INTERNATIONAL MONETARY FUND

# Chinese Housing Market Sentiment Index: A Generative AI Approach and An Application to Monetary Policy Transmission

Kaiji Chen (Emory University) and Yunhui Zhao (IMF)

**WP/24/264**



**2024**
DEC

WORKING PAPER

 WP/24/264

**IMF Working Paper**

Institute for Capacity Development

**Chinese Housing Market Sentiment Index:
A Generative AI Approach and An Application to Monetary Policy Transmission**

**Prepared by Kaiji Chen (Emory University) and Yunhui Zhao (IMF)[1]**

Authorized for distribution by Natan Epstein

December 2024



**ABSTRACT:** We construct a daily Chinese Housing Market Sentiment Index by applying GPT-4o to Chinese news articles. Our method outperforms traditional models in several validation tests, including a test based on a suite of machine learning models. Applying this index to household-level data, we find that after monetary *easing*, an important group of homebuyers (who have a college degree and are aged between 30 and 50) in cities with more optimistic housing sentiment have *lower* responses in non-housing consumption, whereas for homebuyers in other age-education groups, such a pattern does not exist. This suggests that current monetary easing might be more effective in boosting non-housing consumption than in the past for China due to weaker crowding-out effects from pessimistic housing sentiment. The paper also highlights the need for complementary structural reforms to enhance monetary policy transmission in China, a lesson relevant for other similar countries. Methodologically, it offers a tool for monitoring housing sentiment and lays out some principles for applying generative AI models, adaptable to other studies globally.

**RECOMMENDED CITATION:** Chen, Kaiji and Yunhui Zhao. 2024. "Chinese Housing Market Sentiment Index: A Generative AI Approach and An Application to Monetary Policy Transmission," IMF Working Paper No. 24/264, International Monetary Fund, Washington, D.C.

| | |
|---|---|
| JEL Classification Numbers: | C10, C53, E31, E37, R30 |
| Keywords: | Chinese Housing Market Sentiment; Generative AI; Monetary Policy Transmission; Consumption; Crowding-Out |
| Author's E-Mail Address: | kaiji.Chen@emory.edu; YZhao@imf.org |


[1] We thank Jin Zhang (Emory University) for her excellent research assistance. We are also grateful for the helpful discussions with Mehdi Benatiya Andaloussi (all IMF, unless specified otherwise), Tohid Atashbar, Nina Biljanovska, Alessia De Stefani, Chris Evans, Andras Komaromi, Nan Li, Li Liu, Rodolfo Maino, Rui Mano, Jorge Leon Murillo, Lev Ratnovski, Chris Redl, Yannick Timmer (Federal Reserve Board), Stijn Van Nieuwerburgh (Columbia University), Yuanchen Yang, Jiaxiong Yao, participants in the Research Ideas Workshop at the IMF in June 2023 and the FUNDALL seminar at the IMF in October 2024, and especially Junru Guo and Jia He (both from Nankai University).

**Contents**

I. Introduction ... 4

II. Construction of Chinese Housing Market Sentiment Index ... 10

A. Data and Methodology ... 10

B. Principles of Prompt Engineering ... 12

III. Validity Tests and Comparison with Other Models ... 13

A. Comparison with Human Assessments ... 13

B. Sentiment Around Announcements of "Major Housing Policies" ... 22

C. Predicting Housing Prices ... 30

D. A Real-Time Presentation of CHMSI ... 37

IV. An Application to Study China's Monetary Policy Transmission ... 40

A. Research Design ... 40

B. Data and Institutional Background ... 42

C. Empirical Specifications ... 43

D. Empirical Results: Optimistic Regime ... 45

E. Empirical Results: Pessimistic Regime ... 48

F. Empirical Results: Accounting for Potential Endogeneity ... 50

G. Robustness Checks ... 52

H. Discussions ... 53

V. Conclusion and Policy Implications ... 55

References ... 57

Appendices ... 61

# I. INTRODUCTION

Monitoring housing market sentiment is crucial for households, financial institutions, and policymakers in any country with a large housing sector. This is especially the case at the current conjuncture for China, which is suffering from the housing distress following the 2021 Evergrande liquidity crisis. In response, the Chinese authorities have recently eased monetary policy stance to boost domestic consumption and stimulate the economy. For example, the one-year Loan Prime Rate (China's benchmark rate) was cut by 10 basis points in July 2024 and by 20 basis points in September 2024, and the People's Bank of China indicated that it would "increase the intensity of counter-cyclical monetary policy" in November 2024. These raise critical questions: How can we accurately measure housing market sentiment in China? And how does this sentiment affect monetary policy transmission and its impact on consumption in China?

News articles provide a natural opportunity for gauging market sentiment, but traditional methods face challenges in properly capturing the nuanced sentiments expressed in these articles. Artificial intelligence (AI), particularly generative AI-powered Large Language Models (LLMs) such as GPT-4o, holds great potential in addressing these challenges due to its advanced capabilities in processing and understanding complex text data. Unlike keyword-based models, which often miss context and nuance, GPT-4o utilizes deep learning to understand the context and subtleties in language. And compared to some other LLMs such as BERT, which require extensive labeled data for training, GPT-4o can perform zero-shot and few-shot learning,[2] making it more flexible for analyzing diverse and dynamic text data.

In this paper, we apply OpenAI's GPT-4o model to construct a Chinese housing market sentiment index (CHMSI) using news article data. We present multiple versions of the index, each created with different levels of "prompts"—instructions provided to the generative AI models. We summarize eight effective principles for "prompt engineering," a process involving designing and refining input instructions to guide AI models. We find that the housing sentiment index constructed by GPT-4o with "advanced" prompts outperforms those constructed by the traditional keywords-based model and many Chinese LLMs according to all three criteria: First, consistency with

[2] Zero-shot prompting refers to the ability of a model to perform a task without having seen any examples of the task during training, relying instead on the contextual understanding of the prompt alone. Few-shot prompting involves providing the model with a small number of examples (usually one to five) to help it understand the task better before performing it on new data.

human assessments. Using selected testing articles, the housing market sentiment scores generated from GPT-4o model with advanced prompts exhibits the lowest sum of square errors relative to human assessment. Second, the ability to capture market sentiment around major housing policy announcements. Compared with their counterparts generated from other LLMs, the housing sentiment index constructed by GPT-4o exhibits less noise and more conformity with announcements in major housing policies. Third, the ability to predict national housing price trends with machine learning models. We apply different sentiment indices to machine-learning models, commonly employed in macroeconomic forecasting, to predict the year-on-year growth rate of house prices and perform a "horse race" of the models to determine the best performing forecast models. Our optimal forecasting model (the random forest model) using GPT-4o constructed sentiment index achieves a higher forecasting performance than those using the traditional keywords-based approach or the Chinese LLMs. After establishing the validity of the sentiment index constructed by GPT-4o according to the above-mentioned three criteria, we extend our daily sentiment index data to the most recent date where data is available (September 11, 2024). Doing so enables us to trace the housing sentiment during the ongoing housing slump in China as well as following recent major polices. Our up-to-date Chinese housing market sentiment index can well trace the housing sentiment around important changes in housing regulatory policies (such as the "three-red-line" policy) and during housing market distresses (such as those associated with the financial difficulties of Evergrande and Country Garden). And it sheds light on the impact of Chinese government housing stimulus policies in household confidence in housing markets.

We then apply our CHMSI to evaluate the role of the housing sector and the effectiveness of China's monetary easing in stimulating domestic consumption. We employ a household-level dataset between 2013Q3 and 2015Q4, which includes transaction-level credit card expenditures and transaction-level mortgage originations. Our monetary policy shocks are from Chen, Ren, and Zha (2018), constructed using an endogenous switching M2-based monetary policy rule.

Our empirical strategy exploits variations across Chinese cities in exposures to the national-level housing market sentiment. We measure such exposure as the city-level Baidu search index, which serves as a proxy for the attention of a city's potential homebuyers and for other shocks to the city's local housing markets. Specifically, by incorporating Baidu search index normalized by city-level population, we refine the CHMSI to create an "Attention-Adjusted Chinese Housing Market Sentiment Index" (AACHMSI) at the city-level. We classify cities according to AACHMSI across all cities and quarters and compare the responses of non-housing consumption to monetary shocks between households in cities belonging to the top 20th percentile in terms of AACHMSI (i.e., being in an "optimistic regime") and those in cities belonging to the bottom 20th percentile (i.e., being in a "pessimistic regime"). The difference between the average responses of non-housing consumption to monetary policy shocks for households in these two groups of cities provides a creditable estimate of the effects of housing market sentiments on monetary policy transmission into non-housing consumption.

We use a local projection approach to estimate the interactive effects of housing market sentiments on the monetary policy transmission into non-housing consumption. Our local projection results suggest that following monetary easing, the non-housing consumption of homebuyers in cities under *optimistic* housing sentiment increases by *less*, particularly for households with a college degree and aged between 30 and 50. For other age-education groups, however, such a pattern does not exist. Moreover, we find that house prices in cities with more optimistic sentiment increase more in response to monetary policy easing.

Our findings suggest a crowding-out channel for monetary easing on non-housing consumption of existing homeowners, that is: following monetary easing, house prices become higher, encouraging more existing homeowners to trade up for larger houses to reap future capital gains, a mechanism that is particularly relevant when the housing sentiment is optimistic. Since higher house prices increase the downpayments, existing homeowners will reduce their current non-housing consumption. Moreover, higher house prices increase the principals of mortgage loans (hence future principal payments); hence, if the higher principal payments more than offset the lower interest payments, existing homeowners will reduce their future non-housing consumption as well. We interpret our findings by developing a simple analytical model that features the "trade-up" of houses by a group of existing homeowners (i.e., upgrading to larger houses) who have a college degree and are aged between 30 and 50. This group is quantitatively relevant for the entire housing market because, as shown in Chen et al. (2023a), an increase in the number of households who trade up their existing homes contributes to 57.9 percent of the increase in the origination amount of total mortgages, and 61.6 percent of the increase in housing demand.

These findings suggest that monetary easing might be more effective in boosting household consumption at the current conjuncture than in the past for China, as the crowding-out channel may be weaker under the prevailing pessimistic housing sentiment. Moreover, for monetary policy to be effective in boosting household consumption at the current conjuncture, it is important for other policies, such as housing policies and structural reforms, to coordinate with monetary easing to contain housing speculation.

Our paper contributes to four strands of literature. First, it contributes to the emerging literature on applying machine learning and LLMs to economic and financial analyses. On the benefits of applying machine learning, Medeiros et al. (2021) show that in a data-rich environment, the gains of using machine-learning models can be as large as 30 percent in terms of mean squared errors, and that such models can help uncover the main drivers for future inflation.[3] These benefits are confirmed by our paper, where the machine learning-based validation tests also turn out to outperform traditional methods. On the costs of applying machine learning (including broad economic costs), Fuster et al. (2022) find that Black and Hispanic mortgage borrowers in the U.S. are disproportionately less likely to benefit from the introduction of machine learning into default prediction models.[4] As for the benefits and costs of applying *LLMs*, Bartik, Gupta, and Milo (2024) use several LLMs (such as Chat GPT-4 and Claude 3 Opus) to construct a detailed assessment of U.S. zoning regulations; they find that this approach achieves 96 percent accuracy for binary regulatory questions. In addition, Korinek (2023) presents use cases in six areas (such as background research) and argues that economists can gain significantly by taking advantage of LLMs to automate micro-tasks, while also noting the risks of "hallucination" (producing outputs that are inaccurate) and privacy violation. And Chang et al. (2024) provide a comprehensive survey of prompt engineering.

To the best of our knowledge, this paper is one of the first to apply LLMs in household finance and especially in studies on housing and monetary policy transmission. In terms of research questions, rather than using LLMs to predict asset prices or returns (as in most literature), our goal is to use LLMs to analyze unstructured texts from news media to construct housing sentiments, a crucial factor governing households' expectation of house price growth. Although there is growing consensus on the importance of housing market speculation or optimism in the U.S. housing booms and busts of 2000 to 2010, there is limited individual-level data on house price beliefs during the housing boom from 2002 to 2006 that can be linked to measures of speculation at the individual level

[3] Similar insights are documented in Medeiros et al. (2024) considering common factors driving global inflation, Joseph et al. (2024) in the context of U.K. inflation forecasting, as well as the literature that applies Recurrent Neural Networks to inflation forecasting, such as Barkan et al. (2023) and Paranhos (2023).

[4] See also Sadhwani, Giesecke, and Sirignano (2021) and Barbaglia, Manzan, and Tosetti (2023) for the application of machine learning algorithms in predicting mortgage loan defaults.

(Mian and Sufi, 2022).[5] Hence, our approach of using GPT-4o to construct disaggregated housing market sentiments with news article data can be applied to other countries, such as the U.S., thus providing a useful complement to the literature.

Second, our paper contributes to the extensive literature on the effects of monetary policy on household consumption using granular data.[6] For example, Agarwal et al. (2022) show that transmission from mortgage rates to households' non-housing consumption in China is present only for homeowners with unpaid mortgage balances and not for other types of households.[7] In addition, Gerardi, Willen, and Zhang (2023) empirically examine the interactions of mortgage prepayment, race, and monetary policy and find that because White borrowers are more likely to exploit lower interest rates through refinancing, they benefit more from monetary expansions. Many of these studies (with some notable exceptions, such as Gerardi, Willen, and Zhang, 2023) use repeated cross-sectional data (e.g., Cloyne, Ferreira, and Surico, 2020) or impute consumption using administrative data (e.g., Holm, Paul, and Tischbirek, 2021). As noted by Baker et al. (2022) and Cao et al. (2024), since imputed consumption is

[5] After the Global Financial Crisis (GFC), there are household survey data available to measure household beliefs on the housing market. However, no such data is available before then, and our method can be applied to construct similar housing market sentiment before the GFC using, for example, news article data in the U.S.

[6] See, among others, Agarwal, Liu, and Souleles (2007), Agarwal and Qian (2014), Di Maggio et al. (2017), Chen et al. (2023a), Agarwal et al. (2022), and Beraldi and Zhao (2023).

[7] Another view emphasizes the wealth channel of monetary policy transmission. For example, Painter, Yang, and Zhong (2022) empirically show that the marginal propensity to consume out of current wealth in China is low; moreover, households are prohibited from withdrawing housing equity in China, further limiting the wealth channel. Relatedly, although a large number of studies (such as Jarocinski and Smets (2008), Jordà, Schularick, and Taylor (2015), and Williams (2016)) find that monetary policy shocks have significant effects on house prices (and presumably on consumption), some studies show that this is not always the case. For example, Favilukis, Ludvigson, and Van Nieuwerburgh (2017) find that a large foreign capital inflow to an economy would endogenously increase the housing risk premium, substantially offsetting the effect of lower interest rates on house prices. In addition, Aastveit and Anundsen (2022) find that expansionary monetary policy shocks have a larger impact on house prices in supply-inelastic areas, and Albuquerque, Iseringhausen, and Opitz (2024) find that house prices and consumption respond more in supply-inelastic U.S. states. More broadly, regarding the impact of higher house prices on corporate borrowing, the traditional "collateral channel" (i.e., real estate values boost corporate secured borrowing) has been at odds with recent studies. For example, Campello et al. (2022) show that firms infrequently use real estate assets as collateral to secure their borrowing, and that the high systematic risk exposure of collateral assets limits firms' ability to raise secured debt, even when collateral values rise.

constructed as income not saved, it is not possible to identify what type of goods or services consumers acquire. Our transaction-level credit card data, nonetheless, can distinguish between different types of consumption expenditures, and thus we construct non-housing consumption by excluding various housing-related expenses. With this dataset connecting non-housing consumption and mortgage origination at individual levels, we identify a novel crowding-out channel for monetary policy easing to affect non-housing consumption.

Third, our paper contributes to the literature on sentiment analysis in the housing market. Soo (2018) develops the first measures of housing sentiment for 34 cities across the U.S. and finds that housing media sentiment has significant predictive power for future house prices. Our paper is consistent with this finding, and we extend the author's dictionary-based method by more advanced LLM-based methods. Zhou (2018) examines the interaction between housing market sentiment and government intervention in China. The author finds that positive sentiment leads to higher house prices and increased market activity, whereas negative sentiment has the opposite effect. Dong, Hui, and Yi (2021) explore the relationship between housing market sentiment and homeownership in China, concluding that optimistic sentiment boosts homeownership rates by enhancing potential buyers' confidence in future house prices. Cepni and Khorunzhina (2023) construct state-level housing-sentiment indices in the U.S., demonstrating that sentiment significantly influences housing market performance across different states. Furthermore, Cepni (2024) delves into the effects of news media coverage on housing market sentiment across various U.S. states, underscoring the importance of media in molding public perception and market results. Alfano and Guarino (2022) analyze the impact of textual sentiment in news articles on U.K. housing prices, showing that positive news sentiment drives up prices while negative sentiment causes declines. Our paper builds on these studies by using generative AI models to construct a refined and dynamic sentiment index for the Chinese housing market. This index, validated against human assessments, major housing policies, and housing price trends, serves as a robust tool for examining the impact of sentiment on monetary policy transmission and housing consumption.

Finally, our paper contributes methodologically to the literature on how to effectively use generative AI tools. Chen et al. (2023b) lay out foundational principles of prompt engineering and discuss advanced methodologies like chain-of-thought prompting.[8]

[8] Wei et al. (2022) and Zhang et al. (2022) also discuss the Chain-of-Thought prompting and argue that this strategy enhances the performance of LLMs by producing some intermediate reasoning steps.

Barbaglia et al. (2023) provide a systematic study of different guiding principles and offer prompting advice for prompting leading LLMs.[9] Our paper extends these principles by incorporating additional effective strategies, such as requesting the code for replication (thereby enhancing transparency). In addition, Sahoo et al. (2024) provide a systematic survey of prompt engineering in LLMs, discussing the advantages and disadvantages of few-shot versus zero-shot prompting. They argue that although few-shot prompting can improve performance for complex tasks, the selection and composition of prompt examples can significantly influence model behavior. Building on these insights, our paper primarily focuses on zero-shot prompting for sentiment analysis. Our contribution to this literature lies in showcasing the use of generative AI in a novel context and providing rigorous validation tests for the generative AI-constructed indices, including tests that use a suite of machine learning models.

The rest of the paper is organized as follows. Section II constructs Chinese Housing Market Sentiment Index (CHMSI). Section III conducts validity tests of the CHMSI we constructed and compares it with indices constructed by other models. Section IV applies our sentiment index to estimate the role of housing market sentiments on the monetary policy transmission into consumption in China. Section V concludes. The appendices collect some technical details.

# II. CONSTRUCTION OF CHINESE HOUSING MARKET SENTIMENT INDEX

## A. Data and Methodology

We utilize daily news article data from the China Stock Market and Accounting Research (CSMAR) news database to construct the Chinese housing market sentiment index.[10] The dataset starts from January 4, 2012 to September 11, 2024, providing a comprehensive up-to-date view of the housing market sentiment from the news media. The CSMAR database collects news articles related to economics and finance published by major Chinese newspapers, which ensures the content is professional and less biased compared to social media data often used in other studies. This high-quality data source enhances the reliability of our sentiment analysis. First, it provides more professional and curated content, reducing the noise often associated with social media. Second, the CSMAR dataset is updated daily, allowing us to construct a highly timely sentiment

[9] Relatedly, Korinek (2024) illustrates the automated prompt generator offered by Anthropic PBC, a U.S.-based AI startup.

[10] The CSMAR database could be accessed via Wharton WRDS database.

index. This timeliness is crucial for market participants, researchers, and policymakers who require up-to-date information to make informed decisions. Although we use data from 2012 to 2016 for illustration purposes and to manage the financial costs of running the algorithm, the same methodology can be applied to the most recent data to maintain the index's relevance.

Our methodology involves leveraging the sentiment analysis function in GPT-4o, accessed through the OpenAI API. GPT-4o's advanced capabilities in understanding and processing complex text data make it an ideal tool for this task. It offers significant advantages over other sentiment analysis tools, such as keyword-based models and BERT. For the keyword-based approach, we use the dictionary for Chinese finance-related sentiment as in Jiang et al. (2019), which is based on Loughran and McDonald (2011)'s finance/accounting dictionaries. While keyword-based models often miss context and nuance, GPT-4o uses deep learning to capture the subtleties in language. Additionally, GPT-4o supports zero-shot and few-shot learning, making it more flexible and powerful without requiring extensive labeled data.

