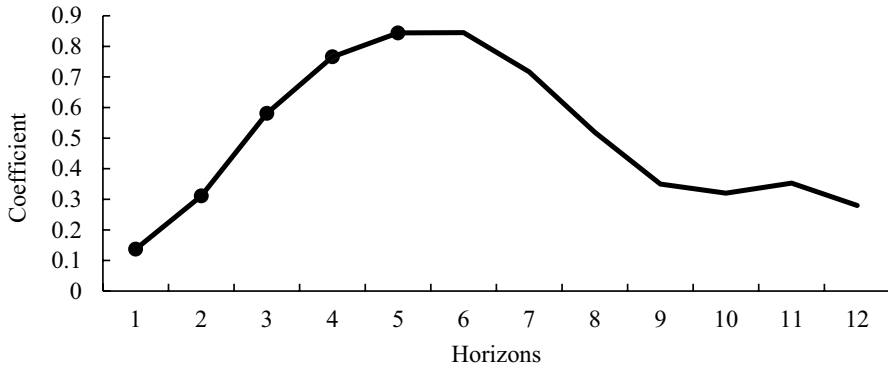
Fig. 2 Sentiments and future multi-horizon returns. *Notes.* This figure depicts the coefficient of sentiment β_b for horizon b ranged from 1 to 12 months in regression (5). The dots on the curve indicates the coefficient β_b is significant and no dots means insignificant

	(1)	(2)	(3)	(4)	(5)
Table 6 Predicting returns by negative and positive sentiments					
Return Horizon	*b=1*	*b=3*	*b=6*	*b=9*	*b=12*
S_{it}^{neg}	0.265***	1.045***	1.746**	1.174	0.939
	(0.084)	(0.322)	(0.748)	(1.176)	(1.201)
S_{it}^{pos}	0.043	0.241	0.197	-0.204	-0.124
	(0.124)	(0.541)	(0.731)	(0.661)	(0.702)
Control_{it}	Yes	Yes	Yes	Yes	Yes
CityFE	Yes	Yes	Yes	Yes	Yes
TimeFE	Yes	Yes	Yes	Yes	Yes
Num.ofcities	18	18	18	18	18
R^2	0.520	0.530	0.519	0.483	0.450
Obs.	780	744	681	608	550

This table reports the results of $R_{i,[t,t+b]} = \alpha_b + \beta_b^{neg} S_{it}^{neg} + \beta_b^{pos} S_{it}^{pos} + \delta_b Control_{it} + \pi_i + \lambda_t + \varepsilon_{it}$. $S_{it}^{pos} = max\{S_{it}^{PLS}, 0\}$, denoting the positive sentiment. $S_{it}^{neg} = min\{S_{it}^{PLS}, 0\}$, denoting the negative sentiment. *Control_{it}* comprises CPI_{it}, $population_{it}$, $unemployment_{it}$, $income_{it}$, $rents_{it}$, PRR_{it}, and their lagged values. City-fixed effects and time-fixed effects are also controlled for. Finally, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. City-clustered standard errors are reported in parentheses

Sentiments and Housing Supply Elasticities

Given housing supply elasticity is one of the key determinants of housing prices (Glaeser et al., 2006; Liu, 2014), we will investigate potential heterogeneous sentiment effects across cities with different housing supply elasticities in this section. The supply elasticity estimates are obtained from Liu et al. (2019), which construct

Table 7 Interactive effects of housing supply elasticities and sentiments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Return Horizon	b=1	b=1	b=3	b=3	b=6	b=6	b=9	b=9
S_{it}^{PLS}	0.137**	0.214**	0.581*	0.851**	0.845	1.205*	0.350	0.880
	(0.062)	(0.077)	(0.316)	(0.344)	(0.525)	(0.583)	(0.667)	(0.708)
$S_{it}^{PLS} \times ElastHigh_i$		-0.361**		-1.253***		-1.718**		-2.441***
		(0.156)		(0.403)		(0.673)		(0.779)
$Control_{it}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$laggedControl_{it}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CityFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num.ofcities	18	18	18	18	18	18	18	18
R^2	0.519	0.524	0.528	0.537	0.516	0.523	0.480	0.496
Obs.	780	780	744	744	681	681	608	608

$ElastHigh_i$ equals 1 if the housing supply elasticity of city i is greater than the 75 percentile of 282 cities estimates in China given by Liu et al. (2019), otherwise equals 0. $Control_{it}$ includes CPI_{it}, $population_{it}$, $unemployment_{it}$, $income_{it}$, $rents_{it}$, and PRR_{it}. We also include the first lags of all the control covariates (denoted by $laggedControl_{it}$). City-fixed effects and time-fixed effects are controlled for in all columns. Finally, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. City-clustered standard errors are reported in parentheses

housing supply elasticities for 282 cities in China by combining natural geographical constraints, cultivated land protection constraints, and floor area ratio regulations of local housing markets. These elasticity measures show housing supply in most Chinese cities is overall inelastic and exhibits cross-sectional heterogeneity.

