International Review of Economics and Finance 97 (2025) 103732

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

The dual role of sentiment on housing prices in China

Xinyu Wang, Zhuangzhi Fang *, Zhenxin Wang

School of Economics, Huazhong University of Science and Technology, Wuhan, 430074, China

ARTICLE INFO

JEL classification:
C32
G41
R21
R31

Keywords:
Housing sentiment
Liquidity
BERT
TVC-VAR

ABSTRACT

This research investigated the role of sentiment on housing prices in China from 2013 to 2022 within a time-varying framework, highlighting the underexplored liquidity channel. We constructed a housing sentiment index by analyzing over 2.44 million social media posts using the Bidirectional Encoder Representations from Transformers (BERT) technique, which is recognized for its advanced capability in processing natural language. With a Time-Varying Coefficient Vector Autoregression (TVC-VAR) model, we estimated the time-varying impulse response of housing prices to sentiment and liquidity shocks, as well as the response of liquidity to sentiment shocks. Our findings revealed that positive public sentiment not only directly pushes housing prices upwards but also indirectly inflates them through an enhanced liquidity, a byproduct of optimism. Notably, the dual role of sentiment becomes more pronounced during the periods with high uncertainty, such as COVID-19 pandemic period. Moreover, our conclusions survived a battery of robustness checks. The results underscore the importance of integrating psychological factors and market dynamics to understand the complexities of real estate markets, particularly in turbulent times.

1. Introduction

Movements in housing markets can not only reflect broad macroeconomic conditions but also significantly influence business cycles (Iacoviello, 2005). The rapid increase of housing prices often lead to bubble formations, the bursting of which can devastate investors and the broad financial system (Cesa-Bianchi, Cespedes, & Rebucci, 2015). The 2008 global financial crisis starkly demonstrates these vulnerabilities (Bernanke, 2008). Hence, the boom and the bust in the housing markets have spurred extensive research.

While substantial studies have focused on advanced economies such as the United States, the dynamics within emerging markets remain underexplored. Notably, housing dynamics in emerging regions exhibit distinct characteristics from those in advanced economies, such as higher and more volatile housing prices inflation (Cesa-Bianchi et al., 2015). Many emerging markets are increasingly vital to global economic growth and fully integrated into the global financial system (Ciarlone, 2011). Disruptions in these markets could provoke repercussions as significant as those triggered by developed countries (Cesa-Bianchi, 2013). Thus, extending existing research to include emerging markets is crucial for understanding the full spectrum of global housing market dynamics and their macroeconomic implications.

Over the past two decades, real estate markets in many emerging economies have demonstrated similar patterns (Ciarlone, 2015). Rapidly soaring housing prices have been a common feature, with annual increases often reaching double digits in both nominal and real terms (Ciarlone, 2011). Post-2008, these markets experienced dramatic reversals, with sharp declines followed by swift recoveries particularly in Asian countries (Ciarlone, 2015). The marked volatility of these markets, characterized by sharp

* Corresponding author.
E-mail address: fangzhuangzhi@hust.edu.cn (Z. Fang).

https://doi.org/10.1016/j.iref.2024.103732
Received 30 May 2024; Received in revised form 21 October 2024; Accepted 15 November 2024
Available online 5 December 2024

rises followed by sudden contractions, suggests speculation at play (Ciarlone, 2015). As such, investigating the drivers behind such dramatic shifts in housing markets of emerging economies becomes imperative.

Traditional economic models, primarily focused on fundamentals and construction costs, fail to fully account for the significant fluctuations observed in housing prices (Case & Shiller, 2003; Duca, Muellbauer, & Murphy, 2021). Recent research has noticed the limitation and increasingly recognized the critical role of sentiment in shaping housing prices (Anastasiou, Kapopoulos, & Zekente, 2021; Shao, Hong, Wang, & Yan, 2023). Furthermore, the real estate market, unlike stock market, is distinguished by higher transaction costs and absence of short-sell mechanism, rendering it particularly susceptible to sentiment (Glaeser & Nathanson, 2015). Given this backdrop, this study aims to explore how sentiment affects housing prices within the volatile emerging economies, seeking to uncover patterns that could benefit policy and investment strategies globally.

Notwithstanding fruitful results existing literature have yielded, several critical aspects of the sentiment-housing price relationship still remain underexplored. Firstly, while the direct impact of sentiment on housing prices is widely acknowledged, detailed investigations into the underlying mechanisms, such as the credit channel examined by Anastasiou et al. (2021), are still very limited. Meanwhile, there is a growing body of literature highlighting the significant place liquidity, the ease at which assets can be transacted, hold in affecting the sentiment-price dynamics (Asriyan, Fuchs, & Green, 2019). Yet, empirical findings pertaining to liquidity-housing price relationship have presented a mixed picture. While a positive correlation between market liquidity and housing prices has been extensively documented (Caplin & Leahy, 2011; Clayton, Ling, & Naranjo, 2009; De Wit, Englund, & Francke, 2013), others reported a negative association (Follain & Velz, 1995; Krainer et al., 2008). Hence, what is the role that liquidity plays in the sentiment-housing prices interplay? Is it an amplifier for an overheated market or a stabilizer to tame the speculative bubble?

Secondly, a large percentage of the prior research have relied on models with constant coefficients, implicitly assuming static relationships among variables (Cox & Ludvigson, 2021; Ling, Ooi, & Le, 2015). However, housing markets globally are deeply intertwined with evolving macroeconomic conditions, policy changes, and significant international events, suggesting a high likelihood of structural changes (Shiller, 2000). This is particular the case in emerging economies, where macroeconomic conditions is more unstable and house price inflation tends to be more volatile than in advanced economies (Cesa-Bianchi et al., 2015; Minetti & Peng, 2013). Consequently, these insights challenge the adequacy of static models and underscore the imperative need for dynamic econometric models.

With these ideas in minds, this paper aims to shed the light on the role of sentiment in shaping housing prices within a time-varying framework, incorporating the liquidity channel simultaneously. Our study drew on observations from China's housing market, which we believe can provide predictive insights applicable to other emerging markets (Shen, Zhao, & Pang, 2024). Specifically, the recent patterns in China's housing market notably parallel those observed in Japan prior to its economic slowdown in the 1990s, characterized by heavy real estate investment, lenient financial conditions, an aging population, and prevalent speculative behavior (Wang, 2017). In light of these similarities, many authoritative economists, including Xiaochuan Zhou, the former president of People's Bank of China, have recommended that China could draw valuable lessons from Japan's historical experiences.[1] Furthermore, shared features, such as swift urbanization, rising household incomes, and evolving financial systems, could significantly shape housing market dynamics in both China and other rapidly developing economies (Cesa-Bianchi et al., 2015; Sufi, 2023). Therefore, the dynamics observed in China's housing market are not only reflective of its internal economic conditions but also provide critical lessons for other emerging markets on similar developmental trajectories. Given this, we focused on the Chinese housing market as our primary case study due to its global significance and representativeness (Fang, Gu, Xiong, & Zhou, 2016).

The quantification of sentiment, however, is thorny, due to its intangible nature (Soo, 2018). Sentiment, which reflects the general beliefs of investors towards a market, encapsulates the aggregate attitudes and speculative tendencies (Baker & Wurgler, 2006). Fortunately, it is well documented that advancements in Natural Language Processing (NLP) have rendered it feasible to effectively capture market sentiment (Algaba, Ardia, Bluteau, Borms, & Boudt, 2020; Zhou, 2018). Among various NLP techniques, Bidirectional Encoder Representations from Transformers (BERT) has been demonstrated to be particularly effective (Devlin, Chang, Lee, & Toutanova, 2018; Gao, Feng, Song, & Wu, 2019; Sousa et al., 2019). BERT's bidirectional architecture, which processes text from both left and right contexts, provides a comprehensive understanding of language, surpassing unidirectional models that only consider preceding text. This capability makes BERT exceptionally suitable for analyzing the nuanced dynamics of housing market sentiment.[2] In light of these, we employed BERT to analyze over 2.44 million *Sina Weibo* posts from 2013 to 2022, thereby constructing a novel housing sentiment index.

Additionally, considering the potential existence of bidirectional and time-varying causal relationships between sentiment, liquidity, and housing price movements, this study employed the Time-Varying Coefficient Vector Autoregression (TVC-VAR) model developed by Gambetti, Pappa, and Canova (2008) and Primiceri (2005) to explore the dynamics among them. The robustness of our findings was further validated through a battery of additional checks, reinforcing the reliability of our findings regarding the complex dynamics of the housing market.

Our contributions to the existing literature are threefold. First, our analysis extended the existing literature on housing sentiment by uncovering the role of the liquidity channel. We revealed the dual effects of sentiment on the housing dynamics: while directly driving housing prices up, positive sentiment stimulates liquidity simultaneously, which in turn further inflates prices. Additionally, our findings complemented the existing literature on liquidity and housing prices by reaffirming the positive correlation between the

[1] See https://cn.nikkei.com/columnviewpoint/viewpoint/55688-2024-05-24-09-26-04.html.

[2] See Appendix A for an overview about BERT model.

two and highlighting liquidity's role as a catalyst in the sentiment-price dynamic. Second, we enriched the literature by incorporating a time-varying perspective to explore the interplay between sentiment, liquidity, and housing prices. Considering the time-varying nature of housing markets amidst rapidly evolving global macroeconomic conditions, we accommodated these shifting relationships and volatilities by using TVC-VAR model. Notably, a major structural shift in the Chinese housing market during the 2020–2022 COVID-19 pandemic was observed, marked by significant economic uncertainty and increased volatility, a phenomenon that static models fail to capture. These findings further highlight the relevance of incorporating time-varying perspective to analysis for housing market. Thirdly, our work constructed a housing sentiment index using BERT, a state-of-the-art NLP method, on a massive dataset of microblogs. To the best of our knowledge, this research marks the first application of BERT to the housing market. In sum, this study not only sheds light on the Chinese housing market but also provides a lesson that can be replicated or adapted to understand housing market dynamics in other emerging economies. The insights gained here are crucial for global economic stakeholders, offering a deeper understanding of how sentiment and liquidity interact in shaping housing markets worldwide.

The paper is organized as follows. Section 2 theoretically explains and models how sentiment impacts on real estate prices through direct and liquidity channel. Section 3 describes the methodology for constructing variables and outlines the model used for empirical analysis. Section 4 presents our main empirical findings and provides a series of robustness checks. Section 5 concludes.