Specifically, we follow these steps to construct the sentiment index:

- Data retrieving and preprocessing. We retrieve daily news articles from the CSMAR database. The collected articles are preprocessed to clean and standardize the text. This step includes removing irrelevant content, normalizing text (e.g., converting to lowercase, removing punctuation), and tokenizing the text for analysis.
- Sentiment analysis with GPT-4o: Using GPT-4o, we analyze the sentiment of each news article. We design prompts that instruct the AI to accurately assess the sentiment. GPT-4o generates sentiment scores based on the context and nuances captured in the articles, ensuring a sophisticated understanding of market sentiment.
- Prompt engineering. We apply eight effective principles of prompt engineering to enhance the performance of the sentiment analysis, as explained subsequently. Section B lays out the details for our principles of prompt engineering. We also set the "temperature parameter" to 0, which is a frequently used method in the machine learning literature to mitigate the replicability concern for LLMs.
- Index construction. We aggregate the daily sentiment scores to construct a comprehensive sentiment index. This index reflects the overall sentiment of the housing market as depicted in the news articles.
- Validation. We validate the constructed sentiment index using three criteria: consistency with human assessments (compare AI-generated sentiment scores with those given by human evaluators), ability to capture market sentiment around major housing policy announcements, and consistency with national housing price evolution.

By following these steps, we ensure that the sentiment index is robust, timely, and reflective of the true market sentiment. The use of GPT-4o for sentiment analysis, combined with effective prompt engineering and rigorous validation, provides a reliable tool for monitoring and analyzing the Chinese housing market.

### B. Principles of Prompt Engineering

Given the crucial role of prompt engineering, we lay out eight principals based on the literature and our own experiments. Specifically:

- *Be specific.* Provide clear and detailed instructions in your prompts. Avoid vague or ambiguous language that may lead to misunderstandings. The more precise your prompt, the more accurate and relevant the AI's response will be. For instance, rather than asking, "What do you think about the housing market?", specify, "Analyze the trends in the Chinese housing market over the past five years and predict future developments based on current data."
- *Incorporate role-playing*. Direct the AI to adopt a specific role or perspective to enhance the relevance of its responses. For example, instruct the AI to "Act as an expert in Chinese housing, macroeconomics, and finance, and provide an assessment of the housing market sentiment in China." This approach helps the AI tailor its response to the context and deliver more insightful analysis.
- *Request replicable code*. When dealing with technical or analytical tasks, request the AI to provide the code or methodology used in its analysis. This ensures transparency and allows for the replication or verification of results. For example, prompt, "Provide the Python code used to calculate the housing market sentiment index based on the given dataset."
- *Iterate and refine.* Employ an iterative approach to refine prompts and improve the quality of the AI's output. Start with a basic prompt and progressively refine it based on the AI's responses. Iteration helps in honing the instructions to achieve the desired level of detail and accuracy. For instance, begin with "Analyze the impact of monetary policy on housing prices" and refine it to "Analyze the impact of recent monetary policy changes on housing prices in major Chinese cities over the past year."
- *Encourage critical evaluation.* Encourage the AI to critically evaluate its responses and consider alternative perspectives. This can be done by including prompts that ask the AI to verify its conclusions or explore different scenarios. For example, prompt, "Assess the impact of rising interest rates on the housing market, and consider if there could be any countervailing factors that might mitigate these effects." In addition, a simple follow-up prompt like "Are you sure?" can often be highly effective.
- *Do not rush.* Explicitly instruct the AI "Don't rush" to allow for sufficient time to process and generate responses. Rushed prompts or expecting immediate

answers can lead to suboptimal results. Ensure that the AI has adequate time and context to understand and respond accurately. This might involve setting realistic expectations for response times and avoiding overly urgent prompts.

- *Apply chain-of-thought*. LLMs are capable of performing detailed reasoning processes by creating steps in between, known as chain-of-thought (CoT) prompting. There are primarily two approaches within CoT prompting. The first one encourages sequential thinking with a simple prompt such as "Let's think step by step." The second one involves crafting detailed examples that pair questions with their respective reasoning sequences to arrive at an answer. The effectiveness of this latter method primarily relies on meticulously creating specific demonstrations. These labor-intensive tasks can be reduced by sequentially soliciting demonstrations with a new prompt "let's think not just step by step, but also one by one," as shown by Zhang et al. (2022).
- *Tip and penalize*. Experience suggests that interestingly, strategies such as artificially offering LLMs a tip if they do well or threatening a penalty if not can improve performance, as also documented in Bsharat, Myrzakhan, and Shen (2023).

By adhering to these principles, researchers can enhance the effectiveness of AI-generated responses, thereby achieving more accurate and insightful outcomes in their analyses, as illustrated in the next section. The prompts we use for the rest of the paper also exemplify the implementation of these principles, and the details are provided in Appendix 1 and Appendix 2.

# III. VALIDITY TESTS AND COMPARISON WITH OTHER MODELS

This section conducts validity tests of the CHMSI we constructed and compares it with indices constructed by other models. We will do this according to three criteria: human assessments of a few testing sentences, sentiment around announcements of "major housing policies," and forecasting power for house price dynamics based on a suite of machine learning models.

## A. Comparison with Human Assessments

Consistent with the literature, comparing with human assessments (or "subjective evaluation") is an essential and intuitive way to gauge the validity of the outputs by LLMs. We (i.e., the authors) first conduct the evaluation based on five sentences we select from the actual news articles in our database: one "benchmark" that clearly has a neutral sentiment based on our manual evaluation, two "positive" sentences that have a clear optimistic tone, and two "negative" sentences that have a clear pessimistic tone. We then conduct the evaluation based on five *articles* selected in a similar way.

Note that although we provide the English versions of these sentences (or titles of news articles) below, we intentionally instruct the LLMs to detect the sentiment directly from the original Chinese text rather than translating first and then detecting the sentiment. Doing so can usually better capture the nuances and subtleties in the original language, as demonstrated by Zhang et al. (2022), among others.

### Testing Results Based on Five *Sentences*

The five sentences we select are (in English):[11]

- **Benchmark Sentence:** The earliest development of Luhuitou Peninsula can be traced back to approximately 20 years ago.
- **Positive Sentence 1:** There are concerns in public opinion that continuous relaxation policies will undermine the achievements of many years of real estate market regulation; such concerns are truly unnecessary.
- **Positive Sentence 2:** Multiple institutions predict that regulatory authorities are likely to continue relaxing real estate policies and introduce stimulus measures to curb the contraction speed of the real estate industry, in order to alleviate the sluggish real estate market.
- **Negative Sentence 1:** Especially in hotspot cities where housing prices have risen too quickly recently, it is not ruled out that more stringent specific regulatory policies will be introduced soon.
- **Negative Sentence 2:** Although overall, the content of this regulation does not exceed market expectations, judging by the operational strength and the central government's determination, the regulation may once again exert long-term pressure on the real estate market and the stock market.

[11] To facilitate replications of our results, we also provide these five sentences in the original language (Chinese): (1) **Benchmark Sentence:** 鹿回头半岛最早的开发可以追溯到大约 20 年前. (2) **Positive Sentence 1:** 有舆论担心，持续的松绑政策会使多年来的楼市调控成果毁于一旦；这种担心实在是不必要. (3) **Positive Sentence 2:** 多家机构预计，监管层很有可能继续放松房地产政策，出台刺激措施遏制房地产业的收缩速度，以缓解低迷的房地产市场. (4) **Negative Sentence 1:** 特别是近期房价上涨过快的热点城市，不排除会尽快出台更加严厉的具体调控政策. (5) **Negative Sentence 2:** 虽然整体而言，此次调控在措施内容上并未超出市场预期，但是从操作力度和中央的决心看，调控有可能再次对地产市场走向和股市形成长期压制.

The validity test and comparison results are presented in Table 1, where the sentiment score ranges from 0 (most pessimistic) to 100 (most optimistic). The keyword-based approach significantly deviates from capturing the tone of the broadly neutral benchmark sentence and misinterprets the overall sentiment of both positive and negative sentences. The performance of the Chinese LLMs is mixed: Two such models (Senta and Qianfan) miss the tones of most of the five sentences by a large margin, which could be due to the bias associated with our extremely small sample of five sentences; but the other four (Kimi, Xinghuo, Baihuan, and Hunyuan) produce accurate assessments of the sentiment. As for GPT-4o, all three versions of its sentiment scores (each corresponding to a different level of the prompt) also accurately capture the tone of the benchmark sentence and the sentiment directions of all other sentences.

**Table 1. Validity Test Against Human Assessments for Selected Sentences**

| | Benchmark | Positive 1 | Positive 2 | Negative 1 | Negative 2 |
|---|---|---|---|---|---|
| Human Assessment | 50 | 70 | 80 | 40 | 30 |
| Keywords-based | 90 | 50 | 36 | 58 | 50 |
| Senta Model | 100 | 1 | 94 | 29 | 38 |
| Qianfan model | 99 | 3 | 89 | 73 | 49 |
| Hunyuan Model | 50 | 60 | 60 | 25 | 35 |
| BaiChuan Model | 65 | 75 | 80 | 40 | 30 |
| Xinghuo Model | 50 | 65 | 70 | 30 | 35 |
| Kimi Assistant | 50 | 65 | 70 | 40 | 25 |
| GPT-4o (basic prompt) | 50 | 40 | 60 | 30 | 30 |
| GPT-4o (intermediate prompt) | 60 | 65 | 70 | 35 | 35 |
| GPT-4o (advanced prompt) | 60 | 70 | 70 | 35 | 35 |

Notes: This table reports the sentiment scores, ranging from 0 (most pessimistic) to 100 (most optimistic), generated by human assessments and various LLMs for each of the five selected sentences from news articles. “Benchmark” refers to the sentence that has a neutral sentiment based on human assessment. “Positive” refers to the sentences with a clear optimistic tone. “Negative” refers to the sentences with a clear pessimistic tone. The sources of this table, and all subsequent tables and figures in this paper, are from the authors.

To more precisely assess the performance of each method and to illustrate the impact of effective prompt engineering, we normalize the sentiment score by the score assigned to the benchmark sentence. This normalization mitigates concerns regarding differing scaling standards across methods in sentiment score assignment. Subsequently, for each method, we calculate the sum of squared errors (SSE) of its sentiment scores relative to human assessments. The larger the deviation of the overall score assignment from the human assessment of the corresponding sentence (after accounting for scaling differences), the poorer the performance of that sentiment analysis method. The results are presented in Table 2.

**Table 2. Validity Test Against Human Assessments for Selected Sentences: Ratio of Sentiment Scores over the Benchmark**

| | Benchmark | Positive 1 | Positive 2 | Negative 1 | Negative 2 | SSE |
|---|---|---|---|---|---|---|
| Human Assessment | 1.00 | 1.40 | 1.60 | 0.80 | 0.60 | 0.00 |
| Keywords-based | 1.00 | 0.56 | 0.40 | 0.64 | 0.56 | 2.18 |
| Senta Model | 1.00 | 0.01 | 0.94 | 0.29 | 0.38 | 2.67 |
| Qianfan model | 1.00 | 0.03 | 0.90 | 0.74 | 0.50 | 2.37 |
| Hunyuan Model | 1.00 | 1.20 | 1.20 | 0.50 | 0.70 | 0.30 |
| BaiChuan Model | 1.00 | 1.15 | 1.23 | 0.62 | 0.46 | 0.25 |
| Xinghuo Model | 1.00 | 1.30 | 1.40 | 0.60 | 0.70 | 0.10 |
| Kimi Assistant | 1.00 | 1.30 | 1.40 | 0.80 | 0.50 | 0.06 |
| GPT-4o (basic prompt) | 1.00 | 0.80 | 1.20 | 0.60 | 0.60 | 0.56 |
| GPT-4o (intermediate prompt) | 1.00 | 1.08 | 1.17 | 0.58 | 0.58 | 0.34 |
| GPT-4o (advanced prompt) | 1.00 | 1.17 | 1.17 | 0.58 | 0.58 | 0.29 |

Notes: This table reports the normalized sentiment score by the score assigned to the benchmark sentences. Each row represents various approaches to generate the sentiment scores. Each column, except the last one, corresponds to the selected sentences from news articles with different tones. The last column reports the sum of squared errors (SSE) of sentiment scores generated by various approaches relative to human assessment.

Table 2 clearly demonstrates the advantages of some Chinese LLMs and the GPT-4o model (particularly when used with an advanced prompt). The first row of Table 2 presents the normalized results from Table 1, indicating that the "Positive 1" and "Positive 2" sentences are more optimistic than the benchmark sentence (by 40 percent and 60 percent, respectively), and that the "Negative 1" and "Negative 2" sentences are more pessimistic.[12] Consistent with Table 1, the keyword-based method displays a high sum of squared errors (SSE) (2.18), indicating significant deviations from human assessments. Again, the use of two Chinese LLMs (Senta and Qianfan) does not improve the sentiment analysis. However, the use of the other six Chinese LLMs and the GPT-4o model with a "basic prompt" dramatically enhances the sentiment analysis. As the quality of the prompt improves, the SSE of the GPT-4o model decreases monotonically, suggesting a gradual improvement in the quality of the sentiment analysis. Note that the GPT-4o model with the advanced prompt is still outperformed by three Chinese LLMs (Kimi, Xinghuo, and Baichuan), a point we will revisit later.

[12] By definition, the SSE of the "human assessment" method is 0 because the deviations are calculated relative to this method.

We now examine *why* the different approaches lead to dramatically different sentiment assessments. Table 3 presents the rationales of the keyword-based approach for its sentiment assessments. This approach, like many traditional sentiment analysis methods, counts the number of positive and negative words based on a predefined dictionary and defines the sentiment score as the (scaled) ratio of the net number of positive words (i.e., the number of positive words minus negative words) over the total number of words. The disadvantage of this method is most evident in the “Positive 2” sentence, where it counts “遏止” (curb) and “收缩” (contraction) as negative words and fails to capture that these words combined in the phrase “curb the contraction” convey a positive sentiment.

**Table 3. Rationales of the Keywords-Based Approach on Sentiment Analysis**

| | NewsContent | posi_num | nega_num | net_posi_num | all_num | net_posi_ratio | senti_score | posi_words | nega_words |
|---|---|---|---|---|---|---|---|---|---|
| Benchmark | 鹿回头半岛最早的开发可以追溯到大约20年前。 | 2 | 0 | 2 | 5 | 0.40 | 90 | 最早#开发 | |
| Positive 1 | 有舆论担心，持续的松绑政策会使多年来的楼市调控成果毁于一旦；这种担心实在是不必要。 | 2 | 2 | 0 | 15 | 0.00 | 50 | 成果#实在 | 担心#担心 |
| Positive 2 | 多家机构预计，监管层很有可能继续放松房地产政策，出台刺激措施遏制房地产业的收缩速度，以缓解低迷的房地产市场。 | 1 | 4 | -3 | 21 | -0.14 | 36 | 缓解 | 刺激#遏制#收缩#低迷 |
| Negative 1 | 特别是近期房价上涨过快的热点城市，不排除会尽快出台更加严厉的具体调控政策。 | 2 | 1 | 1 | 12 | 0.08 | 58 | 上涨#具体 | 严厉 |
| Negative 2 | 虽然整体而言，此次调控在措施内容上并未超出市场预期，但是从操作力度和中央的决心看，调控有可能再次对地产市场走向和股市形成长期压制。 | 1 | 1 | 0 | 19 | 0.00 | 50 | 决心 | 压制 |

Notes: This table reports the number of positive and negative words based on a predefined dictionary and the ratio of the net number of positive words (the number of positives minus that of negatives) over the total number of words.

By contrast, the GPT-4o models with different levels of prompts can capture nuances, logical connections, and subtleties, reflecting not only the direct market outlook but also the indirect implications of government regulations. These nuances are shown in Table 4 for the basic prompt and Table 5 for the advanced prompt. Comparing Table 5 with Table 4 clearly demonstrates that the GPT-4o model with an advanced prompt provides more detailed, rigorous, and clearer reasoning than the basic prompt, explaining the difference in their performance. Moreover, although the GPT-4o model with advanced prompt deviates from the human assessment and assigns a slightly optimistic score to the benchmark sentence (60 versus the human assessment of 50), the rationale given by the model also stimulates human evaluators to think twice about the initial human assessment, which is another advantage of the LLM model.

**Table 4. Rationales of the GPT-4o Model with Basic Prompt on Sentiment Analysis**

| | SentimentScore | Rationale |
|---|---|---|
| **Benchmark** | 50 | The statement is neutral, providing a factual historical context without indicating positive or negative sentiment. |
| **Positive 1** | 40 | Although there are concerns, these concerns are deemed unnecessary, indicating a certain level of confidence in the market. However, the overall sentiment remains cautious. |
| **Positive 2** | 60 | The expectation of regulatory relaxation and stimulus measures indicates a moderately optimistic outlook for mitigating the downturn in the housing market. |
| **Negative 1** | 30 | The statement indicates concern over rapidly rising housing prices and the potential for stricter regulatory policies, suggesting a cautious or pessimistic outlook. |
| **Negative 2** | 30 | Regulatory measures may exert long-term pressure on the real estate market and the stock market, indicating a certain level of pessimism. |

Notes: This table reports the rationale of GPT-4o model with basic prompt for the sentiment scores assigned to each of the five selected sentences from news media.

**Table 5. Rationales of the GPT-4o Model with Advanced Prompt on Sentiment Analysis**

| | SentimentScore | Rationale |
|---|---|---|
| **Benchmark** | 60 | The statement reflects a long history of development, which can imply stability and sustained interest in the area, suggesting a moderately optimistic sentiment. |
| **Positive 1** | 70 | Although there are concerns in public opinion that the relaxation policies may affect the results of market regulation, the article believes these concerns are unnecessary, indicating an optimistic attitude towards the market outlook. |
| **Positive 2** | 70 | It is anticipated that regulatory authorities will relax real estate policies and introduce stimulus measures, indicating that market participants hold an optimistic attitude towards the future recovery of the real estate market. |
| **Negative 1** | 35 | The mention of potential stricter regulatory policies to control the rapid increase in housing prices in hotspot cities suggests a negative sentiment, as it implies that the government may take measures to cool down the market, potentially leading to lower housing prices and reduced profit margins for developers. |
| **Negative 2** | 35 | The article suggests that the regulatory measures, while expected, indicate strong determination from the central authorities, which could exert long-term pressure on the real estate market and stock market, leading to a pessimistic outlook. |

Notes: This table reports the rationale of GPT-4o model with advanced prompt for the sentiment scores assigned to each of the five selected articles from news media.

**Testing Results Based on Five *Articles***

The five articles we select are (in English):[13]

- **Benchmark Article:** The full text of an article published on 2014-09-19, titled *The Beijing Autumn Housing Expo has commenced, with over half of the projects on display being from overseas.*
- **Positive Article 1:** 2015-04-03, titled *The initial effects of policy overlay are becoming apparent, with a 70 percent month-on-month increase in transaction volumes across 30 cities*. This article reflects the impact of the "330 Xin Zheng" (see the next subsection).
- **Positive Article 2:** 2014-10-08, titled *The "loan relaxation" measures have been implemented, and the "lending without considering existing mortgages" policy has exceeded expectations*. This article reflects the impact of the "930 Ren Dai Bu Ren Fang".
- **Negative Article 1:** 2013-03-01, titled *Detailed implementation rules are forthcoming, indicating strong sustained effects following the "National Five Articles."* This article reflects the impact of the "Guo Wu Tiao".
- **Negative Article 2:** 2013-03-04, titled *Real estate regulation needs to refine fiscal and tax measures*. This article also reflects the impact of the "Guo Wu Tiao".

The testing results are presented in Table 6 and Table 7. Unlike the previous case of assessing the sentiment of five *sentences*, the GPT-4o model has the highest performance when assessing the sentiment of five *articles*. Indeed, the sentiment scores produced by GPT-4o models with the intermediate and advanced prompts are most consistent with those by human assessments and outperform the keywords model and all the Chines LLMs. The specific results are presented in Table 6 and Table 7. One possible reason is that compared with these other models, GPT-4o is more capable of capturing the overall context and the linkages across sentences within the article, and this advantage is more pronounced when analyzing articles instead of individual sentences. Another possible reason is that compared with Chinese LLMs, GPT-4o is trained on more extensive data that are less subject to government restrictions.

[13] To facilitate replications of our results, we also provide the titles of these five articles in the original language (Chinese): (1) **Benchmark Article:** 北京秋季房展会开幕 海外项目占比过半. (2) **Positive Article 1:** 政策叠加成效初显 30 城成交面积环比增七成. (3) **Positive Article 2:** 松贷"靴子"落地 "贷清不认房"超预期. (4) **Negative Article 1:** 配套细则呼之欲出 "国五条"后劲足. (5) **Negative Article 2:** 地产调控需细化财税手段.

Note that in this case, the GPT-4o model with a "basic prompt" has a worse performance even than the keywords-based model; but as the quality of the prompt improves, the SSE decreases monotonically and drops below that of the keywords-based model. This underscores further the importance of designing proper prompts while applying the GPT-4o model.

