The Role of Individual Investor Sentiment as a Factor in the Chinese Stock Market

Margaretha Hirsch

Dissertation written under the supervision of Mengdi Gu

Dissertation submitted in partial fulfilment of requirements for the MSc in International Finance at Universidade Católica Portuguesa and for the MSc in International Business at Antai College of Economics and Management, at Shanghai Jiao Tong University, 12.09.2024.

Abstract

Over the past 30 years, the rise of the Chinese economy has elevated the significance of its financial markets. With more data becoming available and regulations improving efficiency, research on multifactor models explaining stock returns in this retail investor-dominated ecosystem becomes feasible. However, the most prominent model, CH-3, employs an extreme filter by excluding the lowest 30% of stocks by market capitalization to reduce the impact of shell value through reverse merger IPOs – a practice dominant in the early 2000s.

This thesis introduces an individual sentiment factor by running a horserace on social media-based proxies. To analyse the performance of the new four-factor CH-G model, it is tested against a replicated CH-3 model across three subsamples, with a focus on including and excluding microcaps.

The CH-G better captures the characteristics of the Chinese stock market in the 2009 to 2023 period, with the sentiment factor proving significant across panels. The model demonstrates higher average adjusted R-squares and notable explanatory power, especially when including all individual stocks instead of excluding the smallest 30%. This supports the argument that excluding microcaps may no longer be appropriate in the evolving Chinese market.

TITLE:	"The Role of Individual Investor Sentiment as a Factor in the Chinese Stock Market"
AUTHOR:	Margaretha Hirsch
KEY WORDS:	Chinese stock returns, multifactor models, CH-3, individual investor sentiment

Resumo

Nos últimos 30 anos, o crescimento da economia chinesa aumentou a relevância de seus mercados financeiros. Com a maior disponibilidade de dados e a melhoria das regulamentações para aumentar a eficiência, torna-se viável a pesquisa de modelos multifatoriais que expliquem os retornos de ações neste ecossistema dominado por investidores de retalho. No entanto, o modelo mais proeminente, CH-3, emprega um filtro extremo excluindo 30% das observações mais baixas do stock de ações por capitalização de mercado, para reduzir o impacto do shell value por meio de IPOs de fusão — uma prática dominante no início dos anos 2000.

Esta tese introduz um fator de perceção individual, realizando uma horserace com proxies baseados em redes sociais. Para analisar o desempenho do novo modelo de quatro fatores CH-G, o mesmo será testado em relação a um modelo CH-3 replicado em três subamostras, com foco na inclusão e exclusão de microcaps.

O CH-G captura melhor as características do mercado de ações chinês no período de 2009 a 2023, com o fator de perceção mostrando-se significativo em todos os painéis. O modelo demonstra médias mais altas de R2 ajustado e impacto explicativo, especialmente ao incluir todas as ações individuais, em vez de excluir as 30% menores, apoiando o argumento de que a exclusão de microcaps pode não ser a mais apropriada num mercado chinês em evolução.

TÍTULO: "The Role of Individual Investor Sentiment as a Factor in the Chinese Stock Market"

AUTOR: Margaretha Hirsch

PALAVRAS-CHAVE: retornos de ações chinesas, modelos multifatoriais, CH-3, sentimento de investidores individuais

Table of Contents

Abbreviations

BW	Baker-Wugler (Index)
CAPM	Capital Asset Pricing Model
CICSI	China Investor Composite Sentiment Index
CNRDS	Chinese Research Data Services Platform
CSMAR	China Stock Market & Accounting Research Database
CSRC	China Securities Regulatory Commission
DSSW	Delong, Shleifer, Summers, and Waldmann
EP	Earnings-per-price
GRS	Gibbons, Ross, and Shanken
IPO	Initial Public Offering
MPT	Modern Portfolio Theory
NYSE	New York Stock Exchange
SSE	Shanghai Stock Exchange
SSECI	Shanghai Stock Exchange Composite Index
SZSE	Shenzhen Stock Exchange
U.S.	United States (of America)

I. Introduction

"Studies of the sentiment of investors are important for two reasons. First, they teach us about biases in the stock market forecasts of investors. Second, they teach us about opportunities to earn extra returns by exploiting those biases."