2. Theoretical framework

The principal objective of this study is to explore the influence of sentiment on real estate prices. In this section, we proposed two theoretical channels through which sentiment may impact on housing prices, laying foundation for the empirical tests in this study.

Specifically, we extended the models proposed by Clayton, MacKinnon, and Peng (2008) and Fisher, Gatzlaff, Geltner, and Haurin (2003, 2004), which define buyers' and sellers' reservation prices based on asset- and ownership-specific attributes, alongside broad market influences. Our enhancement systematically integrated the influence of sentiment on these reservation prices as follows:

$$P_{it}^{b} = \sum a_{jt}^{b} X_{ijt}^{P} + \sum \beta_{kt}^{b} D_{kt} + \gamma_t^{b} S_t^{b}, \tag{1}$$

$$P_{it}^{s} = \sum a_{jt}^{s} X_{ijt}^{P} + \sum \beta_{kt}^{s} D_{kt} + \gamma_t^{s} S_t^{s}, \tag{2}$$

where P_{it}^{b} and P_{it}^{s} represent the reservation prices of buyers and sellers, respectively, for asset i at time t. The vector X_{ijt}^{P} encapsulates asset- and ownership-specific attributes, while D_{kt} includes valuation components that are pervasive across the market. The coefficients β_{kt}^{b} and β_{kt}^{s} measure the impact of these factors on the reservation prices. Sentiment, represented by S_t^{b} for buyers and S_t^{s} for sellers, captures the collective beliefs and speculative tendencies of market participants (Baker & Wurgler, 2006). The terms γ_t^{b} and γ_t^{s} quantify the sensitivity of reservation prices to shifts in market sentiment among buyers and sellers, respectively. Transactions occur when a buyer's reservation price meets or surpasses the seller's price:

$$T_{it}^{*} = P_{it}^{b} - P_{it}^{s} = \sum (\alpha_{jt}^{b} - \alpha_{jt}^{s}) X_{ijt}^{P} + \sum (\beta_{kt}^{b} - \beta_{kt}^{s}) D_{kt} + (\gamma_t^{b} S_t^{b} - \gamma_t^{s} S_t^{s}) \geq 0. \tag{3}$$

2.1. The direct channel

Case and Shiller (2003) argued that investors tends to overreact to unforeseen news events, leading to irrational investment decisions and, subsequently, substantial price variations. Moreover, such behaviors, driven by "animal spirits", would amplify market movements—intensifying rises during bull markets and exacerbating falls during downturns (Burnside, Eichenbaum, & Rebelo, 2016; Caraiani, Gupta, Lau, & Marfatia, 2022). Sentiment may also serve as an unobserved source of information about the housing market. Over-exuberant beliefs can push prices away from fundamentals (Soo, 2018). Hence, the parameter γ^{b} and γ^{s} are expected to be positive. Positive (negative) sentiment, $S_t^{b} > 0$ and $S_t^{s} > 0$ ($S_t^{b} < 0$ and $S_t^{s} < 0$), typically increases (decreases) both P_{it}^{b} and P_{it}^{s}. When both buyers and sellers adjust their reservation prices upward (downward) in responses to positive (negative) sentiment, the actual transaction price would increase (decrease), assuming other factors remain constant. Accordingly, we propose the following hypothesis:

Hypothesis 1. Housing prices increase with an increase in housing market sentiment.

2.2. The indirect channel through liquidity

Sentiment might also impact housing prices indirectly through the liquidity channel. As noted, positive sentiment, $S_t^{b} > 0$ and $S_t^{s} > 0$, typically increases reservation prices for both buyers (P_{it}^{b}) and sellers (P_{it}^{s}). If the sensitivities of buyers and sellers to sentiment (γ_t^{b} and γ_t^{s}, respectively), as well as the magnitude of sentiment changes (ΔS_t^{b} and ΔS_t^{s}), are proportionally aligned, the differential impact on reservation prices may not significantly alter the price gap (T_{it}^{*}). The reality, however, seems at odds with this.

First, sellers typically exhibit loss aversion, a behavioral characteristic that often results in a sluggish response to declining market sentiments, indicating a relatively inelastic response to negative sentiment shifts compared to buyers (Genesove & Mayer, 2001). In contrast, buyers, often more urgent in their transactions, exhibit more heightened sensitivity to market signals and news than liquidity providers (sellers) do (Madhavan, 2000). This higher sensitivity among buyers is compounded by speculative motives and irrational exuberance, as suggested by Case and Shiller (2003), which compels buyers to become more inclined to engage in housing transactions and accelerate their purchasing decisions out of fear that prices will escalate to unaffordable levels. Given these dynamics, it is reasonable to hypothesize that the sensitivity of buyers to sentiment changes is greater than that of sellers ($|\gamma_t^b| > |\gamma_t^s|$).

Second, it is well-established that sellers often lag behind buyers in adjusting their reservation prices due to their reliance on market signals derived from buyer behavior (Clayton et al., 2008; Fisher et al., 2003). That is to say, changes in buyers' sentiment (ΔS^b) generally precede analogous changes in sellers' sentiment (ΔS^s). This delay is partly due to the inherent noise in market signals, which sellers cautiously interpret before modifying their valuations. According to Geltner (1990) and Quan and Quigley (1989), the adjustment of sellers' reservation prices can be mathematically expressed as:

$$P_{it}^s = \lambda_t P_{it}^b + (1-\lambda_t)P_{i,t-1}^s = P_{i,t-1}^s + \lambda_t(P_{it}^b - P_{i,t-1}^s), \tag{4}$$

where λ_t represents the confidence factor, a metric indicating the weight sellers place on new market information. This factor could determine how quickly sellers respond to market changes. As the market provides more consistent signals over time, sellers' valuations gradually converge to those of buyers, causing T_{it}^* to progressively drop below 0 simultaneously. Prior to this alignment, when T_{it}^* remains above 0, transaction volumes are likely to increase ($L_t \uparrow$). Therefore, this misalignment in price adjustments, coupled with buyers' higher sensitivity to sentiment, could enhance market liquidity when sentiment is positive and diminish it when negative.

On the other hand, sentiment could influence the confidence factor, λ_t, by affecting the perceived quality of market signals. It is well-documented that market participants exhibiting positive sentiment are often overconfident, as documented by Bielecki et al. (2004). This overconfidence tends to amplify the perceived reliability of positive market signals, thereby increasing λ_t. Conversely, negative sentiment makes sellers more cautious, slowing down their price adjustments. We capture these dynamics by specifying the confidence factor in the following way:

$$\lambda_t = \delta_0 + \delta_1 noise_t(S_t), \tag{5}$$

where $noise_t$ is inversely related to sentiment ($\partial noise_t/\partial S_t < 0$) and δ_1 is expected to be negative. Thus, positive sentiment reduces noise, effectively raising λ_t and allowing sellers to adjust their reservation prices more rapidly in response to market shifts. This dynamic is captured in the adjustment equation:

$$P_{it}^s = P_{i,t-1}^s + \delta_0 \times (P_{it}^b - P_{i,t-1}^s) + \delta_1 \times noise_t(S_t) \times (P_{it}^b - P_{i,t-1}^s), \tag{6}$$

demonstrating that higher sentiment expedites price adjustments, thereby shortening the duration in which $T_{it}^* > 0$.

Moreover, according to the real option theory posited by Novy-Marx (2009), positive sentiment increases the perceived value of waiting for sellers, as they anticipate further favorable market conditions and thus higher potential returns. This outlook encourages sellers, buoyed by optimistic forecasts, to hold out for higher offers from buyers rather than transacting at current market prices. Conversely, pessimistic sentiment diminishes the perceived value of waiting, implying that the opportunity costs of not transacting are increased. Based on the analysis presented, we have developed two competing, testable hypotheses that reflect alternative theoretical perspectives on the impact of sentiment on liquidity:

Hypothesis 2a. Market liquidity increases with an increase in housing sentiment.

Hypothesis 2b. Market liquidity decreases with an increase in housing sentiment.

Regarding to the relationship between liquidity and housing prices, Clayton et al. (2008) and Fisher et al. (2003) argued that higher liquidity, characterized by greater property trading or transaction volumes, could reduce market noise and accelerate the dissemination of information. Building on this premise, we have refined Eq. (5) to reflect these dynamics:

$$\lambda_t = \delta_0 + \delta_1 noise_t(S_t, L_t), \tag{7}$$

where L_t represents the liquidity at time t and is posited to be inversely associated with $noise_t$ ($\partial noise_t/\partial L_t < 0$). Specifically, an upsurge in liquidity ($L_t \uparrow$) could provide sellers with clearer and more consistent signals about market conditions ($noise_t \downarrow$), such as buyers' increased willingness to pay, thereby raising λ_t and prompting sellers to adjust their asking prices upward ($P_{it}^s \uparrow$) (Novy-Marx, 2009). Furthermore, markets characterized by high liquidity are often subject to behavioral phenomena such as herd behavior and speculative trading (Head, Lloyd-Ellis, & Sun, 2014; Van Dijk & Francke, 2018). These behaviors would amplify buyers' sensitivity to market sentiment (i.e., $L_t \uparrow \rightarrow \gamma_t^b \uparrow$), enhancing their tolerance for higher prices. Additionally, as liquidity escalates ($L_t \uparrow$), so does the propensity for buyers to pay more ($P_{it}^b \uparrow$), influenced by a fear of missing out or the perceived robustness of the market (Head et al., 2014; Van Dijk & Francke, 2018). Consequently, buyers' reservation prices are likely to rise ($P_{it}^b \uparrow$) in response to high liquidity. This response ($P_{it}^b \uparrow$), coupled with the elevated reservation prices set by sellers ($P_{it}^s \uparrow$), drives up the prices at which transactions occur, reinforcing the cycle of price inflation in a buoyant market.

In addition, we proposed to integrate the liquidity of individual properties as a specific attribute within our model. As is suggested by Iwanaga and Hirose (2022, 2023) and Jang (2022), less liquid assets would carry a price discount due to the difficulty in selling them, whereas more liquid assets would command higher prices because the liquidity discount diminishes. We embraced the so-called "liquidity premium" in housing markets to hypothesize a positive relationship between the liquidity of individual properties (L_{ijt}^b for buyers and L_{ijt}^s for sellers) and their respective reservation prices (P_{it}^b and P_{it}^s). We integrated these individual liquidity measures into the vector of property-specific attributes X_{ijt}, thereby treating them as intrinsic characteristics that influence valuation. Empirical studies, such as those by Chordia, Roll, and Subrahmanyam (2000), suggest that the liquidity of individual assets (L_{ijt}) often moves in tandem with market liquidity (L_t), due to common factors influencing liquidity across the market. This co-movement implies that in a more (less) liquid market, individual properties are likely to be easier (harder) to sell. Consequently, both buyers' and sellers' reservation prices (P_{it}^b and P_{it}^s) are expected to increase (decrease), suggesting an overall rise (fall) in housing prices driven by increased (decreased) liquidity. With these ideas in minds, we propose:

Hypothesis 3. Housing prices increase with an increase in housing market liquidity.