The above testing procedure has intentionally normalized the initial level of each method's sentiment score. But as a robustness check and to further account for the level of sentiment score, we calculate the pairwise correlation between the raw scores of the human assessment and those of each method. As shown in Table 8, the sentiment scores constructed by GPT-4o (with intermediate and advanced prompts) have the highest correlations with those by human assessment. In particular, the Senta model now has a poor performance, given that the levels of its sentiment scores are heavily biased towards 100. It is also worth noting that the monotonicity with respect to the level of prompts still holds, that is: as the quality of the prompt improves, the correlation increases monotonically and exceeds that of other models.

**Table 6. Validity Test Against Human Assessments for Selected Articles**

| | Benchmark | Positive 1 | Positive 2 | Negative 1 | Negative 2 |
|---|---|---|---|---|---|
| Human Assessment | 55 | 70 | 80 | 35 | 25 |
| Keywords-Based | 54 | 54 | 51 | 55 | 48 |
| Qianfan Model | 98 | 22 | 27 | 63 | 22 |
| Xinghuo Model | 65 | 40 | 40 | 70 | 45 |
| BaiChuan Model | 70 | 40 | 40 | 30 | 20 |
| Kimi Assistant | 70 | 60 | 45 | 55 | 40 |
| Senta Model | 100 | 94 | 100 | 100 | 100 |
| Hunyuan Model | 60 | 55 | 45 | 50 | 30 |
| GPT-4o (basic prompt) | 60 | 45 | 40 | 40 | 30 |
| GPT-4o (intermediate prompt) | 60 | 55 | 45 | 35 | 35 |
| GPT-4o (advanced prompt) | 60 | 60 | 45 | 35 | 35 |

Notes: This table reports the sentiment scores generated by human assessments and various LLMs for each of the five selected articles from news media. The sentiment score ranges from 0 (most pessimistic) to 100 (most optimistic). "Benchmark" refers to the article that has a neutral sentiment based on human assessment. "Positive" refers to the articles that have a clear optimistic tone. "Negative" refers to the articles that have a clear pessimistic tone.

**Table 7. Validity Test Against Human Assessments for Selected Articles: Ratio of Sentiment Scores over the Benchmark**

| | Benchmark | Positive 1 | Positive 2 | Negative 1 | Negative 2 | SSE |
|---|---|---|---|---|---|---|
| Human Assessment | 1.00 | 1.27 | 1.45 | 0.64 | 0.45 | 0.00 |
| Keywords-Based | 1.00 | 1.00 | 0.94 | 1.02 | 0.89 | 0.67 |
| Qianfan Model | 1.00 | 0.22 | 0.28 | 0.64 | 0.22 | 2.54 |
| Xinghuo Model | 1.00 | 0.62 | 0.62 | 1.08 | 0.69 | 1.39 |
| BaiChuan Model | 1.00 | 0.57 | 0.57 | 0.43 | 0.29 | 1.34 |
| Kimi Assistant | 1.00 | 0.86 | 0.64 | 0.79 | 0.57 | 0.87 |
| Senta Model | 1.00 | 0.94 | 1.00 | 1.00 | 1.00 | 0.74 |
| Hunyuan Model | 1.00 | 0.92 | 0.75 | 0.83 | 0.50 | 0.66 |
| GPT-4o (basic prompt) | 1.00 | 0.75 | 0.67 | 0.67 | 0.50 | 0.90 |
| GPT-4o (intermediate prompt) | 1.00 | 0.92 | 0.75 | 0.58 | 0.58 | 0.64 |
| GPT-4o (advanced prompt) | 1.00 | 1.00 | 0.75 | 0.58 | 0.58 | 0.59 |

Notes: This table reports the normalized sentiment score by the score assigned to the benchmark article. Each row represents various approaches to generate the sentiment scores. Each column, except the last one, corresponds to the selected sentences from news articles with different tones. The last column reports the sum of squared errors (SSE) of sentiment scores generated by various approaches relative to human assessment. "Benchmark" refers to the article that has a neutral sentiment based on manual evaluation. "Positive" refers to the articles that have a clear optimistic tone. "Negative" refers to the articles that have a clear pessimistic tone.

**Table 8. Validity Test Against Human Assessments for Selected Articles: Pairwise Correlation between Raw Scores**

| | Pairwise Correlation between Raw Scores |
|---|---|
| Human Assessment | 1.00 |
| Keywords-Based | 0.26 |
| Qianfan Model | -0.15 |
| Xinghuo Model | -0.47 |
| BaiChuan Model | 0.46 |
| Kimi Assistant | 0.24 |
| Senta Model | -0.41 |
| Hunyuan Model | 0.50 |
| GPT-4o (basic prompt) | 0.40 |
| GPT-4o (intermediate prompt) | 0.63 |
| GPT-4o (advanced prompt) | 0.64 |

Notes: This table reports the pairwise correlation between the raw (i.e., un-normalized) scores of the human assessment and those of each method. A higher correlation suggests a closer linear relationship with the human assessment scores. By definition, the correlation of the human assessment scores with themselves is 1.

### B. Sentiment Around Announcements of "Major Housing Policies"

We manually go through all the housing-related policies in China that were announced by the Chinese government between January 2012 and June 2016. We then identify the "major housing policies" based on the mentions in the CSMAR Chinese news article dataset, as well as the number of searches in Bing.

Ultimately, we identify the following three policies to be the "major housing policies":

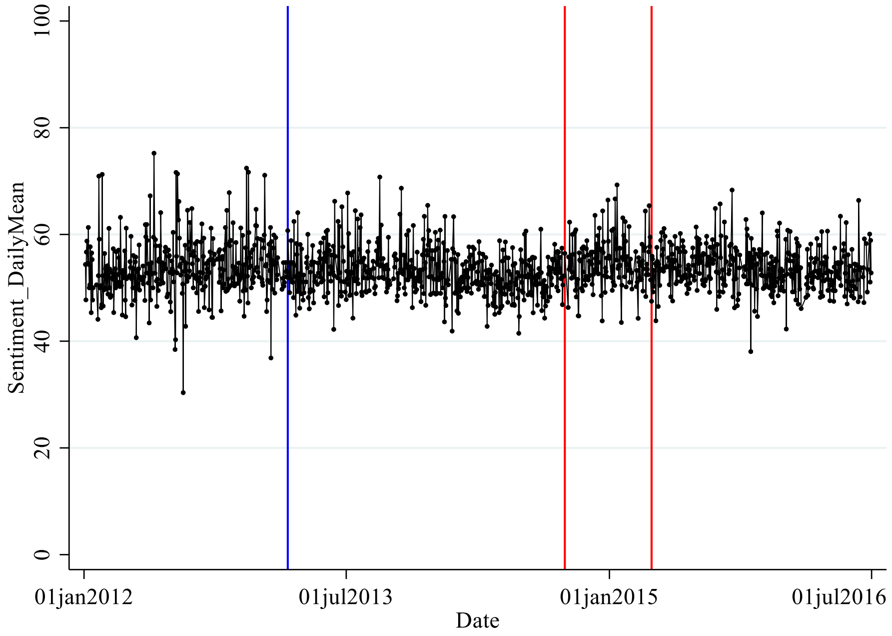
- 2013-03-01: "Guo Wu Tiao". This represents five significant regulatory changes issued by the central government, all of which are tightening regulations.
- 2014-09-30: "930 Ren Dai Bu Ren Fang", which was an easing policy.
- 2015-03-30: "330 Xin Zheng", which means "March 30 New Policy" and was also an easing policy.

We now assess how the sentiment indices constructed by different models evolve around these three major housing policies. To facilitate the subsequent analysis of the monetary policy transmission, these two figures only present the index from January 2012 to June 2016, for which we have the non-housing consumption data.

We first compare the sentiment indices at the daily frequency. The daily average sentiment indices constructed by the keywords-based model, the Hunyuan model, the Senta model, and the GPT-4o model are plotted in Figures 1-4. We choose only two of the six Chinese LLMs presented earlier because the performance of Hunyuan and Senta models in assessing the five testing articles is closest to that of GPT-4o model (see Table 7). Although it is challenging to visualize clear patterns around the major housing policies for all three models, it is evident that the daily sentiment index constructed by GPT-4o is less volatile, especially compared with that constructed by the Senta model. The levels of the daily index are also more sensible, especially when compared with the Senta model that clearly has an optimism bias in the sentiment index it constructs.[14]

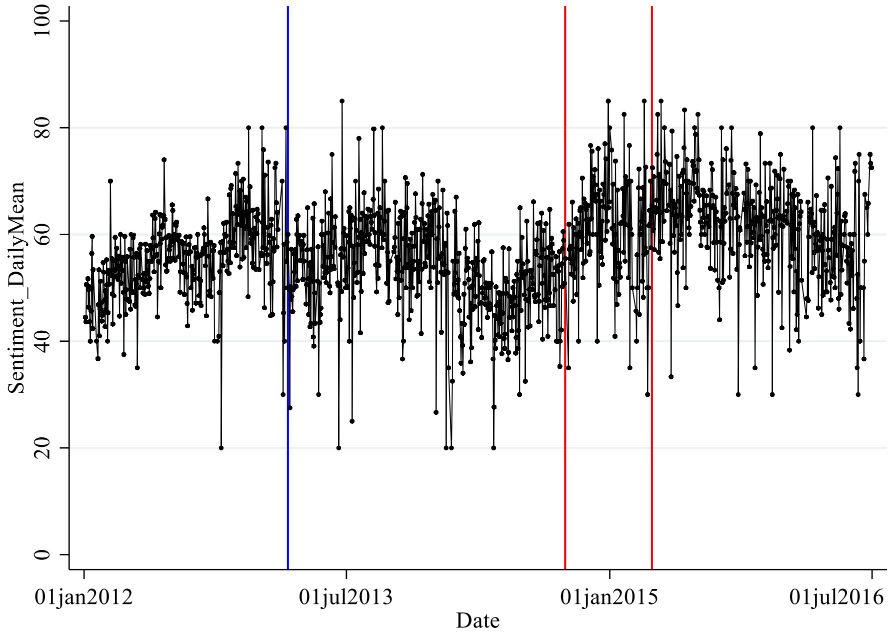
[14] There is no quantitative basis regarding how "major" these housing policies are, and how much we should expect the "correct" housing sentiment to respond. As such, it is challenging to perform a strictly quantitative assessment for this particular validity test. However, the other two complementary validity tests we have conducted are quantitative.

**Figure 1. Daily CHMSI by the *Keywords-Based* Model**



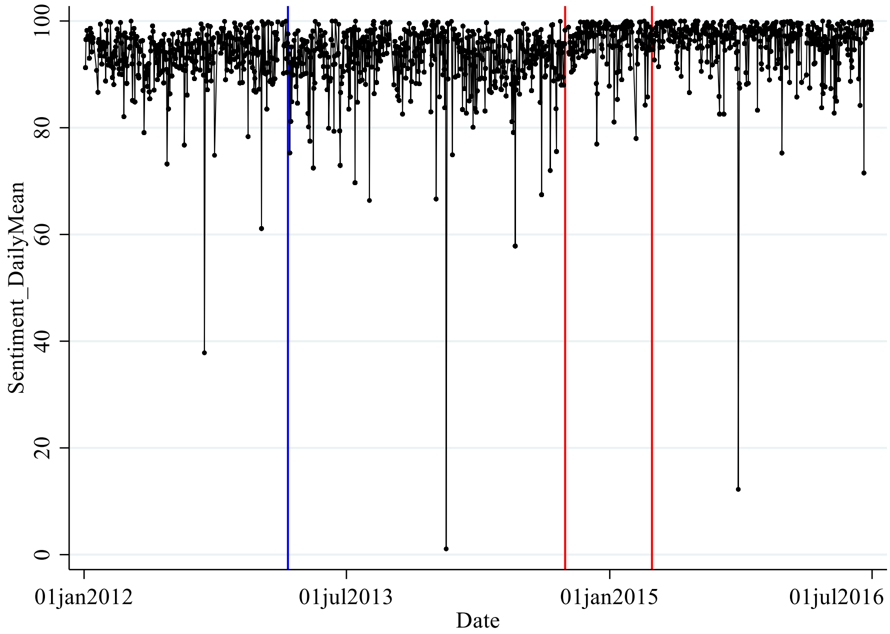

Note: This figure plots the Chinese housing market sentiment index at a daily frequence from January 1, 2012 to June 30, 2016 using keywords-based model. The blue line denotes the date for the five tightening housing regulations “Guo Wu Tiao” issued by the central government (March 1, 2013). The left red line denotes the date for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the date for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

**Figure 2. Daily CHMSI by the Hunyuan Model**



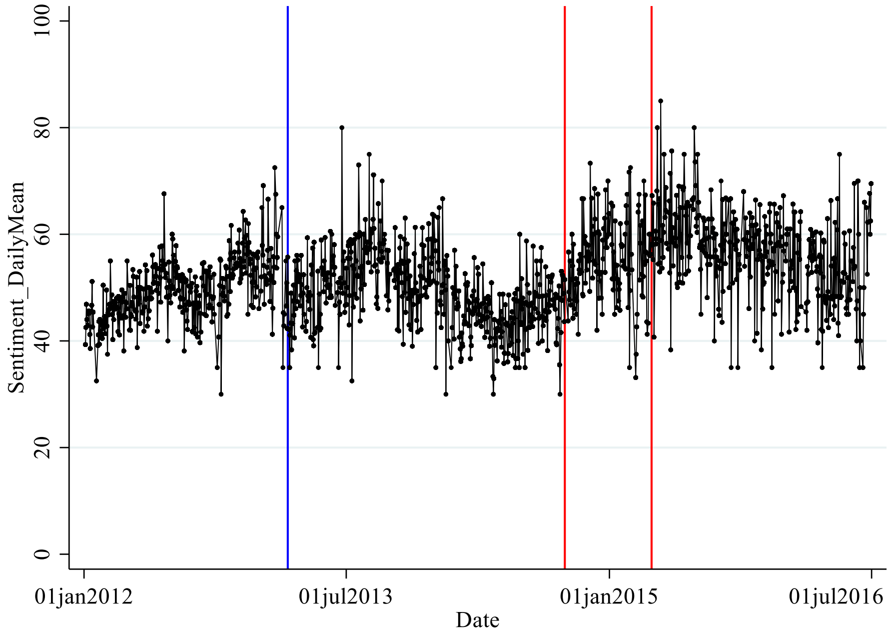

Note: This figure plots the Chinese housing market sentiment index at a daily frequence from January 1, 2012 to June 30, 2016 using Hunyuan model. The blue line denotes the date for the five tightening housing regulations “Guo Wu Tiao” issued by the central government (March 1, 2013). The left red line denotes the date for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the date for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

**Figure 3. Daily CHMSI by the *Senta* Model**




Note: This figure plots the Chinese housing market sentiment index at a daily frequence from January 1, 2012 to June 30, 2016 using Baidu's Senta model. The blue line denotes the date for the five tightening housing regulations "Guo Wu Tiao" issued by the central government (March 1, 2013). The left red line denotes the date for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the date for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

**Figure 4. Daily CHMSI by the GPT-4o Model**

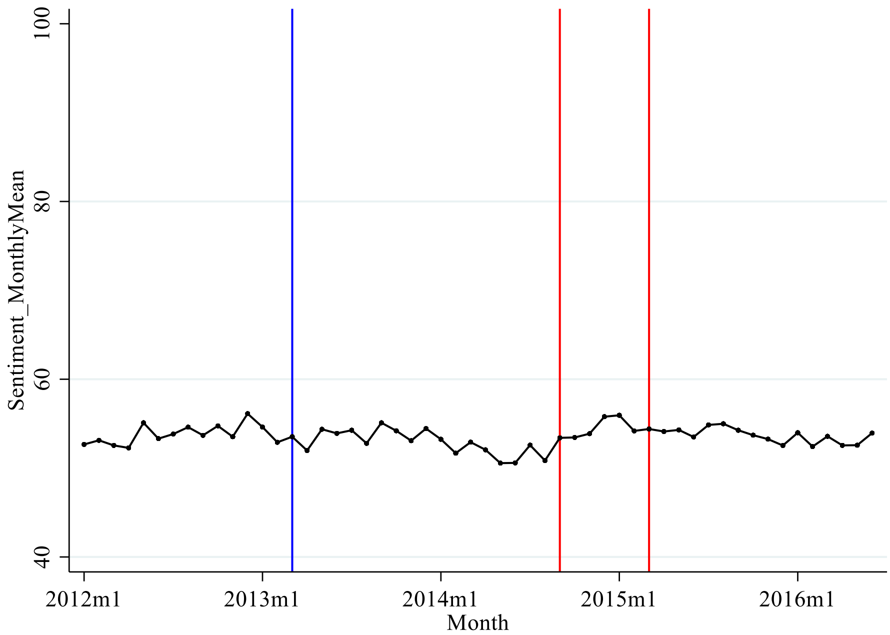



Note: This figure plots the Chinese housing market sentiment index at a daily frequence from January 1, 2012 to June 30, 2016 using GPT-4o model. The blue line denotes the date for the five tightening housing regulations “Guo Wu Tiao” issued by the central government (March 1, 2013). The left red line denotes the date for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the date for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

The comparison of the monthly average sentiment indices is more informative, where the monthly indices are defined as the simple averages of daily indices constructed on all articles during that month. Although none of the four models point to a significant downward shift in housing sentiment after the tightening policy in March 2013, the GPT-4o monthly sentiment index displays a clearer upward shift immediately after the easing policy in September 2014 compared with the other models. Indeed, the keywords-based model (Figure 5) points to a flat sentiment one month after this significantly stimulating policy, whereas the GPT-4o model (Figure 8) immediately detects a more optimistic sentiment. And even though the other two LLMs (Hunyuan and Senta) also capture this pattern, the levels of the GPT-4o-constructed index are more sensible, especially when compared with the Senta model, as noted earlier.

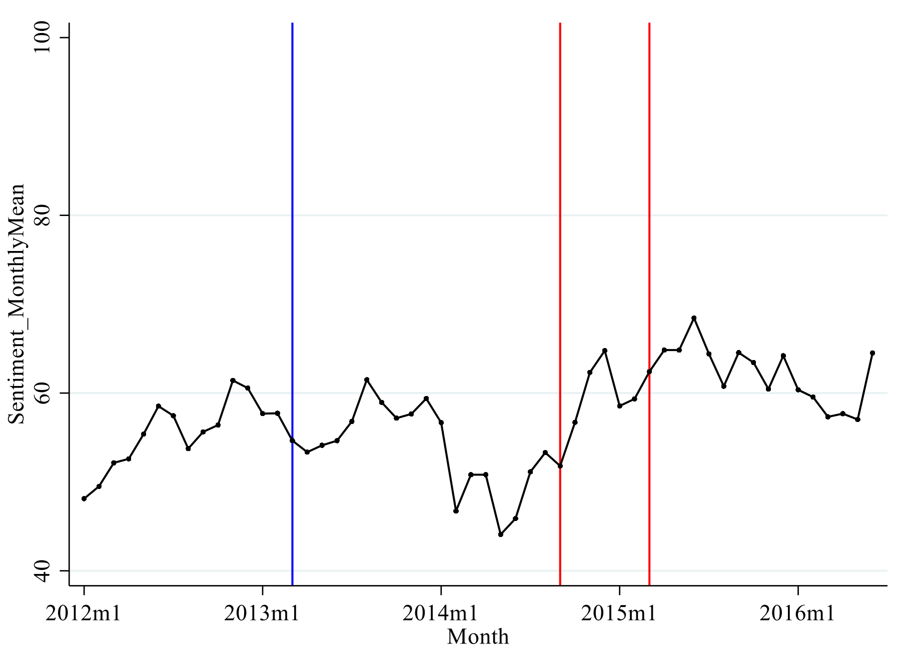
As for the easing policy in March 2015, both the GPT-4o model and the two Chinese LLMs model detect a more optimistic sentiment. This seems to be the correct sentiment based on our own quick assessment after manually reading several news articles after the release of that policy. By contrast, the keywords-based model incorrectly points to a flat or lower sentiment.