(Fisher and Statman 2000)

Most researchers have been focusing on financial market dynamics and asset pricing in the U.S. market due to a long data history and a broad research body (Griffin 2002). However, with the Chinese market emerging and its trading history exceeding 30 years (SSE 2024b; SZSE 2024), research becomes more feasible and exigent. Yet, many scholars rely on Fama-French or similar US-centric models to explain Chinese stock returns or use them as a base for adaption and expansion (Cheung, Hoguet, and Ng 2014; Hou, Xue, and Zhang 2015; Lin 2017; T.-L. Huang 2018; Han and Shi 2022). Additionally, the most prominent asset pricing model comes with an extreme presumption: the CH-3 by Liu et al (2019), simply drops the smallest 30% of firms by market capitalization. The reason behind this is the attempt to exclude shell value contaminated stocks caused by speculations on reverse mergers, which is a practice caused by the strict IPO regulatory environment of the early-2000s.

Indeed – the Chinese stock market is unique not only because of its youth, laws and regulations and fast dynamics, but also because it is dominated by individual investors (X. Han and Li 2017; Carpenter and Whitelaw 2017; X. G. Hu, Pan, and Wang 2021; Ren et al. 2021; Hou, Qiao, and Zhang 2023). This introduces another pillar of research: traditionally, asset pricing models were built on the assumption that market participants behave rationally, making decisions purely based on objective data (Fama 1970; 1991). However, as the field of behavioural finance has evolved, it has become clear that human emotions and biases significantly influence market behaviour (Kahneman and Tversky 1979; Shefrin and Statman 2000). Thus, a growing body of research that integrates sentiment into asset pricing models, recognizing that emotions reflected in financial news, social media, and other sources can drive investor decisions and, consequently, market dynamics (Baker and Wurgler 2006; Ge et al. 2020; Shapiro, Sudhof, and Wilson 2022), with retail investors being especially prone to investing without an underlying rationale (Bu and Pi 2014).

This thesis acknowledges this emerging trend by exploring sentiment as a critical factor in asset pricing models, particularly within the context of the Chinese stock market. In doing so, it acknowledges the unique characteristics of Chinas stock markets, including binding arbitrage constraints and a predominance of local retail investors which differentiates China from its developed counterparts (X. Han and Li 2017). Following the approach of Brown and Cliff (2005), this research assumes that a large subset of investors makes biased asset valuations, these biases are persistent across bourses and time, and the limits to arbitrage hinder the exploitation of asset mispricing.

Moreover, this thesis tests if sentiment directly affects stock returns and hereby aims to evaluate the effectiveness of a sentiment-enhanced factor model that incorporates directly measured individual investor sentiment, derived from text-based analysis on the Guba platform as a fourth factor in explaining stock returns in China, compared to existing factor models.

Hereby, the thesis will follow three research questions:

1) How does excluding the smallest 30% of stocks by market capitalization affect the average adjusted R-squared of the CH-3 and CH-G multifactor models in explaining excess returns on the Chinese stock market?
2) Which Guba sentiment factor, among various constructions, should be added to an augmented CH-3 model for explaining excess returns on the Chinese stock market?
3) For which data subset (all stocks, excluding microcaps, or only microcaps) does the CH-3 or CH-G model provide a better explanation of excess returns in the Chinese stock market?