In Section 4, we shall employ TVC-VAR model to test the above hypotheses based on the data in China, which will be further validated through a series of robustness checks.

3. Materials and methods

This section outlined the methodology and data employed to construct our variables as well as a brief description of the TVC-VAR model, the workhorse of our empirical analysis.

3.1. Housing sentiment index

In this subsection, we developed an indicator aimed at capturing the collective sentiment regarding housing market trends. Recent advancements suggest that measurements of sentiment based on text analysis outperform traditional methods based on surveys, proxies, and internet search indices (Zhou, 2018). Conventional dictionary-based text mining techniques have been criticized for treating words as isolated entities, thereby neglecting the rich contextual dynamics in which they are embedded. BERT advanced previous NLP methods with its deep and bidirectional architecture, vastly exceeding the capabilities of unidirectional models like Long Short-Term Memory (LSTM) (Gao et al., 2019; Sousa et al., 2019). On the other hand, while text mining based on newspapers is subject to the potential bias of news writers (Dougal, Engelberg, Garcia, & Parsons, 2012; Mullainathan & Shleifer, 2005), social media provides a more direct and unfiltered reflection of public sentiment on specific topics (Luca, 2015). Therefore, we employed the BERT model to analyze data from social media, aiming to measure sentiment concerning the housing market.

3.1.1. Collecting and pre-processing data

We sourced a vast corpus of textual data from posts on *Sina Weibo*, one of China's biggest social media platforms.[3] Specifically, we collected all microblogs containing the keyword "house prices" ("Fang Jia" in Chinese) based on *Sina Weibo*'s platform from January 1 2013 to December 31 2022.[4] To ensure the quality and originality of the data, we included only the original microblogs. Then, we extracted its content, posting dates, authors' usernames, and the total number of likes, forwards, and comments. Consequently, this corpus contains more than 2.44 million microblogs.

We then performed meticulous pre-processing on our collected data. Particularly, we eliminated non-textual elements such as links, emoticons, and multimedia to concentrate on the textual content. Traditional Chinese characters were converted to simplified Chinese for consistency. Recognizing the complexity of microblogs, which may contain multiple sentences with varied meanings, we segmented each post into individual sentences. Notably, we retained stop words and punctuation in our breakdown, considering their potential semantic contribution to understand sentiment.

3.1.2. Text classification

In our text classification framework, sentences were categorized into one of three distinct sentiments: irrelevant or neutral, positive, and negative, reflecting the varied attitude of public discourse around housing price trend. Specifically, sentences that neither deliberate on house prices nor convey a clear stance regarding house price trends fall into the irrelevant or neutral category. For example, many microblogs with "house price" are just advertisements. And some express specific concerns or queries about housing prices without indicating a clear trend. We classify these into the irrelevant group. Moreover, we attribute positive (negative) sentiment to sentences explicitly refuting negative (positive) trends.

In our study, we chose RoBERTa-wwm-ext model, developed by the Joint Laboratory of Harbin Institute of Technology (HIT) and iFLYTEK Research (HFL), as the workhorse to classify our corpus (Cui, Che, Liu, Qin, & Yang, 2021).[5] Distinguished from the Chinese BERT model proposed by Google, which analyzes text at the character level, the RoBERTa-wwm-ext model employs HIT Language Technology Platform (LTP) technique, which is particularly advantageous for capturing the nuances of Chinese semantic context, for Chinese Word Segmentation.[6] Moreover, this model is further pre-trained by the whole word mask technique pertaining on the Extended (EXT) dataset.[7] Therefore, we believe that it could serve as a capable tool for our subsequent sentiment analysis.

[3] See https://en.wikipedia.org/wiki/Weibo and https://m.yicai.com/news/101900321.html. *Sina Weibo* possesses more than 605 million active monthly users in the third quarter of 2023.

[4] Though *Sina Weibo* was launched in 2009, we found that microblogs with "housing prices" are very sparse before 2013, which might render enormous noise for our analysis.

[5] See https://github.com/ymcui/Chinese-BERT-wwm/blob/master/README_EN.md for a comprehensive introduction and performance evaluation comparing the RoBERTa-wwm-ext model with others.

[6] See https://ltp.ai/ for more details about HIT LTP technique.

[7] EXT dataset includes Chinese Wikipedia, encyclopedic data, news, and etc., containing a substantial total word count of 5.4 billion words.

Since the RoBERTa-wwm-ext model is pre-trained on general topics, we undertook the fine-tuning process to tailor its performance for sentiment analysis in housing market. Specifically, we randomly selected over 28,600 sentences (i.e., 0.27%) from our dataset and manually labeled them. Then, we used these labeled sentences to fine-tune the RoBERTa-wwm-ext model. We employed the five-fold cross-validation method to evaluate the out-of-sample prediction accuracy of our fine-tuned model, detailed in Appendix B. The results in Table B.1 display the superiority of BERT. To offer more intuition, Table B.2 presents a small part of our sample, exhibiting the original Chinese text alongside its English translation and the sentiment category assigned by the BERT model. The model's precision shines in complex scenarios. For instance, the fourth sentence containing "high housing prices" might be mistakenly classified into positive sentiment in a dictionary-based approach. However, BERT accurately identified the lack of a clear sentiment towards housing price trends in such context. Encouraged by these promising outcomes, we employed the fine-tuned RoBERTa-wwm-ext model to further label the entire corpus of 10.6 million sentences.

3.1.3. Index building

The sentiment index, SI_t, for each month t is computed from the classified sentences, representing the probability of positive sentiment polarity:

$$SI_t = \frac{pos_t}{neg_t + pos_t} \tag{8}$$

where pos_t and neg_t denote the count of positive and negative microblog sentences for month t, respectively. This index ranges between 0 and 1, with values above 0.5 indicating positive market sentiment and below 0.5 indicating negative sentiment.

3.2. Housing prices and market liquidity

Following Hong and Li (2020) and Shao et al. (2023), we adopted the returns on newly-built housing (HR_t) across 70 large and medium-sized cities, released by Chinese National Statistical Bureau, as the proxy for housing market dynamics. We retrieved the month-on-month growth data for newly-built housing prices from the official bureau database.[8] We established January 2011 as the baseline, setting the housing index at 100, and subsequently computed the indices for each city by calculating the cumulative product of the previous month-on-month growth rates. The national composite housing price index was then calculated as the mean of these 70 city-specific indices, serving as the proxy for the national housing prices in China. Monthly housing price return was then computed as: $HR_t = 100 \times ln(P_t/P_{t-1})$, where P_t is the monthly national housing price index.

Given the unavailability of transaction data for these 70 cities, we chose data of transacted units of newly-built commercial resident housing across 30 large and medium-sized cities sourced from the WIND database to align our liquidity and house price indicators as close as possible.[9] The housing transactions and discussions are predominantly concentrated in these major cities in China, ensuring that our indicators accurately reflect national market dynamics (Xie & Chen, 2018).[10] We constructed the market liquidity index, ML_t, based on the trading volumes. Due to the pronounced seasonal effects typical in monthly sales data, we employed the X-12 ARIMA methodology recommended by the U.S. Census Bureau for seasonal adjustment.[11] The monthly market liquidity indicator was calculated as: $ML_t = ln(TV_t/TV_{t-1})$, where TV_t is the monthly transacted units of newly built commercial resident housing in the 30 large and medium-sized cities after adjustment.

3.3. Summary statistics

Following existing real estate literature, we included several control variables, namely the Output Growth Rate (OGR), Inflation Rate (IR), and Real Interest Rate (RIR), with detailed description provided in Appendix C. Our sample spans from January 2013 to December 2022. Table 1 provides a statistical summary. The Augmented Dickey–Fuller (ADF) test has been performed to confirm the stationarity of our time series data, ensuring the variables are suitable for our following VAR analysis.

Moreover, Fig. 1 depicts the sentiment, liquidity, and housing price movements, marked by significant regulatory interventions in China's housing market. Typically, higher sentiment tends to align with increasing liquidity and upward price trends, while the opposite holds true. Sentiment was high in 2013 but plummeted swiftly following the release of the "New National Five Regulations" in February 2013. Policy shifts beginning in 2014 corresponded with a resurgence in sentiment, peaking in 2015 concurrent with property inflation. The "Shanghai Nine Regulations" in March 2016 marked the start of tightening measures, resulting in subdued market confidence throughout 2017. In contrast, late 2018 experienced a resurgence in confidence, coinciding with periods of monetary easing, while 2019 observed more subdued sentiment during times of enhanced regulatory measures. The onset of COVID-19 in 2020 temporarily altered regulatory focus, coinciding with short-lived sentiment changes. However, subsequent measures, such as the "Three Red Lines" requirements and centralizing land supply in pilot areas, saw sentiment gradually reach its lowest point. Overall, it appears that our constructed sentiment index likely reflected people's beliefs and responses to market events over the past decade.

[8] See http://data.stats.gov.cn/ for more details about the 70 large and medium-sized cities.

[9] The 30 large and medium-sized cities in China include Beijing, Shanghai, Guangzhou, Shenzhen, Chengdu, Dalian, Fuzhou, Hangzhou, Harbin, Changchun, Changsha, Nanchang, Nanjing, Qingdao, Suzhou, Wuhan, Xiamen, Anqing, Baotou, Dongguan, Foshan, Huizhou, Jiangyin, Kunming, Lanzhou, Nanning, Shaoguan, Shijiazhuang, Wuxi, Yangzhou, and Yueyang. For more detailed introduction, please refer to WIND database.

[10] See https://finance.sina.cn/china/gncj/2023-02-24/detail-imyhuuvv9200314.d.html.

[11] See https://www.census.gov/content/dam/Census/library/working-papers/1998/adrm/jbes98.pdf for more details about the X-12 ARIMA methodology.

Table 1
Statistics summary of variables.