**Figure 5. Monthly CHMSI by the *Keywords-Based* Model**



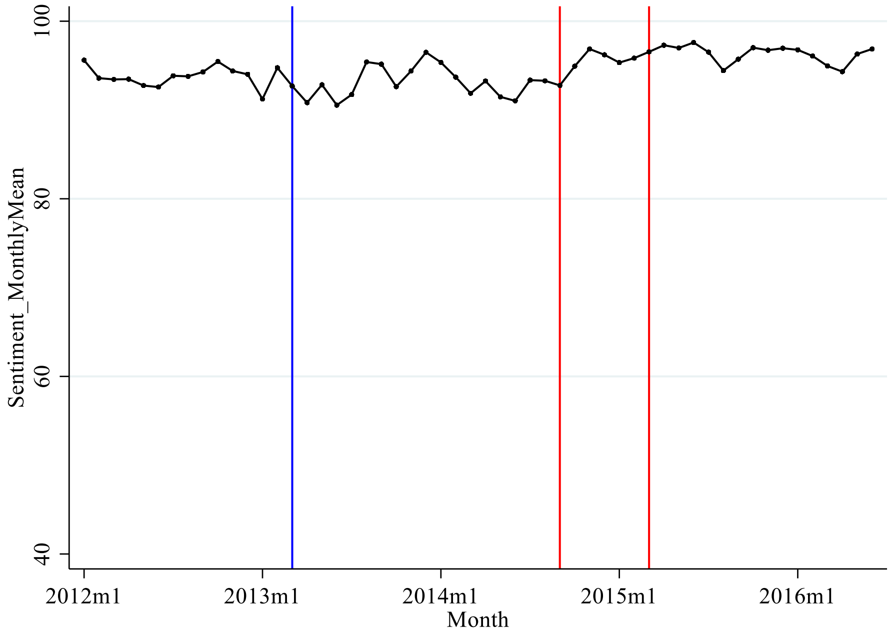

Note: This figure plots the Chinese housing market sentiment index at a quarterly frequence from January 1, 2012 to June 30, 2016 using keywords-based model. The blue line denotes the quarter for the five tightening housing regulations "Guo Wu Tiao" issued by the central government (March 1, 2013). The left red line denotes the quarter for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the quarter for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

**Figure 6. Monthly CHMSI by the *Hunyuan* Model**



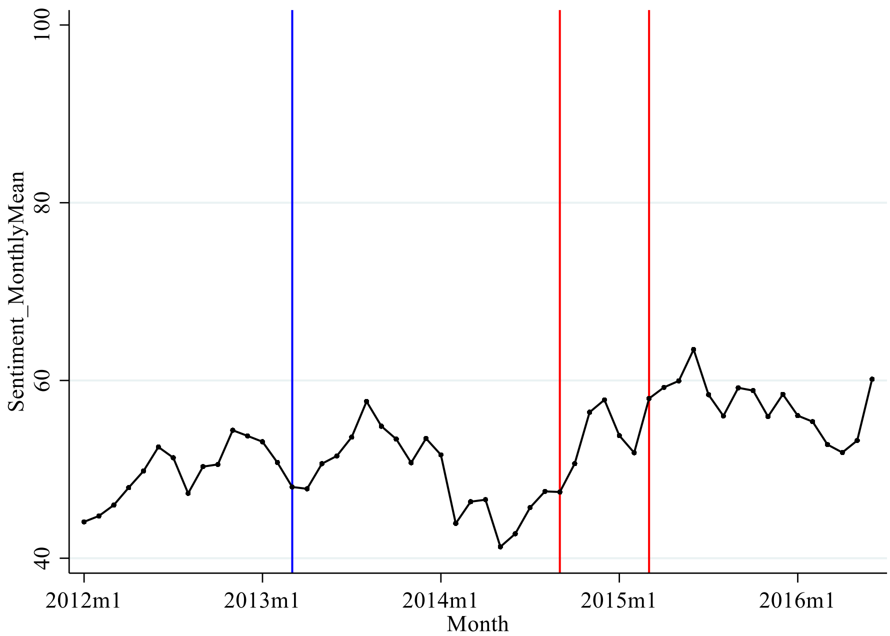

Note: This figure plots the Chinese housing market sentiment index at a quarterly frequence from January 1, 2012 to June 30, 2016 using Hunyuan model. The blue line denotes the quarter for the five tightening housing regulations “Guo Wu Tiao” issued by the central government (March 1, 2013). The left red line denotes the quarter for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the quarter for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

**Figure 7. Monthly CHMSI by the *Senta* Model**




Note: This figure plots the Chinese housing market sentiment index at a quarterly frequence from January 1, 2012 to June 30, 2016 using Baidu's Senta model. The blue line denotes the quarter for the five tightening housing regulations "Guo Wu Tiao" issued by the central government (March 1, 2013). The left red line denotes the quarter for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the quarter for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

**Figure 8. Monthly CHMSI by the GPT-4o Model**




Note: This figure plots the Chinese housing market sentiment index at a quarterly frequence from January 1, 2012 to June 30, 2016 using GPT-4o model. The blue line denotes the quarter for the five tightening housing regulations "Guo Wu Tiao" issued by the central government (March 1, 2013). The left red line denotes the quarter for the relaxed mortgage policy non-primary houses (September 30, 2014). The right red line denotes the quarter for relaxed mortgage policy on both primary and secondary homes (March 30, 2015).

## C. Predicting Housing Prices

We now relate the constructed CHMSI to the actual national-level housing price growth at a monthly frequency, for each of the four models (keywords-based model, Hunyuan model, Senta model, and GPT-4o model).[15] We choose only two of the six Chinese LLMs presented earlier again because the performance of Hunyuan and Senta models in assessing the five testing articles is closest to that of GPT-4o model. To this end, we use the sentiment index to forecast house price growth and compare the forecasting power among the four indices. To allow for flexible functional forms and capture the complex nonlinear relationship between sentiment and house price, we use *machine learning* models for this purpose, similar to Liu, Yang, and Zhao (2022). Despite the small number of predictors, using machine learning models can still better capture the

[15] The housing price data are obtained from the CEIC database.

potential nonlinearities among them, and nonlinearities between them and the outcome variable, as established in IMF (2021). Although this comparison is only suggestive because housing prices are highly endogenous and affected by many factors beyond housing market sentiment, it could still shed some light on the validity of the constructed housing market sentiment index.

Specifically, our primary objective is to assess the predictive power of different sentiment indices on the Chinese housing market. To this end, we incorporate machine-learning models, commonly employed in macroeconomic forecasting, to predict the year-on-year growth rate of house prices (HPG_YOY). These models include Random Forest, ElasticNet, Lasso, and XGBoost, and we compare their performance with a traditional reduced-form VAR model. The comparison criterion is the out-of-sample root mean squared errors (RMSE) and Mean Absolute Error (MAE), both being frequently used in the forecasting evaluation literature.[16]

Our analysis is conducted in several steps. First, we preprocess and merge sentiment data with house price data. Each of the three sentiment indices is averaged monthly and merged with the house price growth data. We then perform the "horse race" of models to determine the best-performing forecasting model. The models considered are VAR, Lasso, ElasticNet, XGBoost, and Random Forest. We also include the results of an AR(1) model as the benchmark (with a constant and the lagged house price growth as predictors), following the practice in the literature. We employ an expanding window approach for cross-validation, which sequentially increases the training sample by one month. This approach ensures that the model is trained on past data and validated on future data, which is crucial to prevent future information from being leaked to the past.

The data from January 2011 to June 2015 are used for training and cross-validation. More precisely, we use the sentiment data from January 2011 to June 2015 and the house price growth data from January 2012 to June 2015; we do so because we also allow for the lagged sentiment to predict the current house price growth (up to lags of 12 months). More precisely, to predict the year-on-year growth rate of house price in the current month, we use the lagged year-on-year growth rates of house price (up to

[16] RMSE is the square root of the average of the squared errors between the forecasted values and the actual values. MAE is the average of the absolute errors between the forecasted values and the actual values. By definition, RMSE is more sensitive to outliers because squaring the errors before averaging them gives higher weight to larger errors. By contrast, MAE is less sensitive to outliers because it treats all errors linearly. Hence, RMSE is preferred in situations where larger errors need to be penalized more heavily, such as in financial forecasting or risk assessment.

lags of 12 months), the lagged sentiment index (up to lags of 12 months), and the sentiment index in the current month. As for inclusion of lags, for VAR, the number of lags is optimally selected using AIC and BIC; for Lasso and ElasticNet, they include all 12 lags initially, but the regularization process effectively selects the relevant lags by shrinking some coefficients to zero, which acts as a form of optimal lag selection; for XGBoost, it also includes all 12 lags as features, but its tree-based mechanism inherently identifies and uses the most important lags during the splitting process. And for Random Forest, it includes all 12 lags as features, but it does not "shrink" coefficients or explicitly discard features. Instead, it evaluates all provided lags and uses its inherent feature importance mechanism to weight the importance of each lag.

For each machine learning model, we tune hyperparameters using grid search with time-series cross-validation to minimize the average RMSE in the training set. We then compute the out-of-sample forecasting performance metrics of these models using data from July 2015 to June 2016.

The predicted house price growth rates from various models are presented in Figures 9-13, for the keywords-constructed sentiment, Hunyuan model-constructed sentiment, Senta model-constructed sentiment, the GPT-4o-constructed sentiment, and the case without any sentiment index, respectively. We then calculate the RMSE for each model to assess its forecasting performance. Table 9 summarizes the performance results.

**Table 9. Performance of Out-of-Sample Forecasts across Models and Sentiment Indices**

| Model | RMSE | MAE | Sentiment |
|---|---|---|---|
| AR(1) | 7.16 | 6.50 | Keywords |
| VAR | 4.83 | 3.85 | Keywords |
| Lasso | 3.70 | 3.04 | Keywords |
| ElasticNet | 3.64 | 2.96 | Keywords |
| RandomForest | 3.40 | 2.90 | Keywords |
| XGBoost | 6.75 | 6.20 | Keywords |
| AR(1) | 7.16 | 6.50 | Hunyuan |
| VAR | 4.54 | 3.61 | Hunyuan |
| Lasso | 3.70 | 3.04 | Hunyuan |
| ElasticNet | 3.64 | 2.96 | Hunyuan |
| RandomForest | 4.10 | 3.37 | Hunyuan |
| XGBoost | 6.70 | 6.10 | Hunyuan |
| AR(1) | 7.16 | 6.50 | Senta |
| VAR | 11.34 | 10.46 | Senta |
| Lasso | 3.70 | 3.04 | Senta |
| ElasticNet | 3.64 | 2.96 | Senta |
| RandomForest | 3.76 | 3.14 | Senta |
| XGBoost | 6.78 | 6.05 | Senta |
| AR(1) | 7.16 | 6.50 | GPT |
| VAR | 18.85 | 16.28 | GPT |
| Lasso | 3.63 | 3.13 | GPT |
| ElasticNet | 3.59 | 3.08 | GPT |
| RandomForest | 3.39 | 2.87 | GPT |
| XGBoost | 6.38 | 5.90 | GPT |
| AR(1) | 7.16 | 6.50 | No Sentiment |
| VAR | 4.51 | 3.91 | No Sentiment |
| Lasso | 3.63 | 3.13 | No Sentiment |
| ElasticNet | 3.59 | 3.08 | No Sentiment |
| RandomForest | 3.77 | 3.16 | No Sentiment |
| XGBoost | 6.30 | 5.90 | No Sentiment |

Note: This table reports the out-of-sample root mean square errors (RMSE) across various forecasting models, each predicting the year-on-year growth rate of house prices by lagged house prices growth rate, and contemporaneous and lagged housing sentiment indices constructed using models listed in the right column. The model results without using any sentiment index are also presented as a benchmark.

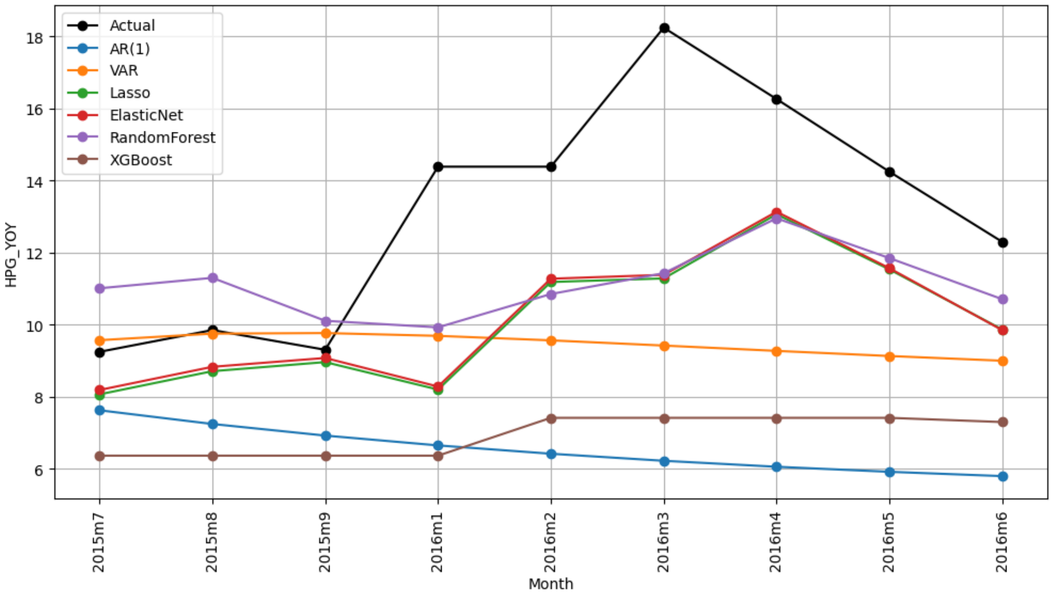
As shown in Table 9, while using the keywords-based sentiment index, the best performing forecasting model is the random forest model (marked in red), which has the lowest RMSE (3.40) among all models. And while using the Hunyuan-constructed sentiment index, the best performing forecasting model is the elastic net model, with an

RMSE of 3.64. The same is true while using the Senta-constructed sentiment index.[17] Finally, while using the GPT-4o sentiment index, the best performing forecasting model is the random forest model, with a lower RMSE of 3.39, which is also lower than the RMSE of the best performing benchmark model that does not use any sentiment index. Therefore, the *optimal* forecasting model using the GPT-4o constructed sentiment index achieves a *better* forecasting performance than those using the keywords-based sentiment index and the other two LLMs-constructed sentiment index (Hunyuan and Senta). Moreover, using the GPT-4o constructed sentiment index also helps enhance the forecasting performance relative to the case without using any sentiment index. The same conclusion holds when evaluating forecasting performance using the MAE measure.

Note that across all sentiment indices and forecasting models, all machine learning models outperform the AR(1) model, and almost all machine learning models outperform the corresponding VAR model, highlighting the forecasting advantage of machine learning models relative to traditional models (such as the VAR model). This advantage likely reflects machine learning models' flexibility and their ability to capture complex nonlinearities, as widely established in the literature.

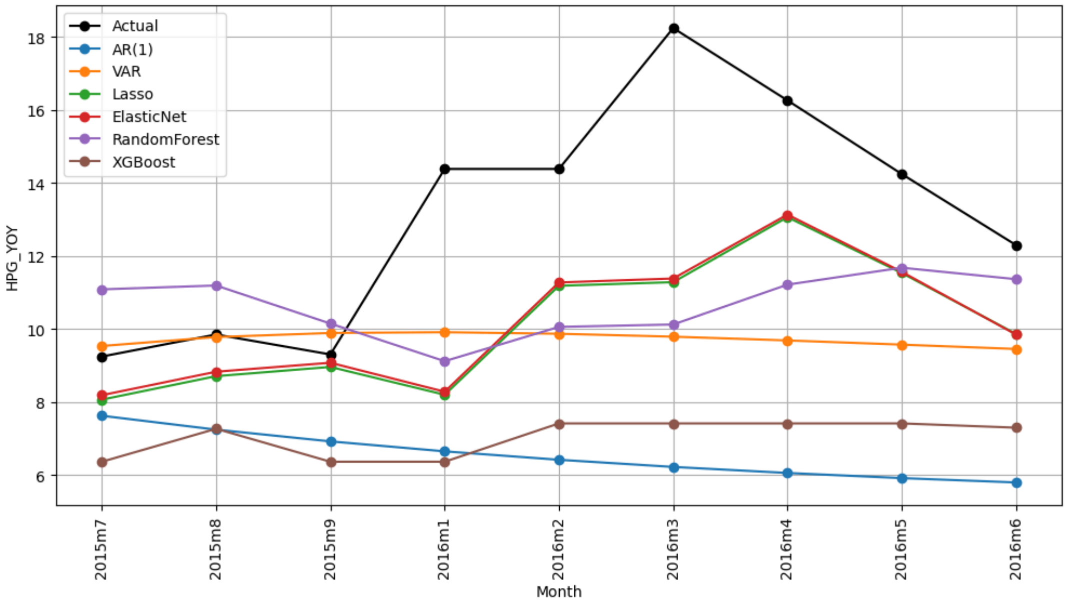
[17] Note that it is possible for two different models to have the same forecasting performance.

**Figure 9. Out-of-Sample Forecasts of House Price Growth Using the Keywords Sentiment**



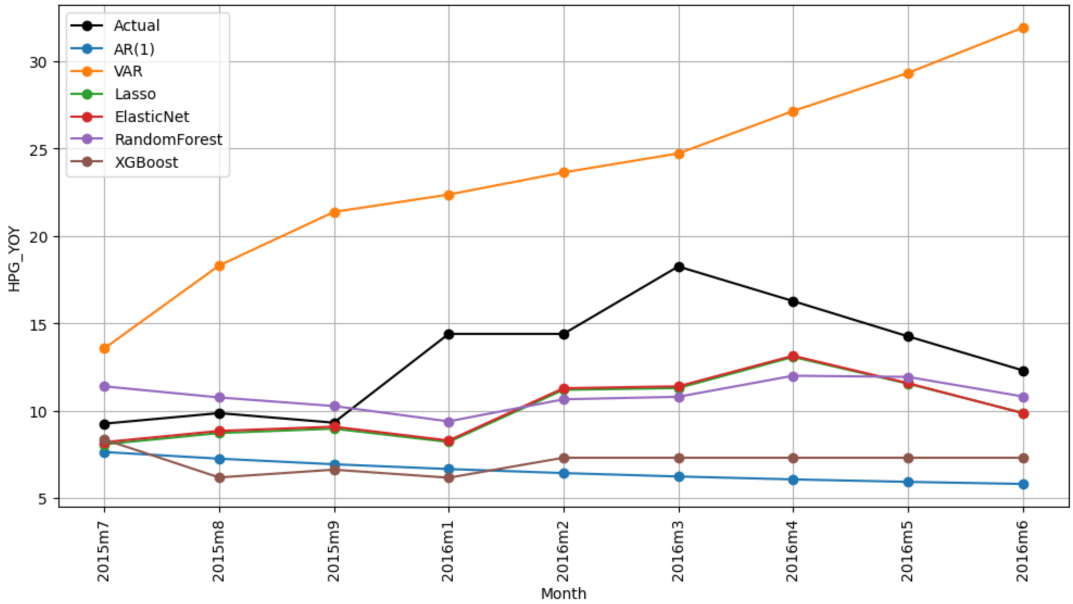

Notes: This figure plots the out-of-sample forecast of house price growth rate from various forecasting models using keywords-based sentiment index.

**Figure 10. Out-of-Sample Forecasts of House Price Growth Using the Hunyuan Sentiment**



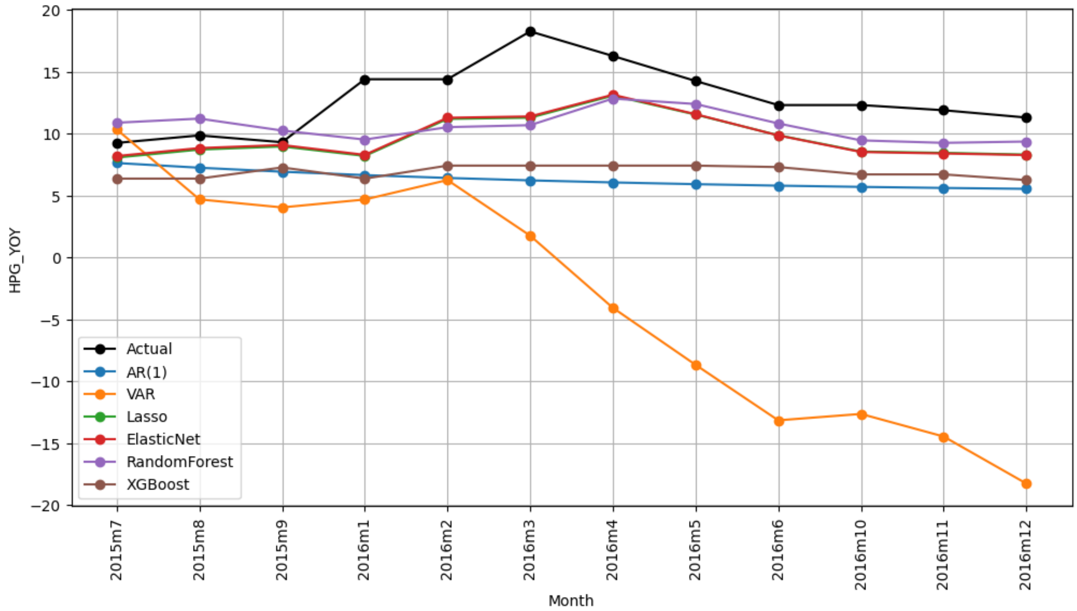

Notes: This figure plots the out-of-sample forecast of house price growth rate from various forecasting models using the Hunyuan sentiment index.