Based on these elasticity estimates, we divide cities into two groups, one with a relatively elastic housing supply and the other with a relatively inelastic housing supply. More specifically, we define a dummy variable, $ElastHigh_i$, taking the value of 1 if the elasticity of city i is greater than the 75 percentile of elasticities of 282 cities given by Liu et al. (2019), and the value of zero otherwise. In our sample, Shanghai, Xiamen, and Hefei belong to the supply elastic group ($ElastHigh_i = 1$). Note that Shanghai has a relatively high supply elasticity, which could be related to its looser land use regulations and flat terrain in the area (Liu et al., 2019).

Then an interactive term of S_{it}^{PLS} and $ElastHigh_i$ is included in Eq. (5). The results are given in Table 7. For b=1, the coefficient on the interactive term is significantly negative, which means that the local housing sentiment has a stronger impact on returns for cities with relatively inelastic housing supply than cities with a more elastic housing supply. Also, the heterogeneous sentiment effects are persistent even for longer horizons (b=3, 6, 9). This significant heterogeneous sentiment impacts on housing returns of cities with varying supply elasticities are consistent with findings in Zheng et al. (2016), which show that the confidence index based on Google search also has a larger impact on housing appreciations of cities with relatively inelastic housing supply.

The finding that local housing supply elasticities play an important role in the sentiment effects highlights the necessity of considering local market conditions in

the analysis of sentiment analysis. For cities with a much more elastic housing supply, the impacts of market sentiments on returns are much smaller or even negligible. This may be because that cities with elastic housing supply could adjust housing supply easily when expecting a warming-up market sentiment. However, the supply elasticity is not time-invariant. According to Liu et al. (2019), natural geographical constraints, cultivated land protection constraints, and floor area ratio regulations are the three main factors in determining local supply elasticities. While the natural physical landscape is hard to change and land protection constraints are not allowed to loosen, local official governments could loosen regulations on floor area ratios in land development to alleviate the impacts of housing sentiments.

Predicting Sentiments with Returns

So far, we have shown that the transaction-based local housing sentiment has a positive effect on future housing returns, exhibits the pattern of short-run underreaction and long-run overreaction, and has an asymmetry impact on housing returns. In addition, the market sentiment effects are much stronger and more significant for cities with relatively inelastic housing supply.

Intuitively, sentiment may be affected by past housing returns. Soo (2018) shows that past price appreciations predict higher media sentiment for the U.S. housing markets. In this section, we investigate the determinants of housing sentiments by running the regression as follows:

$$S_{i,t+1}^{PLS} = \alpha + \beta R_{it} + \gamma Control_{it} + \delta S_{it}^{PLS} + \mu_i + \lambda_t + \varepsilon_{it}, \tag{6}$$

where $S_{i,t+1}^{PLS}$ denotes the sentiment of city i in month $t + 1$, R_{it} denotes the housing return of city i in month t. We control for $S_{it}^{PLS}, rents_{it}, PRR_{it}, CPI_{it}, population_{it}, unemployment_{it}, income_{it}$ (and their first lags), city-fixed effects as well as time-fixed effects in Eq. (6).

Table 8 reports the estimation results. Columns (1)-(4) show that the housing return R_{it} is positively correlated with future sentiment $S_{i,t+1}^{PLS}$, consistent with Soo (2018). Higher past housing returns will foster higher market sentiments. Results of Column (4) indicate a significant positive serial correlation in housing sentiments. These results indicate that the expectations of Chinese real estate investors are backward-looking and there exists a significant feedback effect between housing returns and market sentiments. This feedback effect could result in a pricing-sentiment spiral, as defined by Ling et al. (2015), in which the dynamic interplay between sentiment and house price appreciations can create a self-reinforcing spiral. The pricing-sentiment spiral could potentially extend the length and magnitude of housing market cycles in China and hence largely exaggerates the ongoing housing market fever in China.

Moreover, we find negative but insignificant effects of rent growth rates ($rents_{it}$) and price-rent ratios (PRR_{it}) on local housing sentiments as shown in Table 8. To further explore how persistently the past accumulated returns can push up local sentiments, we replace R_{it} with $R_{i,[t-b,t]}$ in Eq. (6). The last three columns of Table 8 report the estimates for 3-month, 6-month, and 9-month return horizons, respectively. The results show that only the past 1-month return significantly predicts higher market sentiments,

Table 8 Predicting sentiments with past returns

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Return Horizon	$b=1$	$b=1$	$b=1$	$b=1$	$b=3$	$b=6$	$b=9$
$R_{i,[t-b,t]}$	0.195***	0.185***	0.155**	0.041*	0.009	0.005	-0.000
	(0.046)	(0.048)	(0.063)	(0.023)	(0.008)	(0.005)	(0.003)
$rents_{it}$		-0.001	0.016	-0.010	0.002	0.001	-0.004
		(0.009)	(0.037)	(0.017)	(0.020)	(0.022)	(0.020)
PRR_{it}		0.085	0.698	-0.095	0.199	0.183	0.048
		(0.123)	(0.858)	(0.334)	(0.380)	(0.441)	(0.408)
S_{it}^{PLS}				0.825***	0.824***	0.821***	0.833***
				(0.034)	(0.039)	(0.044)	(0.045)
$Control_{it}$	No	Yes	Yes	Yes	Yes	Yes	Yes
$laggedControl_{it}$	No	No	Yes	Yes	Yes	Yes	Yes
CityFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TimeFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num.ofcities	18	18	18	18	18	18	18
R^2	0.163	0.191	0.220	0.751	0.749	0.738	0.729
Obs.	813	797	793	780	776	759	705