	Mean	Std	Min	Max	Skewness	Excess Kurtosis	ADF
SI	0.35	0.10	0.17	0.57	0.24	−0.89	−3.01**
ML	0.00	0.16	−0.87	0.78	−0.16	12.08	−8.58***
HR	0.34	0.51	−1.17	1.95	−0.21	0.80	−3.25**
OGR	0.60	1.74	−9.23	8.41	−0.07	14.73	−5.72***
IR	0.15	0.50	−1.21	1.59	0.15	0.39	−3.15**
RIR	1.71	0.80	−0.10	3.90	0.46	−0.23	−5.05***

* The skewness and kurtosis of normal distribution are 0 and 3, respectively. ***, **, * indicate the statistical significance level of 10%, 5%, and 1%, respectively.

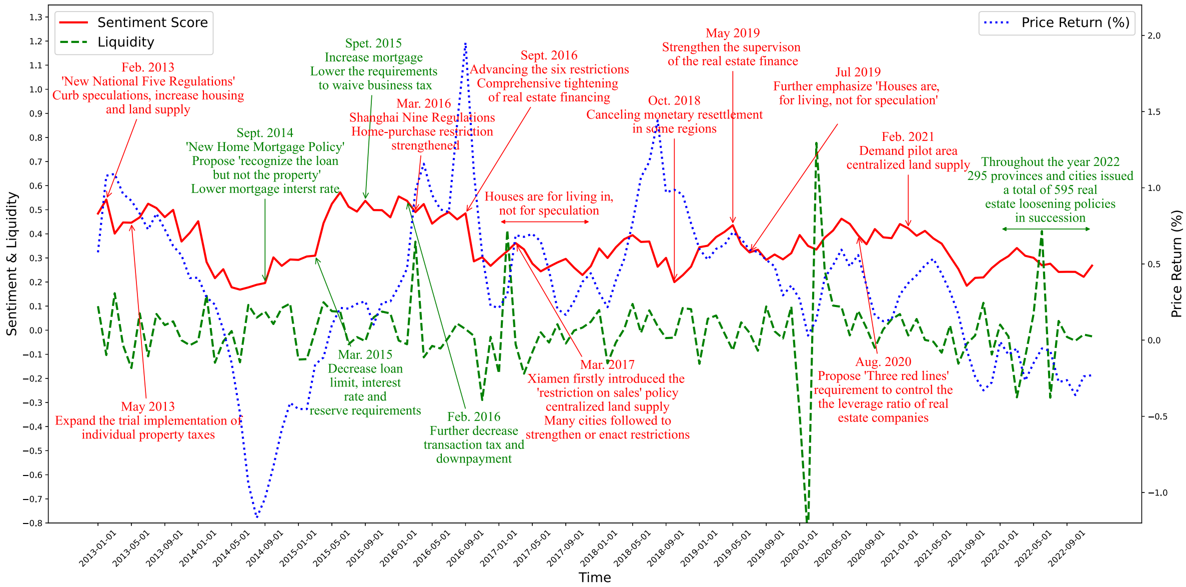

Fig. 1. The sentiment index, market liquidity, and housing price variations.

3.4. Empirical model

3.4.1. VAR model with time-varying coefficients

According to Del Negro and Primiceri (2015) and Primiceri (2005), the reduced-form VAR with time-varying coefficients could be written as:

$$x_t = A_{0,t} + A_{1,t}x_{t-1} + A_{2,t}x_{t-2} + \cdots + A_{p,t}x_{t-p} + u_t \tag{9}$$

where x_t represents a vector of observed endogenous variables. It includes housing sentiment index (SI), housing market liquidity (ML), housing prices (HR), output growth rate (OCG), inflation rate (IR), and real interest rate (RIR) in this study. $A_{0,t}$ is a vector of time-varying intercepts, and $A_{i,t}$ (for $i = 1, \ldots, p$) are matrices of time-varying coefficients. The lag order p is determined by information criteria. The vector of reduced-form innovations u_t follows a white noise Gaussian process with a mean of zero and a covariance matrix Σ_t.

We assume u_t is a linear transformation of the underlying structural shocks ε_t: $u_t \equiv S_t\varepsilon_t$, where $E\{\varepsilon_t\varepsilon_t'\} = I$ and $E\{\varepsilon_t\varepsilon_{t-k}'\} = 0$ for all t and $k = 1, 2, 3, \ldots$. Additionally, we assume $S_tS_t' = \Sigma_t$. The corresponding structural VAR is therefore given by:

$$S_t^{-1}x_t = S_t^{-1}A_{0,t} + S_t^{-1}A_{1,t}x_{t-1} + S_t^{-1}A_{2,t}x_{t-2} + \cdots + S_t^{-1}A_{p,t}x_{t-p} + \varepsilon_t. \tag{10}$$

3.4.2. Bayesian VAR estimates

Following Gali and Gambetti (2015), we denote $\theta_t = vec(A_t')$ where $A_t = [A_{0,t}, A_{1,t}, \ldots, A_{p,t}]$. θ_t is assumed to evolve over time according to the process:

$$\theta_t = \theta_{t-1} + \omega_t, \tag{11}$$

where ω_t is a Gaussian white noise process with zero mean and constant covariance Ω, independent of u_t at all leads and lags.

The time variation of Σ_t is modeled as $\Sigma_t = F_t D_t F_t'$, where F_t is lower triangular, with ones on the main diagonal, and D_t is a diagonal matrix. Let σ_t be the vector containing the diagonal elements of $D_t^{1/2}$ and $\phi_{i,t}$ a column vector with the non-zero elements of the $(i+1)$-th row of F_t^{-1} with $i = 1, 2, \ldots, 5$. We assume:

$$log\sigma_t = log\sigma_{t-1} + \zeta_t, \tag{12}$$

$$\phi_{i,t} = \phi_{i,t-1} + \nu_{i,t}, \tag{13}$$

where ζ_t and $\nu_{i,t}$ are white noise Gaussian processes with zero mean and covariance matrices Ξ and Ψ_i, respectively. We assume that $\nu_{i,t}$ is independent of $\nu_{j,t}$, for $j \neq i$, and that ω_t, ε_t, ζ_t and $\nu_{i,t}$ (for $i = 1, \ldots, 6$) are mutually uncorrelated at all leads and lags.

This study employed Cholesky decomposition of Σ_t. The variables' ordering is ranked based on the economic rationale. Particularly, GDP has been posited first due to its foundational influence on all subsequent variables as a broad indicator of economic activity. The inflation rate follows, reflecting its status as a primary concern for monetary policy. Next, the real interest rate is placed, representing monetary policy responses that typically adjust after observing changes in output and inflation, yet precede variables of our interest. Housing prices come after these macroeconomic indicators, acknowledging that housing markets react to general economic conditions and monetary policy. As posited by Caplin and Leahy (2011) and Soo (2018), housing prices exhibit a sluggish response to liquidity, while liquidity, in turn, demonstrates a delayed reaction to sentiment. Given these, we have the order of the observed endogenous variables: $x_t = [OGR_t, IR_t, RIR_t, HR_t, ML_t, SI_t]'$.

Following Gali and Gambetti (2015) and Primiceri (2005), Bayesian methods with the Gibbs sampling are employed to estimate the model with time-varying coefficients. Specifically, we use the first $\tau = 48$ samples to calibrate the initial state. To avoid losing too many data points, we started our first draw from $\tau/2 + 1$. Then, we performed 22,000 draws, while discarding the first 20,000 to ensure the convergence.[12]

4. Empirical results

This section presented the time-varying impulse responses estimated from our baseline TVC-VAR model first. Then, we discussed the outcomes of robustness checks.

4.1. Time-varying impulse responses

We employed impulse response analysis to investigate the time-varying effect of one variable on the other variable in the VAR model. Following Gali and Gambetti (2015), we integrated an additional dimension for a better illustration of the time-varying dynamic nature inherent in our Bayesian VAR model's impulse responses. Figs. 2 and 3 display the impulse responses of housing prices and liquidity to a one-standard-deviation sentiment shock over a 24-month horizon, which is inspired by the results in Soo (2018), from 2015 to 2022, respectively. Fig. 4 demonstrates the corresponding response of housing prices to the shock from liquidity.

It can be observed from Fig. 2 that a one standard deviation shock to market sentiment caused consistently positive responses in housing prices though the duration and magnitude of these responses vary over time, supporting our Hypothesis 1. Specifically, the duration of responses in housing prices did not variate much over time. They typically peaked in the seventh month (approximately 0.6%) and gradually declined afterwards while still remained positive, echoing the findings of Ling et al. (2015) and Soo (2018). These responses generally converged to zero after 20 months, reflecting a profound influence of sentiment, in line with the observations of Hui, Dong, Jia, and Lam (2017). This pattern may be attributable to factors such as high transaction cost and lengthy transaction times, which introduce inertia into the market. Consequently, sentiment-driven effects tend to manifest persistently over extended periods.

Regarding their patterns over time, sentiment's impacts initially declined from 2015, and reached a trough in September 2016. For instance, the housing price response was only 0.467% in September 2016 with a seven-month lag. This decline coincides with a period of stringent regulation in China, including the "Shanghai Nine Regulations" and the directive principle that "Houses are for living in, not for speculation", which likely temper speculative fervor and moderate sentiment's impact. Following this period, the responses of housing prices displayed a mild increase till mid-2019, followed by a brief decline until early-2020. Thereafter, a sharp escalation in response magnitude can be observed. This variation pattern coincides with significant global and domestic shifts, such as the deceleration of China's economic growth, the Sino-US trade war, and the substantial economic disruptions caused by the COVID-19 pandemic. These events have introduced heightened uncertainty, which likely magnified the role of sentiment as a primary driver in the housing market dynamics (Burnside et al., 2016).

Fig. 3 lends credence to our Hypothesis 2a, with liquidity's responses to sentiment shocks persistently positive. In comparison to housing prices' responses to sentiment shocks, the liquidity's were relatively short-lived, indicating its quicker adjustment in the market. Liquidity peaked in responses with a 2-month lag — suggesting the frictions in the matching process — and then reverted to base levels after 7 months. This might be attributed to the fact that the increased housing demand driven by sentiment does not inherently alter fundamental demand–supply relationship in a lasting manner (Shiller, 2000).

[12] See Gali and Gambetti (2015) for more estimation technical details.

