Journal of the Asia Pacific Economy

ISSN: 1354-7860 (Print) 1469-9648 (Online) Journal homepage: www.tandfonline.com/journals/rjap20

The performance of Chinese real estate firms: media concerns and special treatment status

Kris Boudt & Xianda Liang

To cite this article: Kris Boudt & Xianda Liang (13 Nov 2025): The performance of Chinese real estate firms: media concerns and special treatment status, Journal of the Asia Pacific Economy, DOI: 10.1080/13547860.2025.2584029

To link to this article: https://doi.org/10.1080/13547860.2025.2584029

JOURNAL OF THE ASIA PACIFIC ECONOMY
https://doi.org/10.1080/13547860.2025.2584029

The performance of Chinese real estate firms: media concerns and special treatment status

Kris Boudt[a] and Xianda Liang[b]

[a]Department of Economics, Ghent University, Ghent, Belgium; [b]Faculty of Economic and Social Sciences, Vrije Universiteit Brussel, Brussels, Belgium

ABSTRACT
The proportion of Chinese listed real estate firms under special treatment (ST) is an important but lagging financial risk indicator for investing in Chinese real estate funds. In this paper, we propose to use media concerns about real estate as an early warning indicator of financial risk. Using regression analysis, we confirm their predictive value for the proportion of ST real estate firms in China. Furthermore, we show the financial value of the news-based early warning indicator for a risk on/off tactical asset allocation portfolio invested in the CSI 300 Real Estate Index.

ARTICLE HISTORY
Received 27 January 2025
Accepted 30 September 2025

KEYWORDS
China; financial health; news analytics; prediction; real estate; special treatment; tactical asset allocation

1. Introduction

The real estate industry has been a pillar industry of China's economy for many years. While bank loans are the main source of funding, it has become possible to directly invest in real estate firms through the capital market. Several of these firms are in financial distress, as indicated by their classification as 'Special Treatment' by the Shanghai and Shenzhen stock exchanges. They classify a listed firm as a special treatment (ST) firm if the firm experiences financial troubles such as the audited net asset being negative in the recent fiscal year, the interruption of the firm's operations due to a damaging lawsuit, or the blocking of the main bank account. From 2009 to 2024, the percentage of ST real estate firms varies between 1% and 13%, indicating a strong time variation in the financial distress of the Chinese real estate sector. Unfortunately, the percentage of ST firms itself is a lagging indicator, due to the delay between the events causing financial distress and the official announcement of the ST status of the firm. Therefore, an outstanding question is whether there is more timely and useful information in news data to characterize the current financial conditions of the Chinese real estate sector.

CONTACT Xianda Liang xianda.liang@vub.be Faculty of Economic and Social Sciences, Vrije Universiteit Brussel, Bd de la Plaine 5, 1050 Ixelles, Brussels, Belgium.

In this paper, we propose to monitor the attention and concerns about real estate in news data as an early warning signal for financial distress. We study their predictive relationship for the proportion of Chinese listed real estate firms that have the ST status. The underlying rationale is that some of the events causing the ST status are likely to be covered in the news prior to the ST status announcement of the firms. This pertains to both macro-economic (business cycle and policy) and firm specific news driving the likelihood of ST status changes of firms. We also test the information value of media concerns about real estate for investment analysis, since Ardia et al. (2023) show that, for climate change news, there is a strong association between media concerns and stock market reaction. The most likely economic channel is that investors use news data to revise their cashflow expectations and discount factors when investing in real estate. Formally, Nimark and Pitschner (2019) prove that investors can reduce the entropy of their posterior beliefs by delegating their information choice to specialized news providers through a news selection framework. Ball-Rokeach and DeFleur (1976) argue that as societies grow complex, mass media will take on more unique information functions and can shape the attitude of stakeholders toward an event. The agenda-setting theory (McCombs and Shaw 1972) also suggests that the media influence the attitude of stakeholders toward the real estate sector in terms of the attention devoted to the sector.

We propose two monthly news-based indices, a media attention index and a real estate concern index, to systematically capture relevant information in news articles for gains in predicting how the financial situation evolves in the Chinese real estate sector. We use articles published by *Economic Daily*, *Guangming Daily*, *China Youth Daily* and *Beijing Daily* to construct the indices. The sample of newspaper articles used in the study are collected from the websites and a database of the four main newspapers in China in the period January 2009–June 2024. This period is of interest because the Chinese real estate market developed at an increasing speed during this time, and the Chinese government issued many policies to exert control on the real estate market over this period. The sample collected from the above four sources contains 1345394 articles written in Chinese from January 2009 to June 2024. Furthermore, we use a keyword-based filter and estimate a Latent Dirichlet Allocation (LDA) model to obtain a refined corpus containing the articles that are related to the Chinese real estate sector. To construct the media attention index, we calculate the frequency of the keywords for each real estate article and aggregate the frequencies by taking the average across months and news outlets. For the real estate concern index, we set the article-level index at the product of the media attention and the proportion of negative words to all the polarized words (positive and negative words) in the article. Next, we build a linear regression framework to solve the question as to whether there is incremental value in the news-based indices for the proportion of Chinese listed real estate firms with ST status. We first estimate a benchmark linear model with only the lagged values of the proportion of ST real estate firms, and control variables. Then we fit models with the lagged values of the media indicators and their shocks being taken into account and compare their goodness of fit with the benchmark model. The result implies that the news-based indices do add value to the forecast of the ST proportion series.

Lastly, we examine the financial value of the concern index and its shock by applying them to tactical asset allocation (TAA) processes of the CSI 300 Real Estate Index. The proposed strategy follows a safety-first approach by investing in real estate only when there is a low level of media concern about real estate and no shock that increases the concern. The proposed strategy beats significantly a buy-and-hold and a stop-loss strategy.

The remainder of the paper is organized as follows. We review the literature which is related to our research in Section 2. In Section 3, the definition of ST firms is given, and a chronology of the development of the Chinese real estate industry is reviewed. Section 4 presents the data and methodology used. In Section 5, we build a linear regression framework to test whether there is incremental value embedded in the news-based indices. Section 6 reports the results of a robustness analysis. Section 7 proposes the tactical asset allocation strategy using the media concern index and its shock. In Section 8, we draw our conclusions.

2. Literature review

Our research is related to an increasing body of literature developing news-based indicators. First, there are studies measuring uncertainty with news-based indicators. Baker, Bloom, and Davis (2016) construct an economic policy uncertainty (EPU) index based on newspaper coverage frequency for 12 economies. They find that policy uncertainty raises stock price volatility and reduces investment and employment in sectors that are sensitive to policy at the firm level, while at the macro level, policy uncertainty is a sign of decline in output, investment and employment in the United States. Huang and Luk (2020) also create a monthly EPU index for China in the period 2000–2018 using multiple local newspapers. Their policy uncertainty also foreshadows declines in employment, equity price and output. Furthermore, Yu et al. (2021) study the relationship between the carbon emissions of China's manufacturing firms and a modified Chinese EPU index at the provincial level, which is constructed using daily newspapers of 31 provinces in China. They also provide an analysis of channels through which the EPU can affect firm emission intensity. Manela and Moreira (2017) develop a measure of uncertainty based on front-page articles in the *Wall Street Journal*, in an attempt to find the implied volatility in the 1980s. Their index peaks at world wars, fiscal crises and stock market crashes, but they capture only the uncertainty from the news. Caldara and Iacoviello (2022) create news-based geopolitical risk indices in both daily and monthly frequency at the global, industry and firm levels. Their indices capture disastrous events like the two world wars and Cuban missile crisis, and they find that higher geopolitical risk foreshadows lower investment and employment and is associated with higher disaster probability and larger downside risks to GDP growth.

In the field of real estate, studies using text-based indicators are focusing on predicting house prices and real estate market performance through sentiment or tone. First, most studies have focused on US real estate market. Soo (2018) provides real-time measures of local housing sentiment across 34 major cities in the US and confirms that the sentiment has significant predictive power for future house prices, using local housing news. Bork, Møller, and Pedersen (2020) propose a housing sentiment index using partial least square based on household survey responses to questions about house buying conditions. The index tracks the future growth rate of house prices better than other macroeconomic variables. This also holds in the long term and at the country level. Nowak and Smith (2017) incorporate the data of MLS remarks into a hedonic pricing model. Using penalized regression techniques, they find that the text data decreases pricing error by more than 25%. Beracha, Lang, and Hausler (2019) examine the predictive ability of news-based sentiment in the performance of the private commercial real estate market in the US. Koelbl (2020) documents a positive relationship between the sentiment embedded in the Management Discussion and Analysis of US Real Estate Investment Trusts (REITs) and firm performance in the future. Paulus, Koelbl, and Schaefers (2022) aim to explain the underpricing puzzle in the initial public offering of REITs in the US with the uncertainty of Form S-11, while no statistically significant results have been documented. Ruscheinsky et al. (2018b) capture news sentiment that is relevant to U.S. securitized real estate markets and document a significant relationship of the sentiment with future REIT market movements. Moreover, Hausler, Ruscheinsky, and Lang (2018) derive sentiment measures of professional news headlines published by S&P Global Market Intelligence database *via* support vector networks and investigate the relationship between the sentiment measures and the U.S. securitized and direct commercial real estate market.

Second, several studies have also investigated the use of text-based indicators in the context of Chinese real estate market. Huang (2020) develops a lexicon-based measure of housing sentiment for 24 cities in China by parsing through newspaper articles from 2006 to 2007. The measure is found to have strong predictive power for future house prices. Additionally, Huang builds another sentiment index using a random forest algorithm, which improves the accuracy of the benchmark logistic model from 0.88 to 0.96. Gao and Zhao (2018) extract media influence through newspaper articles and construct a buyer confidence index on the housing market of Guangzhou, a metropolitan area in the south of China. They find that buyer confidence is significantly and positively correlated with local house price. Li, Wang, and Liu (2022) construct a housing market sentiment index for 30 large and medium-sized cities in China using posts on Sina Weibo and explore the heterogeneity of the spatial effect of market sentiment on housing price from the view of geographical region. Zhu et al. (2023) also utilize posts on Sina Weibo to construct future and past sentiment indices that capture people's prior beliefs and posterior feelings, taking advantage of textual analysis techniques from deep learning.

In addition to the US and China, a small but growing number of research has explored applications of text-based indicators across other countries. Park et al. (2016) quantify topic weights and tone polarity using the textual analysis method. They find that compared to a model using only macroeconomic variables, the addition of text-based variables improves the prediction of the price and quantity of the housing market in Seoul. Walker (2016) test the influence of real-estate news published by Financial Times on real estate stock market of the UK. Plössl, Just, and Wehrheim (2021) study the news coverage and sentiment of real estate trends in

Germany and their relationship with residential price (Plössl and Just 2024) and total returns of asset classes of residential, office and retail (Plössl, Paulus, and Just 2023), using topic modeling. Ruscheinsky, Lang, and Schäfers (2018a) create a sentiment lexicon for German real estate sector and yield robust relationship between extracted sentiment measure and the residential real estate market in Germany. Thus, the existing literature using media indicators pays most of their attention to the evolution of house prices and real estate market performance, and less attention is paid to the gains in forecasting the financial health of the real estate market based and its application to investments when using media concerns. Our study aims to fill this research gap with a focus on the proportion of listed real estate firms with ST status.

3. Definition of ST firms and a chronology of real estate industry development in China

In this section, we first define firms under ST status, which forms the basis of the dependent variable in this study. We then review the development of the Chinese real estate market and summarize market and policy events regarding real estate in the period 2009–2024. We also link the narrative discussion of these events to the evolution of the business climate index for the Chinese real estate industry.