**Figure 11. Out-of-Sample Forecasts of House Price Growth Using the Senta Sentiment**



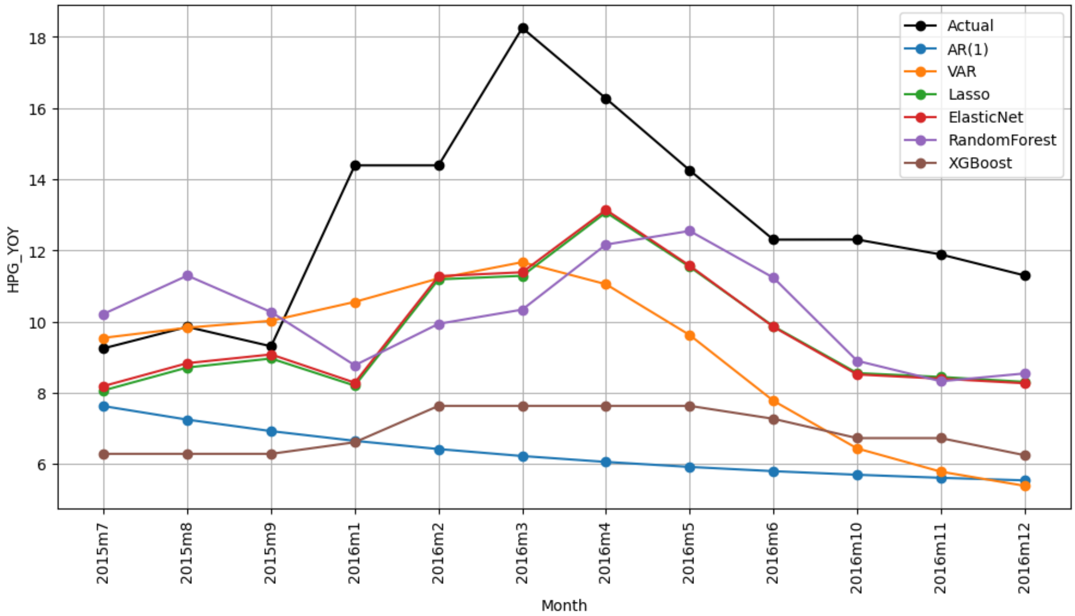

Notes: This figure plots the out-of-sample forecast of house price growth rate from various forecasting models using the Senta sentiment index.

**Figure 12. Out-of-Sample Forecasts of House Price Growth Using the GPT-4o Sentiment**




Notes: This figure plots the out-of-sample forecast of house price growth rate from various forecasting models using the GPT-4o sentiment index.

**Figure 13. Out-of-Sample Forecasts of House Price Growth without Using Any Sentiment Index**




Notes: This figure plots the out-of-sample forecast of house price growth rate from various forecasting models without using any sentiment index. In this case, the VAR model degenerates to an AR(p) model, where the number of lags p is optimally selected using the AIC.

### D. A Real-Time Presentation of CHMSI

After establishing the validity of the sentiment index constructed by GPT-4o according to the aforementioned three criteria, we now plot the monthly average index from January 4, 2012 to the most recent date (September 11, 2024).[18] Doing so enables us to trace the housing sentiment during the ongoing housing slump in China, as well as some other recent major policies. Specifically, we examine the following events/policies:

- March 2013 (blue): "Guo Wu Tiao". This represents five significant regulatory changes issued by the central government, all of which are tightening regulations.

[18] In total, it takes about 16 hours for GPT-4o (operated on a high-performing A100 Nvidia GPU) to construct the daily sentiment index from January 4, 2012 to September 11, 2024. The total number of tokens is 159,734,855, and the total financial cost is $804.64 (excluding the relatively low costs of testing the code on a small-scale basis). Equivalently, the financial cost per million tokens is $5.04. The operations of other LLMs are free of charge, although they achieve a lower performance than GPT-4o, as established in our paper. Going forward, updating the index using the new data would be cheap because one does not need to rerun the algorithm for the historical data.

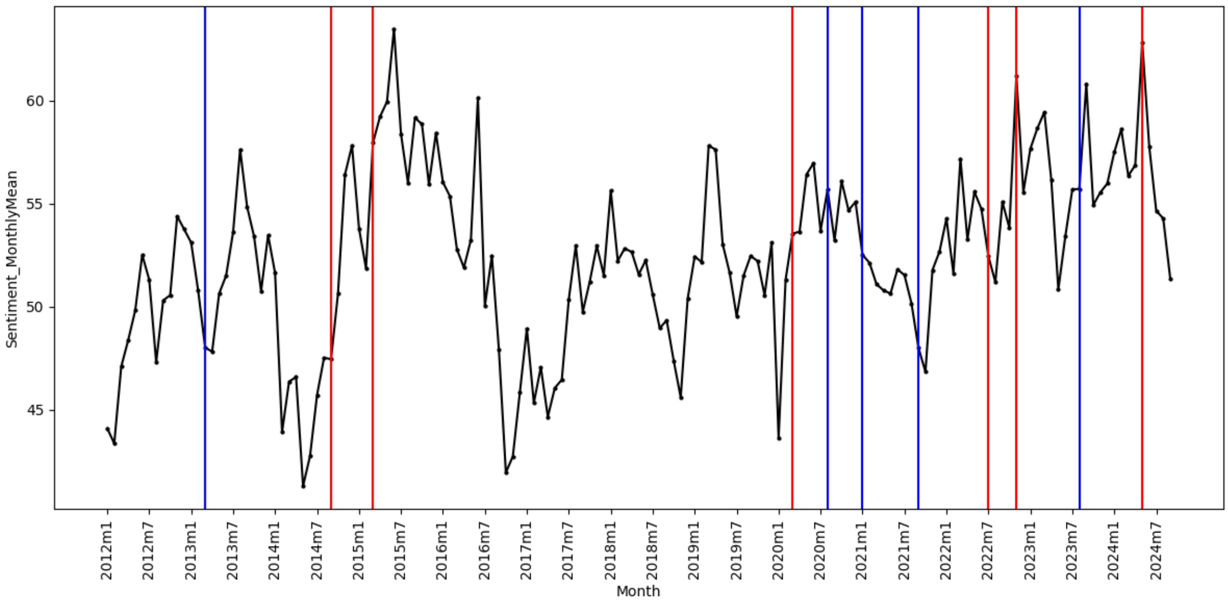
- September 2014 (red): "930 Ren Dai Bu Ren Fang", which was an easing policy.
- March 2015 (red): "330 Xin Zheng", which means "March 30 New Policy" and was also an easing policy.
- March 2020 (red): The early stage of the COVID lock-down, where the massive household savings may have benefited the housing market, especially given the limited options of non-housing investment vehicles in China (see Bayoumi and Zhao, 2020).[19]
- August 2020 (blue): The release of the "Three Red Lines" policy, that is: (1) liability-to-asset ratio < 70 percent; (2) net gearing ratio < 100 percent; (3) cash-to-short-term-debt ratio > 1.
- January 2021 (blue): The implementation of a new "draconian" rule (with a new concentration management system) to limit banks' property related lending to their capitalization based on a five-tier grade.
- September 2021 (blue): In summer of 2021, payments due on Evergrande's debt (in the hundreds of billions of dollars) resulted in the Evergrande liquidity crisis, leading to a drop in many stock market indices on September 20, 2021 (source: *Financial Times*, September 20, 2021).
- July 2022 (red): In July 2022, amid mounting concerns over the completion of unfinished houses, the Chinese Communist Party's Politico Bureau announced the policy of "ensure the delivery of houses and safeguard the interests of homebuyers" ("Bao Jiao Lou, Wen Min Sheng").
- November 2022 (red): In November 2022, the Chinese government came up with a comprehensive policy package consisting of sixteen financial policies (called "Jin Rong Shi Liu Tiao").
- August 2023 (blue): Country Garden, the largest private property developer in China (generally believed to be one of the healthiest developer), warned of a large net loss for the first half of 2023 due to loss on property projects and declining profit margins (source: Sohu).
- May 2024 (red): "517 Xin Zheng", meaning "May 17 New Policy". On May 17, 2024, China's central bank issued three significant documents to stimulate the housing market, detailing that: The down payment ratio for first homes will be reduced to a historic low, the interest rate on provident fund loans will be lowered by 0.25 basis points, and the interest rate floor for first and second homes will be abolished.

The sentiment evolution is plotted in Figure 14, where the first three vertical lines are the same as those in Figure 4. The early stage of the COVID lock-down (March 2020) is marked with a sharp rise in housing sentiment, consistent with the observations in

[19] Many other countries also experienced rapid house price increases at the beginning of Covid, such as the U.S. (as in Zhao, 2020) and a broad range of other countries (as in Biljanovska, Fu, and Igan, 2023, which covers 40 countries).

Bayoumi and Zhao (2020). This optimistic sentiment reversed two months after the release of the dramatic "Three Red Lines" policy in August 2020, reinforced by the implementation of the new "draconian" rule in January 2021. The pessimistic sentiment almost reached the bottom in September 2021, when Evergrande failed to make its debt payments. And the default of Country Garden in August 2023 reversed the recovering trend of the housing sentiment.

**Figure 14. Monthly CHMSI by the *GPT-4o* Model from 2012m1 to 2024m9**




Notes: This figure plots the real-time Chinese housing sentiment index constructed using GPT-4o model with advanced prompts. The blue lines refer to the quarters in which a tightening housing policy was announced, or housing markets were distressed due to defaults by giant real estate developers. The red lines refer to the quarters with a loosening housing policy announcement or a positive housing market development.

Interestingly, housing sentiment did not monotonically decline since the distress of Evergrande in September 2021. There are two possible explanations. First, this possibly reflects investors' expectations that the government will intervene to support the market and the impact of the actual government support. For example, in July 2022, the central government announced massive stimulating policies, and sentiment indeed started rising one month later. And a similar pattern appeared after the government announced the sixteen financial policies in November 2022.

The second potential explanation is that there might have been a regime switch in housing-related media coverage in China. After the Evergrande incident in September 2021, such coverage is more selective, with a more positive tone. For example, on July

31, 2023, when Country Garden (the largest private property developer in China) experienced a financing difficulty, only one Chinese news article in our database covered it (with a sentiment score of 30). All the other Chinese news articles covered positive developments in the industry, most of which receive a sentiment score of 70 or above. As a result, the average sentiment score of all housing-related Chinese news articles on July 31, 2023 is still high, despite the significance of the Country Garden issue on that day. However, despite this potential challenge, our sentiment index can still be a useful tool for comparing sentiments within a given "regime". Moreover, our analyses in subsequent sections use the sentiment index before June 2015, before the aforementioned regime switch.

But one feature is clear: since the Evergrande distress, the sentiment does remain volatile, reflecting a high degree of uncertainties about when the market will bottom out. Such uncertainties could potentially explain why the turnaround of sentiment after the July 2022 and November 2022 easing policies did not sustain, and why the recent round of housing stimulus in May 2024 did not lift up the sentiment.

# IV. AN APPLICATION TO STUDY CHINA'S MONETARY POLICY TRANSMISSION

## A. Research Design

In this section, we use the constructed housing market sentiment index to explore empirically to what extent housing market sentiment affects the transmission of monetary policy into household consumption. Our empirical strategy exploits variations across Chinese cities in their exposures to the national-level housing market sentiment as measured by the city-level Baidu search index, which is a proxy for the attention of a city's potential homebuyers and for other shocks to the city's local housing markets.[20]

To this end, we first construct "Attention-Adjusted Chinese Housing Market Sentiment Index" (AACHMSI) as the product of our nation-level CHMSI and the city-level Baidu

[20] An alternative approach is to directly construct the city-level sentiment index. However, this is not feasible due to the lack of comprehensive city-level data on news articles, given that the coverage of news articles by local news agencies is limited and uneven over time. Meanwhile, using of the national sentiment index, we could better capture the housing policy shocks, which usually occur at the national level.

search index.[21] That is:

$$AACHMSI_{c,t} = CHMSI_t \times BaiduSearchIndex_{c,t}$$

There are two reasons for this adjustment. First, while the CHMSI is available at the national level, city-level variations are necessary to conduct the local projection analysis. Second, even if all cities experience the same level of optimism, households in different cities may be "exposed" differently to that level of optimism. For instance, if the entire country is experiencing high optimism about the housing market during one quarter, but there are significantly more housing-related Baidu searches in one city (such as Hangzhou) during the same quarter, it suggests that households in that city are paying much more attention to the housing market and are therefore more "exposed" to the optimism. This city-level Baidu search index also captures city-specific housing-related shocks in a given quarter, such as rumors that a new economic development zone will be established in a city. A similar approach is used by some other studies, such as Da, Engelberg, and Gao (2015) and Gao, Ren, and Zhang (2020).

To study how housing market sentiment affects the monetary policy transmission into non-housing consumption, we classify cities according to AACHMSI across all combinations of (City, Quarter) and compare the responses of non-housing consumption to monetary shocks between households in (City, Quarter) combinations belonging to the top 20th percentile in terms of AACHMSI (being in an "optimistic *regime*") and those in (City, Quarter) combinations belonging to the bottom 20th percentile (being in a "pessimistic regime"). The difference between the average responses of non-housing consumption to monetary policy shocks for households in these two groups provides a creditable estimate of the effects of housing market sentiments on monetary policy transmission into consumption.

To account for the potential heterogeneity of household responses, we categorize households into age-education groups. Those with a college degree and above are used as a proxy for the high-income group and those with a high school diploma and below serve as a proxy for the low-income group. Young households are measured as those with household head aged 18-30, while the old households are measured as those with household head aged 30-50.[22] Since potential homebuyers differ in their income levels

[21] The Baidu searches are normalized by the city's population to mitigate the impact of mechanical effects due to population size and the associated higher internet usage.

22 Old households in this paper refer to the middle-aged households in reality. Relative to those aged 18-30 below or aged 30-50, our sample contains few observations for households above age 50 who have mortgage originations. Therefore, in this paper, we only compare the

and housing tenure status (i.e., renters versus existing homeowners), we expect the effects of housing market sentiments on monetary policy transmission into consumption to be different across age-education groups.

## B. Data and Institutional Background

Our main data source for non-durable consumption and mortgage originations is a representative household-level dataset of China, which covers all 31 provinces and municipalities. We only focus on credit card holders with mortgages, and we exclude individuals whose marital status is divorced and those over the age of 50. We aggregate the consumption data at individual-quarter level, with each individual's credit card spending within each fiscal quarter being calculated as the sum of their consumptions during the given quarter. The final data set we obtain contains 8,617 individuals, with a sample period of 2013Q3-2015Q4.[23] The summary statistics of our sample are shown in Table 10.

**Table 10. Summary Statistics**

| Variable | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| *Quarterly Consumption* | 50,649 | 17,145.011 | 41,949.407 | 0 | 274,125.00 |
| *MPS* | 50,649 | -0.002 | 0.007 | -0.017 | 0.010 |
| *AACHMSI* | 50,649 | 1.110 | 14.444 | -46.030 | 38.462 |
| *Optimistic* | 50,649 | 0.195 | 0.396 | 0 | 1 |
| *Age* | 50,649 | 33.372 | 6.803 | 18 | 50 |
| *Young* | 50,649 | 0.336 | 0.472 | 0 | 1 |
| *High-educated* | 50,649 | 0.399 | 0.490 | 0 | 1 |
| *Male* | 50,649 | 0.640 | 0.480 | 0 | 1 |
| *Married* | 50,649 | 0.685 | 0.464 | 0 | 1 |

Notes: This table reports the summary statistics of our sample. The sample period covers the third quarter of 2013 to the fourth quarter of 2015. *Quarterly Consumption* measures the non-housing expenditure by consumers in each quarter. *MPS* is defined as the actual M2 growth minus the predicted M2 growth. *AACHMSI* is the product of our nation-level CHMSI and the city-level Baidu search index. *Optimistic* is a dummy variable that equals one if the (city, quarter) for the mortgage borrower is in the top 20th percentile across all cities and quarters in terms of its AACHMSI. *Age* is the age at which a borrower enters the sample. *Young* is a dummy variable that equals one if the borrower aged 18-30, and zero otherwise. *High-educated* is a dummy variable that equals one if the borrower has a bachelor's degree or above, and zero otherwise. *Male* is a dummy variable that equals one if the borrower is a man, and zero otherwise. *Married* is a dummy variable that equals one if the borrower is married, and zero otherwise.

---

responses of consumption to monetary policy easing between two age groups, 18-30 and 30-50.

[23] Since the consumption records are only available for June 2013 in 2013Q2, we exclude this quarter from our sample.

The monetary policy shocks are proxied by a frequently-used series in the literature, defined as the actual M2 growth minus the predicted M2 growth (2000Q1–2017Q4), *a la* Chen et al. (2018). The "predicted M2 growth" is estimated using an endogenous switching M2-based monetary policy rule, accounting for the historical asymmetric relationship between M2 growth and the deviation of GDP growth from target. This estimated de facto rule aligns well with China's institutional background and has a closer fit than that of the Taylor rule for the U.S. data.[24]

We would like to provide some institutional background about China's monetary policy transmission. China's monetary policy framework is complex, with multiple objectives, targets, and instruments in place, as noted by Das and Song (2022). It is widely recognized that the transmission from M2 to mortgage rates is limited due to China's interest rate regulations and guidance. Moreover, China's monetary policy framework was largely quantity-based before 2018Q1 and more interest rate-based since then (see, for example, Amstad, Sun, and Xiong, 2020). There are also limits to market determination of mortgage rates, despite mortgage rates being fully floating. Moreover, the transmission from mortgage rates to households' non-housing consumption is present only for homeowners with unpaid mortgage balances and not for other types of households (Agarwal et al., 2022).

As for the construction of the CHMSI and the attention-adjusted CHMSI, we use a text database consisting of Chinese articles in major Chinese news outlets from the CSMAR database. We also interact it with as the city-level Baidu search index, which is provided by Baidu and used by some other studies (such as Lang et al., 2021, and Gao, Li, and Wang, 2023).

### C. Empirical Specifications

We employ the local projections *a la* Jordà et al. (2015) to estimate the effects of housing market sentiment on the response of consumption to monetary easing. This approach will trace down the dynamic effects of the monetary policy shock on the non-housing consumption for households who have bought houses in the quarter right after the shock realization. Specifically:

[24] For an alternative measure of monetary policy shocks for China, see Das and Song (2022), who construct monetary policy shocks based on daily changes in interest rate swap rates around the date of monetary policy events.

$$LnC_{i,c,\ t_i^0+k} = \alpha_0 + \beta^k MPS_{\ t_i^0-1} + \gamma^k \left(MPS_{\ t_i^0-1} \times Optimistic_{c,\,t_i^0-1}\right) + \eta^k Optimistic_{c,\,t_i^0-1}$$

$$+\xi^k \left(MPS_{\ t_i^0-1} \times Moderate_{c,\,t_i^0-1}\right) + \omega^k Moderate_{c,\,t_i^0-1}$$

$$+\alpha_1 City_c + \alpha_2 Q_{\ t_i^0+k} + \alpha_3 \left(City_c \times Q_{\ t_i^0+k}\right) + u_{i,c,\,t_i^0+k}$$

In the above specification, $t_i^0$ is the quarter in which household i purchases the house (this can be either a first-time homebuyer or a repeated homebuyer). $LnC_{i,\,t_i^0+k}$ is the natural log of real non-durable consumption of household i, $k$ quarters after the house purchase. $MPS_{t_i^0-1}$ is monetary policy shock one quarter *before* the house purchase time. The reason for using one quarter before instead of the current quarter is that following monetary easing, it takes time for households to submit mortgage application and for mortgage application to be approved by banks. $Optimistic_{c,\,t_i^0-1}$ is a dummy variable indicating whether the city of household i at the time of monetary policy shock (which is one quarter before household i purchases its house) is an "optimistic regime". Specifically, the "optimistic regime" is defined as being in the top 20th percentile across all cities and quarters in terms of its "Attention-Adjusted Chinese Housing Market Sentiment Index" (AACHMSI). $Q_{\ t_i^0+k}$ is the year-quarter fixed effect (such as the 2014Q2 and 2014Q3 dummies), capturing (for example) the time-varying business cycle factors, rather than the simple seasonality. $City_c \times Q_{\ t_i^0+k}$ is the interaction of the city and year-quarter dummies, aimed at capturing common factors that drive both the non-durable consumption and (for example) the housing sentiment in a given city during a given year-quarter. Note that the $Optimistic_{c,\,t_i^0-1}$ variable still has household-level variations because our empirical strategy examines the impact of a monetary policy shock (more precisely, the change of money supply) one quarter before each household's house purchase decision. For example, suppose Household 1 in the city of Hangzhou bought a house in 2014Q2, and Household 2 in Hangzhou bought a house in 2014Q4; and suppose (Hangzhou, 2014Q1) is in an optimistic regime, but (Hangzhou, 2014Q3) is not, then the same variable $Optimistic_{c,\,t_i^0-1}$ equals 1 for Household 1 and 0 for Household 2.