Dependent variable is $S_{i,t+1}^{PLS}$, the sentiment of city i in month $t + 1$; $R_{i,[t-b,t]}$ denotes the cumulative returns (in percentage) of the past 1 month ($b = 1$), 3 months ($b = 3$), 6 months ($b = 6$), and 9 months ($b = 9$), respectively. $Control_{it}$ include PRR_{it}, $rents_{it}$ CPI_{it}, $population_{it}$, $unemployment_{it}$, and $income_{it}$. We also include the first lags of all the control covariates (denoted by $laggedControl_{it}$). City-fixed effects and time-fixed effects are also controlled for in the regression. Standard errors clustered at the city level are in parentheses. Finally, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

while past cumulative returns over longer horizons (i.e., 3-month, 6-month, and 9-month) seem to exert no significant impacts on housing sentiments. The most recent past returns (1-month) appear to exert the greatest impact on housing sentiment, similar to findings in Soo (2018).

Robustness Checks

In this section, we conduct several robustness checks on our main regression results. Firstly, to remove the potential shock of COVID-19 on the Chinese real estate market, we exclude the sample period after November 2019 and re-estimate our main regression. Secondly, we use alternative approaches (**scaled PCA** and **PCA**) to construct the sentiment index and re-estimate the main regression. Thirdly, we substitute an alternative proxy of sentiment, the new listings of second-hand properties, for turnover and then re-construct our sentiment index and re-estimate the main regression. Finally, to confirm the reliability of the **recursive look-ahead-bias-free**

implementation of the **PLS** approach, we conduct a placebo test following Kelly and Pruitt (2013).

Pre-Covid-19 Sub-sample Estimation

COVID-19 broke out in November 2019 and was considered to have caused a huge negative shock to the Chinese economy and especially to the real estate market. To remove the potential effect of COVID-19, we discard the data after November 2019 and re-estimate our main empirical regression. Using the sample from January 2016 to October 2019, Panel A of Table 12 in Appendix B shows that a one-standard-deviation increase in local housing sentiment is associated with a future monthly return increase of approximately 0.14%, which is almost the same as that from the full sample estimation in "Predicting Returns with Local Housing Sentiments" section.

Alternative Approaches of Sentiment Construction: scaled PCA and PCA

To further prove the robustness of the predictability of local housing sentiment on future returns, we consider two alternative sentiment construction methods, i.e., the **scaled PCA** by Huang et al. (2022) and the traditional **PCA** method. For both alternative methods, we first follow the first four steps as described in "Housing Sentiment Index" section to get the final sentiment proxies (**Median-Intv**, **Turnover**, and **SMB**). And then, for the **PCA** method, we extract the first principal component of these three proxies to derive the **PCA** housing sentiment S_{it}^{PCA}.

For the **scaled PCA** method, after we get the final sentiment proxies following the first four steps, for month t, we run the following regression:

$$Return_s = \delta_{i,t} + \gamma_{i,t} proxy_{i,s-1} + u_{i,s}, s \leq t, (7).$$

by using information only up to month t. We then apply **PCA** to scaled proxies $(\widehat{\gamma}_{1,t}\textbf{\textit{MedianIntv}}, \widehat{\gamma}_{2,t}\textbf{\textit{Turnover}}, \widehat{\gamma}_{3,t}\textbf{\textit{SMB}})$ to extract the first component as the housing sentiment S_t^{sPCA}.[12] By recursively estimating the above predictive regression (7) and applying **PCA** to scaled proxies with an expanding window scheme, we obtain a recursive **look-ahead-bias-free scaled PCA** estimate of the housing sentiment sequence $\{S_t^{sPCA}, t = t_0, \ldots, T\}$. Like the recursive implementation of **PLS**, we use the first twelve months as the initial training sample and the constructed **scaled PCA** housing sentiment sequences are associated with one-year shorter sample periods.

As the correlation matrix in Table 9 of Appendix A shows, the new alternative **scaled PCA** sentiment index S_{it}^{sPCA} is significantly positively correlated with the original sentiment S_{it}^{PLS} (with a correlation coefficient of 0.491). The PCA sentiment S_{it}^{PCA} is also significantly positively correlated with S_{it}^{PLS}, but with a much smaller

---

[12] The PCA procedure of the scaled PCA analysis is also conducted recursively with an expanding window scheme. In other words, for month t, the PCA analysis is applied to $(\widehat{\gamma}_{1,t}\textbf{\textit{MedianIntv}}, \widehat{\gamma}_{2,t}\textbf{\textit{Turnover}}, \widehat{\gamma}_{3,t}\textbf{\textit{SMB}})$ with proxy sequences ending at month t.

correlation coefficient of 0.233. We re-estimate our main regression by replacing the sentiment estimates with S_{it}^{PCA} and S_{it}^{sPCA} with results summarized in Panel B (**scaled PCA**) and Panel C (**PCA**) of Table 12 in Appendix B, respectively. As results show, housing sentiment indices derived from both alternative methods still exert a significant impact on future housing returns, although with smaller and less significant impacts than that from the **PLS** method.

To briefly sum up, the predictability of the local housing sentiment we document is robust to the choice of the sentiment index construction method. However, our results do suggest that the **PLS** method seems to do a better job at constructing a predictive sentiment index than **scaled PCA** and **PCA** in our research setting.