3.1. Definition of special treatment firms

The China Securities Regulatory Commission (CSRC) introduced a special 'delisting mechanism' in 1998 in an attempt to improve the governance of Chinese listed companies and protect investor interests. China's two stock exchanges, the Shanghai and Shenzhen exchanges, began to label some listed firms as ST firms. A firm is classified as an ST firm if there is a certain abnormality in its financial status, which may cause investors facing difficulties in judging the prospects of the firm and harm investor interests. Normally, the ST prefix consists of ST and *ST caps. A firm with a *ST cap is likely to be delisted in the future due to serious problems with its operation, such as continuous losses, while the ST caps indicate abnormalities in the operation of a firm, such as the main bank account of the firm being blocked. In this study, real estate firms with an ST or *ST cap are regarded as financial abnormal and referred to ST firms in the rest of the study.

Typically, according to the regulations of the Shanghai[1] and Shenzhen[2] exchanges, a firm is labeled as an ST firm if one of the following conditions is satisfied: (1) the lower of the audited total profit, net profit, or net profit after deducting non-recurring gains and losses in the recent fiscal year is negative, and its revenue is less than 300 million Yuan simultaneously (*ST); (2) the audited net asset is negative in the recent fiscal year (*ST); (3) the auditors issue a negative opinion or state that they are unable to issue an opinion on a firm's financial report in the recent fiscal year (*ST or ST); (4) non-operating funds take up more than 5% of the net asset or its amount exceed 10 million Yuan (ST); (5) the lower of the net profit before and after deducting non recurring gains and losses has been negative for the past three consecutive fiscal years, and the audit report for the most recent fiscal year shows uncertainty in the company's ability to continue (ST); (6) a firm's operations have been stopped, and it is unlikely for it to be restored within three months due to accidents like a natural disaster or the firm is involved in a damaging lawsuit or arbitration (ST); (7) the main bank account of the firm is blocked (ST). Apart from these conditions, a firm is also given an ST cap if the firm breaks the law or has other problematic operations, such as illegal information disclosure and serious dishonest practices.

There are also studies linking the ST cap to financial distress at the firm level. For example, Kim, Ma, and Zhou (2016) study the probability that ST firms survive their financial distress and restore to the normal listing status using a Cox proportional hazards model. Zhang, Altman, and Yen (2010) develop a Z-score model to identify firms that have potential distress, in which the ST cap is used to mark firms whose performance is unsatisfactory. Cheng and Jiang (2023) and Zhou et al. (2022) explore the issue of failure recurrence in Chinese ST firms after they overcome their financial distress. Li et al. (2021) study the ability of corporate governance measures to predict financial distress of firms. Tao et al. (2017) also discover a relationship between the political connection of financial distressed firms (ST firms) and the subsidies they receive from the government. Therefore, the ST status of a firm indicates that the firm is undergoing financial distress. We use the monthly proportion of ST real estate firms in China to proxy the financial situation of the Chinese real estate sector. Information on ST real estate firms is obtained from the RESSET database. According to the information collected, there are 38 ST real estate firms in the period January 2009–June 2024. We report the basic information of these firms—namely, their names, the period they are in ST status and the average market capitalization (weight) across their listing period in Table 1.

To the best of our knowledge, we are the first to study the aggregate time series evolution of the share of listed real estate firms with the ST status. We consider two weighting approaches: equal-weighting and market capitalization weighting. The advantage of equal-weighting is that the resulting monthly time series equals the proportion of all listed real estate firms with the ST status. It is computed by taking first the count of ST real estate firms and then dividing it by the number of all listed real estate firms. Another advantage of equal-weighting is that the resulting time series can be interpreted as the monthly probability that a listed real estate firms on the Shanghai and Shenzhen exchanges has the ST status. An alternative weighting approach is to use market capitalization weighting. This approach, considered in the robustness section, has the advantage that it gives more weight to larger real estate firms and is thus more informative about the potential market impact of the ST labeling decision.

In Figure 1, we report both the equally-weighted and market capitalization weighted time series with relevant events in the Chinese real estate sector. We find that market capitalization weighted proportion is on average lower than the equally-weighted proportion. This is consistent with Table 1 where we see that firms in Special Treatment tend to have a relatively low market capitalization. This is expected since these firms are in distress and thus have lower equity value and less prospects to grow. We do find a high degree of co-movement between the two proportion

Table 1. List of ST real estate firms in China: start and end months of their ST periods and the corresponding average market capitalization and equal weights.

Name	ST period	Market cap (MC weight vs. equal weight)(MC in billion RMB)
Zhongrun Resources Investment Corp. Ltd	2009.06–2010.03	6.453 (0.16% *vs.* 0.87%)
Rongan Property Corp. Ltd	2009.09–2010.03	9.812 (0.41% *vs.* 0.74%)
Hna Investment Group Ltd	2021.05–2023.06	5.069 (0.36% *vs.* 0.74%)
Jinke Property Group Ltd	2010.04–2011.07 2024.04–2024.06	20.38 (1.39% *vs.* 0.74%)
CCCG Real Estate Co. Ltd	2009.06–2011.03	4.708 (0.33% *vs.* 0.74%)
Greattown Holdings Ltd	2011.10–2012.02	12.13 (0.78% *vs.* 0.74%)
Shanghai Guijiu Co. Ltd	2017.03–2021.05	2.965 (0.21% *vs.* 0.75%)
Jiangsu Phoenix Property Investment Co. Ltd	2010.03–2010.04	4.743 (0.33% *vs.* 0.74%)
Hua Yuan Property Co. Ltd	2011.08–2024.06	6.427 (0.42% *vs.* 0.74%)
Jinan High-tech Development Co. Ltd	2019.04–2020.06	4.549 (0.33% *vs.* 0.74%)
Shenzhen Quanxinhao Co. Ltd	2009.01–2010.07 2021.04–2022.06 2023.05–2024.06	2.467 (0.16% *vs.* 0.78%)
Neoglory Prosperity Inc	2018.12–2022.05	12.23 (0.62% *vs.* 0.81%)
CRED Holding Co. Ltd	2011.08–2012.07 2021.04–2022.04	4.066 (0.25% *vs.* 0.75%)
Tianjin Troila Information Technology Co. Ltd	2020.04–2022.04	5.752 (0.40% *vs.* 0.74%)
Guangzhou Yuetai Group Co. Ltd	2021.05–2022.05 2023.05–2023.06	5.718 (0.40% *vs.* 0.74%)
Fujian Oriental Silver Star Investment Co. Ltd	2011.08–2012.04	0.930 (0.14% *vs.* 1.05%)
Hainan Airport Infrastructure Co. Ltd	2021.02–2022.09	36.18 (2.10% *vs.* 0.83%)
Shanghai Shimao Co. LTD	2023.05–2024.05	14.40 (1.03% *vs.* 0.74%)
ChengDu Hi-Tech Development Co. Ltd	2009.01–2009.04	1.906 (0.27% *vs.* 0.91%)
Zhonghong Holding Co. Ltd	2010.02–2010.03	11.15 (0.83% *vs.* 0.78%)
Shenzhen Wongtee International Enterprise Co. Ltd	2012.04–2014.02	6.428 (0.41% *vs.* 0.78%)
CASIN Real Estate Development Group Co. Ltd	2012.04–2013.03	4.027 (0.29% *vs.* 0.74%)
China Enterprise Company Limited	2016.03–2017.02	15.53 (1.11% *vs.* 0.74%)
Tianjin Jinbin Development Co. Ltd	2019.03–2020.05	5.810 (0.40% *vs.* 0.74%)
Beijing Huaye Capital Holdings Coo. Ltd	2019.04–2019.11	8.470 (0.63% *vs.* 0.75%)
Jiangsu Dagang Co. Ltd	2020.04–2021.05	4.549 (0.26% *vs.* 0.77%)
Yunnan Metropolitan Real Estate Development Co. Ltd	2021.04–2023.04	6.354 (0.45% *vs.* 0.74%)
Tahoe Group Co. Ltd	2022.05–2023.06	14.45 (0.91% *vs.* 0.75%)
Sichuan Languang Development Co. Ltd	2023.04–2023.05	1.295 (0.10% *vs.* 0.95%)
Myhome Real Estate Development Group Co. Ltd	2023.04–2023.05	7.155 (0.50% *vs.* 0.74%)
Aoyuan Beauty Valley Technology Co. Ltd	2023.05–2024.06	6.251 (0.35% *vs.* 0.78%)
Macrolink Culturaltainment Development Co. Ltd	2023.05–2024.04	9.536 (0.35% *vs.* 0.74%)
Yango Group Co. Ltd	2023.05–2023.06	17.92 (1.16% *vs.* 0.74%)
Soyea Technology Co. Ltd	2023.05–2024.06	2.902 (0.19% *vs.* 0.78%)
China Calxon Group Co. Ltd	2023.04–2023.05	9.846 (0.67% *vs.* 0.74%)
Sundy Land Investment Co. Ltd	2023.05–2023.06	5.322 (0.33% *vs.* 0.74%)
Jiangsu Zhongnan Construction Group Co. Ltd	2024.04–2024.05	19.66 (1.20% *vs.* 0.83%)
Dima Holdings Co. Ltd	2024.05–2024.06	9.351 (0.56% *vs.* 0.76%)

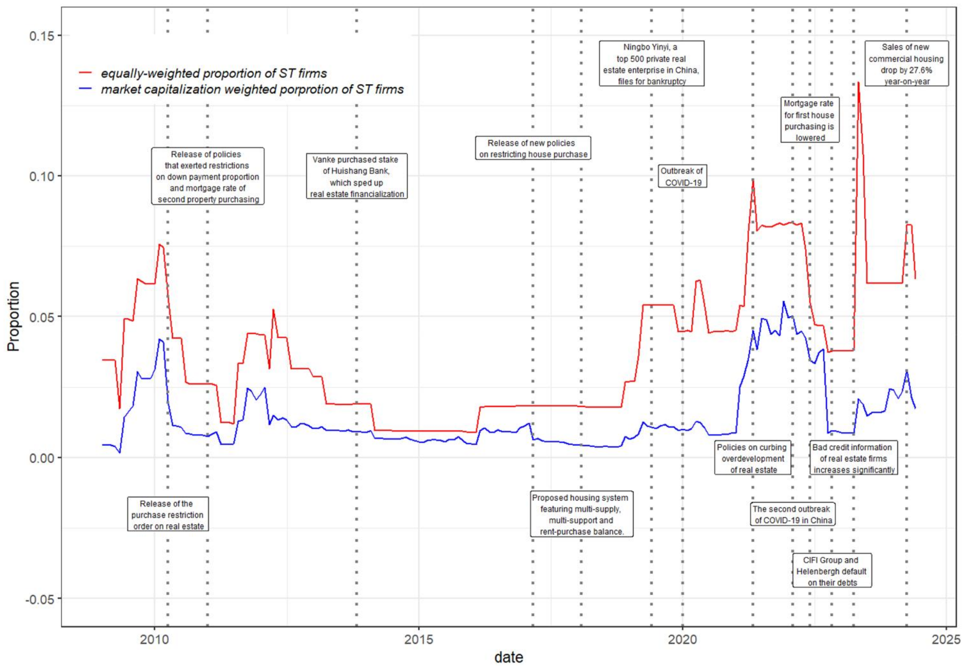

Figure 1. Monthly evolution of the proportion of ST real estate companies in China.

series, with a correlation of 0.79. We choose equally-weighting for our main analysis because of the advantage of interpretation as probability. The upward spikes in both series at the beginning of 2010 capture the release of the purchase restriction order on real estate. At the beginning of 2020, COVID-19 broke out, leading to increasing trends of the equally-weighted and market capitalization weighted proportion series in the future. Lastly, in the first quarter of 2023, the amount of bad credit information, most of which regards tax arrears, increased by 13% quarter-on-quarter, leading to rises in both series.

3.2. Chronology of real estate industry development in China

Before 1998, the real estate market in China had not yet formed because there was a welfare housing allocation policy stipulating that houses must be allocated to people by their work unit at that time. This policy was not abolished until 1998 when a new policy, named Decision on Deepening the Reform of Urban Housing System, was issued by the State Council of China. This new policy turned the old housing allocation system into a market system. To acquire a house, people need to buy one themselves rather than wait for allocation of one. Therefore, it was the issue of the new policy in 1998 that set off the pace of the Chinese real estate market. In the period January 2009–June 2024, the Chinese real estate market was developing at a high speed under various macro-control policies from the Chinese government. Fang (2018) states that in this period, real estate developers continued to expand their market scale, and the market share of the top 100 real estate companies increased. The credit volume in the real estate industry grew simultaneously, which also brought

Table 2. Main market and policy events for Chinese listed real estate over the period 2009–2024.