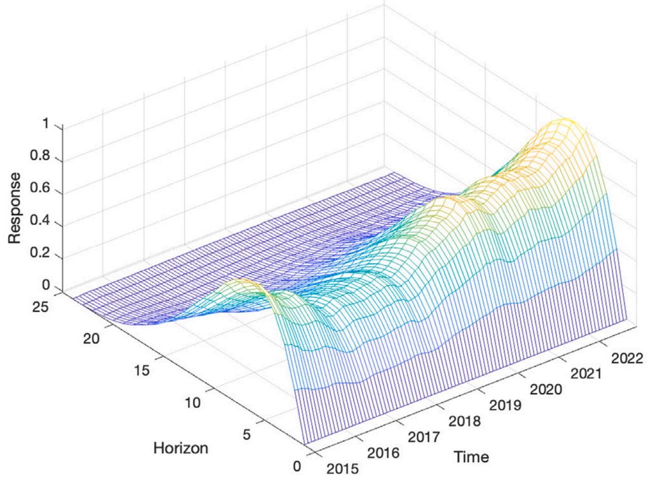

Fig. 2. Sentiment shock to prices: Baseline case.

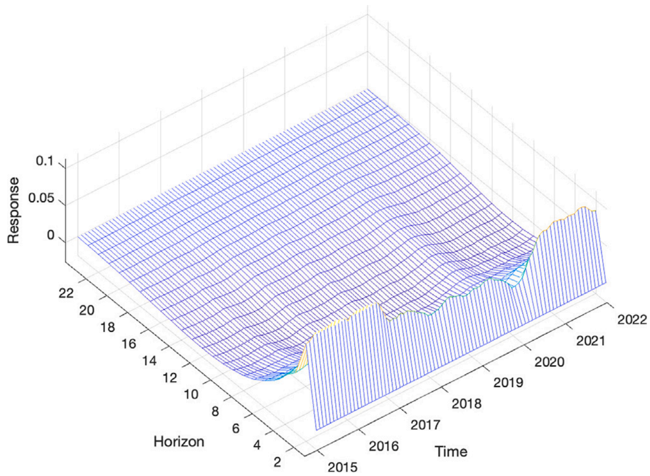

Fig. 3. Sentiment shock to liquidity: Baseline case.

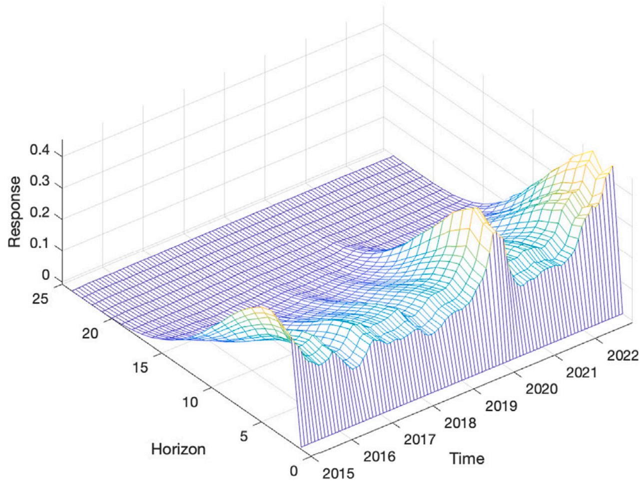

Fig. 4. Liquidity shock to prices: Baseline case.

Concerning the magnitude, the liquidity's maximum responses to sentiment shocks, ranging from 0.039% to 0.126% at the 2-month lag, were weaker than those observed in housing prices, ranging from 0.467% to 1.011% with a 7-month lag. When it comes to the time-varying characteristics of the liquidity responses, two approximate hump-shaped patterns can be observed. The first pattern spans from 2015 to 2016, and the second aligns with the COVID-19 pandemic period from 2020 to 2022. Between the years 2017 to 2019, the trajectory remained relatively stable, lingering around 0.08% at the 2-month lag. As previously mentioned, we conjectured that the heightened sensitivity between 2020 to 2022 might be attributed to the economic slowdown and the prevailing uncertainty in that period. The resultant economic anxieties likely made market participants particularly over-react to sentiment, heightening the liquidity's sensitivity to psychological factors. Moreover, housing is traditionally viewed as a robust investment in China (Wu & Tidwell, 2015). This perception likely fuel housing investments during volatile periods to hedge risk (Eraslan, 2016). Such surges in demand might manifest as an enhanced sensitivity of liquidity to sentiment.

Fig. 4 presents the persistent positive response of housing prices to a one-unit standard deviation shock of liquidity, which substantiates Hypothesis 3. This positive association aligns with the findings in stock market (Iwanaga & Hirose, 2022, 2023; Jang, 2022). Moreover, housing prices typically peaked in their response around four to five months later following a liquidity shock. Interestingly, expanded observations on Figs. 2 & 3 reveal that liquidity reacts to sentiment shocks with its peak effect materializing within two to three months, while the most pronounced impact of sentiment on housing prices emerges after approximately seven to eight months. The alignment of these peak timings robustly supports the hypothesis that liquidity acts as a crucial intermediary between sentiment and housing prices. This sequence of effects highlights the intricate role of liquidity in mediating the transmission of sentiment onto housing market dynamics, emphasizing its pivotal function in this complex interplay. Furthermore, the duration of these responses exhibited temporal variation: from 2015 to 2018, responses generally dissipated after 12 to 14 months, extending to 19 months in 2019 and 18 months from 2020 to 2022. This temporal extension slightly deviated from stock market behaviors, where the influence of liquidity is often short-lived, typically dissipating within a few months (Jang, 2022). Due to the high subjectivity involved in valuing housing assets, discrepancies between perceived and actual values might lead to a persistent positive association between liquidity and housing prices (Lam & Hui, 2018).

Moreover, two hump-shaped patterns are discernible in Fig. 4. In particular, until mid-2018, responses remained relatively stable at approximately 0.02% with a four-month lag. However, responses intensified sharply starting in mid-2018, peaking at 0.429% with a four-month lag, and kept gradually declining before December 2019. This pattern was followed by another pronounced half-hump from 2020 to 2022, peaking at 0.455% in December 2022 with the same 4-month lag, coinciding with the COVID-19 pandemic. The impacts of liquidity from 2016 to 2019 were relatively muted.

In sum, sentiment exerts a dual effect on housing prices: a direct positive influence and an indirect positive effect via liquidity. A plausible interpretation is that optimistic sentiment directly inflates prices by fostering "irrational exuberance" and facilitating transactions at elevated prices (Case & Shiller, 2003). Concurrently, enhanced liquidity, spurred by positive sentiment, facilitates higher prices by encouraging speculative trading. This, in turn, induces herding behavior, which further escalates housing prices. Thus, increased liquidity serves as an additional mechanism through which sentiment can intensify its impact on housing market dynamics. Moreover, significant time-varying patterns have been detected. Both the direct sentiment-price dynamic and the mediating role of liquidity become particularly pronounced during periods of heightened uncertainty and volatility, such as COVID-19 pandemic era, highlighting the importance of adopting a time-varying perspective in analyzing these relationships.

4.2. Robustness checks

To assess the robustness of our results in Section 4.1, we conducted a comprehensive series of robustness checks. Firstly, we substituted the original sentiment with the weighted one according to their influence. Subsequently, we replaced the specification of housing prices by the average housing prices across 100 cities and another GDP-weighted housing prices across 30 large- and medium-sized cities. Lastly, we applied our analysis to China's second-hand housing market.

We firstly transitioned from using an equally-weighted sentiment series to the one weighted by influence. Previously, when aggregating the classification results of microblog sentences into sentiment indexes, each sentence was treated with equal weight. However, the engagement metrics associated with a microblog, such as the number of reposts, comments, and likes, can signify its influential power on social media. These metrics were therefore computed and subsequently employed as weights in our construction of the sentiment index. The adjusted results, depicted in Figs. D.1-D.3, are broadly consistent with our baseline findings. The minor divergence is that the effect of sentiment on liquidity and housing prices is less pronounced during COVID-19 pandemic period. This is expected. Notably, posts that garner high engagement are often authored by online influencers who, during times of crisis such as a pandemic, tend to stabilize public sentiment rather than exacerbate panic in China. This tendency can lead to an underestimation of the impact of crises on emotional responses, fostering a more optimistic outlook during such periods. Consequently, this may lead to results that depart from those obtained using an equally-weighted sentiment index, which might more accurately reflect the sentiments among the general populace.

To further validate our findings, we conducted another robustness check by replacing the composite housing prices index of newly-built housing across 70 large and medium-sized cities by the average housing prices across 100 cities, sourced from the WIND database.[13] Compared with the commonly used housing prices index from the 70 cities, the data across 100 cities could offer a broader coverage and potentially a more comprehensive representation of China's national housing market dynamics. Inspections

[13] For more compilation details, please refer to the introduction in WIND database.

to Figs. D.4–D.6 revealed that the results are largely consistent with our initial findings. The minor difference is that all the responses became more significant during 2015–2019 period. This variation is plausible given that, in contrast to the 70-city index, which predominantly includes large and medium-sized cities, the 100-city average encompasses a higher proportion of less developed cities. These cities tends to undergo rapid urbanization, a factor that likely enhances their sensitivity to market sentiment.

Moreover, a subsequent robustness test involved substituting the composite housing prices index with a weighted average index across 30 large and medium-sized cities, which are identical to those used in the liquidity indicator construction. Here, the annual GDP of each city served as the weighting factor. The outcomes, illustrated in Figs. D.7–D.9, largely align with our primary analysis, though the responses were observed to be more transient compared to the baseline scenario. This variation can be logically explained by the new index's exclusion of many medium-sized cities included in the baseline index. A plausible explanation is that the larger cities within the 30-city index are equipped with more robust regulatory frameworks, diminishing the overall sentiment's impact compared to medium-sized cities.

In our baseline analysis and the previous three robustness checks, we focused on the newly-built housing market. Given that the second-hand housing market also plays a critical role in the overall dynamics of China's housing market, we shifted our analysis to consider liquidity and housing price indicators specific to second-hand houses. We downloaded listing volume and price indices for second-hand houses provided from the China Real Estate Association.[14] Due to data availability constraints, these time series commence from January 2015. We calculated monthly liquidity and housing prices indicators as follows: $AML_t = (V_t - V_{t-1})/V_{t-1}$ and $AHR_t = 100 * (PI_t - PI_{t-1})/PI_{t-1}$, where V_t and PI_t represent the listing volume and price indices for second-hand houses in month t.[15] The results from 2017 to 2022 shown in Figs. D.10–D.12 confirm our conclusions are valid even in the second-hand housing market, indicating that our findings can be generalized across the entire housing market. The observed patterns in these figures bear strong resemblances to our baseline results, albeit with some minor variations. For instance, liquidity responses to sentiment shocks in the second-hand market are more volatile and pronounced than those in the newly-built housing market, potentially due to fewer regulatory constraints in the repeated sale market. Additionally, the responses of second-hand housing prices to liquidity shocks began to decline at the beginning of 2022, while they remained significant in the newly-built housing market. This divergence is reasonable, as ongoing economic downturns exert a more pronounced effect on the second-hand housing market, which typically requires more immediate capital input.