Alternative Proxy: New listings

Lowry (2003) suggests that the number of initial public offerings (IPOs) reflects stock market sentiment. Similarly, in the housing market, a larger number of new listings could indicate a higher market sentiment. In this section, we replace the raw proxy **$Turnover_{it}$** with **$NewListings_{it}$**, the number of new listings at month t in city i.[13] Then we apply these proxy cleaning procedures to **MedianIntv**, **NewListings**, and **SMB**, and conduct the **recursive look-ahead-bias-free** implementation of **PLS** to construct a **look-ahead-bias-free** sentiment index $S_{it}^{PLS'}$ for each city in our sample.

As the correlation matrix in Table 9 of Appendix A shows, **NewListings** is significantly positively correlated with **Turnover** (with a correlation coefficient of 0.118) but negatively correlated with **MedianIntv** (with a coefficient of -0.122). And the new alternative sentiment index $S_{it}^{PLS'}$ is significantly positively correlated with the original S_{it}^{PLS} (with a correlation coefficient of 0.466). We display the main regression results using this alternative sentiment index $S_{it}^{PLS'}$ in Panel D of Table 12 in Appendix B. The coefficient on $S_{it}^{PLS'}$ (0.178 and significant at 5% level) is close to 0.137 from the main estimation in Column (6) of Table 4. This result suggests that our findings on the significant predictability of housing sentiment on housing returns are robust with the alternative proxy **NewListings**.

Placebo Test

Following Kelly and Pruitt (2013), we conduct a placebo test to confirm that our estimation procedure does not generate sentiment predictability mechanically. Intuitively, we can test our method by using simulated random sentiment known to have no true return forecasting ability. In detail, we conduct the placebo test as follows. Firstly, we generate three random series from the standard

---

[13] We thank an anonymous referee for the suggestion on the **NewListings** proxy.

normal distribution (with zero mean and unit variance) as "fake" raw proxies for each city. Secondly, we then conduct the five-steps approach as described in "Housing Sentiment Index" section to these randomly generated proxy series to construct a "placebo" **look-ahead-bias-free PLS** sentiment index $S_{it}^{Placebo}$. Finally, we examine the predictability of this "placebo" sentiment on future returns. If we find no significant predictive power of this "placebo" sentiment, we could provide evidence that the **look-ahead-bias-free** implementation **of PLS** does not yield **"false predictability"** mechanically. As the results in Table 13 of Appendix C show, this "placebo" sentiment exhibits no consistent predictability on future housing returns.[14] This test confirms that the recursive **look-ahead-bias-free** implementation of **PLS** does not mechanically yield a sentiment with significant predictability.

Conclusions

This paper investigates the relationship between local housing sentiments and returns in Chinese housing markets. We construct monthly city-level sentiment indices for 18 Chinese cities from January 2016 to October 2020 by using a massive second-hand transaction dataset through a recursive **look-ahead-bias-free** implementation of the **PLS** method. These local housing sentiment indices are based on two housing market liquidity proxies and a small-house return premium measure. Empirically, we find that local housing sentiments can significantly predict future housing market returns. The sentiment impact is comparable with estimates for the U.S. housing market in the literature, implying that the Chinese housing markets are also susceptible to irrational sentiments even under a stringent regulatory environment.

Furthermore, a salient short-run underreaction and long-run overreaction pattern of sentiment effects is documented. Further analysis shows that local housing sentiment impacts are asymmetric, and housing returns in cities with relatively inelastic housing supply are more sensitive to local market sentiments. Last but not the least, we show that the expectations of Chinese real estate investors are backward-looking and there exists a significant feedback effect between housing returns and market sentiments.

Our major findings are robust to alternative sentiment construction methods and alternative sentiment proxy choice, and consistent for the sub-sample before COVID-19. Our empirical analysis enriches the literature on the role of sentiments in the housing market and provides better insights into the housing market dynamics. This paper's findings can provide references for policymakers to stabilize and improve the functioning of the housing market.

---

[14] Except for Column (3). In Column (3) of Table 13 of Appendix C, the "placebo" sentiment is significant at the 10% level. However, this significance is not consistent as it becomes insignificant in the preferred regression of Column (6).

Appendix A

Table 9 Correlation of sentiments, proxies, and returns

	S_t^{PLS}	S_t^{SPCA}	S_t^{PCA}	$S_{it}^{PLS'}$	$MedianIntv_t$	$Turnover_t$	SMB_t	$NewListings_t$	R_t
S_t^{PLS}	1.000								
S_t^{SPCA}	0.491***	1.000							
S_t^{PCA}	0.233***	0.234***	1.000						
$S_t^{PLS'}$	0.466***	0.420***	0.372***	1.000					
$MedianIntv_t$	-0.212***	-0.297***	-0.066**	-0.134***	1.000				
$Turnover_t$	0.293***	0.167***	0.115***	0.163***	-0.575***	1.000			
SMB_t	0.001	0.022	0.064**	0.030	-0.018	-0.032	1.000		
$NewListings_t$	-0.037	0.003	-0.033	-0.039	-0.122***	0.118***	-0.015	1.000	
R_t	0.327***	0.151***	0.211***	0.365***	-0.233***	0.458***	0.013	0.065**	1.000