Market/policy events	Time point	Description
1. Release of purchase restriction order	April 2010	A policy named Ten Articles of the New Country was issued by the State Council of China to restrict real estate purchasing, which leads to a decrease in house demand.
2. Release of restrictions on second property purchasing	January 2011	A policy named Eight Articles of the New Country was issued by the State Council of China to restrict the down payment proportion and mortgage rate of second property purchasing.
3. Vanke purchased 3 billion HKD of stakes in Huishang Bank	October 2013	Vanke, a real estate enterprise, purchased 3 billion HKD of Huishang Bank's stakes, becoming the largest shareholder. This can be regarded as a sign that real estate firms in China began to enter the banking industry.
4. New set of restricting policies on house purchase	March 2017	New policies focusing on raising down payment proportion were released in many cities in China to subdue house demand.
5. A new housing system proposed by MHURD	February 2018	The Ministry of Housing and Urban-Rural Development (MHURD) has proposed a housing system featuring multi-subject supply, multi-channel support and simultaneous development of rental and purchase.
6. Ningbo Yinyi files for bankruptcy	June 2019	Ningbo Yinyi, a real estate enterprise among China's top 500 private enterprises, filed for bankruptcy in June 2019. The bankruptcy reason given by Cai (2019) is a huge debt.
7. Outbreak of COVID-19	January 2020	The global economy is undergoing a fierce impact due to this unprecedented pandemic. But the impact on the Chinese real estate industry seems to be limited.
8. Tightening policies on real estate finance	May 2021	Tightening policies on real estate finance (e.g. raising interest rates) to curb the real estate industry from overdeveloping.
9. The second outbreak of COVID-19	February 2022	The restriction brought about by the pandemic forced sales center and agencies of real estate to suspend operations, nearly freezing the market.
10. Mortgage rate for the first house purchasing is lowered	May 2022	A newly released house credit policy adjusts the mortgage rate for the first house to a lower level.
11. Two real estate firms defaulted on their debts	November 2022	Two Chinese firms, Helenbergh China Holding Limited and CIFI Group, defaulted on their debts, due to insufficient profitability and liquidity pressure.
12. Bad credit information of real estate firms increases	March 2023	The China Real Estate Industry Association reports 39472 pieces of information in bad credits, most of which are about tax arrears. The number increases by 13% quarter-on-quarter.
13. New commercial housing sales drop	March 2024	The sales of new commercial housing drops by 27.6% year-on-year in the first quarter of 2024.

about concerns toward the industry. We summarize policy and market events regarding the Chinese real estate sector in Table 2.

The National Bureau of Statistics compiles a business climate index for the Chinese real estate industry—namely the National Real Estate Business Climate (NREBC) index. This index measures the prosperity of the Chinese real estate market, aggregating data from all real estate firms, regardless of their listed status. There are 8 indicators used for the compilation of the NREBC, namely real estate investment, sources of funds, land transfer, developed land area, sales price and vacancy rate of commercial housing and floor space of new construction and buildings completed. The weight of every indicator is determined by expert scoring, and the seasonal factors are removed. The NREBC provides a holistic snapshot of the Chinese real estate

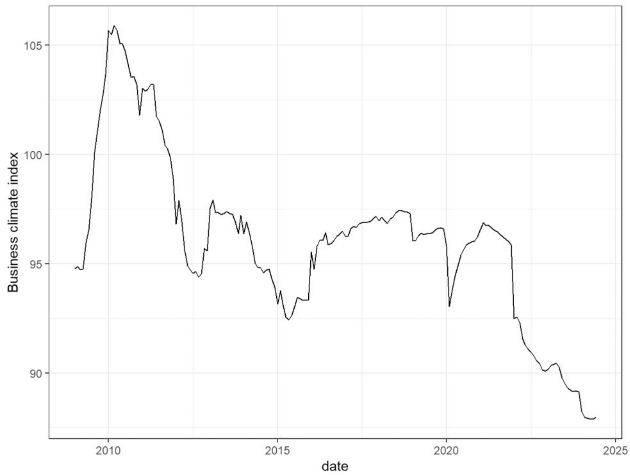

Figure 2. Monthly evolution of the business climate index in China.

sector, which is critical for assessing economic risks and guiding regulatory decision. The monthly evolution of the index is shown in Figure 2. The downward trend after 2010 resulted from the issue of a purchase restriction order on real estate, and the outbreak of COVID-19 accounts for the downward peak at the beginning of 2020. After 2022, the Chinese government relaxes the quarantine measures, which lead to the second outbreak of COVID-19. Thus, the index continues to drop after 2022.

4. Data and methodology

4.1. Real estate news corpus

We select *Economic Daily*, *Guangming Daily*, *Beijing Daily* and *China Youth Daily* as the sources of news. The reasons are as follows. First, we require the newspaper data to be available over long time span starting in January 2009. This excludes recently established media like Cailian Press and newspapers with limited online availability, such as China Real Estate Business and Nanfang Daily. Thus, we restrict our selection to the four newspapers mentioned above. Furthermore, *Economic Daily* is an important information channel for economic policies issued by the Chinese government. As the real estate industry has a strong relationship with the macroeconomic situation, it is necessary to take macroeconomic information into account. Taking *Economic Daily* as the source will not miss economic information that exerts influence on the real estate industry. *Guangming Daily* is a comprehensive newspaper run by the Central Committee of the Chinese Communist Party. The major reader group of this newspaper is intellectuals including government officers, workers in state-run universities and so on. From this newspaper, we can capture the points of view and tone of the Central Committee on the real estate industry in China as relevant factors. *Beijing Daily* has the largest circulation in Beijing, focusing on the economics, politics and culture of the city. As Beijing is the capital of China, most of the listed real estate companies in China have a branch in Beijing. Therefore, the situation of these companies is closely related to the economic environment in Beijing. To capture this, we

Table 3. Keywords to select real estate articles and their English translation.

Filters	English translation
房地产	Real estate
房产	
不动产	
房市	Real estate market
楼市	
房价	House price
商品房	Commercial housing
住宅	House
住房	
房屋	
房源	Housing resources
楼盘	Properties for sale
房企	Real estate firms
房地产业	Real estate industry
房地产商	Real estate developer
不动产商	
房屋出租	Property rental
房屋租赁	

added *Beijing Daily* to the source list. *China Youth Daily* is a comprehensive newspaper for the youth in China, whose credibility ranks top three among Chinese news media in the past ten years. The main reader group of this newspaper includes young staff, students in universities, staff in state-owned firms and so on. Therefore, this newspaper captures the view of the young generation in China.

We retrieve the newspaper articles from *Economic Daily*, *Guangming Daily*, *Beijing Daily* and *China Youth Daily* for the period January 2009–June 2024. The articles from the sources are available on the official websites of *Economic Daily*, *Guangming Daily*, *Beijing Daily* and *China Youth Daily* and the database of *Beijing Daily*. The total number of articles is 1345394.

We then query the corpus through 3 steps to identify real estate-related articles and to obtain a refined real estate news corpus. First, we apply a keyword filter to remove the articles that are irrelevant to the real estate sector, retaining 90829 articles in the corpus. The list of keywords and the English translation are presented in Table 3. Second, we further reduce the sample through a set of geographical words to ensure that the remaining articles are about the mainland of China. This step reduces our selection of articles to 88462. Third, Shapiro, Sudhof, and Wilson (2022) note that computational sentiment for articles with less than 200 words tends to be noisy. As a typical Chinese word has two characters, we remove the articles that are shorter than 400 characters, which further narrows down the corpus size to 81481.

In Table 4, we report the total number of real estate articles, the total number of news articles and the percentage of real estate articles published by the sources in the sample. Among the four sources, *Economic Daily* published the most in terms of percentage (6.8%), and *Beijing Daily* published the most in terms of number (26244) of real estate articles. On the contrary, *Guangming Daily* published the least in terms of percentage (4.3%), and *China Youth Daily* published the least in terms of number (15779).

Figure 3 visualizes the evolution of the monthly proportion that real estate articles take up for each source. The contribution of each newspaper is proportional to the width of bands with different colors. The figure further illustrates that *Economic*

Table 4. Corpus of real estate news articles.

	Guangming Daily	*Economic Daily*	*China Youth Daily*	*Beijing Daily*
Total number of articles	362751	336486	214633	431524
Number of real estate articles	15779	22794	14228	26244
Percentage	4.3%	6.8%	6.6%	6.1%
Source	https://epaper.gmw.cn/gmrb/html/2021-04/19/nbs.D110000gmrb_01.htm	http://paper.ce.cn/jjrb/html/2021-04/19/node_2.htm	http://zqb.cyol.com/html/2021-04/19/nbs.D110000zgqnb_01.htm	The database of *Beijing Daily* office

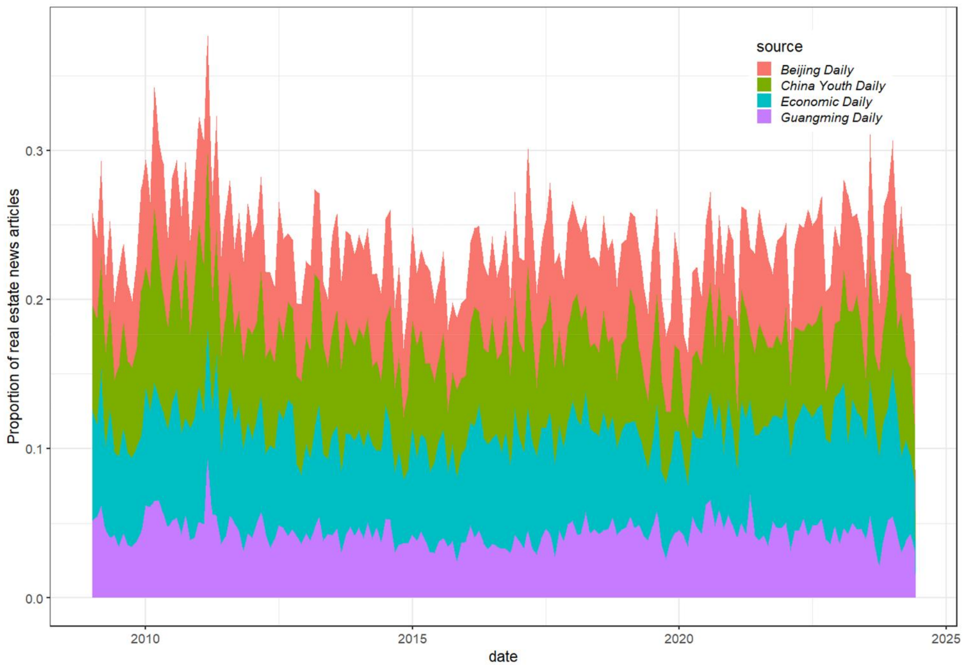

Figure 3. Monthly proportion of real estate news articles.

Daily, *China Youth Daily* and *Beijing Daily* published a larger proportion of real estate news than *Guangming Daily* did at most of the time points in the period. At the beginning of 2010, the proportion of all four sources peaked upward simultaneously, especially for *China Youth Daily*, as the result of shocks arose from the release of the purchase restriction policy on houses. Moreover, the percentage of real estate news is varying for all four newspapers, which highlights that standardization in the aggregation process is necessary.