These findings attest to the robustness of our conclusions, highlighting sentiment's substantial dual role in housing market dynamics which also influenced by shifting market conditions and regulatory landscapes.

5. Conclusion

In this study, we explored the mechanisms through which sentiment impacts housing prices within a time-varying perspective, highlighting the liquidity channel. We employed BERT to construct a sentiment index based on more than 2.44 million microblogs and applied TVC-VAR model to Chinese data from 2013 to 2022. Our analysis acknowledges the potential dual effects of sentiment on housing prices. Firstly, positive sentiment directly drives prices upward. Secondly, positive sentiment indirectly inflates prices through enhanced liquidity. Moreover, these effects are notably amplified during periods of high uncertainty, such as the COVID-19 pandemic. Our findings remained robust across different sentiment and housing price series included in the model and held true when applied to the second-hand housing market.

The implications of our research are manifold. First, we advocate for the development of regulatory frameworks that prioritize transparency and limit speculative activities, such as tiered property tax rates to discourage property hoarding, stricter lending criteria for investment properties, and incentives for affordable housing development. Additionally, implementing rental market regulations and utilizing technology such as blockchain for transaction transparency may enhance market stability. Moreover, the establishment of a real-time monitoring system for housing market sentiment could serve as a global model for capturing the cyclical nature of housing markets. Such systems, coupled with our recommended regulatory measures, can facilitate informed policy discussions aimed at maintaining market stability across various economic contexts, particularly during global crises such as pandemics. Our findings, which reveal a time-varying relationship among key market variables, support the need for adaptive regulatory measures. These measures should be agile enough to evolve in response to changing market conditions, ensuring resilience against unpredictable macroeconomic shocks. By adopting these approaches, policymakers worldwide can not only stabilize the housing market but also bolster the broader financial system, contributing to sustained economic stability on an international scale.

However, we also acknowledge some limitations for our research which remains for future extensions. Currently, our analysis was conducted at the national level. Future research could aim to develop sentiment indices at the city level and explore potential heterogeneous effects of sentiment across different urban areas. Such detailed insights could further refine policy interventions tailored to specific market conditions.

CRediT authorship contribution statement

Xinyu Wang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Zhuangzhi Fang:** Conceptualization, Writing – review & editing. **Zhenxin Wang:** Data curation, Validation, Writing – review & editing.

[14] See http://www.fangchan.com/ and https://research.ke.com/ for more details.

[15] We multiplied the AHR series by 100 to ensure that the values are not too small.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to Brian Lucey (the editor) and two anonymous referees for their constructive comments. Xinyu Wang acknowledges the support of the China Scholarship Council program (Project ID: 202306160084). We also appreciate the excellent research assistance provided by Tan Jun Rui.

Appendix A. Introduction of BERT

Recent advancements in Natural Language Processing (NLP) have been profoundly influenced by the introduction of pre-trained language models, constituting a significant paradigm shift. Among the most notable developments are Google's BERT and OpenAI's GPT series, which are characterized by their initial unsupervised training on vast corpuses of textual data, followed by fine-tuning on smaller, task-specific labeled datasets. This methodology has enhanced performance across a range of NLP tasks, including text classification, sequence labeling, and machine reading comprehension.

The BERT model, particularly, undergoes a sophisticated two-phase training regimen. During the pre-training phase, BERT is designed to capture linguistic patterns at the character, word, sentence, and inter-sentence levels through a variety of tasks. The Masked Language Model (MLM) and Next Sentence Prediction (NSP) are foundational to this phase, fostering the model's deep contextual and relational insights. MLM obscures segments of the text, encouraging the model to predict the hidden elements in a manner reminiscent of a cloze test, thereby enhancing its bidirectional contextual understanding. Concurrently, NSP challenges the model to determine the sequential coherence between sentence pairs, enriching its comprehension of textual relationships.

Fundamentally based on the Transformer architecture, BERT leverages the multi-head self-attention mechanism, setting it apart from traditional neural network models like CNNs and RNNs. This structure enables unmediated interactions between any pair of words within a sentence, addressing the challenges of dependency parsing and promoting parallel computational efficiency. Input to the network consists of token representation vectors—the aggregation of word embeddings, position embeddings, and, for certain tasks, sentence embeddings. Position embeddings provide crucial information about token positioning, a critical aspect not inherently accounted for in Transformer architectures. Sentence embeddings are utilized selectively, based on the task's contextual requirements.

The multi-head self-attention mechanism is integral to BERT's design, allowing it to parse complex semantic structures within natural language by generating token representations as adaptive weighted sums of all token representations in a sentence. This ensures that context remains confined to the input sentences. The architecture's sequential attention layer processing culminates with the final model output, derived from the last attention layer.

A unique aspect of BERT is the [CLS] token, which precedes input sentences. For text classification tasks, the representation of this token after fine-tuning is used to inform the softmax classification layer. This innovative methodology highlights BERT's critical contribution to NLP, redefining standards for language model training and application with unprecedented accuracy and efficiency.

Appendix B. Classification accuracy

To assess the classifiers' out-of-sample prediction accuracy, we utilized five-fold cross-validation, a widely adopted method. This entails partitioning the labeled corpus into five equally sized subsets, with each subset constituting one-fifth of the entire corpus. The classifiers are trained five times in succession, each time using a different fold as the testing set and the remaining four as the training set. After each training session, the classifiers predict the outcomes in the corresponding test set. The average prediction accuracies across the five folds serve as a measure of the classifiers' performance. To mitigate overfitting, early-stopping techniques were used. Table B.1 presents the BERT model's accuracies, detailing performance across training and test datasets, as well as the averaged results. For comparative analysis, we also included results from the LSTM model employed by (Zhu et al. 2023). The LSTM, a variant of the Recurrent Neural Network (RNN), sequentially encodes semantic information of words within a sentence but relies on unidirectional architecture, conditioning outputs solely on preceding words (left context). According to Table B.2, our BERT model, achieving an impressive average out-of-sample accuracy of 85.33\%, significantly outperform the LSTM's 75.73\%, demonstrating its effectiveness in following text classification. This aligns with findings within the NLP field (Devlin et al. 2018; Sousa et al. 2019). The BERT model's ability to capture bidirectional contextual information may contribute to its superior performance. Specifically, we employed the model which trained on the former four folds to classify our entire corpus, motivated by its best performance.

Table B.1
Comparison of accuracy on training and test data.

Algorithm	Data type	First fold	Second fold	Third fold	Fourth fold	Fifth fold	Average
LSTM	Training Data	75.64%	75.78%	75.90%	75.84%	75.50%	75.75%
	Test Data	76.11%	75.53%	75.06%	75.32%	76.66%	75.73%
BERT	Training Data	93.65%	96.02%	93.31%	93.31%	93.32%	93.92%
	Test Data	85.17%	84.81%	83.70%	85.14%	86.31%	85.33%

Table B.2
Results of examples of classification.

Microblog sentences in Chinese	Microblog sentences in English	Category
我预计 2021 年，中国住房市场以稳中趋升为主，大部分城市仍可能较快上涨。	I predict a generally stable yet rising trend in China's 2021housing market, with many cities likely to see quick prices increases.	Positive
从这个细节能看得出出来，如同我们年初的预测，今年房价地价还是会被压制住。	This detail suggests, as we predicted at the start of the year, that housing and land prices will continue to be suppressed this year.	Negative
你对于房价的心态是？	What is your attitude toward housing prices	Irrelevant
因为高房价带来的暴利，吸引了大量资本进入房地产行业，这阻碍了其他产业的发展和创造就业岗位，严重制约实体经济的发展	High housing prices, attracting substantial capital into the real estate sector, hinder the development of other industries and job creation, severely constraining the growth of the real economy.	Irrelevant

Appendix C. Details of control variables

This subsection shall provide a detailed description of the control variables' construction, namely *GDP*, inflation rates (*IR*), and real interest rates (*RIR*). The Consumer Price Index (*CPI*) on a month-over-month basis is adopted as a proxy for *IR*, while the Real GDP growth rate proxies for *OGR*. The Real Interest Rate is calculated by adjusting the nominal fixed-term deposit rate for the *CPI* growth rate. Specifically, we re-calibrated the month-over-month *CPI* using January 2013 as the base period. Due to the unavailability of monthly *GDP* figures, we applied the quadratic-match sum method to interpolate quarterly *GDP* data into monthly values, which were then seasonally adjusted. The nominal *GDP* was deflated using the CPI index to derive the Real *GDP*. Additionally, the deflated regular fixed-term deposit one-year interest rate served as the proxy for the Real Interest Rate. The primary data sources for this analysis were China's National Statistical Bureau and the CSMAR database. Their detailed constructions are as follows:

$$\text{Output Growth Rate} = ln\left(\frac{\text{Real GDP}_t}{\text{Real GDP}_{t-1}}\right), \tag{C.1}$$

$$\text{Inflation Rate} = ln\left(\frac{CPI_t}{CPI_{t-1}}\right) \times 100, \tag{C.2}$$

$$\text{Real Interest Rate} = \text{Nominal Interest Rate} - \text{CPI Growth Rate}. \tag{C.3}$$

Appendix D. Additional results

See Figs. D.1–D.12.

Data availability

Data will be made available on request.

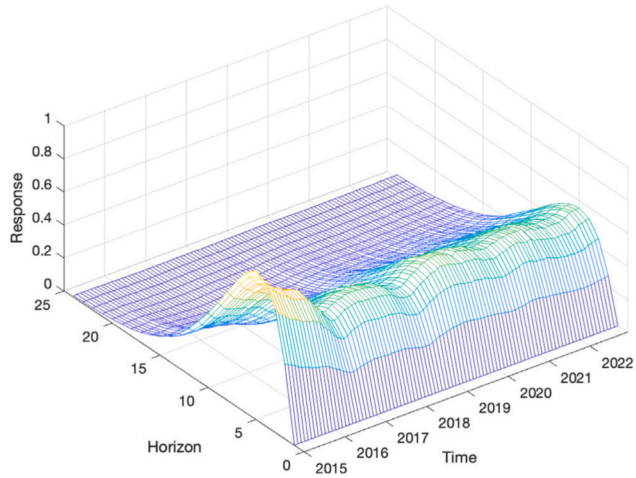

Fig. D.1. Sentiment shock to prices: Weighted sentiment.