S_t^{PLS} stands for the sentiment index from the PLS method, while S_t^{sPCA} stands for the sentiment index from the scaled PCA method, S_t^{PCA} stands for the sentiment index from the PCA method, and $S_{it}^{PLS'}$ stands for the sentiment index based on the alternative proxy **NewListings**. Significance levels of 10%, 5%, 1% are marked by *, **, and ***, respectively

Table 10 Correlation of sentiment index with confidence indices

Panel A: Beijing								
	CC_{BJ}	CE_{BJ}	HB_{CN}	IC_{CN_F}	IC_{CN_P}	IC_{CN}	CC_{CN}	CE_{CN}
Correlation	0.460	0.472	-0.042	0.205	0.006	0.058	-0.262	-0.252
P-Value	0.114	0.104	0.799	0.447	0.983	0.758	0.103	0.117
Panel B: Shanghai								
	CC_{SH}	CE_{SH}	HB_{CN}	IC_{CN_F}	IC_{CN_P}	IC_{CN}	CC_{CN}	CE_{CN}
Correlation	0.145	0.036	0.344**	-0.126	-0.432**	-0.561***	-0.105	-0.115
P-Value	0.636	0.908	0.017	0.557	0.035	0.000	0.480	0.436

HB_{CN} is the Housing Boom Index of China; IC_{CN_F} is the Investor Confidence Index for Domestic Economic Fundamentals; IC_{CN_P} is the Investor Confidence Index for Economic Policy; IC_{CN} is the Investor Confidence Index of China; CC_{CN} is the Consumer Confidence Index of China; CE_{CN} is the Consumer Expectations Index of China. In Panel A, CC_{BJ} is the Consumer Confidence Index of Beijing; CE_{BJ} is the Consumer Expectation Index of Beijing; In Panel B, CC_{SH} is the Consumer Confidence Index of Shanghai; CE_{SH} is the Consumer Expectation Index of Shanghai. All these data are sourced from the National Bureau of Statistics and Wind. Significance level of 10%, 5%, 1% are marked by *, **, and ***, respectively

Table 11 Predicting future returns by local housing sentiments

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
S_{it}^{PLS}	0.451**	0.447**	0.522***	0.296**	0.113*	0.137**
	(0.176)	(0.177)	(0.174)	(0.120)	(0.060)	(0.062)
$Autumn_t$	-0.991***	-1.025***	-0.922***	-0.638***	-0.619***	
	(0.197)	(0.202)	(0.186)	(0.176)	(0.152)	
$Spring_t$	0.240	0.212	0.280	0.289**	-0.191	
	(0.155)	(0.143)	(0.185)	(0.136)	(0.153)	
$Default_t$		-1.363*	-1.756**	-3.401***	-2.502***	
		(0.754)	(0.757)	(1.001)	(0.822)	
PMI_t		0.028	0.038	-0.013	-0.046	
		(0.038)	(0.036)	(0.033)	(0.033)	
$M2G_t$		0.095	0.114	0.328***	0.178	
		(0.074)	(0.068)	(0.098)	(0.103)	
CPI_{it}			0.029	-0.085	-0.232**	-0.364**
			(0.103)	(0.093)	(0.105)	(0.152)
$Population_{it}$			-0.216	0.098	0.159	0.103
			(0.205)	(0.573)	(0.331)	(0.374)
$Unemployment_{it}$			0.649	-3.950*	-3.420**	-2.367**
			(0.778)	(2.024)	(1.264)	(1.077)
$Income_{it}$			0.875	3.139	2.383	1.660
			(0.602)	(2.735)	(1.823)	(1.696)
$Rents_{it}$			0.025	0.318***	0.011	0.014
			(0.019)	(0.059)	(0.031)	(0.035)
PRR_{it}			-0.715**	7.555***	-0.414	-0.294
			(0.275)	(1.359)	(0.876)	(0.989)
$Default_{t-1}$				2.185	1.602*	
				(1.257)	(0.897)	
PMI_{t-1}				0.059*	0.028	
				(0.030)	(0.030)	
$M2G_{t-1}$				0.329***	0.169**	
				(0.078)	(0.060)	
$CPI_{i,t-1}$				0.025	-0.003	-0.093
				(0.080)	(0.074)	(0.128)
$Population_{i,t-1}$				-0.255	-0.286	-0.232
				(0.574)	(0.348)	(0.408)
$Unemployment_{i,t-1}$				4.785*	3.697**	2.568*
				(2.581)	(1.501)	(1.299)
$Income_{i,t-1}$				-2.777	-2.321	-1.673
				(2.752)	(1.868)	(1.728)
$Rents_{i,t-1}$				0.010	-0.004	-0.000
				(0.013)	(0.015)	(0.012)

Table 11 (continued)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
$PRR_{i,t-1}$				-8.350***	-0.192	-0.367
				(1.376)	(0.810)	(0.919)
$R_{i,t}$					0.606***	0.598***
					(0.067)	(0.068)
Constant	0.779***	0.048	-0.142	-1.318	2.960	1.465
	(0.061)	(1.599)	(2.272)	(1.975)	(1.780)	(1.417)
CityFE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
TimeFE	*No*	*No*	*No*	*No*	*No*	*Yes*
Num.ofcities	18	18	18	18	18	18
R^2	0.114	0.129	0.174	0.352	0.494	0.519
Obs.	799	799	786	780	780	780