To obtain a better overview of what the reduced selection of articles discusses, we estimate an LDA model of Blei, Ng, and Jordan (2003) with 30 topics. LDA is an unsupervised machine-learning algorithm that infers latent topics among a set of articles. It produces a vector of topic prevalence for each article. We label the topics manually by looking at the ten most probable keywords for each topic and reading the 10 most relevant articles under each topic. The topics, their keywords and their prevalence are reported in Appendix. The prevalence of a certain topic in this study is calculated as its topic weight in the whole corpus—namely, the sum of the prevalence of every article divided by the total number of articles in the corpus. We then find the topics regarding the real estate industry, which in this part are 2, 10, 12, 17 and 24. We subsequently remove the articles that are irrelevant to these topics, decreasing the number of articles to 80571. Lastly, we eliminate the articles that are highly associated with false positives, such as advertisement, anniversaries and articles regarding COVID-19. The final corpus size is 65672.

4.2. Real estate news indices

We now turn to the question about how to map the obtained set of real estate news articles into a time series that will be informative about the evolution of the proportion of ST real estate firms in China. The simplest indicator is a count-based indicator that captures the media attention to real estate. Such attention-based indicators belong to the first generation of popular media-based indicators of economic conditions. Popular examples are the recession attention index (Boudt et al. 2023) and the EPU index (Baker, Bloom, and Davis 2016). A new generation of the media indices augment the media attention indicators with article sentiment leading to an index of concern (Ardia et al. 2023). In this part, the two monthly news-based indices are constructed using the refined news corpus—namely, the real estate media attention index and real estate concern index.

4.2.1. Monthly media attention index for real estate industry (MARE)

Media attention refers to how often a certain thing is reported by the media in a certain period. We take three steps to construct the media attention index:

1. We calculate the article-level media attention as the frequency of the keywords in Table 3.

$$ARE_{s,m,i} = \frac{n_{s,m,i}}{N_{s,m,i}}, \tag{4.1}$$

   where s refers to a news source, m represents for a certain month in the entire period, and i is the ith article of the source s in the month m. $n_{s,m,i}$ and $N_{s,m,i}$ are the numbers of the keywords and total words in the ith article, respectively. $ARE_{s,m,i}$ is the article-level media attention of the ith article of the source s in the month m.
2. Calculate the monthly media attention for a news source. We take the average of the article-level media attention of the articles published by the source s in the month m.

$$MARE_{s,m} = \sum_{i=1}^{I_{s,m}} \frac{ARE_{s,m,i}}{I_{s,m}}, \tag{4.2}$$

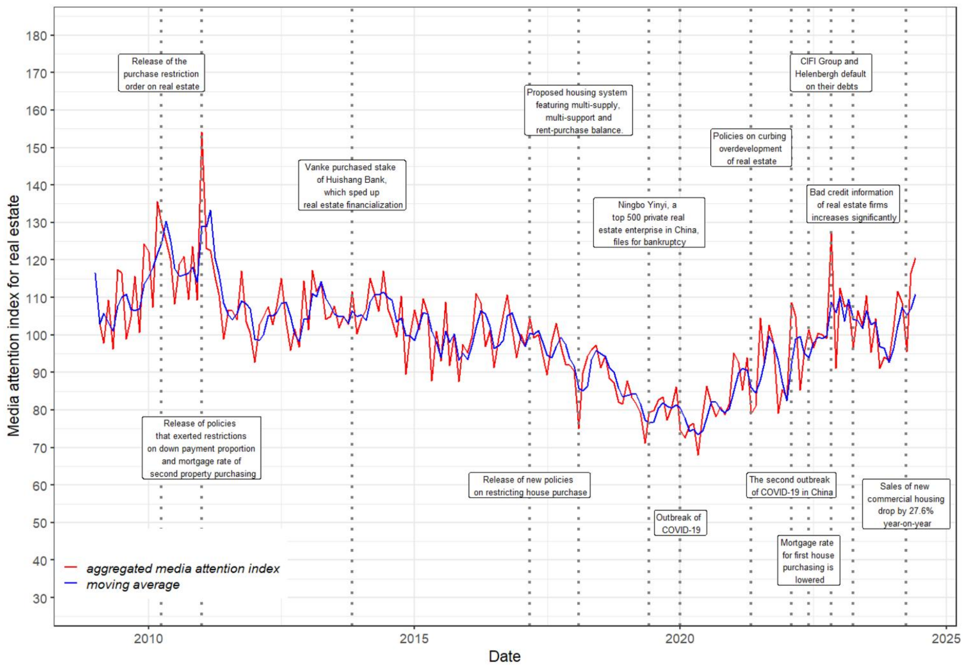

Figure 4. Monthly media attention index for real estate in China and its 90-day moving average.

where $MARE_{s,m}$ is the monthly media attention of the source s in the month m, and $I_{s,m}$ is the total number of the articles.

3. Follow the methods of standardization and normalization proposed by Baker, Bloom, and Davis (2016) to deal with source heterogeneity. Furthermore, we apply the square root function to capture the fact that more real estate articles increase media attention and concern at a decreasing rate, as Ardia et al. (2023) do. Ten real estate articles are unlikely to increase the indices ten times. One explanation for this is the 'echo chamber' phenomenon proposed by Flaxman, Goel, and Rao (2016), pointing out that people are inclined toward the news that is in line with their views. Finally, we normalize the index to a mean of 100.

$$ARE_m = \sqrt{\frac{1}{4}\sum_{s=1}^{4}\frac{MARE_{s,m}}{\sigma_s}}, \tag{4.3}$$

$$MARE_m = \frac{100 \times ARE_m}{\sum_{m=1}^{M} ARE_m / M}, \tag{4.4}$$

where M is the total number of months in the entire period, σ_s is the standard deviation of the monthly media attention index of source s, ARE_m is the square root of the monthly averaged index across the four news outlets, and $MARE_m$ is the normalized index.

The evolution of the media attention index and its 90– day moving average are shown in Figure 4. The media attention index peaks in 2010 and 2011 when the purchase restriction and finance restriction order on real estate were released, respectively. The direct target of the two orders is real estate, prompting the real estate industry to receive more attention from the media. Meanwhile, the upward peak at around the beginning of 2020 coincided with the COVID-19 shock, and an increasing number of bad credit information in the first quarter of 2023 attracted more attention from news media to the real estate sector.

4.2.2. Monthly real estate concern index (MCORE)

The goal of this index is to capture concerns toward the Chinese real estate industry. To define concern, we use the definition proposed by Ardia et al. (2023)—that is, the perception of risk and related consequences associated with this risk. Based on this definition, we calculate the real estate concern index at article level through the formula below.

$$CORE_{s,m,i} = ARE_{s,m,i} \times \frac{neg_{s,m,i}}{pos_{s,m,i} + neg_{s,m,i}}, \tag{4.5}$$

where s denotes a news certain source, m represents a certain month in the period, $ARE_{s,m,i}$ is the attention index at article level for the ith article in the month m of the source s, and $pos_{s,m,i}$ and $neg_{s,m,i}$ represents the number of positive and negative words in the ith article, respectively, in the month m from the source s.

The attention index reflects the amount of discussion on the real estate industry, and the second ratio reflects risk perception, as a larger number of negative words in real estate articles means there might be risk events pertaining to the real estate sector looming.

To identify polarized words in an article, we create a sentiment lexicon for the Chinese real estate sector—namely, a list that consists of pairs of polarized words and their associated sentiment score, using the word-embedding method of Borms et al. (2021). First, we define sentiment in this study as Algaba et al. (2020) do: the disposition of an entity toward an entity, expressed *via* a certain medium. The disposition should have a measurable polarity or semantic orientation. Second, we manually select two groups of seed words which are positive and negative. Then we expand the two groups of words through the word-embedding technique—that is, find the words that have high cosine similarity (have a similar meaning) to the seed words and add them to the two groups. After combining the two fully expanded groups of words, we obtain the sentiment lexicon for the real estate sector. Additionally, we combine our real estate lexicon with a financial sentiment lexicon created by Yao et al. (2021) to further detect polarized words about finance since finance also plays a key role in the operation of real estate firms. However, raw counts of polarized words may give rise to error since the polarity of a word can be changed by a valence-shifting word, which exerts a semantic impact on polarized words. A common example is 'not good', which will be interpreted positively if the presence of the negator 'not' is not taken into consideration. Thus, to further incorporate valence-shifting words, we use the valence-shifting cluster approach proposed by Ardia et al. (2021) instead. This cluster approach searches for valence-shifting words in a cluster of a maximum of four words before and two words after a polarized word.[3] Lastly, the numbers of positive and negative word clusters are counted.[4]

Next, we follow the same workflow as we did for the media attention index to standardize, aggregate and normalize the concern index to a mean of 100:

$$MCORE_{s,m} = \sum_{i=1}^{I_{s,m}} \frac{CORE_{s,m,i}}{I_{s,m}} \tag{4.6}$$

$$CORE_m = \sqrt{\frac{1}{4}\sum_{s=1}^{4} \frac{MCORE_{s,m}}{\sigma_s}}, \tag{4.7}$$

$$MCORE_m = \frac{100 \times CORE_m}{\sum_{m=1}^{M} CORE_m / M}, \tag{4.8}$$

where $MCORE_{s,m}$ $CORE_m$, $MCORE_m$, M and $I_{s,m}$ represent the monthly concern for the source s in the month m, the square root of the monthly average concern across the sources, the standardized monthly concern, the total number of months in the entire period and the total number of articles published by the source s in the month m, respectively.

The evolution of the index and its 90− day moving average are shown in Figure 5. The concern index peaks upward, for example, at the beginning of 2010 because the

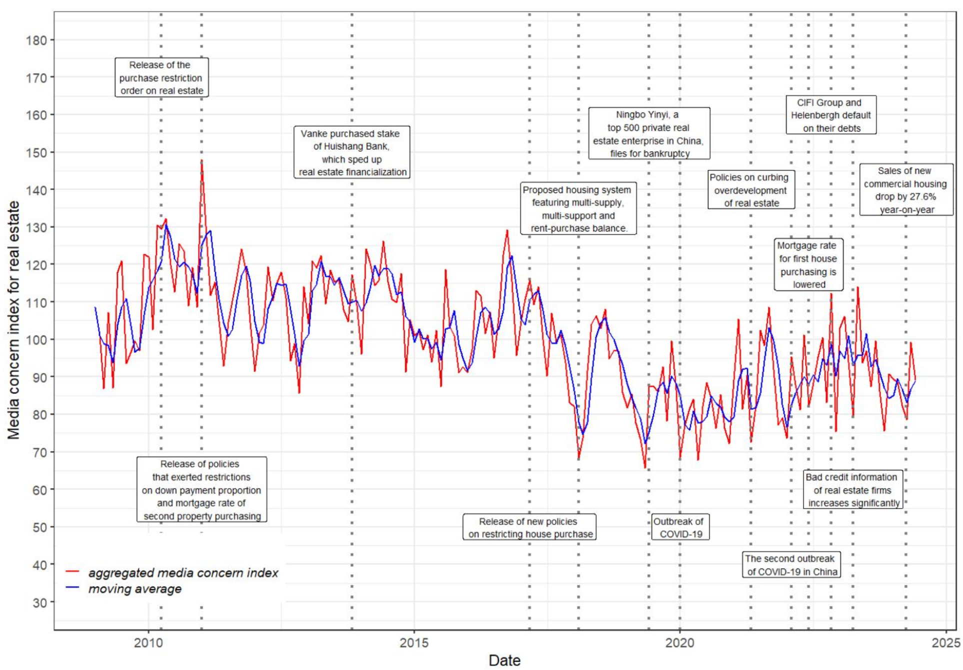

Figure 5. Monthly media-based real estate concern index for China and its 90-day moving average.

release of restriction orders on house purchases triggered more perception of risk in the Chinese real estate industry. Similarly, the outbreak of COVID-19 spurs an increasing trend after 2020 because more risks are perceived at the time of COVID-19. Additionally, the concern index captures the positive effect brought about by a downward adjustment of mortgage rate for the first house with a downward peak in 2022.