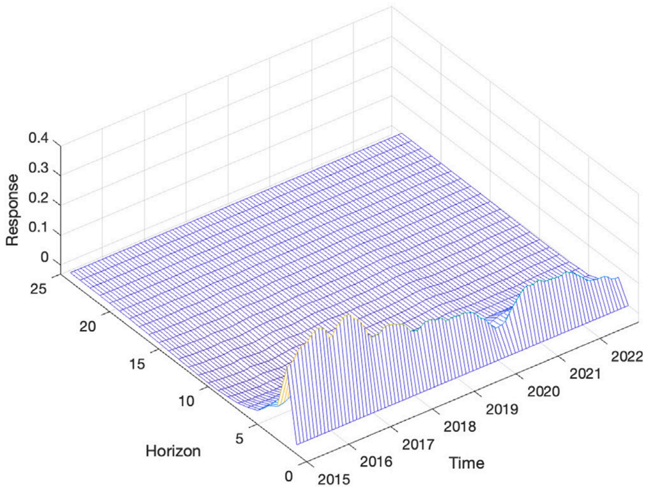

Fig. D.2. Sentiment shock to liquidity: Weighted sentiment.

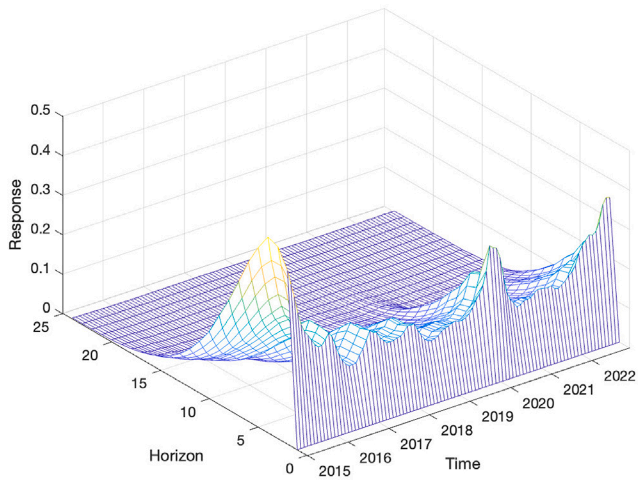

Fig. D.3. Liquidity shock to prices: Weighted sentiment.

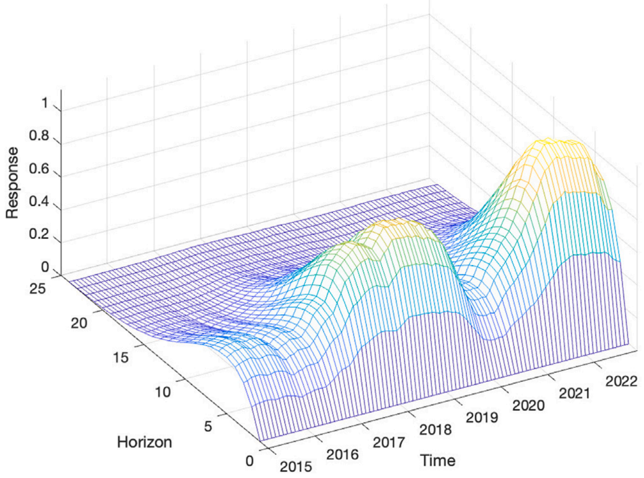

Fig. D.4. Sentiment shock to prices: Alternative housing prices.

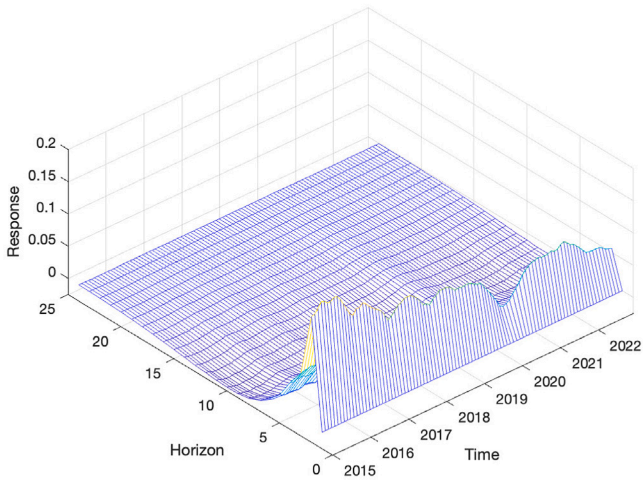

Fig. D.5. Sentiment shock to liquidity: Alternative housing prices.

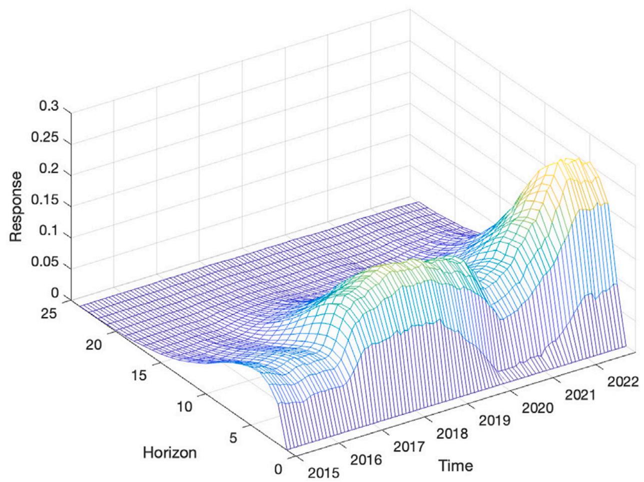

Fig. D.6. Liquidity shock to prices: Alternative housing prices.

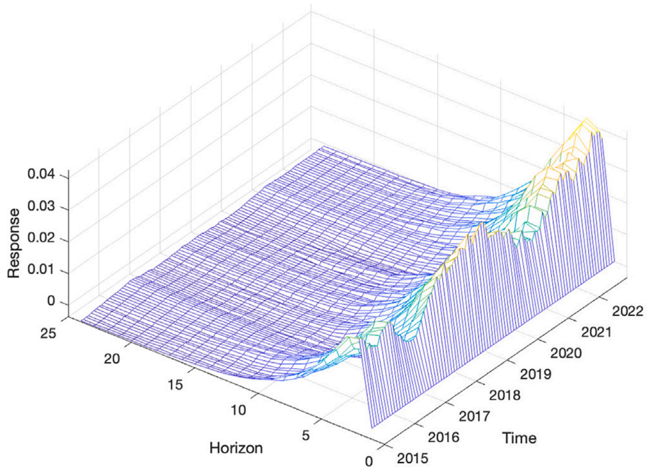

Fig. D.7. Sentiment shock to prices: Weighted average housing prices.

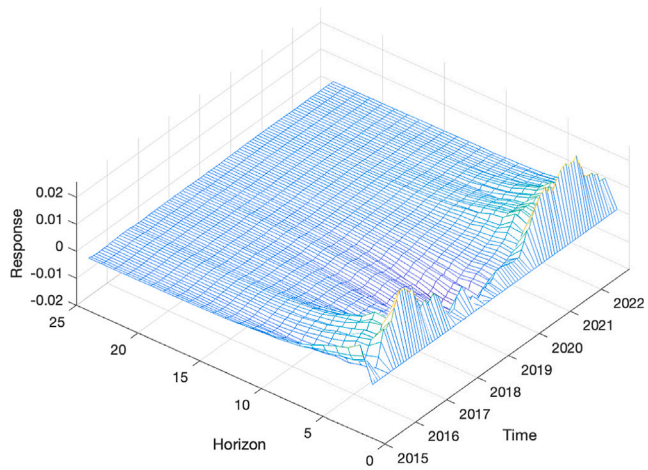

Fig. D.8. Sentiment shock to liquidity: Weighted average housing prices.

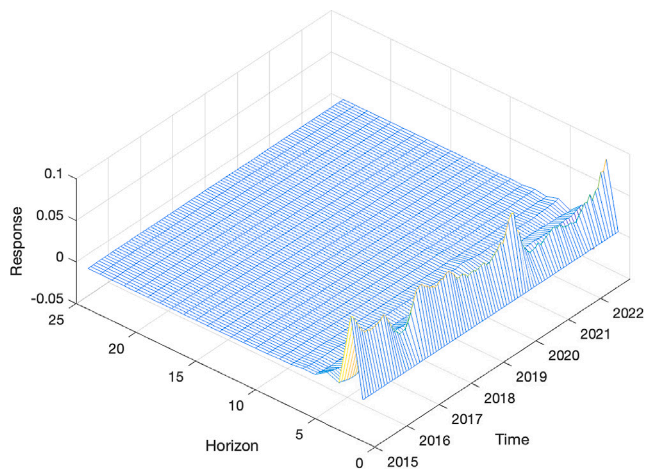

Fig. D.9. Liquidity shock to prices: Weighted average housing prices.

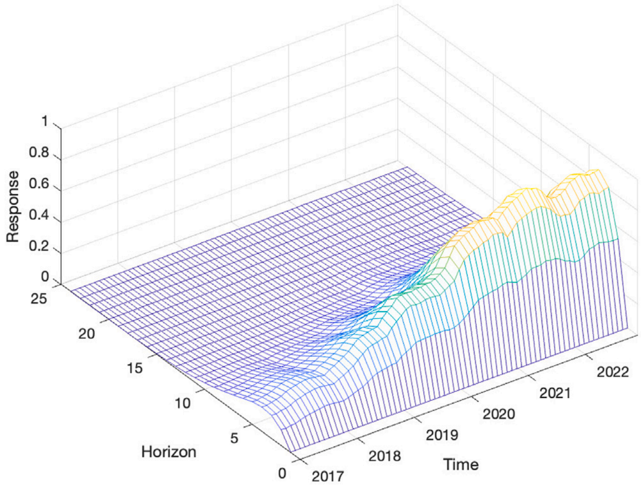

Fig. D.10. Sentiment shock to prices: Second-hand case.