To answer these questions, this thesis is structured into seven sections. Firstly, the following chapters provide an overview of the Chinese financial markets, including an overview of its history, regulatory particularities and the typical Chinese investor. Secondly, the theory of empirical asset pricing models and their adaptations to the Chinese market will be introduced. Thirdly, sentiment as a measurement in asset pricing is discussed, and a shortlist of suitable proxies is selected. Based on previous research and data availability, sentiment will be proxied by a measure derived from the social media platform Guba by Eastmoney.com and the sentiment-inclusive four-factor model introduced in this thesis is accordingly termed CH-G. Before proceeding with the quantitative analysis of the proposed multifactor model against the existing CH-3 model, an analysis roadmap is laid out through testable hypotheses and a methodology overview. Next, the data sampling process is detailed, and the construction of the CH-3 and CH-G models is explained. The research results begin with replicating the CH-3 model, including an analysis of the microcap exclusion criteria. Subsequently, four different Guba sentiment factors are evaluated to select the best fit for the Chinese market. The CH-G model is then constructed and compared to the CH-3 model. Finally, 5x5 portfolios are used to test the explanatory power of both models through Gibbons, Ross, and Shanken (GRS) tests. To conclude the quantitative section, three robustness tests are conducted: one testing the choice of Guba as a sentiment proxy, another using a turnover-inclusive model (CH-4), and the last examining a different period using daily data. Finally, limitations and directions for future research are discussed before the conclusion.

II. Research Background

In the following, the nature of the Chinese stock market is explored in two chapters, one focusing on the stock markets and one on the investors.

2.1 The Chinese Financial Market

Starting in 1978, China initiated economic reforms to transition its centrally planned economy to a market-oriented economy (Chan, Fung, and Thapa 2007), whereby a lack of regulations and protective laws fuelled the vast development in the 2000s (Allen, Qian, and Qian 2005). A tangible example of delayed regulatory updates is the reversed pecking order in capital-raising methods: Chinese companies often prioritize retained earnings for capital but commonly issue equity over taking on long-term debt. This tendency is partly due to state-owned banks' limitations in providing long-term loans in the early 2000s, and more significantly because of weak enforcement of laws that protect equity stakeholders (Chen 2004; Chan, Fung, and Thapa 2007). Since then, China has undergone a number of regulatory transformations (see Chapter 2.2.1).
Today, China is the second-largest economy by GDP (O’Neill 2023) and plays a crucial role in the global financial landscape (Carpenter, Lu, and Whitelaw 2021), evolving from an isolated financial market to one of the world's most significant.

2.1.1 History and Development of the Chinese Financial Market

The origins of the Chinese stock market lie in Shanghai, where stock trading commenced as early as June 1866. However, disruptions ensued on December 8, 1941, as a result of the Japanese invasion, with the discontinuation of trading in 1949. Individual stock data is available in the period 1871 to 1940 (Fan 2021; SSE 2024b). Today, the People’s Republic of China trades on three distinct bourses: the Shanghai Stock Exchange (SSE), the Shenzhen Stock Exchange (SZSE), and the Beijing Stock Exchange (BSE), established in November 1990, December 1990 and September 2021, respectively

(Darrat and Zhong 2000; Chen, Chong, and She 2014; SSE 2024b; SZSE 2024; BSE 2024). Shanghai today lists over 2170 stocks (SSE 2023) and is home to the Science and Technology Innovation Board (SSE STAR Market) since 2018 (SSE 2024a). Shenzhen's equivalent is the ChiNext board, which lists stocks of the high-tech and emerging industries, while two small and medium enterprise boards have been established in both, Shenzhen and Beijing (Carpenter and Whitelaw 2017; SZSE 2024; BSE 2024).

The mainland Chinese stock markets offer A- and B-shares. A-shares are listed and traded in Renminbi (RMB) and were formerly only available to mainland Chinese citizens (Darrat and Zhong 2000). With programs like the *Qualified Foreign Institutional Investor* program, the *Renminbi Qualified Foreign Investor* program and the *Stock Connect* programs between Hong Kong and mainland China, A-shares are available to both Chinese and foreign investors since 2002 (Carpenter and Whitelaw 2017; LSEG 2023). B-shares are listed in foreign currencies, namely U.S. dollars on SSE and Hong Kong dollars on SZSE. Before 2001, B-shares were only open to foreign investors (Darrat and Zhong 2000; Chen 2004).