4.3. Shocks in real estate news indices

Graphical analysis and unit root tests show the persistence of the indices. To extract the information in the media indices, we therefore recommend to use time series methods to extract the shocks. We use the workhorse ARMA(p,q) time series model for this applied to the media indices taken in first differences. Specifically, to extract shocks from the media attention index and concern index, we select the best-fit time-series model through three information criteria (AIC, AICc, BIC). The results show that ARMA (1,1) is the best-fit models for the changes in both the media attention index and the concern index. We thus fit the two models with the two news-based indices to obtain the residuals ϵ as the shocks in the two indices. The equations are presented below:

$$\Delta MARE_t = \alpha + \beta \Delta MARE_{t-1} + \theta \epsilon_{t-1} + \epsilon_t, \tag{4.9}$$

$$\Delta MCORE_t = \alpha + \beta \Delta MCORE_{t-1} + \theta \epsilon_{t-1} + \epsilon_t. \tag{4.10}$$

The ARMA (1,1) model for the media attention index has estimated coefficients (p-value) for β of −0.002 (0.984) and θ of −0.730 (0.001). For the model of the concern index, it has estimated coefficients for β of 0.276 (0.008) and θ of −0.884 (0.001). Finally, we perform a cumulative periodogram white-noise test for the residuals produced from the two ARMA models (Hu, Kutan, and Sun 2018). The results show that we cannot reject the hypothesis that the residuals are white noise (the p-values are 0.919 and 0.892 for the media attention index and concern index, respectively), meaning that the ARMA (1,1) model fit the two indices well. The series of residuals obtained in this step will be the measurement of the shocks.

5. Predictive value of the news-based indices for Chinese real estate special treatment

In this section, we proceed to the question of whether the news-based indices provide incremental value on the financial situation of Chinese real estate sector. To solve this question, we fit and compare benchmark models and media-indicator-included models that predict the current ST proportion with and without control variables. The benchmark models use only lag values of ST series and macroeconomic variables as control, while the media-indicator-included models further contain the news-based indices and their shocks. We hypothesize that more accurate forecasts can be obtained through the models with the media indicators than through the benchmark models. This can be reflected by an increase in the goodness of fit measured by adjusted R^2 and F-test that compares the fit of the media-indicator-included models with that of the benchmark models. The variables used are as follows.

5.1. Dependent variable

Our target variable is the change in the monthly proportion of ST real estate firms in China, calculated as the number of ST real estate firms divided by the total number of listed real estate firms. This variable reflects to some extent the evolution in the financial environment through the proportion of real estate firms in financial distress.

5.2. Main independent variables

The monthly change in media attention index, the concern index and their shocks. The indices represent the level of media attention and concern expressed *via* the media on the Chinese real estate sector, and the shocks reflect the unexpected change in media attention and media concern. Given the property of ST status, we use 1–24 months of lags of the indices and their shocks as the main independent variables. We also take the lags of the ST series into account to control the influence of past values of the ST series.

5.3. Control variables

In our analysis, we also incorporate five control variables. To capture the overall state of the real estate sector, we use the business climate index for real estate (NREBC). For extracting the influence originating from money supply, we consider China's broad money supply (M2). Regarding the stock market information, we include the monthly return of the CSI 300 Real Estate Index, which is a comprehensive stock index for the real estate sector, and the monthly standard deviation of its daily return. Lastly, to account for long-term capital cost, we take into account the monthly yield to maturity of five-year treasure bonds. The data for these control variables are sourced from the CEIC and RESSET databases. In order to prevent spurious regression, we take the first-order difference of the natural logarithm of all the control variables except the monthly return of CSI 300 Real Estate Index, as the monthly return represents price changes of the index. The results of the augmented Dickey-Fuller (ADF) test verify that all the variables in the models are stationary at the 1% level.

In addition, we perform a correlation analysis for the control and dependent variables. The results are reported in Table 5, where *NREBC*, *M*2, *ret*, *std* and *tbond* represent the business climate index, the broad money supply, the monthly return of CSI 300 Real Estate Index, the standard deviation of daily return of CSI 300 Real Estate Index and the monthly yield to maturity of 5-year treasure bond. It can be noted that the correlations of the change of ST ratio with all the control variables are not significant, indicating limited explanatory power of the control variables.

Table 5. The pairwise correlations between the dependent and control variables.

	ΔST	$\Delta\log(NREBC)$	$\Delta\log(M2)$	ret	$\Delta\log(std)$	$\Delta\log(tbond)$
ΔST	1	2.3%	−4.1%	3.1%	−5.7%	−4.3%
$\Delta\log(NREBC)$	—	1	1.4%	−12.9%*	−10.1%	15.8%**
$\Delta\log(M2)$	—	—	1	19.5%***	5.4%	−1.6%
ret	—	—	—	1	14.0%*	14.8%**
$\Delta\log(std)$	—	—	—	—	1	7.3%
$\Delta\log(tbond)$	—	—	—	—	—	1

*, **, ***denote the significance levels at 10%, 5% and 1%, respectively.

Next, we estimate two benchmark models: the first one uses only the lagged values of the ST series, and the second one further consider the control variables:

$$\Delta ST_t = \alpha + \sum_{i=1}^{24} \beta_i \Delta ST_{t-i} + \epsilon_t, \tag{5.1}$$

$$\Delta ST_t = \alpha + \sum_{i=1}^{24} \beta_i \Delta ST_{t-i} + \gamma' controls_t + \epsilon_t, \tag{5.2}$$

where *controls* represents the five control variables mentioned above.

To find the incremental information embedded in the two media indices and their shocks, we incorporate the indices and shocks into the benchmark models, where *media* represents change in the media attention index, change in the concern index and their shocks.

$$\Delta ST_t = \alpha + \sum_{i=1}^{24} \alpha_i media_{t-i} + \sum_{j=1}^{24} \beta_j \Delta ST_{t-j} + \epsilon_t, \tag{5.3}$$

$$\Delta ST_t = \alpha + \sum_{i=1}^{24} \alpha_i media_{t-i} + \sum_{j=1}^{24} \beta_j \Delta ST_{t-j} + \gamma' controls_t + \epsilon_t. \tag{5.4}$$

We report the comparison of the goodness of fit between the benchmark model and media-indicator-included models in Table 6. It can be noted that the adjusted R^2 increases from 10.48% to 46.89% and 43.66% after incorporating the media attention index and its shock into the benchmark model with the control variables, respectively. For the model without the control variables, the adjusted R^2 increases from 12.28% to 45.82% (42.28%) for the media attention index (and its shock). For the cases of the concern index and its shock, the adjusted R^2 rises from 10.48% to 42.79% and 42.94% with the control variables in the benchmark model, respectively. For the model without the control variables, the adjusted R^2 surges to 41.22% and 42.05%. From the results of the F-test, we also observe that ignoring the news-based indices and their shocks significantly decreases the fit of all models with media indicators, regardless of the inclusion of the control variables. These results show that the news-based indices are capable of providing incremental value in predicting the proportion of ST real estate firms in China.

Table 6. Incremental information of news-based indices on ST proportion series.

Models	News indices	Goodness-of-fit statistics			
		Adj. R^2 (%)	RSS (%)	F-test	p-value
Panel A: With controls					
Benchmark model		10.48	1.291	—	—
Models with media indicators	Attention	46.89	0.626	4.743	9.99×10^{-9}
	Attention shock	43.66	0.664	4.214	1.27×10^{-7}
	Concern	42.79	0.674	4.083	2.42×10^{-7}
	Concern shock	42.94	0.672	4.016	2.17×10^{-7}
Panel B: Without controls					
Benchmark model		12.28	1.314	—	—
Models with media indicators	Attention	45.82	0.668	4.507	2.34×10^{-8}
	Attention shock	42.28	0.712	3.945	3.88×10^{-7}
	Concern	41.22	0.725	3.790	8.50×10^{-7}
	Concern shock	42.05	0.715	3.911	4.59×10^{-7}

On the whole, the adjusted R^2 of the benchmark model is significantly increased after incorporating the media indicators into the model. These gains justify the usefulness of the media indicators in forecasting the financial situation in the Chinese real estate sector.

6. Robustness analyses

The previous section has shown the predictive value of the proposed real estate media attention and concern indices for the proportion of ST real estate firms in China. In this section, we analyze the robustness of this finding to choices in the construction of the variables used.

6.1. Robustness to choice of news sources

We compare the aggregated real estate media attention index with the corresponding index per source in Figure 6. It can be noted that there is large co-movement, and a similar plot is obtained for the real estate concern index. In Panel A of Table 7, we report the regression result using the real estate media indices per source. For *Economic Daily*, *Guangming Daily* and *Beijing Daily*, all four media indicators lead to the same conclusion, reflected by the significant results of the F-tests and increased adjusted R^2. However, in the case of *China Youth Daily*, the concern shock is the solely robust media indicator. Overall, this analysis validates to use a combination of news sources.

6.2. Robustness to the choice of lexicon

In this part, we first change the lexicon used in the calculation of the concern index to a lexicon created by Jiang et al. (2019). Based on this lexicon, we recalculate the concern index and use this new index as the independent variable in the linear regression framework to find out whether the conclusion is still robust. We compare the goodness of fit between the benchmark model and the model using the new concern index and report the results in Panel B of Table 7. From the results, the concern

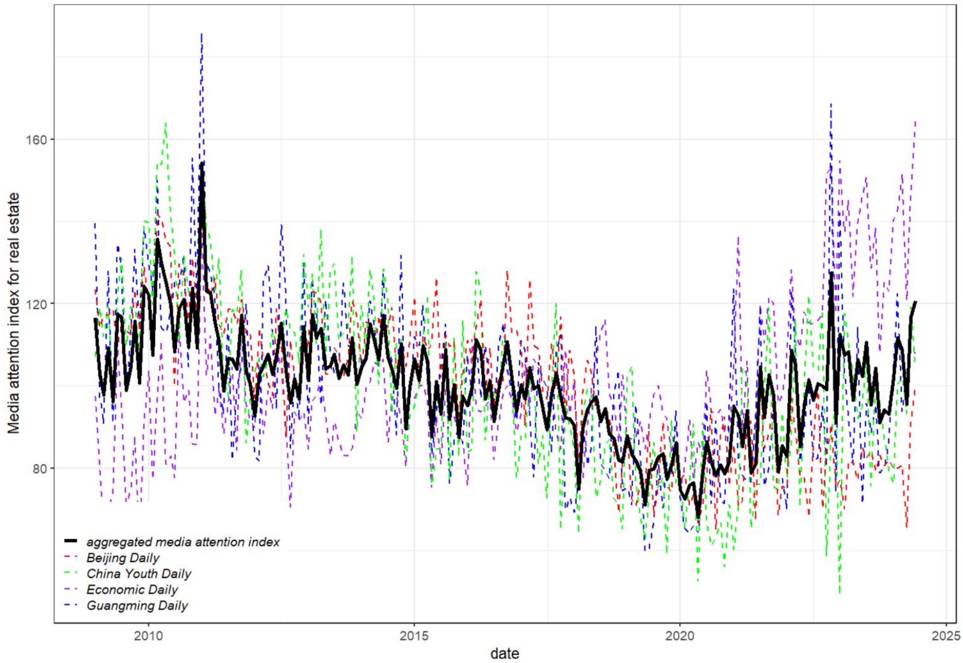

Figure 6. The aggregated media attention index vs. media attention index per source.

Table 7. Sensitivity of goodness-of-fit statistics to the choice of news sources, lexicons and window size.