Controlled covariates include seasonality dummies (*Autumn$_t$* and *Spring$_t$*), macroeconomic covariates (*Macros$_t$*), and fundamental controls (*Fundamentals$_{it}$*). *Macros$_t$* contains *PMI$_t$* (Purchasing Managers Index), *M2G$_t$* (the growth rate of M2), and *Default$_t$* (the yield spread between AA-grade corporate bonds and AAA-grade corporate bonds). *Fundamentals$_{it}$* contains *CPI$_{it}$* (the growth rate of the consumer price index), *population$_{it}$* (the growth of population), *unemployment$_{it}$* (the unemployment rate), *income$_{it}$* (the growth rate of the average income of on-the-job employees), *rents$_{it}$* (the growth rate of the average rents per square meter per month), and *PRR$_{it}$* (the house price to rent ratio). City-fixed effects are controlled for in each column. Time-fixed effects are controlled for in Column (6). In the last three columns, we also include the first lagged values of *Macros$_t$* and *Fundamentals$_{it}$*. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. City-clustered standard errors are reported in parentheses

Appendix B

Table 12 Robustness Checks: Predicting future returns by local housing sentiments

	(1)	(2)	(3)	(4)	(5)	(6)
	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
Panel A. Pre-COVID-19 estimation (for January 2016-October 2019)						
S_{it}^{PLS}	0.519**	0.483**	0.576***	0.339**	0.117*	0.142**
	(0.205)	(0.211)	(0.195)	(0.123)	(0.062)	(0.062)
$R_{i,t}$					0.581***	0.590***
					(0.072)	(0.066)
$Macros_t$	No	Yes	Yes	Yes	Yes	No
$Fundamentals_{it}$	No	No	Yes	Yes	Yes	Yes
$laggedControl_{it}$	No	No	No	Yes	Yes	Yes
$CityFE$	Yes	Yes	Yes	Yes	Yes	Yes
$TimeFE$	No	No	No	No	No	Yes
R^2	0.128	0.173	0.263	0.440	0.556	0.584
$Obs.$	606	606	594	589	589	589
Panel B. Alternative method: Scaled PCA						
S_{it}^{sPCA}	0.176*	0.182*	0.257**	0.168**	0.072	0.092*
	(0.089)	(0.090)	(0.113)	(0.079)	(0.049)	(0.045)
$R_{i,t}$					0.614***	0.606***
					(0.062)	(0.062)
$Macros_t$	No	Yes	Yes	Yes	Yes	No
$Fundamentals_{it}$	No	No	Yes	Yes	Yes	Yes
$laggedControl_{it}$	No	No	No	Yes	Yes	Yes
$CityFE$	Yes	Yes	Yes	Yes	Yes	Yes
$TimeFE$	No	No	No	No	No	Yes
R^2	0.073	0.091	0.130	0.347	0.497	0.523
$Obs.$	804	804	790	784	784	784
Panel C. Alternative method: PCA						
S_{it}^{PCA}	0.313*	0.308*	0.428**	0.254**	0.103*	0.100*
	(0.160)	(0.159)	(0.168)	(0.106)	(0.052)	(0.052)
$R_{i,t}$					0.590***	0.582***
					(0.068)	(0.068)
$Macros_t$	No	Yes	Yes	Yes	Yes	No
$Fundamentals_{it}$	No	No	Yes	Yes	Yes	Yes
$laggedControl_{it}$	No	No	No	Yes	Yes	Yes
$CityFE$	Yes	Yes	Yes	Yes	Yes	Yes
$TimeFE$	No	No	No	No	No	Yes
R^2	0.072	0.077	0.175	0.361	0.494	0.517
$Obs.$	964	964	919	907	907	907

Table 12 (continued)

	(1) $R_{i,t+1}$	(2) $R_{i,t+1}$	(3) $R_{i,t+1}$	(4) $R_{i,t+1}$	(5) $R_{i,t+1}$	(6) $R_{i,t+1}$
Panel D. Alternative proxy:*MedianIntv*, *NewListings,* and *SMB*						
$S_{it}^{PLS'}$	0.535***	0.534***	0.582***	0.373***	0.161**	0.178**
	(0.144)	(0.146)	(0.152)	(0.101)	(0.075)	(0.073)
$R_{i,t}$					0.590***	0.580***
Macros$_t$	No	Yes	Yes	Yes	Yes	No
Fundamentals$_{it}$	No	No	Yes	Yes	Yes	Yes
laggedControl$_{it}$	No	No	No	Yes	Yes	Yes
CityFE	Yes	Yes	Yes	Yes	Yes	Yes
TimeFE	No	No	No	No	No	Yes
R^2	0.138	0.154	0.195	0.366	0.496	0.522
Obs.	800	800	788	782	782	782

S_{it}^{sPCA} stands for the sentiment from scaled PCA. S_{it}^{PCA} is the sentiment from PCA. $S_{it}^{PLS'}$ is the PLS sentiment based on **MedianIntv, NewListings**, and **SMB**. *Control$_{it}$* contains both of *Macros$_t$* and *Fundamentals$_{it}$*. *Macros$_t$* contains PMI_t, $M2G_t$, and *Default$_t$*. *Fundamentals$_{it}$* contains CPI_{it}, *population$_{it}$*, *unemployment$_{it}$*, *income$_{it}$*, *rents$_{it}$*, and PRR_{it}. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. City-clustered standard errors are reported in parentheses