Next, we employ the following empirical specifications to estimate the effects of housing market sentiments on the response of house prices to monetary easing

$$HPI_{c,\,t+k} = \tilde{\alpha}_0 + \tilde{\gamma}^k\left(MPS_{t-1} \times Optimistic_{c,\,t-1}\right) + \tilde{\xi}^k\left(MPS_{t-1} \times Moderate_{c,\,t-1}\right)$$

$$+\tilde{\alpha}_1 City_c + \tilde{\alpha}_2 Q_{t+k} + u_{c,\,t+k}$$

In the above specification, the subscript $c$ denotes City $c$; $Optimistic_{c,\,t-1}$ and $Moderate_{c,\,t-1}$ indicate City $c$ during the one quarter before the monetary policy shock is in the aforementioned optimistic and moderate housing sentiment regimes, respectively. Note that different from the above non-housing consumption regression (which is at the household level), the house price regression is at the city level because we only have city-level house price variations. However, similar with the non-housing consumption regression, the house price regression also tracks the response to the monetary policy shock over time, as well as how the initial housing market sentiment regime (i.e., during the one quarter before the monetary policy shock) affects that response. In addition, because our house price data $Optimistic_{i,\,t_i^o-1}$ is at the (city, quarter) level, we cannot control for the interaction term $City_c \times Q_{t+k}$; and because we have controlled for city and quarter fixed effects, we do not include the $MPS_{t-1}$, $Optimistic_{c,\,t-1}$, and $Moderate_{c,\,t-1}$ as separate regressors. For example, because all cities in the same quarter share the same values of $MPS_{t-1}$, including both the $MPS_{t-1}$ variable and the city dummy will result in perfect collinearity.

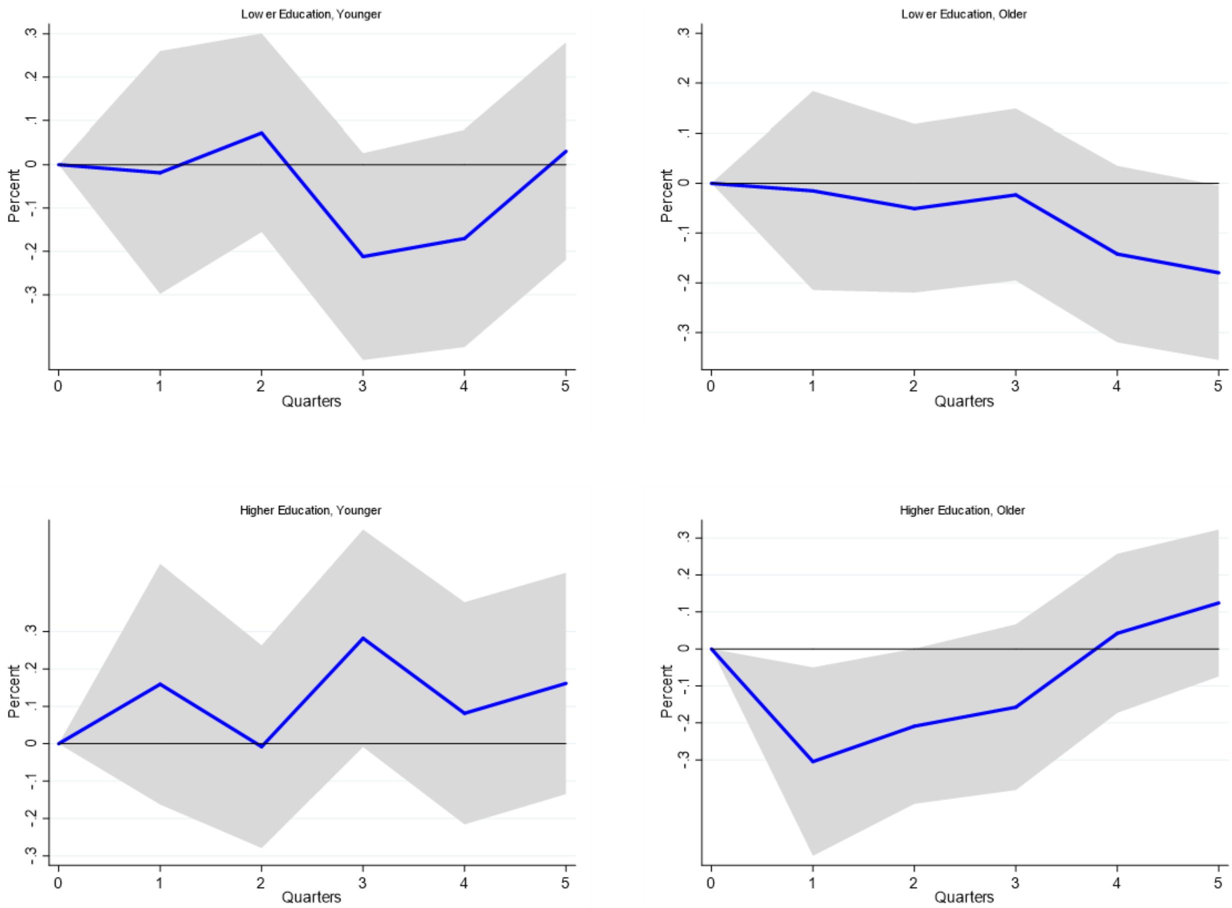
## D. Empirical Results: Optimistic Regime

We find that monetary policy transmission to consumption in China appears weaker under more optimistic sentiment about the housing market. Specifically, as shown in Figure 15, following monetary easing shocks, households' non-housing consumption increases less in cities with more optimistic housing sentiment during the house-purchase period (t=1) and one quarter after (t=2). This is particularly the case for the (High Education, Old) group (i.e., with a college degree and aged 30 to 50). Therefore, following monetary easing, aggregate non-housing consumption will not increase as much, or may even decrease if the consumption-dampening effect of these households dominates the consumption-increasing effect of other households through, for example, the standard intertemporal substitution channel (lower opportunity costs of current consumption), increased economic activities in housing-related sectors, and higher local government revenues through land sales.

One explanation for this result is the "trade-up" effect, similar to Chen et al. (2023a). Specifically, with more optimistic housing sentiment, the expected return from investing in houses increases. This would motivate existing homeowners to sell and buy larger houses (“trade-up”), resulting in higher down payments, higher future debt-servicing costs, and lower non-housing consumption. Households of high education tend to have higher incomes, and “old” households tend to have lower balances of existing mortgages (since they have serviced their existing mortgages longer than “young” existing homeowners). These would make the trade-up more *feasible* for the (High Education, Old) existing homeowners than for other types of existing homeowners. And the lower balances of existing mortgages mean that the (High Education, Old) existing homeowners will need to make a lower amount of upfront payment (note that existing homeowners need to pay off the legacy mortgages before switching to a larger house), making the trade-up more *desirable* for them than for other types of existing homeowners.

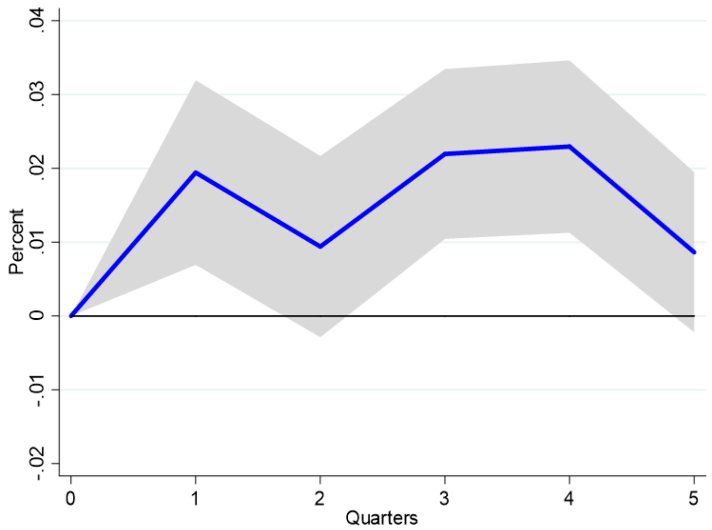
We then examine the role of housing sentiment for the monetary policy transmission into house prices, which is key for the above-mentioned crowding-out channel to work. We conjecture that house prices will increase more for more optimistic cities in response to monetary policy easing. Our empirical results are consistent with this conjecture. Indeed, following monetary easing shocks, house prices increase in cities under an optimistic housing market sentiment; moreover, this effect is larger and more persistent than that under a pessimistic sentiment (Figure 16). Recall that the wealth effect of house price increases on consumption is muted in China (e.g., because Chinese homeowners cannot take home equity loans), further limiting the increase in aggregate consumption following monetary easing.

**Figure 15. Consumption Responses to Monetary Easing under An Optimistic Regime**




Notes: This figure plots the difference of average responses of non-housing consumption to monetary policy shocks for mortgage borrowers in the top 20th percentile of AACHMSI as compared with those in the bottom 20th percentile of AACHMSI (for brevity, we refer to this as "Responses… under An Optimistic Regime" in the title of the figure; the same comment applies to similar figures). The top left panel refers to households with a high-school diploma and below and age 18-30. The total right panel refers to individuals with a high-school diploma and below and age 30-50. The bottom left panel refers to individuals with a college degree and above and age 18-30 and the bottom right panel refers to individuals with a college degree and above and age 30-50. The shaded areas are the 90 percent confidence intervals.

**Figure 16. House Price Response to Monetary Easing under An Optimistic Regime**




Notes: This figure plots the difference of average responses of non-housing consumption to monetary policy shocks for higher educated older mortgage borrowers residing in cities in the top 20th percentile of AACHMSI as compared with those residing in cities in the bottom 20th percentile of AACHMSI. The shaded areas are the 90 percent confidence intervals.

## E. Empirical Results: Pessimistic Regime

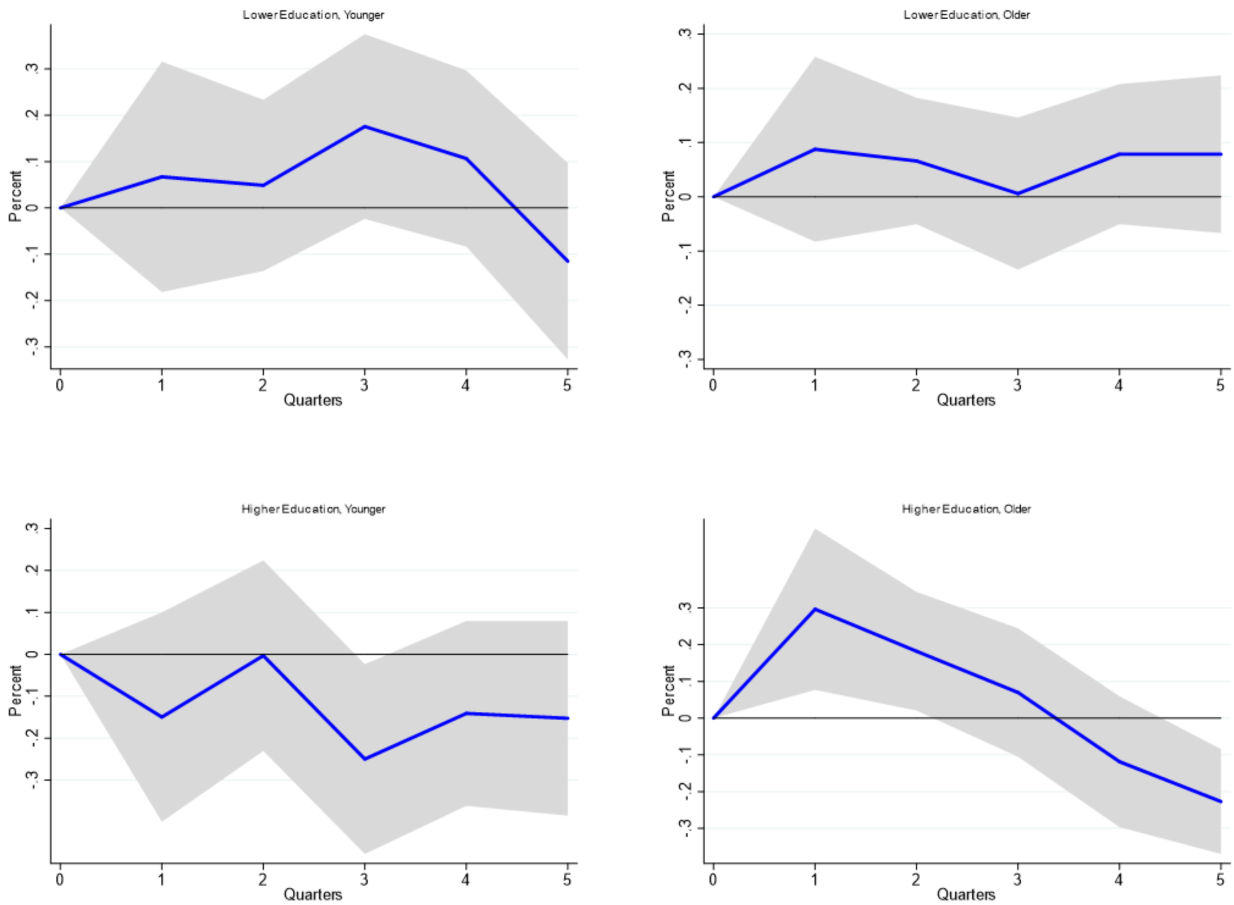
We now conduct a closely related empirical test to examine how the non-housing consumption and house prices respond under a pessimistic regime. Doing so can directly shed light on the monetary policy effectiveness in China at the current conjuncture.

The empirical strategy is similar to that for the optimistic regime. To implement it, we replace the $Optimistic_{i,\, t_i^0-1}$ dummy by the $Pessiimistic_{i,\, t_i^0-1}$ dummy, which equals 1 if the city of Household $i$ at the time of monetary policy shock has an Attention-Adjusted Chinese Housing Market Sentiment Index that lies in the bottom 20th percentile across cities and quarters.

The results are shown in Figures 17-18, which suggest that the crowding-out effects of housing on non-housing consumption under an optimistic regime are *no longer* present under a pessimistic regime. Specifically, as shown in Figure 17, following monetary easing shocks and for three of the four household groups, the responses in households' non-housing consumption in cities with pessimistic housing sentiment do not differ significantly from those in cities with non-pessimistic sentiment, either during the house-purchase period (t=1) or in subsequent quarters. Similarly, the house price response to monetary easing under the pessimistic regime does not differ significantly from the non-pessimistic regime (in the house-purchase period) and is more *negative* with pessimistic housing sentiment during all three subsequent quarters after the house purchase (Figure 18).

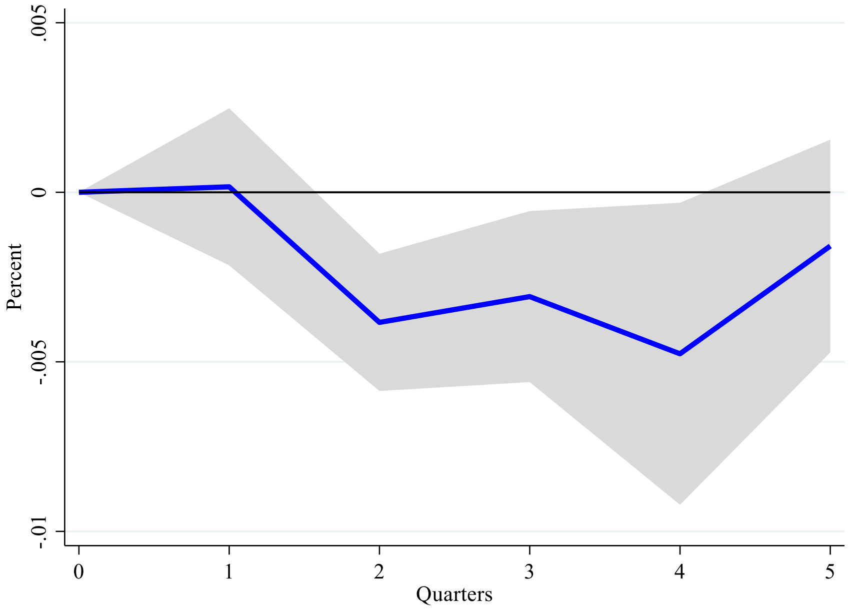
In fact, for the (High Education, Old) group (i.e., with a college degree and aged 30 to 50), the non-housing consumption response to monetary easing is more positive with pessimistic housing sentiment than with non-pessimistic sentiment (Figure 17). These findings are consistent with the trade-up channel explained above: with "pessimistic" housing sentiment (such as the emphasis on housing as living and not as speculation), the expected return from investing in houses decreases; this lowers existing homeowners' incentives to trade up for larger houses , and the associated savings from the lower down payments (as the new houses are smaller) and future debt servicing costs would boost non-housing consumption.

**Figure 17. Consumption Responses to Monetary Easing under A Pessimistic Regime**



Notes: This figure plots the difference of average responses of non-housing consumption to monetary policy shocks for higher educated older mortgage borrowers residing in cities in the bottom 20th percentile of AACHMSI as compared with those residing in cities in the top 20th percentile of AACHMSI. The shaded areas are the 90 percent confidence intervals.

**Figure 18. House Price Response to Monetary Easing under A Pessimistic Regime**




Notes: This figure plots the difference of average responses of non-housing consumption to monetary policy shocks for higher educated older mortgage borrowers residing in cities in the bottom 20th percentile of AACHMSI as compared with those residing in cities in the top 20th percentile of AACHMSI. The shaded areas are the 90 percent confidence intervals.

### F. Empirical Results: Accounting for Potential Endogeneity

One may argue that the construction of the attention-adjusted CHMSI has a potential issue of endogeneity or double counting: the city-level search intensity (attention) may have already reflected the nation-level sentiment, and thus including both the nation-level $CHMSI_t$ and the city-level $BaiduSearchIndex_{c,t}$ may have double counted the impact of the nation-level sentiment.

To account for this, we adopt the following two-step procedure. In Step 1, we run the following regression at the daily level to isolate the impact of the nation-level sentiment on the city-level search intensity (attention):

$$BaiduSearchIndex_{c,t} = \gamma_0 + \gamma_1 CHMSI_t + v_{c,t}$$

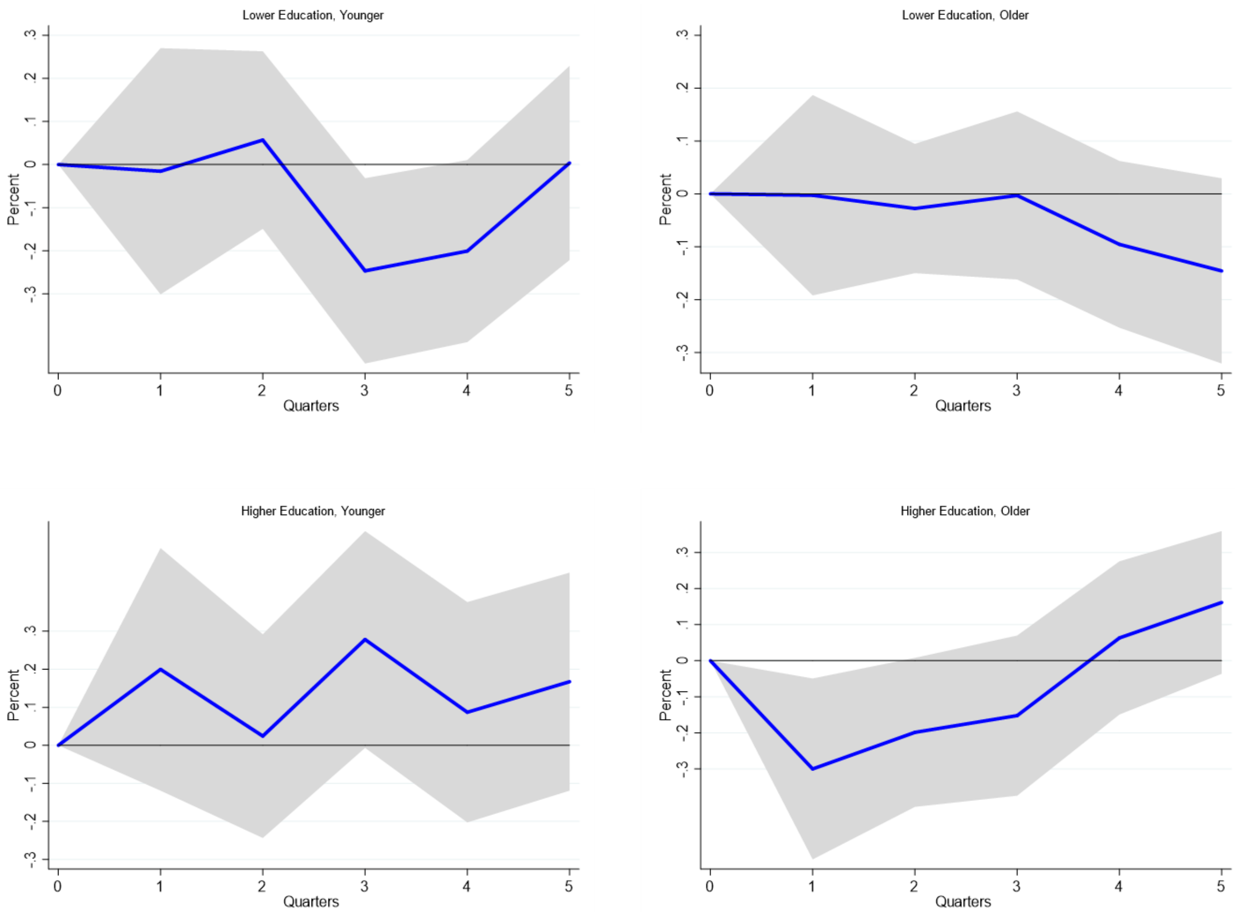
And in Step 2, we use the residual from the above regression to construct the attention-adjusted CHMSI (AACHMSI) at the daily level:

$$AACHMSI_{c,t} = CHMSI_t \times \hat{v}_{c,t}$$

We then aggregate the daily AACHMSI at the quarterly level, rank the quarterly AACHMSI indices across cities and quarters to define the optimistic, moderate, and pessimistic sentiment regimes accordingly, and conduct local projections for non-housing consumption as before.