		Goodness-of-fit statistics			
Sources	News indices	Adj. R^2 (%)	RSS (%)	F-test	p-value
Benchmark model	—	10.48	1.291	—	—
Panel A: Choice of news sources					
Economic Daily	Attention	24.43	0.890	2.009	0.008
	Attention shock	25.09	0.882	2.065	0.006
	Concern	26.61	0.865	2.200	0.003
	Concern shock	32.81	0.792	2.814	1.46×10^{-4}
Guangming Daily	Attention	37.43	0.737	3.351	9.47×10^{-6}
	Attention shock	37.17	0.740	3.319	1.11×10^{-5}
	Concern	21.38	0.926	1.757	0.027
	Concern shock	27.14	0.859	2.249	0.003
China Youth Daily	Attention	16.12	0.988	1.367	0.142
	Attention shock	13.99	1.014	1.223	0.240
	Concern	16.16	0.988	1.370	0.140
	Concern shock	20.66	0.935	1.701	0.035
Beijing Daily	Attention	27.61	0.853	2.292	0.002
	Attention shock	22.12	0.918	1.816	0.021
	Concern	28.88	0.838	2.413	0.001
	Concern shock	28.15	0.847	2.342	0.002
Panel B: Choice of lexicon					
Concerns based on new lexicon	Concern	42.27	0.680	4.006	3.55×10^{-7}
	Concern shock	40.87	0.697	3.805	9.60×10^{-7}
Panel C: Choice of window size					
Models with 3-month moving averaged media indicators	Attention	44.88	0.649	4.407	4.99×10^{-8}
	Attention shock	44.75	0.651	4.387	5.51×10^{-8}
	Concern	45.45	0.643	4.500	3.19×10^{-8}
	Concern shock	44.07	0.659	4.279	9.29×10^{-8}

index based on the new lexicon still increases the accuracy of the benchmark model at 1%-level significance, indicating that the conclusion is robust to the alternative choice of lexicon.

6.3. Robustness to smoothing

The text-based indicators exhibit a high degree of volatility. A smoother indicator can be obtained through moving average approach. We take a 3-month (90-day) moving average of the media attention and concern indices to mitigate the volatility of the indices. The evolution of the smoothed indicators is shown in Figures 4 and 5. The missing data in the news-based indices produced by the moving average approach are replaced with the original monthly data. Using the new smoothed indicators, similar results are yielded: the conclusion of accuracy gains is robust to smoothing. We report the results in Panel C of Table 7.

6.4. Robustness to the choice of topics

In this part, we test the robustness of indices under certain real estate topics given by the LDA model in Section 4.1. To calculate the media indices under a certain topic x, we first calculate the topic-weighted media indices at article level, which are set at the product of the prevalence of topic x and the original article-level media indices in Section 4.2:

$$ARE_{s,m,i,x} = prevalence_{s,m,i,x} \times ARE_{s,m,i}, \tag{6.1}$$

$$CORE_{s,m,i,x} = prevalence_{s,m,i,x} \times CORE_{s,m,i}, \tag{6.2}$$

where $prevalence_{s,m,i,x}$ represents for the prevalence of topic x for the ith article published by the source s in the month m, while $ARE_{s,m,i}$ and $CORE_{s,m,i}$ stand for the original media attention index and the concern index at article level, respectively. For the processes of aggregation and standardization, we use the same workflow as the one in Section 4.2.

We report the goodness-of-fit statistics of the models with the topic-specific media attention index and concern index in Table 8. It can be observed that the indicators under the topics of supply-side reform, finance and affordable housing are all informative in predicting the ST proportion series, as the models containing the indicators outperform the benchmark model significantly. Moreover, the media attention indicators under the topic of real estate market are robust, whereas the concern indicators are not since higher accuracy is not achieved after incorporating the concern indicators into the benchmark model. Similarly, all the media indicators weighted by the topic regarding the economy have limited performance in the forecast of the ST series. This result suggests that the noise in predicting the dynamics of the ST series mainly stemmed from articles under the topics of China's economy.

Table 8. Sensitivity of goodness-of-fit statistics to the choice of topics.

		Goodness-of-fit statistics			
Topics	News indices	Adj. R^2 (%)	RSS (%)	F-test	p-value
Benchmark model		10.48	1.291	—	—
Supply-side reform	Attention	34.07	0.777	2.953	7.19×10^{-5}
	Attention shock	34.28	0.774	2.977	6.38×10^{-5}
	Concern	18.03	0.966	1.503	0.082
	Concern shock	18.64	0.959	1.548	0.068
Finance	Attention	20.24	0.940	1.668	0.040
	Attention shock	20.71	0.934	1.705	0.034
	Concern	19.60	0.947	1.620	0.050
	Concern shock	20.09	0.942	1.656	0.043
China's economy	Attention	12.24	1.034	1.110	0.346
	Attention shock	13.07	1.024	1.163	0.293
	Concern	4.80	1.122	0.675	0.866
	Concern shock	5.93	1.109	0.736	0.804
Affordable housing	Attention	38.55	0.724	3.493	4.60×10^{-6}
	Attention shock	36.40	0.749	3.225	1.80×10^{-5}
	Concern	37.32	0.739	3.338	1.01×10^{-5}
	Concern shock	35.95	0.755	3.171	2.36×10^{-5}
Real estate market	Attention	21.56	0.924	1.771	0.025
	Attention shock	22.22	0.916	1.825	0.020
	Concern	11.50	1.043	1.063	0.397
	Concern shock	12.44	1.032	1.122	0.333

6.5. Robustness to the definition of the dependent variable

In this part, we change the dependent variable, the ST proportion series, to a monthly series of market capitalization weighted proportion of ST real estate firms in China. We calculate the series with the following equation:

$$MCST_t = \frac{\sum_{i=1}^{N_t} MarCap_{i,t} \times ST_{i,t}}{\sum_{i=1}^{N_t} MarCap_{i,t}}, \tag{6.3}$$

where $MCST_t$ is the proportion of the market capitalization that ST real estate firms take up at time t. $MarCap_{i,t}$ is the market capitalization of real estate firm i at time t. $ST_{i,t}$ equals 1 when the firm is in ST status at time t; otherwise, it equals 0. N_t represents the total number of listed real estate firms at time t. The advantage of the market capitalization weighted series over the equal-weighting approach is that it reflects better the financial market impact. In contrast, equal-weighting has the advantage that it can be interpreted as the probability that a listed real estate firm has the ST label in a given month.

We redo the regression workflow using the series of market capitalization proportions and report the results in Table 9. The conclusion of accuracy improvement using the media indicators is robust to the prediction of the market capitalization of ST real estate firms, with a decrease of at least 26% in the residual sums of square when adding the media indicators to the regression.

Table 9. Sensitivity of goodness-of-fit statistics to changing the dependent variable to the market capitalization weighted proportion of real estate firms with ST label on the Shenzhen and Shanghai exchanges.

		Goodness-of-fit statistics			
Models	News indices	Adj. R^2 (%)	RSS (%)	F-test	p-value
Benchmark model		15.01	0.198	—	—
Models with media indicators	Attention	23.90	0.145	1.638	0.046
	Attention shock	25.16	0.142	1.740	0.029
	Concern	23.75	0.145	1.626	0.049
	Concern shock	26.87	0.139	1.885	0.015

7. Financial evaluation of the real estate media indices

Investing in Chinese real estate stocks is risky. Over our period of analysis, the main Chinese real estate stock index, the CSI 300 Real Estate Index, has first increased by 276.4% (from 3077 to 11304), and then decreased by 294.6% (from 11304 to 2865). In this section, we study whether the media concerns about real estate are informative or not to decide on investing in or disinvesting from Chinese real estate stocks. This risk on/risk off approach is called tactical asset allocation (TAA). For example, Hull, Qiao, and Bakosova (2017) transform the values of a forecast model into a market-timing strategy to obtain higher returns. We believe that the environment of the Chinese real estate market will become worse if the values of the concern index and its shock exceed a certain threshold due to the negative relationship between the market condition and the media concerns. Thus, we recommend that investors focus on risk-free assets to avoid losses when the values of media concerns increase. We expect higher returns in the process where the concern index and its shock function as indicators of trends in the Chinese real estate stock market. Furthermore, we do not use the media attention index because media attention does not have a definite tone on the real estate sector. Although negative events tend to be more covered by news media, as pointed out by Soroka and McAdams (2015) and Lowry (2008), it is still possible that positive events in the real estate sector are reported in newspapers, which means the market environment becomes better simultaneously with an increase of the media attention.

7.1. CSI 300 real estate index

We select the CSI 300 Real Estate Index as the target real estate stock index. The index, a subindex of the CSI 300 index, consists of all samples of real estate stocks with large capitalization and good liquidity in the CSI 300 index. The CSI 300 investment universe is adjusted semiannually, in which stocks that violate laws or regulations or have significant issues with their financial statements are removed. Consequently, the CSI 300 Real Estate Index is able to mirror the performance of the real estate firms listed on the Shanghai and Shenzhen exchanges. The entire period that we are interested in is January 2009–June 2024, and we download the return data for the CSI 300 Real Estate index from the RESSET database.

7.2. Market trend indicators

We use the concern index and its shock to capture trends in the Chinese real estate stock market. However, in Section 4.3, we quantify the shock of concern as the error terms of the ARMA (1,1) model, which is estimated using the data in the entire period. To avoid forward-looking bias, we implement an expanding window estimation approach. Our first estimation sample covers the period 2009–2012. Then, the model undergoes monthly re-estimation as new data becomes available. We repeat this step until the entire period is covered. Furthermore, to obtain enhanced robustness against potential model instability, we also use a rolling window framework of 36 months, which systematically discards the earliest data when integrating new observations, to obtain the regression estimates and the corresponding residuals.

To track the market trends, we follow the logic of moving-average-based trading systems. For example, Faber (2007) proposes to reduce risks through a market-timing model, in which the buy and sell strategies are based on the 10-month moving average of price. Thus, in our study, the threshold for the concern index is set at its moving average of 6 months:

$$k_{CORE,t} = \frac{1}{6}\sum_{s=t-5}^{t} CORE_s, \quad t \geq 6 \tag{7.1}$$

where $k_{CORE,t}$ represents the threshold for the concern index in the tth month, and $CORE_s$ stands for the values of the concern index in the sth month. t takes the value from 6 to the total number of months throughout the period. For the shock of concern, the threshold is set at 0 to capture the case that unexpected changes increase the media concern.

7.3. The tactical asset allocation process

In this part, we report the TAA process of the CSI 300 Real Estate Index, using the concern index and its shock to track market trends. We first calculate a cumulative buy-and-hold and a stop-loss wealth series of the investment in the real estate stock index as benchmarks. The buy-and-hold wealth is calculated as follows:

$$BW_n = \prod_{t=1}^{n}(1 + R_t), \tag{7.2}$$

where BW_n is the cumulative wealth in the nth month, and R_t is the monthly return of the CSI 300 Real Estate Index in the tth month. For the stop-loss strategy, we apply a rule that yearly resets the floor of the cumulative wealth as a loss threshold. Studies have documented the validity of such stop-loss rules, such as Kaminski and Lo (2014). Our rule assumes that investors will stop investing in the CSI 300 Real Estate Index next month if the cumulative wealth of this month is lower than 90% of its initial value in the year. Moreover, at the start of each year, the investment automatically resumes, and the loss threshold is reset based on the starting wealth of the new year.

$$StopLoss_t = \begin{cases} 1, & w_t \geq 0.9 \times w_t^* \\ 0, & otherwise, \end{cases} \tag{7.3}$$

where $StopLoss_t$ represents the decision made by investors based on the stop-loss strategy, and w_t^* stands for the wealth at the beginning of the year for time t. Lastly, we calculate the cumulative wealth series of the stop-loss strategy:

$$SLW_n = \prod_{t=2}^{n}(1 + StopLoss_{t-1} \times R_t). \tag{7.4}$$

Second, we incorporate the concern index and its shock into the TAA process to monitor the trends in the real estate market. Our trading strategy is that investors continue to invest in the real estate stock index when the values of the concern index and its shock are not greater than their threshold k simultaneously, due to the negative relationship between the market trends and media concern. Otherwise, investors will invest in risk-free asset, and the corresponding return will be set at 0.

$$MarCond_t = \begin{cases} 1, & CORE_t \leq k_{CORE,t} \ and \ shock_t \leq 0 \\ 0, & otherwise, \end{cases} \tag{7.5}$$