Appendix C

Table 13 Placebo Test: Predicting future returns by a "placebo" sentiment

	(1) $R_{i,t+1}$	(2) $R_{i,t+1}$	(3) $R_{i,t+1}$	(4) $R_{i,t+1}$	(5) $R_{i,t+1}$	(6) $R_{i,t+1}$
$S_{it}^{Placebo}$	0.168	0.157	0.188*	0.118	-0.020	-0.014
	(0.101)	(0.101)	(0.095)	(0.070)	(0.054)	(0.050)
Autumn$_t$	-1.004***	-1.083***	-0.953***	-0.641***	-0.627***	-0.474
	(0.176)	(0.187)	(0.174)	(0.173)	(0.154)	(0.491)
Spring$_t$	0.278*	0.201	0.287	0.256*	-0.226	-0.527
	(0.154)	(0.143)	(0.171)	(0.126)	(0.149)	(0.537)
$R_{i,t}$					0.627***	0.620***
					(0.066)	(0.065)
Macros$_t$	No	Yes	Yes	Yes	Yes	No
Fundamentals$_{it}$	No	No	Yes	Yes	Yes	Yes
laggedControl$_{it}$	No	No	No	Yes	Yes	Yes
CityFE	Yes	Yes	Yes	Yes	Yes	Yes
TimeFE	No	No	No	No	No	Yes
Num.ofcities	18	18	18	18	18	18
R^2	0.067	0.083	0.113	0.339	0.495	0.519
Obs.	798	798	785	779	779	779

$S_{it}^{Placebo}$ stands for a "placebo" sentiment based on three randomly generated sentiment proxies using the recursive look-ahead-bias-free implementation of PLS. *Control$_{it}$* contains both *Fundamentals$_{it}$* and *Macros$_t$*. *Macros$_t$* refers to macroeconomic variables including PMI_t, $M2G_t$, and *Default$_t$*. *Fundamentals$_{it}$* contains CPI_{it}, *population$_{it}$*, *unemployment$_{it}$*, *income$_{it}$*, *rents$_{it}$*, and PRR_{it}. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively. City-clustered standard errors are reported in parentheses

Declarations

Conflicts of Interest All authors declare no in this manuscript.

References

Anastasiou, D., Kapopoulos, P., Zekente, K.-M. (2021). Sentimental shocks and house prices. *The Journal of Real Estate Finance and Economics*, 1–29.

Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets, 7*(3), 271–299.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance, 61*(4), 1645–1680.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives, 21*(2), 129–152.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics, 49*(3), 307–343.

Bork, L., Møller, S. V., & Pedersen, T. Q. (2020). A new index of housing sentiment. *Management Science, 66*(4), 1563–1583.

Case, K. E., & Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity, 2003*(2), 299–362.

Case, K. E., Shiller, R. J., Thompson, A. (2012). *What have they been thinking? Home buyer behavior in hot and cold markets*. National Bureau of Economic Research.

Chauvet, M., Gabriel, S., & Lutz, C. (2016). Mortgage default risk: New evidence from internet search queries. *Journal of Urban Economics, 96*, 91–111.

Chen, K., & Wen, Y. (2017). The great housing boom of china. *American Economic Journal: Macroeconomics, 9*(2), 73–114.

Chow, G. C., & Niu, L. (2015). Housing prices in urban china as determined by demand and supply. *Pacific Economic Review, 20*(1), 1–16.

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.

Clayton, J., MacKinnon, G. (2000). Explaining the discount to nav in reit pricing: Noise or information? Available at SSRN 258268.

Clayton, J., MacKinnon, G. (2002). Departures from nav in reit pricing: The private real estate cycle, the value of liquidity and investor sentiment. *Real Estate Research Institute, Working Paper*.

Clayton, J., MacKinnon, G., & Peng, L. (2008). Time variation of liquidity in the private real estate market: An empirical investigation. *Journal of Real Estate Research, 30*(2), 125–160.

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

Daniel, K., Hirshleifer, D., Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance*, *53*(6), 1839–1885.

Das, P. K., Freybote, J., & Marcato, G. (2015). An investigation into sentiment-induced institutional trading behavior and asset pricing in the reit market. *The Journal of Real Estate Finance and Economics, 51*(2), 160–189.

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.

Fang, H., Gu, Q., Xiong, W., & Zhou, L.-A. (2016). Demystifying the Chinese housing boom. *NBER Macroeconomics Annual, 30*(1), 105–166.

Ferreira, F., & Gyourko, J. (2012). Heterogeneity in neighborhood-level price growth in the united states, 1993–2009. *American Economic Review, 102*(3), 134–140.

Fu, Y. M., & Qian, W. L. (2014). Speculators and price overreaction in the housing market. *Real Estate Economics, 42*(4), 977–1007.

Gallimore, P., & Gray, A. (2002). The role of investor sentiment in property investment decisions. *Journal of Property Research, 19*(2), 111–120.

Glaeser, E., Huang, W., Ma, Y., & Shleifer, A. (2017). A real estate boom with Chinese characteristics. *Journal of Economic Perspectives, 31*(1), 93–116.