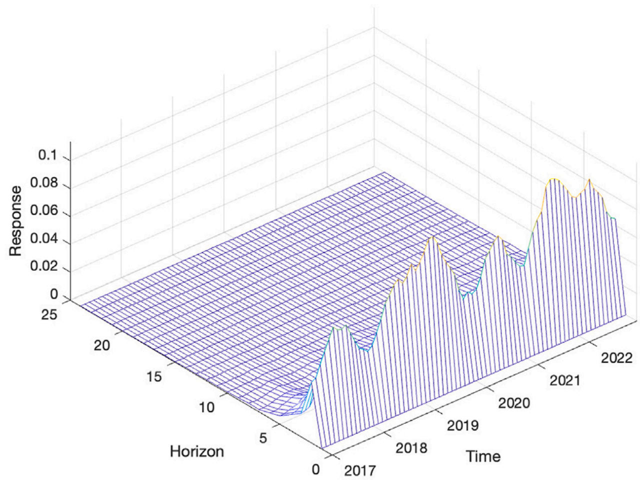

Fig. D.11. Sentiment shock to liquidity: Second-hand case.

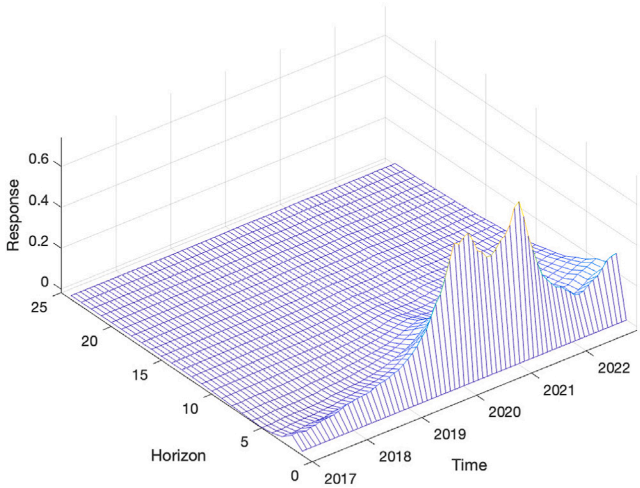

Fig. D.12. Liquidity shock to prices: Second-hand case.

References

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys*, *34*(3), 512–547.

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

Asriyan, V., Fuchs, W., & Green, B. (2019). Liquidity sentiments. *American Economic Review*, *109*(11), 3813–3848.

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

Bernanke, B. S. (2008). *Housing, mortgage markets, and foreclosures*. Remarks at The Federal Reserve System Conference on Housing and Mortgage Markets, Washington, D.C.

Bielecki, T. R., Björk, T., Jeanblanc, M., Rutkowski, M., Scheinkman, J. A., Xiong, W., et al. (2004). Heterogeneous beliefs, speculation and trading in financial markets. In *Paris-Princeton lectures on mathematical finance 2003*, (pp. 217–250). Springer.

Burnside, C., Eichenbaum, M., & Rebelo, S. (2016). Understanding booms and busts in housing markets. *Journal of Political Economy*, *124*(4), 1088–1147.

Caplin, A., & Leahy, J. (2011). Trading frictions and house price dynamics. *Journal of Money, Credit and Banking*, *43*, 283–303.

Caraiani, P., Gupta, R., Lau, C. K. M., & Marfatia, H. A. (2022). Effects of conventional and unconventional monetary policy shocks on housing prices in the United States: The role of sentiment. *Journal of Behavioral Finance*, *23*(3), 241–261.

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

Cesa-Bianchi, A. (2013). Housing cycles and macroeconomic fluctuations: A global perspective. *Journal of International Money and Finance*, *37*, 215–238.

Cesa-Bianchi, A., Cespedes, L. F., & Rebucci, A. (2015). Global liquidity, house prices, and the macroeconomy: Evidence from advanced and emerging economies. *Journal of Money, Credit and Banking*, *47*(S1), 301–335.

Chordia, T., Roll, R., & Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics*, *56*(1), 3–28.

Ciarlone, A. (2011). Housing wealth effect in emerging economies. *Emerging Markets Review*, *12*(4), 399–417.

Ciarlone, A. (2015). House price cycles in emerging economies. *Studies in Economics and Finance*, *32*(1), 17–52.

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*, 5–37.

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

Cox, J., & Ludvigson, S. C. (2021). Drivers of the great housing boom-bust: Credit conditions, beliefs, or both? *Real Estate Economics*, *49*(3), 843–875.

Cui, Y., Che, W., Liu, T., Qin, B., & Yang, Z. (2021). Pre-training with whole word masking for Chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 3504–3514.

De Wit, E. R., Englund, P., & Francke, M. K. (2013). Price and transaction volume in the Dutch housing market. *Regional Science and Urban Economic*, *43*(2), 220–241.

Del Negro, M., & Primiceri, G. E. (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *The Review of Economic Studies*, *82*(4), 1342–1345.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv preprint arXiv:1810.04805.

Dougal, C., Engelberg, J., Garcia, D., & Parsons, C. A. (2012). Journalists and the stock market. *The Review of Financial Studies*, *25*(3), 639–679.

Duca, J. V., Muellbauer, J., & Murphy, A. (2021). What drives house price cycles? International experience and policy issues. *Journal of Economic Literature*, *59*(3), 773–864.

Eraslan, S. (2016). Safe-haven demand for housing in London. *Economic Modelling*, *58*, 482–493.

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

Fisher, J., Gatzlaff, D., Geltner, D., & Haurin, D. (2003). Controlling for the impact of variable liquidity in commercial real estate price indices. *Real Estate Economics*, *31*(2), 269–303.

Fisher, J., Gatzlaff, D., Geltner, D., & Haurin, D. (2004). An analysis of the determinants of transaction frequency of institutional commercial real estate investment property. *Real Estate Economics*, *32*(2), 239–264.

Follain, J. R., & Velz, O. T. (1995). Incorporating the number of existing home sales into a structural model of the market for owner-occupied housing. *Journal of Housing Economics*, *4*(2), 93–117.

Gali, J., & Gambetti, L. (2015). The effects of monetary policy on stock market bubbles: Some evidence. *American Economic Journal: Macroeconomics*, *7*(1), 233–257.

Gambetti, L., Pappa, E., & Canova, F. (2008). The structural dynamics of US output and inflation: what explains the changes? *Journal of Money, Credit and Banking*, *40*(2–3), 369–388.

Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-dependent sentiment classification with BERT. *IEEE Access*, *7*, 154290–154299.

Geltner, D. (1990). Return risk and cash flow risk with long-term riskless leases in commercial real estate. *Real Estate Economics*, *18*(4), 377–402.

Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics*, *116*(4), 1233–1260.

Glaeser, E. L., & Nathanson, C. G. (2015). Housing bubbles. In *Handbook of regional and urban economics (Vol. 5)* (pp. 701–751). Elsevier.

Head, A., Lloyd-Ellis, H., & Sun, H. (2014). Search, liquidity, and the dynamics of house prices and construction. *American Economic Review*, *104*(4), 1172–1210.

Hong, Y., & Li, Y. (2020). Housing prices and investor sentiment dynamics: Evidence from China using a wavelet approach. *Finance Research Letters*, *35*, Article 101300.

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.

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review*, *95*(3), 739–764.

Iwanaga, Y., & Hirose, T. (2022). Liquidity shock and stock returns in the Japanese equity market. *Pacific-Basin Finance Journal*, *75*, Article 101849.

Iwanaga, Y., & Hirose, T. (2023). Liquidity changes and decomposition in the Japanese equity market. *Pacific-Basin Finance Journal*, *81*, Article 102115.

Jang, J. (2022). What drives stock market underreaction to liquidity shocks? Evidence from Korea. *Asia-Pacific Journal of Financial Studies*, *51*(1), 44–80.

Krainer, J., et al. (2008). Falling house prices and rising time on the market. *FRBSF Economic Letter*.

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.

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.

Luca, M. (2015). User-generated content and social media. In *Handbook of media economics (Vol. 1)* (pp. 563–592). Elsevier.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, *3*(3), 205–258.

Minetti, R., & Peng, T. (2013). Lending constraints, real estate prices and business cycles in emerging economies. *Journal of Economic Dynamics & Control*, *37*(12), 2397–2416.

Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review*, *95*(4), 1031–1053.

Novy-Marx, R. (2009). Hot and cold markets. *Real Estate Economics*, *37*(1), 1–22.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*(3), 821–852.

Quan, D. C., & Quigley, J. M. (1989). Inferring an investment return series for real estate from observations on sales. *Real Estate Economics*, *17*(2), 218–230.
Shao, J., Hong, J., Wang, X., & Yan, X. (2023). The relationship between social media sentiment and house prices in China: Evidence from text mining and wavelet analysis. *Finance Research Letters*, *57*, Article 104212.
Shen, S., Zhao, Y., & Pang, J. (2024). Local housing market sentiments and returns: Evidence from China. *The Journal of Real Estate Finance and Economics*, *68*(3), 488–522.
Shiller, R. C. (2000). Irrational exuberance. *Philosophy and Public Policy Quarterly*, *20*(1), 18–23.
Soo, C. K. (2018). Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies*, *31*(10), 3689–3719.
Sousa, M. G., Sakiyama, K., de Souza Rodrigues, L., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019). BERT for stock market sentiment analysis. In *2019 IEEE 31st international conference on tools with artificial intelligence* (pp. 1597–1601). IEEE.
Sufi, A. (2023). *Housing, household debt, and the business cycle: An application to China and Korea*: *Technical report*, National Bureau of Economic Research.
Van Dijk, D. W., & Francke, M. K. (2018). Internet search behavior, liquidity and prices in the housing market. *Real Estate Economics*, *46*(2), 368–403.
Wang, Z. (2017). Comparative analysis of real estate bubble in China and Japan. In *3rd international conference on economics, management, law and education* (pp. 195–199). Atlantis Press.
Wu, Y., & Tidwell, A. (2015). Inflation-hedging properties of regional Chinese real estate market: Evidence from 35 cities in China. *Applied Economics*, *47*(60), 6580–6598.
Xie, S., & Chen, J. (2018). Beyond homeownership: Housing conditions, housing support and rural migrant urban settlement intentions in China. *Cities*, *78*, 76–86.
Zhou, G. (2018). Measuring investor sentiment. *Annual Review of Financial Economics*, *10*, 239–259.