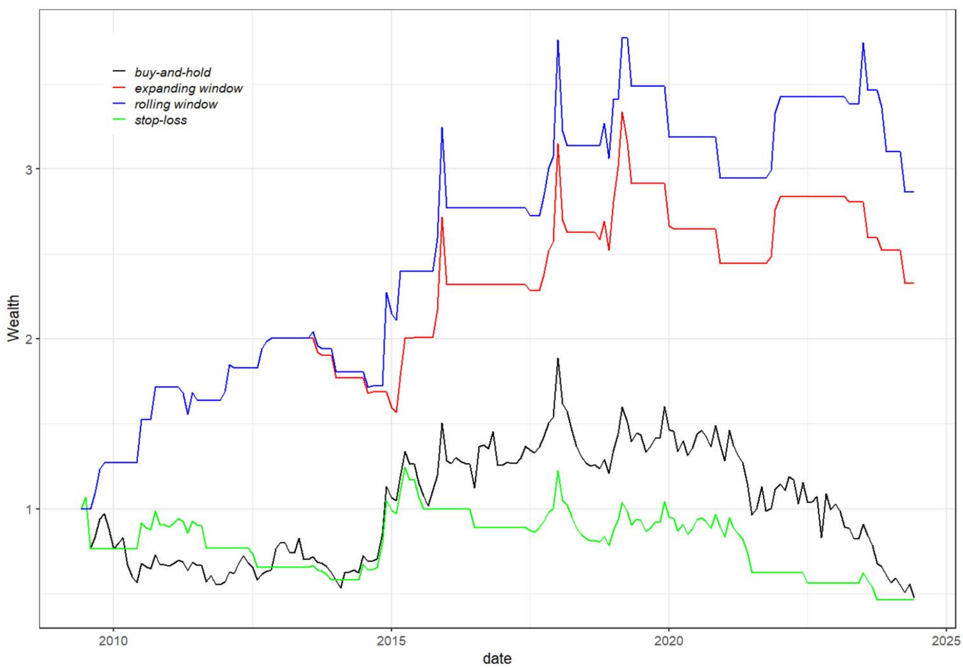

Figure 7. Wealth evolution for the buy-and-hold (black line), stop-loss (green line) and tactical asset allocation (red and blue lines) investments in the CSI 300 real estate index.

where $MarCond_t$ reflects whether an investor will invest in the real estate index or not in the tth month.

We then calculate the cumulative wealth of TAA under the monitoring of the concern index and its shock and plot the evolution of the buy-and-hold, stop-loss and TAA wealth of \$1 in Figure 7.

$$TW_n = \prod_{t=2}^{n}(1 + MarCond_{t-1} \times R_t), \tag{7.6}$$

where TW_n stands for the cumulative wealth of TAA in the nth month.

We find that both the buy-and-hold and stop-loss wealth declines from \$1 to \$0.47 from 2009 to June of 2024, which clearly demonstrates the ineffectiveness of the stop-loss strategy. Regarding the TAA processes, the wealth utilizing the expanding window strategy rises from \$1 to \$2.33, significantly outperforming the buy-and-hold and stop-loss strategies in terms of final wealth. Furthermore, the rolling window strategy yields a final wealth of \$2.86, further increasing the final wealth by 23% compared to the expanding window strategy. A distinct difference between the two TAA wealth series and the benchmarks emerges in 2020 and afterward, following the outbreak of COVID-19. The CSI 300 Real Estate Index has suffered a drawdown over 60% since the start of 2020, a loss that has not been offset by the stop-loss strategy. Conversely, large drawdowns have not occurred in the TAA wealth series due to the safeguard of the concern index and its shock. Likewise, the buy-and-hold and stop-loss wealth shows decreasing trends from the beginning of the period to 2014, while the TAA wealth series are on the rise with the monitoring of the concern index and its shock. These results justify the financial value embedded in the concern index and its shock.

8. Conclusion

The Chinese real estate industry, as a pillar industry of the economy, has attracted numerous investments. While investing in Chinese real estate firms offers an opportunity to diversify portfolios, it also comes with a high time-varying investment risk. An ex post indicator of financial distress in the listed Chinese real estate sector is the proportion of firms with the status of Special Treatment (ST). Firms receive this status because of financial distress manifested in their annual report or because of events such as the interruption of operational activities due to a lawsuit or losing access to the bank account. Many of the events that drive the likelihood of ST status are covered in news articles, both at the macro-economic (business cycle, policy) and firm specific level. Hence, news articles can be regarded as a potentially more timely source of information. In this paper, we empirically test whether news articles provide incremental value for the prediction of the financial situation of the Chinese real estate sector using textual analysis and a linear regression framework. Furthermore, we evaluate the financial value of media concerns about real estate for investors through a tactical asset allocation investment analysis.

Our dependent variable of interest is the evolution of the proportion of real estate firms that are under special treatment status. We are interested in this variable because the proportion reflects the financial distress in the real estate sector. We confirm that over the period January 2009–June 2024, the ST proportion peaks when real estate events occur—for example, at the beginning of 2010 when the purchase restriction order on real estate was released. As a leading indicator for the proportion of ST real estate firms, we introduce two types of news-based indicators, using news articles published in the *Economic Daily*, *Guangming Daily*, *China Youth Daily* and *Beijing Daily*. The first index tracks media attention devoted to real estate, while the second one monitors concerns about real estate.

Next, we test whether the media attention index, the concern index and their shocks provide incremental value in forecasting the current ST proportion by a linear regression framework with certain macroeconomic variables and lags of the ST series being taken into account. The shocks are extracted from the ARMA (1,1) models for the indices. We first estimate a benchmark model using only the lag values of the ST series and the control variables and the models which further include the indices and their shocks. We then compare the goodness of fit between the models and discover that including the news-based indices and their shocks in the benchmark model increases the goodness of fit significantly. The results are robust to the checks in Section 6. Therefore, we can confirm that the media indicators have incremental value in the forecast of the financial situation, indicating that news articles can be regarded as a timely source of financial information with incremental value for the Chinese real estate sector.

Lastly, we examine the financial value of the concern index and its shock to investors by applying them to the TAA processes of the CSI 300 Real Estate Index. We document that TAA wealth grows from 1 to 2.33 and 2.86 under the monitoring of the concern index and its shock, using expanding and rolling window strategies, respectively. However, the wealth invested in the buy-and-hold and stop-loss strategies drops from 1 to about 0.47, which is significantly lower than the TAA wealth. Therefore, we confirm that the concern index and the shock have the potential to be useful indicators of risk on/off investment regimes in the Chinese real estate stock market.

Overall, the proposed news-based Chinese real estate attention and concern indices can be seen as coincident indicators of the financial situation in the Chinese real estate sector. As such, they can be a potentially useful tool for a timely interpretation of the financial situation in the Chinese real estate sector and a new viewpoint for bank managers to gain insights. In addition, the concern index and its shock also have the potential to guide tactical asset allocation when investing in listed Chinese real estates. A promising direction for further research is to use the proposed indicators in a machine learning context with big data and to update their values at a higher frequency, such as weekly and daily, to obtain a more timely forecast. Moreover, application to other sectors, especially the financial sectors, would also be interesting. We leave these questions to future studies.

Notes

1. https://www.sse.com.cn/lawandrules/sselawsrules/stocks/mainipo/c/c_20250425_10777756.shtml

2. https://docs.static.szse.cn/www/lawrules/rule/stock/supervision/mb/W020250425730780478873.pdf
3. Similarly, to identify valence-shifting words, we use a lexicon constructed through the same workflow.
4. As suggested by a reviewer, an alternative approach is to use LLM sentiment method. To validate our analysis, we randomly select 1000 articles from the corpus. DeepSeek, a LLM developed by a Chinese firm named Hangzhou DeepSeek Artificial Intelligence, executes the task of counting positive and negative words in each article. We calculate the Spearman correlation of 0.72 between the concern scores derived by lexicon-based method and DeepSeek. This result indicates the consistency between our lexicon-based methods and the LLM sentiment method.

Acknowledgements

We thank participants of the 2023 Belgian Financial Research Forum, David Ardia, Lieven Baele, the editor (Jian Yu) and two anonymous reviewers for helpful feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the China Scholarship Council (#202007650028) and the Research Foundation Flanders (#G014420N).

Notes on contributors

Kris Boudt is professor at Ghent University, Vrije Universiteit Brussel and Amsterdam, and founder CEO of Sentometrics, a university spinoff specialized in news media monitoring. He is a guest lecturer at Sichuan University. He obtained his PhD from KU Leuven in 2008. His research has been published in International Journal of Forecasting, Journal of Econometrics, Management Science and Review of Finance, among others.

Xianda Liang is a PhD candidate at Vrije Universiteit Brussel. He received the B.S. degree in science and the M.S. degree in economics from Renmin University of China in 2017 and 2020, and a M.S. degree in management from Vrije Universiteit Brussel. His research interests include econometrics, financial risk management and natural language processing.

ORCID

Kris Boudt http://orcid.org/0000-0002-1000-5142
Xianda Liang http://orcid.org/0000-0003-1568-8744

References

Algaba, A., D. Ardia, K. Bluteau, S. Borms, and K. Boudt. 2020. “Econometrics Meets Sentiment: An Overview of Methodology and Applications.” *Journal of Economic Surveys* 34 (3): 512–547. https://doi.org/10.1111/joes.12370.

Ardia, D., K. Bluteau, S. Borms, and K. Boudt. 2021. "The R Package Sentometrics to Compute, Aggregate and Predict with Textual Sentiment." *Journal of Statistical Software* 99 (2): 1–40. https://doi.org/10.18637/jss.v099.i02.
Ardia, D., K. Bluteau, K. Boudt, and K. Inghelbrecht. 2023. "Climate Change Concerns and the Performance of Green vs. Brown Stocks." *Management Science* 69 (12): 7607–7632. https://doi.org/10.1287/mnsc.2022.4636.
Baker, S. R., N. Bloom, and S. J. Davis. 2016. "Measuring Economic Policy Uncertainty." *The Quarterly Journal of Economics* 131 (4): 1593–1636. https://doi.org/10.1093/qje/qjw024.
Ball-Rokeach, S. J., and M. L. DeFleur. 1976. "A Dependency Model of Mass-Media Effects." *Communication Research* 3 (1): 3–21. https://doi.org/10.1177/009365027600300101.
Beracha, E., M. Lang, and J. Hausler. 2019. "On the Relationship between Market Sentiment and Commercial Real Estate Performance—A Textual Analysis Examination." *Journal of Real Estate Research* 41 (4): 605–638. https://doi.org/10.22300/0896-5803.41.4.605.
Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022.
Bork, L., S. V. Møller, and T. Q. Pedersen. 2020. "A New Index of Housing Sentiment." *Management Science* 66 (4): 1563–1583. https://doi.org/10.1287/mnsc.2018.3258.
Borms, S., K. Boudt, F. Van Holle, and J. Willems. 2021. "Semi-Supervised Text Mining for Monitoring the News about the ESG Performance of Companies." In *Data Science for Economics and Finance: Methodologies and Applications*, edited by Sergio Consoli, Diego Reforgiato Recupero, and Michaela Saisana, 217–239. Cham: Springer International Publishing.
Boudt, K., A. De Block, G. Langenus, and P. Reusens. 2023. "Nowcasting GDP through the Lens of Economic States." NBB Working Paper.
Cai, J. 2019. "又一百强房企倒下! 地产业洗牌加速 [Another Top 100 Real Estate Enterprise Fall: The Acceleration of Real Estate Reshuffle]." *Real Estate Guide* 7, 44-47.
Caldara, D., and M. Iacoviello. 2022. "Measuring Geopolitical Risk." *American Economic Review* 112 (4): 1194–1225. https://doi.org/10.1257/aer.20191823.
Cheng, W., and Y. Jiang. 2023. "Falling in the Same Predicament Again? A Network Embeddedness Perspective of Organizational Failure Recidivism in Special Treatment Firms." *Management Decision* 61 (3): 815–837. https://doi.org/10.1108/MD-11-2021-1426.
Faber, M. 2007. "A Quantitative Approach to Tactical Asset Allocation." *The Journal of Wealth Management* 9: 69–79
Fang, H. 2018. "中国房地产发展与物业服务创新历程 [Real Estate Development and Property Service Innovation in China]." *China Property Management* 10: 26–27.
Flaxman, S., S. Goel, and J. M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80 (S1): 298–320. https://doi.org/10.1093/poq/nfw006.
Gao, Q., and T. Zhao. 2018. "The Influence of Home Buyer Sentiment on Chinese Housing Prices––Based on Media Text Mining." *International Journal of Economics and Finance* 10 (9): 145. https://doi.org/10.5539/ijef.v10n9p145.
Hausler, J., J. Ruscheinsky, and M. Lang. 2018. "News-Based Sentiment Analysis in Real Estate: A Machine Learning Approach." *Journal of Property Research* 35 (4): 344–371. https://doi.org/10.1080/09599916.2018.1551923.
Hu, Z., A. M. Kutan, and P.-W. Sun. 2018. "Is US Economic Policy Uncertainty Priced in China's A-Shares Market? Evidence from Market, Industry, and Individual Stocks." *International Review of Financial Analysis* 57: 207–220. https://doi.org/10.1016/j.irfa.2018.03.015.
Huang, Y. 2020. "Market Sentiments and the Housing Markets." PhD diss. Virginia Tech.
Huang, Y., and P. Luk. 2020. "Measuring Economic Policy Uncertainty in China." *China Economic Review* 59: 101367. https://doi.org/10.1016/j.chieco.2019.101367.
Hull, B., X. Qiao, and P. Bakosova. 2017. "Return Predictability and Market-Timing: A One-Month Model." *Journal of Investment Management* 17 (2019): 47–64.