After correcting for the potential double counting issue, our main findings still hold. Specifically, as shown in Figure 19, following monetary easing shocks, households' non-housing consumption increases less in cities with more optimistic housing sentiment during the house-purchase period (t=1) and one quarter after (t=2). This is particularly the case for the (High Education, Old) group (i.e., with a college degree and aged 30 to 50). Hence, as in the case without considering the potential double counting issue, aggregate non-housing consumption following monetary easing will not increase as much, or may even decrease if the consumption-dampening effect of these households dominates the standard consumption-increasing effect of other households.

**Figure 19. Correcting for Double Counting: Consumption Responses to Monetary Easing under An Optimistic Regime**



Notes: This figure plots the difference of average responses of non-housing consumption to monetary policy shocks for mortgage borrowers in the top 20th percentile of AACHMSI as compared with those in the bottom 20th percentile of AACHMSI, after correcting for the potential double counting issue. The top left panel refers to households with a high-school diploma and below and age 18-30. The total right panel refers to individuals with a high-school diploma and below and age 30-50. The bottom left panel refers to individuals with a college degree and above and age 18-30 and the bottom right panel refers to individuals with a college degree and above and age 30-50. The shaded areas are the 90 percent confidence intervals.

### G. Robustness Checks

We conduct several robustness checks to validate our findings. First, we employ a 25th-percentile threshold to define the optimistic regime, allowing for a broader set of cities to be classified as optimistic based on the attention-adjusted CHMSI. The results, presented in Appendix Figures 1-2, are consistent with those obtained under a 20th-percentile threshold. Specifically, following monetary easing shocks, non-housing consumption increases are less pronounced in cities exhibiting higher optimistic housing sentiment during the house-purchase quarter and the subsequent quarter, especially for the (High Education, Old) demographic group. Intuitively, because the 25th-percentile threshold encompasses a broader set of cities, the sentiment difference between cities within the optimistic regime and those outside it is reduced, leading to a smaller difference in consumption responses—though still statistically significant—compared with the 20th-percentile scenario. Moreover, following monetary easing shocks, house prices rise more substantially in cities characterized by an optimistic housing market than in those with a pessimistic market.

Second, we employ a 15th-percentile threshold to define the optimistic regime. The results, detailed in Appendix Figures 3-4, also align with those obtained under the 20th-percentile threshold. Additionally, because the 15th-percentile threshold includes a smaller set of cities, the sentiment contrast between cities within the optimistic regime and others is magnified, consequently enlarging the difference in consumption responses relative to the 25th-percentile case.

Third, we utilize household-level observations (alongside a 20th-percentile threshold) to define the optimistic regime. Here, we rank sentiment index values at the household rather than the city level while defining the optimistic regime. This is to account for the potential variability in the number of households across cities. Although this imbalance issue is not severe in our dataset, we nonetheless perform this robustness check. The results for non-housing consumption responses, as shown in Appendix Figure 5, mirror those obtained when defining the optimistic regime at the city level. Note that the house price responses remain unchanged when we define the optimistic regime using the household-level observations, hence the exclusion of the house price figure from this specific robustness check.

Fourth, while considering the potential double counting issue around the Baidu search index, we further allow different cities to be affected differently by the nation-level sentiment. That is, in the Step-1 regression, we include the interaction term between the city dummy and the nation-level CHMSI index:

$$BaiduSearchIndex_{c,t} = \gamma_0 + \gamma_1 CHMSI_t + \gamma_2 City_c \times CHMSI_t + v_{c,t}$$

We then use the residual from this regression to construct the attention-adjusted CHMSI and proceed accordingly as in the main empirical section (that is, we use the top 20th percentile and the city-level observations to define the optimistic regime). The results for non-housing consumption responses, as shown in Appendix Figure 6, once again confirm those obtained without including the interaction term $City_c \times CHMSI_t$.[25]

## H. Discussions

The preceding section shows that for high-educated mortgage borrowers aged 30-50, higher housing market sentiments dampen the response of non-housing consumption to monetary policy shocks, while for other age-education groups, the effects of housing market sentiments are insignificant. In this section, we provide a plausible theory to explain this pattern.

The literature emphasizes two effects of monetary policy easing on consumption. The first effect is the substitution effects, under which a decrease in interest rate reduces the incentive for savings and increases non-housing consumption. The second effect is the wealth effect, under which a decrease in policy rate would increase house prices, and thus the net worth of existing homeowners. Under both effects, a decrease in interest rates increases non-housing consumption.

In this paper, we propose another channel for monetary policy easing to affect household consumption, especially for those who trade up their existing homes after monetary easing. We call this channel the *crowding-out* channel, in which monetary easing and optimistic housing sentiment induce existing homeowners to trade up their houses , reducing their non-housing consumption. The intuition is that, following

[25] When correcting for the potential double counting issue, we have conducted other robustness checks, such as using different percentiles and/or household-level observations to define the housing sentiment regimes. The results are not reported due to space constraint, but they are available upon request.

monetary easing, particularly if housing sentiment is favorable, house prices tend to rise, encouraging more existing homeowners to trade up their primary homes for further capital gains, thereby reducing their non-housing consumption, all else being equal.

We rationalize this mechanism in a two-period economy with housing purchase. A formal description of the model is presented in the Online Appendix. In our model, households differ in initial housing and mortgage positions. A renter has no initial housing stock and no outstanding mortgage debt. In contrast, an existing homeowner, who is a potential buyer of a larger primary house, has some initial stock of housing and outstanding mortgage at the beginning of period 1. There is no saving technology apart from housing. The only borrowing allowed is mortgage, which is subject to a maximum loan-to-value ratio. The mortgage is long-term so that if an existing homeowner keeps his existing home, he can pay the mortgage balance at the end of period 2. However, if a homeowner sells his existing home in period 1, he needs to pay off the outstanding mortgage in full.

Accordingly, current renters decide whether to buy or rent a house. Buying a house provides more housing utility services than renting the house of the same size but incurs mortgage interest payment in period 2. For renters who decide to purchase houses (i.e., for first-time homebuyers), their period-1 consumption still follows the conventional wisdom, that is, an easing of monetary policy, by reducing the mortgage interest rate, increases consumption by those first-time home buyers.

An existing homeowner chooses between staying or trading up their current home. Trading up occurs once two conditions are met. First, it is feasible to trade up, which is satisfied when the homeowner's period-1 income is sufficiently high to ensure a non-negative consumption. Second, it is optimal to trade up, which can be met when the homeowner is above a certain age (so that he has been making mortgage payments for a long time) and thus the outstanding mortgage balance is sufficiently low.

Accordingly, a reduction in interest rates following monetary easing increases the probability for existing homeowners to trade up their homes via two effects. The first one is to directly reduce future mortgage interest payments. The other is an indirect effect, where monetary easing increases housing prices and thus the net capital gain of trading up to a larger house and selling it afterwards. Importantly, an optimistic housing sentiment and the associated higher expected return for housing purchase makes it more likely for existing homeowners to trade up their homes in response to a marginal reduction in interest rates. Since households who trade up their homes reduce current consumption for intertemporal substitution, this implies that monetary policy easing would reduce the *average* consumption of existing homeowners.

Our sample contains four age-education groups: young (old) and high (low) educated households who originate mortgage. Note that within each age-education group, there are two types of households in our sample, renters who become first-time homebuyers and existing homeowners who trade up their houses. An existing homeowner who trades up his home reduces his consumption, while this is not the case for a first-time homebuyer. Hence, the response of average consumption for each age-education group to monetary policy easing depends crucially on whether a reduction in interest rates increases the share of existing homeowners who trade up in the total mortgage borrowers, which in turns depends on the extent to which monetary policy easing increases the probability of trading up.

Compared with other age-education groups, old high-educated homeowners tend to have higher initial incomes and lower outstanding mortgage balances. Thus, upon a reduction in interest rates, the probability of trading up tends to increase more for old highly educated homeowners. Therefore, according to our theory, for old and high-educated households, the crowding out effect of monetary easing tends to dominate the other two effects that crowd in non-housing consumption. This is especially the case when households have higher expected return for housing purchase, which we proxy using housing sentiment index. In contrast, the insignificance of the effects of housing market sentiments on the monetary policy transmission for other age-education groups suggests that the crowding-out effect is largely offset by the other two channels, as households in the other age-education groups tend to have lower income or larger outstanding mortgage.

## V. CONCLUSION AND POLICY IMPLICATIONS

This paper presents a novel approach to constructing a CHMSI using generative AI models, specifically GPT-4o, to analyze news articles from the CSMAR database. By incorporating Baidu search data normalized by population, we refine this index to create an Attention-Adjusted Chinese Housing Market Sentiment Index (AACHMSI) at the city level. Our methodology leverages the advanced capabilities of GPT-4o for nuanced sentiment analysis, significantly outperforming traditional models (such as the keyword-based model) and many Chinese LLMs (such as the Hunyuan and Senta models).

Moreover, this paper outlines eight well-defined principles for prompt engineering, along with practical advice for researchers. Prompt engineering has become an essential method for enhancing the capabilities of LLMs. These principles optimize the advantages of rapidly evolving generative AI models while mitigating common issues. They not only improve the development of the CHMSI but also have broader applications for researchers working with generative AI across various domains.

Our analysis reveals that monetary easing has a muted effect on households' non-housing consumption in China, particularly in cities with higher housing market optimism. This phenomenon can be explained through the "crowding-out channel," where a higher probability of trading up their homes by existing homeowners, driven by optimism, crowds out non-housing consumption. These findings underscore the limitations of monetary policy in stimulating aggregate consumption in the context of an optimistic housing market and highlight the necessity of complementary structural reforms.

Our paper highlights two significant implications for the design of the macroeconomic policy package in China, which may also be applicable to other countries facing similar circumstances:

First, monetary easing may be more effective at boosting household consumption when housing exuberance sentiment is contained. Due to a shift in housing sentiment resulting from China's ongoing housing slump, the house price channel—and consequently, the crowding out of non-housing consumption—might be weaker than in previous periods. Even though our results are obtained using the M2-based historical monetary policy shocks, they could still suggest that monetary easing implemented via lower policy rates may now be more potent in enhancing consumption relative to the past, given that the crowding out channel of housing on non-housing consumption highlighted in our paper still applies to the monetary easing implemented through rate cuts.

Second, it is imperative to strengthen the monetary policy framework and deepen structural reforms. This includes phasing out deposit/lending rate guidance, as recommended in IMF (2024b). Furthermore, structural reforms designed to weaken the house price channel—such as enhancing social safety nets, providing alternative investment opportunities, and promoting financial market development—are crucial to mitigate the crowding out of non-housing consumption and thus further enhance the effectiveness of monetary policy stimulus.

For future research, there are several avenues to explore. For example, applying our methodology to other countries with significant housing market dynamics could provide comparative insights and enhance the generalizability of our findings. In addition, integrating additional data sources, such as social media and other online platforms, could further refine the sentiment index and provide a more comprehensive view of housing market sentiment.

## REFERENCES


Agarwal, S., Y. Deng, Q. Gu, J. He, W. Qian, and Y. Ren. 2022. Mortgage debt, hand-to-mouth households, and monetary policy transmission. *Review of Finance* 26: 487-520.

Aastveit, K. A., and A. Anundsen. 2022. Asymmetric effects of monetary policy in regional housing markets. American Economic Journal: Macroeconomics 14 (4): 499–529.

Agarwal, S., C. Liu, and N. S. Souleles. 2007. The reaction of consumer spending and debt to tax rebates—evidence from consumer credit data. *Journal of Political Economy* 115: 986-1019.

Agarwal, S., and W. Qian. 2014. Consumption and debt response to unanticipated income shocks: Evidence from a natural experiment in Singapore. *American Economic Review* 104: 4205-30.

Albuquerque, B., M. Iseringhausen, and F. Opitz. 2024. The housing supply channel of monetary policy. IMF Working Paper 2024/23.

Alfano, V., and M. Guarino. 2022. A word to the wise analyzing the impact of textual strategies in determining house pricing. *Journal of Housing Research* 31: 88-112.

Amstad, M., G. Sun, and W. Xiong. 2020. *The Handbook of China's Financial System.* Princeton University Press.

Baker, S. R., L. Kueng, S. Meyer, and M. Pagel. 2022. Consumption imputation errors in administrative data. *Review of Financial Studies* 35: 3021-59.

Barbaglia, L., S. Manzan, and E. Tosetti. 2023. Forecasting loan default in Europe with machine learning. *Journal of Financial Econometrics* 21: 569-96.

Barkan, O., J. Benchimol, I. Caspi, E. Cohen, A. Hammer, and N. Koenigstein. 2023. Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting* 39: 1145–1162.

Bartik, A., A. Gupta, and D. Milo. 2023. The costs of housing regulation: Evidence from generative regulatory measurement. *Available at SSRN.*

Bayoumi, T., and Y. Zhao. 2020. Incomplete financial markets and the booming housing sector in China. IMF Working Paper 2020/265.

Beraldi, F., and Y. Zhao. 2023. *The pricing-out phenomenon in the U.S. housing market*. IMF Working Paper 2023/1.

Biljanovska, N., C. Fu, and D.O. Igan. 2023. Housing affordability: A new dataset. IMF Working Paper 23/247.

Bsharat, S. M., A. Myrzakhan, and Z. Shen. 2023. Principled instructions are all you need for questioning llama-1/2, gpt-3.5/4. *arXiv preprint arXiv:2312.16171*.

Campello, M., R. Connolly, G. Kankanhalli, and E. Steiner. 2022. Do real estate values boost corporate borrowing? Evidence from contract-level data. *Journal of Financial Economics* 144 (2): 611-644.

Cao, J., C. Cui, V. Dinger, M. B. Holm, and S. Kang. 2024. Identifying the depreciation rate of durables from marginal spending responses. *Journal of Money, Credit and Banking*.

Cepni, O. 2024. Fifty Shades of the US States: News Media Coverage and Predictability of House Prices. *Available at SSRN 4712829*.

Cepni, O., and N. Khorunzhina. 2023. Geography of Housing Sentiment over Business Cycles. *Available at SSRN 4350783*.

Chang, K., S. Xu, C. Wang, Y. Luo, T. Xiao, and J. Zhu. 2024. Efficient prompting methods for large language models: A survey. arXiv:2404.01077.

Chen, K., J. Ren, and T. Zha. 2018. The nexus of monetary policy and shadow banking in China. *American Economic Review* 108: 3891-936.

Chen, K., Q. Wang, T. Xu, and T. Zha. 2023a. Aggregate and distributional impacts of LTV policy: Evidence from China's micro data.

Chen, B., Z. Zhang, N. Langrené, and S. Zhu. 2023b. Unleashing the potential of prompt engineering in Large Language Models: A comprehensive review. *arXiv preprint arXiv:2310.14735*.

Cloyne, J., C. Ferreira, and P. Surico. 2020. Monetary policy when households have debt: new evidence on the transmission mechanism. *Review of Economic Studies* 87: 102-29.

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

Das, S., and W. Song. 2022. Monetary policy transmission and policy coordination in China. IMF Working Paper 2022/74.

Di Maggio, M., A. Kermani, B. J. Keys, T. Piskorski, R. Ramcharan, A. Seru, and V. Yao. 2017. Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging. *American Economic Review* 107: 3550-88.

Dong, Z., E. C. Hui, and D. Yi. 2021. Housing market sentiment and homeownership. *Journal of Housing and the Built Environment* 36: 29-46.

Favilukis, J., S. C. Ludvigson, and S. Van Nieuwerburgh. 2017. The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium. *Journal of Political Economy* 125: 140-223.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2022. Predictably unequal? The effects of machine learning on credit markets. *Journal of Finance* 77: 5-47.

Gao, Y., Y. Li, and Y. Wang. 2023. The dynamic interaction between investor attention and green security market: an empirical study based on Baidu index. *China Finance Review International* 13: 79-101.

Gao, Z., H. Ren, and B. Zhang. 2020. Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis* 55: 549-80.

Gerardi, K., P. Willen, and D. Zhang. 2023. Mortgage prepayment, race, and monetary policy. *Journal of Financial Economics* 147(3): 498-524.

Holm, M. B., P. Paul, and A. Tischbirek. 2021. The transmission of monetary policy under the microscope. *Journal of Political Economy* 129: 2861-904.

IMF. 2021. How to assess country risk: The vulnerability exercise approach using machine learning. Technical Notes and Manuals No. 2021/003.

Jarocinski, M., and F. Smets. 2008. House prices and the stance of monetary policy.

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. Manager sentiment and stock returns. *Journal of Financial Economics* 132: 126–149.

Jordà, Ò., M. Schularick, and A. M. Taylor. 2015. Betting the house. *Journal of International Economics* 96: S2-S18.

Joseph, A., G. Potjagailo, C. Chakraborty, and G. Kapetanios. 2024. Forecasting UK inflation bottom up. *International Journal of Forecasting* 40 (4): 1521-1538.

Korinek, A. 2023. Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature* 61: 1281-317.

Korinek, A. 2024. LLMs level up—better, faster, cheaper: June 2024 update to section 3 of ‘generative AI for economic research: Use cases and implications for economists,’ published in the Journal of Economic Literature 61 (4).

Lang, Q., J. Wang, F. Ma, D. Huang, and M. W. M. Ismail. 2021. Is Baidu index really powerful to predict the Chinese stock market volatility? New evidence from the internet information. *China Finance Review International* 13: 263-84.

Liu, Y., D. Yang, and Y. Zhao. 2022. *Housing Boom and Headline Inflation: Insights from Machine Learning*. IMF Working Paper 2022/151.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. Journal of Finance 66: 35-65.

Medeiros, M. C., G. F. Vasconcelos, Á. Veiga, and E. Zilberman. 2021. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics* 39: 98-119.

Mian, A. and A. Sufi. 2022. Credit Supply and Housing Speculation, *Review of Financial Studies*, 35, 680-719.

Mian, A., L. Straub, and A. Sufi. 2021. Indebted demand. *Quarterly Journal of Economics* 136: 2243-307.

Painter, G., X. Yang, and N. Zhong. 2022. Housing wealth as precautionary saving: Evidence from urban China. *Journal of Financial and Quantitative Analysis* 57: 761-89.

Paranhos, L. 2023. Predicting inflation with Recurrent Neural Networks. Bank of England Working Paper.

Sadhwani, A., K. Giesecke, and J. Sirignano. 2021. Deep learning for mortgage risk. *Journal of Financial Econometrics* 19: 313-68.

Sahoo, P., A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha. 2024. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.

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

Wei, J., X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, and D. Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35: 24824-37.

Williams, J. C. 2016. Measuring the effects of monetary policy on house prices and the economy. *BIS paper*.

Zhang, Z., A. Zhang, M. Li, and A. Smola. 2022. Automatic chain of thought prompting in large language models. *arXiv preprint arXiv:2210.03493*.

Zhao, Y. 2020. U.S. Housing market during COVID-19: Aggregate and distributional evidence. *COVID Economics* 50: 113-154.

Zhou, Z. 2018. Housing market sentiment and intervention effectiveness: Evidence from China. *Emerging Markets Review* 35: 91-110.

# APPENDICES

## Appendix 1. The "Basic Prompt" Used in the GPT-4o Model

For the following article, do not translate. Just detect the sentiment on the Chinese housing market, with a sentiment score of 0 being most pessimistic and 100 being most optimistic. Explain your sentiment analysis methodology in detail and provide your rationale. While doing the sentiment analysis, set your temperature value to 0.

## Appendix 2. The "Intermediate Prompt" and "Advanced Prompt" Used in the GPT-4o Model

These prompts implement most of the eight principles for effective prompt engineering. Note that the red color denotes the additions from the "Basic Prompt" to the intermediate and advanced prompts. In addition, the differences between the intermediate prompt and the advanced prompt are the last two instructions, which implement the "Chain-of-Thought" and the "tip and penalize" principles (marked in green).