Glaeser, E. L., Gyourko, J., & Saiz, A. (2008). Housing supply and housing bubbles. *Journal of Urban Economics, 64*(2), 198–217.

Glaeser, E. L., Gyourko, J., & Saks, R. E. (2006). Urban growth and housing supply. *Journal of Economic Geography, 6*(1), 71–89.

Glaeser, E. L., & Ward, B. A. (2009). The causes and consequences of land use regulation: Evidence from greater Boston. *Journal of Urban Economics, 65*(3), 265–278.

Green, R. K., Malpezzi, S., & Mayo, S. K. (2005). Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources. *American Economic Review, 95*(2), 334–339.

Han, L. (2013). Understanding the puzzling risk-return relationship for housing. *The Review of Financial Studies, 26*(4), 877–928.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance, 54*(6), 2143–2184.

Huang, D. S., Jiang, F. W., Li, K. P., Tong, G. S., & Zhou, G. F. (2022). Scaled PCA: A new approach to dimension reduction. *Management Science, 68*(3), 1678–1695.

Huang, D. S., Jiang, F. W., Tu, J., & Zhou, G. F. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies, 28*(3), 791–837.

Hui, E. C., Dong, Z., & Jia, S. (2018). How do housing price and sentiment affect consumption distribution in china? *Habitat International, 77*, 99–109.

Hui, E. C. M., Dong, Z., Jia, S., & Lam, C. H. L. (2017). How does sentiment affect returns of urban housing? *Habitat International, 64*, 71–84.

Hui, E. C. M., & Wang, Z. Y. (2014). Market sentiment in private housing market. *Habitat International, 44*, 375–385.

Jin, C., Soydemir, G., & Tidwell, A. (2014). The us housing market and the pricing of risk: Fundamental analysis and market sentiment. *Journal of Real Estate Research, 36*(2), 187–220.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review, 95*(1), 161–182.

Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance, 68*(5), 1721–1756.

Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics, 186*(2), 294–316.

Keynes, J. M. (1936). The supply of gold. *The Economic Journal, 46*(183), 412–418.

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.

Letdin, M., Sirmans, S., Sirmans, G. S. (2022). Betting against the sentiment in reit nav premiums. *The Journal of Real Estate Finance and Economics, 64*(4), 590–614.

Lin, C. Y., Rahman, H., & Yung, K. (2009). Investor sentiment and reit returns. *The Journal of Real Estate Finance and Economics, 39*(4), 450–471.

Ling, D. C., Naranjo, A., & Scheick, B. (2014). Investor sentiment, limits to arbitrage and private market returns. *Real Estate Economics, 42*(3), 531–577.

Ling, D. C., Ooi, J. T., & Le, T. T. (2015). Explaining house price dynamics: Isolating the role of nonfundamentals. *Journal of Money, Credit and Banking, 47*(S1), 87–125.

Liu, X. (2014). Housing supply elasticity, determinant factors and housing price performance of urban china. *Finance & Trade Economics, 35*(4), 125–137.

Liu, X., Du, C., & Li, S. (2019). Natural geographical constraints, land use regulations and china's housing supply elasticity. *The Journal of Economic Research, 54*(4), 99–105.

Lowry, M. (2003). Why does IPO volume fluctuate so much? *Journal of Financial Economics, 67*(1), 3–40.

Miller, N., & Pandher, G. (2008). Idiosyncratic volatility and the housing market. *Journal of Housing Research, 17*(1), 13–32.

Pigou, A. C. (1927). The laws of diminishing and increasing cost. *The Economic Journal, 37*(146), 188–197.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics, 125*(3), 1253–1296.

Shiller, R. C. (2000). Irrational exuberance. *Philosophy and Public Policy Quarterly, 20*(1), 18–23.

Shiller, R. J. (2010). *Animal spirits*. Princeton University Press Princeton.

Soo, C. K. (2018). Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies, 31*(10), 3689–3719.

Tan, Y., Wang, Z., Zhang, Q. H. (2020). Land-use regulation and the intensive margin of housing supply. *Journal of Urban Economics, 115*, 103199.

Wang, S., Chan, S. H., & Xu, B. (2012). The estimation and determinants of the price elasticity of housing supply: Evidence from china. *Journal of Real Estate Research, 34*(3), 311–344.

Wang, Z., & Zhang, Q. (2014). Fundamental factors in the housing markets of china. *Journal of Housing Economics, 25*, 53–61.

Wei, S.-J., Zhang, X., Liu, Y. (2012). *Status competition and housing prices*. National Bureau of Economic Research.

Wu, J., Deng, Y. H., & Liu, H. Y. (2014). House price index construction in the nascent housing market: The case of china. *Journal of Real Estate Finance and Economics, 48*(3), 522–545.

Wu, J., Gyourko, J., & Deng, Y. (2016). Evaluating the risk of Chinese housing markets: What we know and what we need to know. *China Economic Review, 39*, 91–114.

Zheng, S., Sun, W., & Kahn, M. E. (2016). Investor confidence as a determinant of china's urban housing market dynamics. *Real Estate Economics, 44*(4), 814–845.

Zhou, Z. Y. (2018). Housing market sentiment and intervention effectiveness: Evidence from china. *Emerging Markets Review, 35*, 91–110.