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. "Manager Sentiment and Stock Returns." *Journal of Financial Economics* 132 (1): 126–149. https://doi.org/10.1016/j.jfineco.2018.10.001.

Kaminski, K. M., and A. W. Lo. 2014. "When Do Stop-Loss Rules Stop Losses?" *Journal of Financial Markets* 18: 234–254. https://doi.org/10.1016/j.finmar.2013.07.001.

Kim, M. H.-Y., S. Ma, and Y. A. Zhou. 2016. "Survival Prediction of Distressed Firms: Evidence from the Chinese Special Treatment Firms." *Journal of the Asia Pacific Economy* 21 (3): 418–443. https://doi.org/10.1080/13547860.2016.1176645.

Koelbl, M. 2020. "Is the MD&A of US Reits Informative? A Textual Sentiment Study." *Journal of Property Investment & Finance* 38 (3): 181–201. https://doi.org/10.1108/JPIF-12-2019-0149.

Li, J., Y. Wang, and C. Liu. 2022. "Spatial Effect of Market Sentiment on Housing Price: Evidence from Social Media Data in China." *International Journal of Strategic Property Management* 26 (1): 72–85. https://doi.org/10.3846/ijspm.2022.16255.

Li, Z., J. Crook, G. Andreeva, and Y. Tang. 2021. "Predicting the Risk of Financial Distress Using Corporate Governance Measures." *Pacific-Basin Finance Journal* 68: 101334. https://doi.org/10.1016/j.pacfin.2020.101334.

Lowry, D. T. 2008. "Network TV News Framing of Good vs. Bad Economic News under Democrat and Republican Presidents: A Lexical Analysis of Political Bias." *Journalism & Mass Communication Quarterly* 85 (3): 483–498. https://doi.org/10.1177/107769900808500301.

Manela, A., and A. Moreira. 2017. "News Implied Volatility and Disaster Concerns." *Journal of Financial Economics* 123 (1): 137–162. https://doi.org/10.1016/j.jfineco.2016.01.032.

McCombs, M. E., and D. L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *Public Opinion Quarterly* 36 (2): 176–187. https://doi.org/10.1086/267990.

Nimark, K. P., and S. Pitschner. 2019. "News Media and Delegated Information Choice." *Journal of Economic Theory* 181: 160–196. https://doi.org/10.1016/j.jet.2019.02.001.

Nowak, A., and P. Smith. 2017. "Textual Analysis in Real Estate." *Journal of Applied Econometrics* 32 (4): 896–918. https://doi.org/10.1002/jae.2550.

Park, M. J., D. Choi, I. C. Moon, and S. Park. 2016. "Does News Affect the Housing Market?" Available at SSRN 2754426.

Paulus, N. M., M. Koelbl, and W. Schaefers. 2022. "Can Textual Analysis Solve the Underpricing Puzzle? A US Reit Study." *Journal of Property Investment & Finance* 40 (6): 548–570. https://doi.org/10.1108/JPIF-06-2021-0052.

Plössl, F., and T. Just. 2024. "News Coverage vs Sentiment: Evaluating German Residential Real Estate Markets." *International Journal of Housing Markets and Analysis* 17 (2): 395–417.

Plössl, F., T. Just, and L. Wehrheim. 2021. "Cyclicity of Real Estate-Related Trends: Topic Modelling and Sentiment Analysis on German Real Estate News." *Journal of European Real Estate Research* 14 (3): 381–400.

Plössl, F., N. M. Paulus, and T. Just. 2023. "Trade vs. Daily Press: The Role of News Coverage and Sentiment in Real Estate Market Analysis." *Journal of Property Research* 40 (4): 333–364.

Ruscheinsky, J. R., M. Lang, and W. Schäfers. 2018a. "Real Estate Media Sentiment through Textual Analysis." *Journal of Property Investment & Finance* 36 (5): 410–428. https://doi.org/10.1108/JPIF-07-2017-0050.

Ruscheinsky, J., K. Kandlbinder, W. Schaefers, M. Dietzel, and K. Rochdi. 2018b. "Predicting Real Estate Market Movements: The First Textual Analysis-Based-Sentiment Application in Germany." Working Paper.

Shapiro, A. H., M. Sudhof, and D. J. Wilson. 2022. "Measuring News Sentiment." *Journal of Econometrics* 228 (2): 221–243. https://doi.org/10.1016/j.jeconom.2020.07.053.

Soo, C. K. 2018. "Quantifying Sentiment with News Media across Local Housing Markets." *The Review of Financial Studies* 31 (10): 3689–3719. https://doi.org/10.1093/rfs/hhy036.

Soroka, S., and S. McAdams. 2015. "News, Politics, and Negativity." *Political Communication* 32 (1): 1–22. https://doi.org/10.1080/10584609.2014.881942.
Tao, Q., Y. Sun, Y. Zhu, and X. Yang. 2017. "Political Connections and Government Subsidies: Evidence from Financially Distressed Firms in China." *Emerging Markets Finance and Trade* 53 (8): 1854–1868. https://doi.org/10.1080/1540496X.2017.1332592.
Walker, C. B. 2016. "The Direction of Media Influence: Real-Estate News and the Stock Market." *Journal of Behavioral and Experimental Finance* 10: 20–31. https://doi.org/10.1016/j.jbef.2016.02.001.
Yao, J., X. Feng, Z. Wang, R. Ji, and W. Zhang. 2021. "语调, 情绪及市场影响: 基于金融情绪词典 [Tone, Sentiment and Market Influence Based on Financial Sentiment Lexicon]." *Journal of Management Sciences in China* 24 (5): 26–46.
Yu, J., X. Shi, D. Guo, and L. Yang. 2021. "Economic Policy Uncertainty (EPU) and Firm Carbon Emissions: Evidence Using a China Provincial EPU Index." *Energy Economics* 94: 105071. https://doi.org/10.1016/j.eneco.2020.105071.
Zhang, L., E. I. Altman, and J. Yen. 2010. "Corporate Financial Distress Diagnosis Model and Application in Credit Rating for Listing Firms in China." *Frontiers of Computer Science in China* 4 (2): 220–236. https://doi.org/10.1007/s11704-010-0505-5.
Zhou, F., L. Fu, Z. Li, and J. Xu. 2022. "The Recurrence of Financial Distress: A Survival Analysis." *International Journal of Forecasting* 38 (3): 1100–1115. https://doi.org/10.1016/j.ijforecast.2021.12.005.
Zhu, E., J. Wu, H. Liu, and K. Li. 2023. "A Sentiment Index of the Housing Market in China: Text Mining of Narratives on Social Media." *The Journal of Real Estate Finance and Economics* 66 (1): 77–118. https://doi.org/10.1007/s11146-022-09900-5.

Appendix. Topics discussed in real estate articles

Table A1 exhibits the LDA topic distribution of the refined corpus used in our study. We also report 10 most probable keywords and topic prevalence for each topic. The topic prevalence is calculated as the topic weight of each topic in the whole corpus: the sum of the prevalence of every article divided by the total number of articles.

Table A1. Topics discussed in real estate articles.

Topic	Keywords	Prevalence
1. Government finance	Government, local, firm, policy, fund, tax, finance, budget, expense, income	0.023
2. Supply-side reform	Economy, development, reform, increase, the country, consumption, investment, policy, market, accelerate	0.044
3. Online shopping	Consumption, service, COVID-19, platform, data, information, the internet, consumer, network, product	0.027
4. Green building	Building, garbage, project, energy conservation, green, construct, technology, energy, pollution, build	0.026
5. Real estate lawsuit	Law, court, case, house, contract, firm, perform, property, judiciary, relocation	0.037
6. Politics	Development, the people, Jinping Xi, China, the party, reform, chairman, construction, socialism, society	0.033
7. Global economy	China, the US, country, economy, world, international, global, cooperation, Hong Kong, development	0.029
8. Social equality	Society, government, China, the country, economy, local, research, investigation, process, influence	0.052
9. Business	firm, company, group, investment, program, market, industry, government, limited company, China	0.032
10. Finance	Loan, finance, bank, risk, financing, fund, asset, market, interest rate, support	0.035
11. Technology industry	Development, industry, innovation, technology, construction, talent, service, economy, technology	0.039

(*continued*)

Table A1. Continued.

Topic	Keywords	Prevalence
12. China's economy	Growth, year-on-year, investment, growth rate, economy, decrease, percentage, month, price, industry	0.040
13. Travel	Travel, ecology, development, construction, tourist, feature, city, culture, protect, park	0.024
14. Chinese ethos	Tibet, work, nationality, Xinjiang, the mass, the people, troop, Jiabao Wen, the country, spirit	0.021
15. Student	Teenagers, the talent, education, school, student, start business, job, job hunting, teacher, university	0.027
16. System reform	Construction, reform, system, management, advance, society, improve, mechanism, build, development	0.039
17. Affordable housing	House, home, build, program, rent, guarantee, housing rental, affordable family, land	0.038
18. Natural disaster	Earthquake, rebuild, disaster area, the mass, disaster, rescue, meet emergency, suffer, recovery, work	0.034
19. Administration	Work, cadre, leader, enforce, build, supervise, committee, report, conference, suggestion	0.031
20. Support the poor	Rural area, villager, farmer, countryside, poverty, support the poor, agriculture, get rid of poverty, land, development	0.032
21. Legal affair	Limited company, announcement, deliver, by law, as of, regard, book, current, judge, notification	0.008
22. Neighborhood	Neighborhood, resident, owner, property, remake, street, elevator, community, house, residence	0.042
23. Company affair	Department, register, information, handle, service, unit, examine and approve, accumulation fund, firm, apply	0.034
24. Real estate market	Real estate, market, city, house price, price, home, policy, regulation, residence, Beijing	0.039
25. Art	Culture, China, movie, life, art, shanghai, story, composition, compose, spirit	0.027
26. Culture	Culture, history, building, protect, tradition, antique, hutong, ruins, Beijing, China	0.027
27. Life	Child, life, the aged, house, time, money, job, home, find, raise	0.069
28. Urbanization	Development, population, job hunting, city, society, rural area, urbanization, city and countryside, income, migrant worker	0.030
29. Community service	Community, service, the mass, provide for the aged, work, activity, organization, grass-root unit, society, solve	0.030
30. Urban development	City, Beijing, construction, layout, center, program, the capital, the city of Beijing, development, area	0.033