Act as an expert in sentiment analysis, macroeconomics, housing markets, especially in the Chinese housing markets. Apply the advanced NLP capacities of your own GPT-4 engine to do sentiment analysis in this chatbot, without using external NLP tools. For the following article, do not translate. Just detect the sentiment on the Chinese housing market, with a sentiment score of 0 being most pessimistic and 100 being most optimistic. Explain your sentiment analysis methodology in detail and provide your rationale. While doing the sentiment analysis, set your temperature value to 0.

Don't use any external library like NLTK; don't process the file by any NLP technique, library, or basket of keywords. Just read and detect the sentiment by yourself (i.e., GPT-4), so apply the complex and advanced capabilities associated with GPT-4 to deeply understand the context, the inferences, etc. and then distill the sentiment.

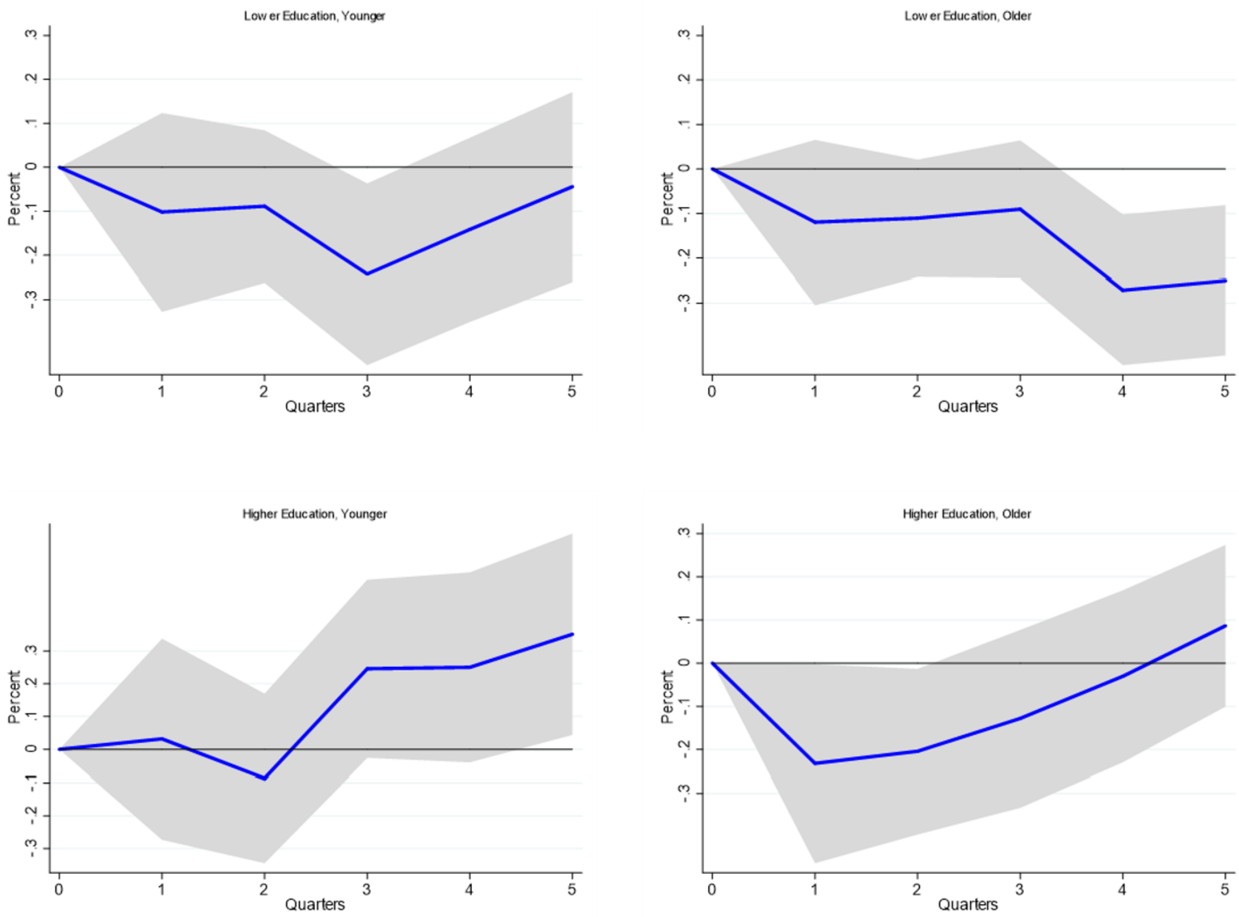
More requirements are:

1. Do NOT just look at keywords; instead, read through each article as a whole, understand the **context**, and then give your overall sentiment score.
2. If the article conveys that **regulators** will tighten their regulation on the housing market, for example, to reduce the overly-high profit margin of real estate developers, then it might decrease the sentiment because it might mean that the

regulators will want to lower the housing price? That's just one possibility, but please make your own assessment based on your full capabilities to account for the context, nuances, logical inferences, etc., and focusing on the (direct and implied) sentiment of market participants towards the housing market in the Mainland of China.

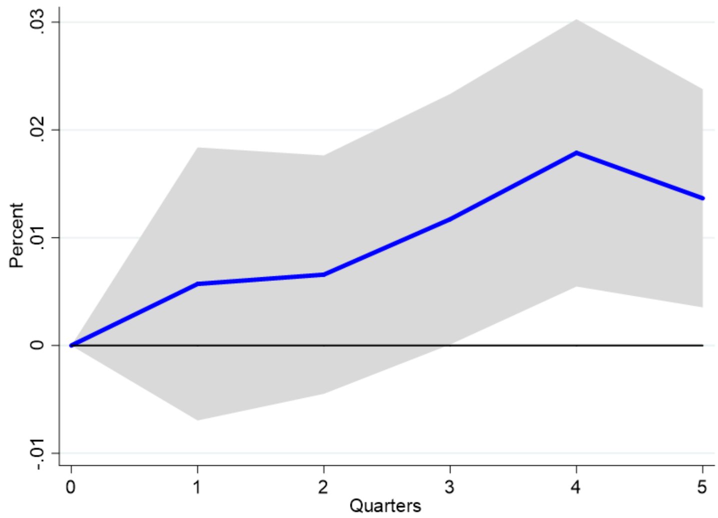
3. Please apply the **logics** very carefully. For example, if the land price is very high, then because it's part of the house price, it would indicate that house price would be high, meaning it's quite optimistic (all else being equal).
4. Explain your sentiment analysis methodology in detail and write down the Python code that I could use to replicate your results through the API, involving something like this: response = client.chat.completions.create(model="gpt-4o"…); while doing so, put all Python codes in one cell.
5. Please generate more **variations** in your sentiment scores, and do NOT make them too close to 50. Instead, make the range of the sentiment score to be at least 25 (e.g., from 40 to 65). But while doing so, please stay truthful to the original content. I just want you to apply your advanced capabilities to process and understand the nuances, contexts, logics, inferences, etc. and detect more variations in the sentiments.
6. Take your time to do the task properly.
7. Let's think not just step by step, but also one by one.
8. I will tip you US$ 1000 if you do the job well but will fine you US$ 1000 if you don't.

## Appendix Figure 1. Consumption Responses to Monetary Easing under An Optimistic Regime with A 25-Percentile Threshold



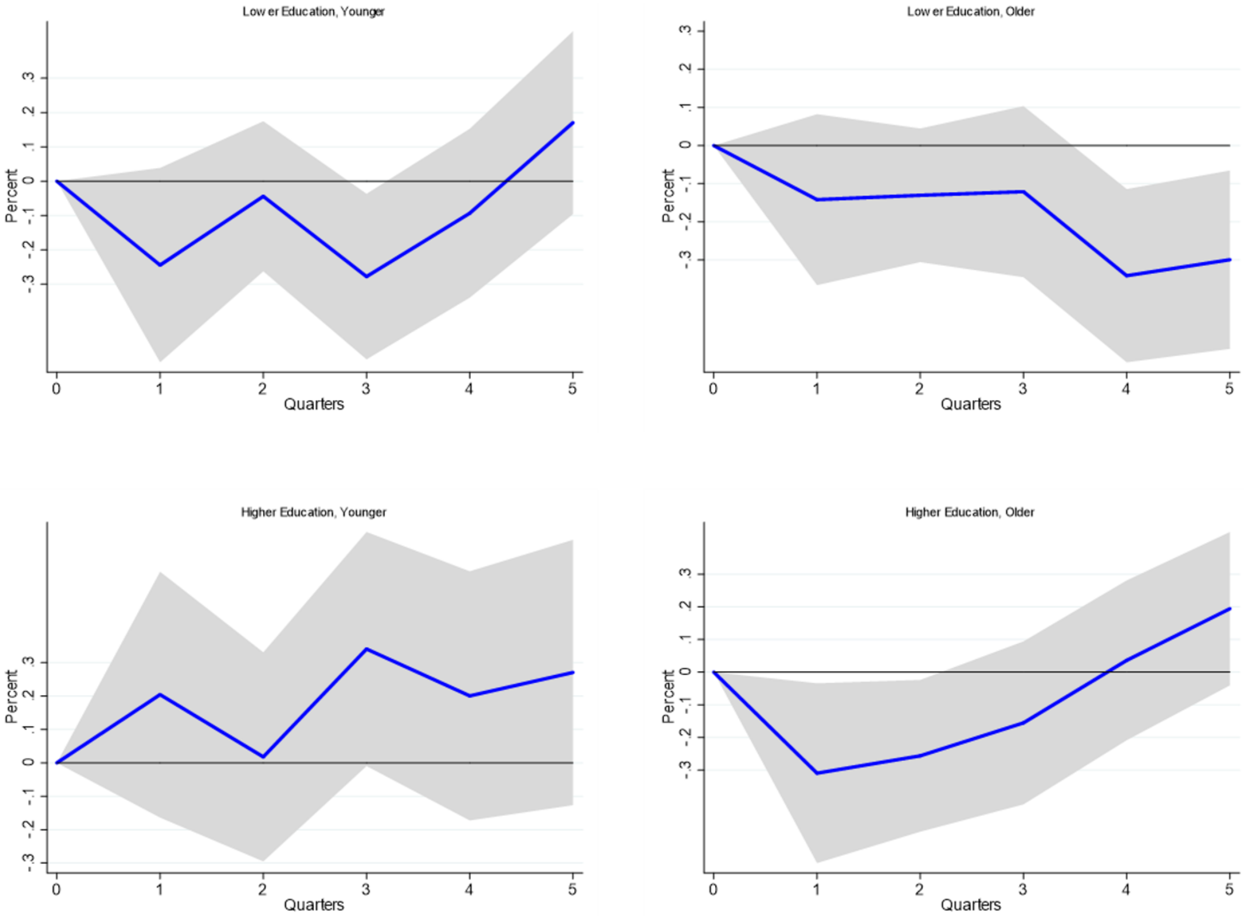

Notes: This figure plots the difference of average responses of non-housing consumption to monetary policy shocks for mortgage borrowers in the top 25th percentile of AACHMSI as compared with those in the bottom 25th percentile of AACHMSI. The top left panel refers to households with a high-school diploma and below and age 18-30. The total right panel refers to individuals with a high-school diploma and below and age 30-50. The bottom left panel refers to individuals with a college degree and above and age 18-30 and the bottom right panel refers to individuals with a college degree and above and age 30-50. The shaded areas are the 90 percent confidence intervals.

## Appendix Figure 2. House Price Response to Monetary Easing under An Optimistic Regime with A 25-Percentile Threshold



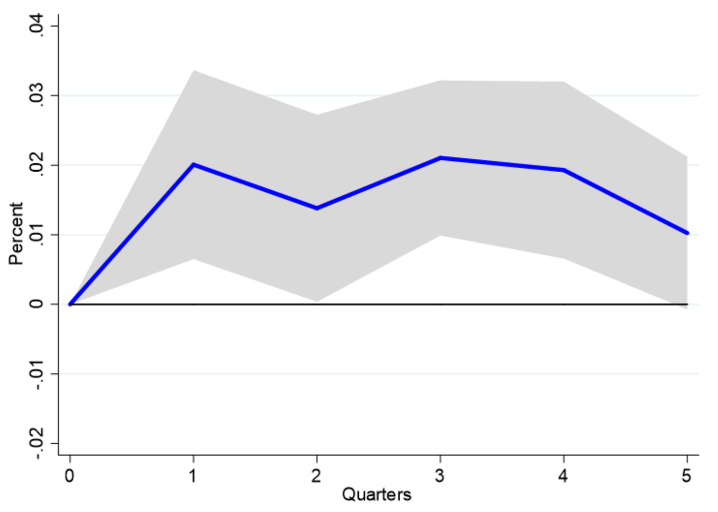

Notes: This figure plots the difference of average responses of non-housing consumption to monetary policy shocks for higher educated older mortgage borrowers residing in cities in the top 25th percentile of AACHMSI as compared with those residing in cities in the bottom 25th percentile of AACHMSI. The shaded areas are the 90 percent confidence intervals.

## Appendix Figure 3. Consumption Responses to Monetary Easing under An Optimistic Regime with A 15-Percentile Threshold



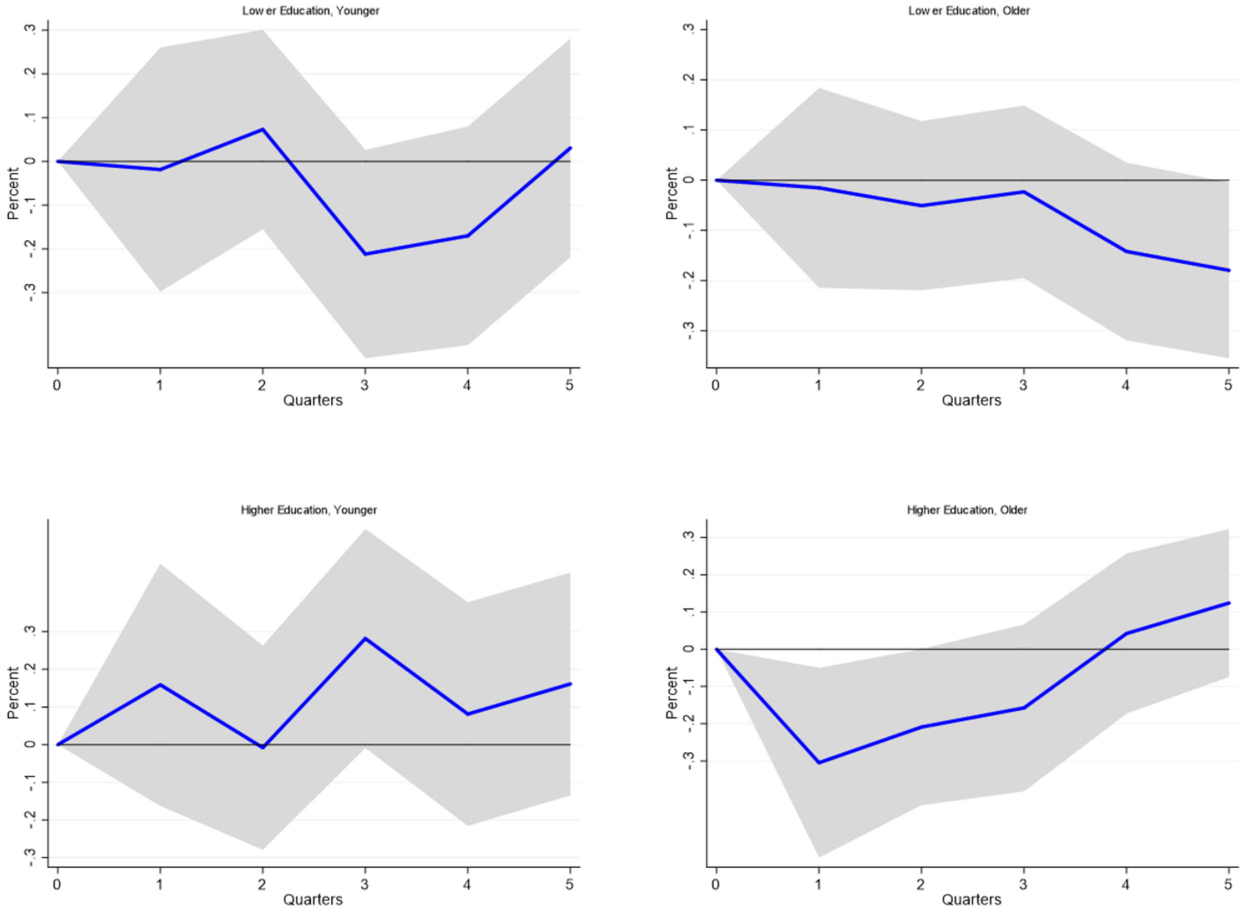

Notes: The shaded areas are the 90 percent confidence intervals.

## Appendix Figure 4. House Price Response to Monetary Easing under An Optimistic Regime with A 15-Percentile Threshold



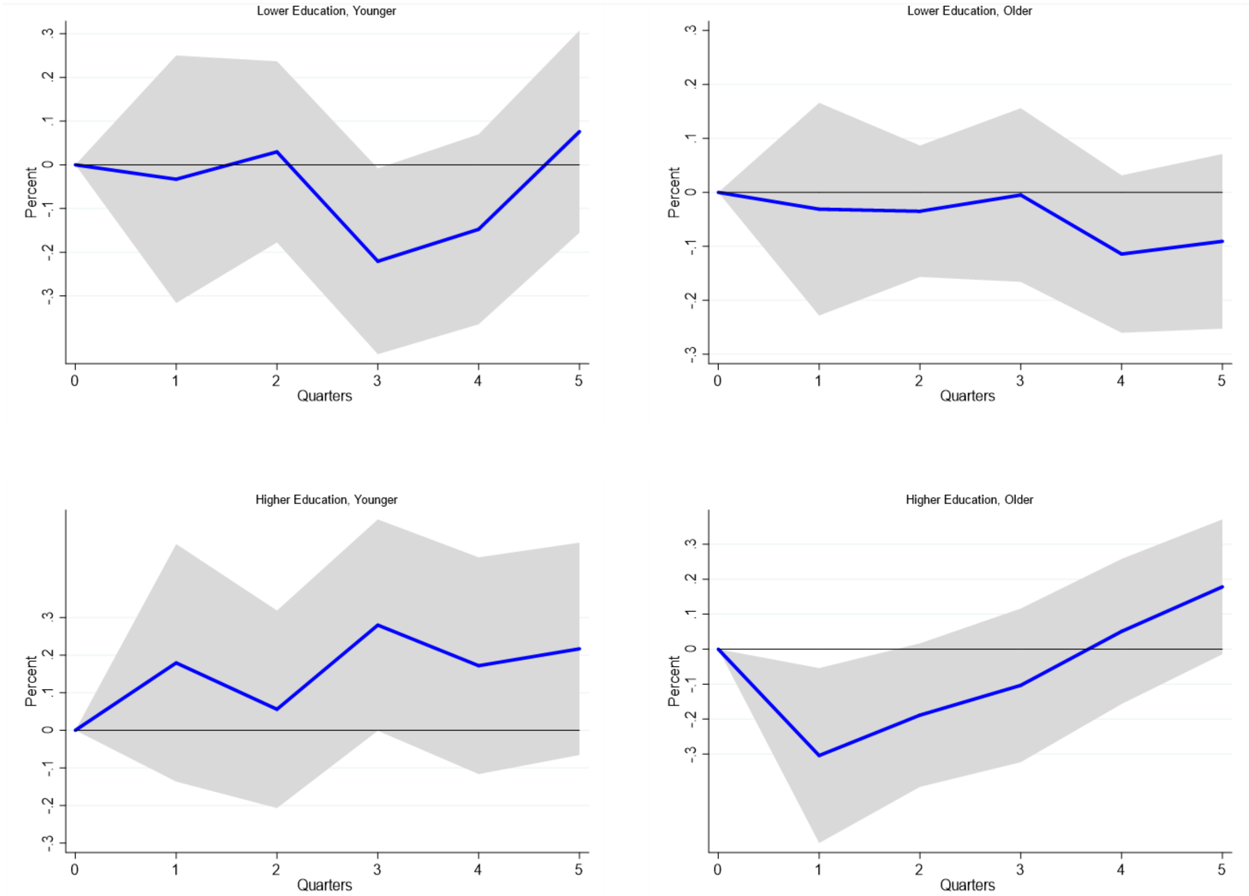

Notes: The shaded areas are the 90 percent confidence intervals.

**Appendix Figure 5. Consumption Responses to Monetary Easing under An Optimistic Regime with A 20-Percentile Threshold and Household-Level Observations**




Notes: The shaded areas are the 90 percent confidence intervals.

**Appendix Figure 6. Correcting for Double Counting and Allowing for City-Specific Responses of Baidu Search Index (20-Percentile and City-Level Observations)**




Notes: The shaded areas are the 90 percent confidence intervals.