As shown in Figure 1, China's stock markets have experienced several crashes over the years, which contrast China's GDP growth in the same period, hinting at potential decoupling of (macroeconomic) fundamentals and stock market performance (Yao and Luo 2009; Liang and Willett 2015). In the early 2000s, China's banking landscape was characterized by non-performing loans (NPL). As per government decision and with coherent (financial) support, the four largest state-owned banks were reformed through NPL unloading and capital injections, leading to the public listing of three out of the four by 2006. This state-backing, as well as post-IPO stock price increases and further state-owned company IPOs spurred domestic and international investment optimism, resulting in a non-sustainable, four-fold increase of the Shanghai Stock Exchange Composite Index (SSECI) between early 2006 and autumn 2007. The index then sharply decreased to its lowest point in October 2008, further exacerbated by the impacts of the global financial crisis (Yao and Luo 2009; Cheema, Man, and Szulczyk 2020). The post-2008 market was characterized by political reforms, fiscal stimuli and an accommodative monetary policy, including lower interest rates and loosened capital reserve thresholds

(YHuang, Miao, and Wang 2019; Chen and Haga 2021). This increased the SSECI by more than 150% in a year, starting July 2014, before spiralling into another market crash in 2015 (Huang, Miao, and Wang 2019). With market stabilization in mind, the government appointed the so-called *National Team* in June 2015, which's bail-out investments amounted to 6% of the total market capitalization and impacted approximately 50% of all listed stocks between June and September 2015. Additional measures were taken to restrict direct market participation and ban large share divestments, suspend IPOs and prohibit short selling (Brunnermeier, Sockin, and Xiong 2020; Huang, Miao, and Wang 2019).

Figure 1 Cumulative value-weighted market excess mean returns of all A-shares of the Shenzhen and Shanghai stock exchange, January 2000 – April 2024, data retrieved from CSMAR database.

After a slow upward trend during 2016 and 2017, China exhibited a bearish market in 2018. In the following year, the Chinese stock market gained strongly until May, fluctuating immensely before reaching the historically third highest peak at the end of 2021. Since then, the market has been showing a downward trend.

Nevertheless, mainland China is ranked second in total market capitalization of listed domestic companies (The World Bank 2022a), with Chinese companies listed in Hong

Kong excluded. Coherently, China is being ranked third when comparing stocks traded in relation to GDP – a measure used to exclude non-tradable shares – only preceded by Hong Kong SAR and the Republic of Korea (The World Bank 2022b), highlighting the market's importance in the global financial landscape.

2.1.2 Laws and Regulations of the Chinese Financial Market

In the past, the Chinese regulatory bodies have taken measures repetitively to stabilize the financial market and bring down short-term volatility (Hess 2010). Besides traditional tools such as interest rate changes and bank reserve requirements, other market mechanisms include high margin requirements (State Council of the People's Republic of China 2023), a t+1 trading system, prohibiting non-qualified investors (e.g. retail investors) from buying and selling the same stocks on the same day, and a daily limit corridor on stock price changes, which's surpassing result in trading halts. Until today, regulations on both short selling and IPOs are generally strict and both are occasionally suspended (Xiong and Yu 2011; Carpenter and Whitelaw 2017; Brunnermeier, Sockin, and Xiong 2020; Guo et al. 2022; Tang 2023).

Short Selling Restrictions

Although the legal ban on short selling has been lifted in China in 2005 (Hess 2010), shorting has only been made available to qualified investors through a pilot-scheme including 90 initial stocks after March 30th, 2010 (Gu, Kang, and Xu 2018; Ma, Anderson, and Marshall 2018). These stocks were continuously revised and carefully chosen in accordance with size, liquidity, and volatility qualifications, resulting in a list of approximately 280 stocks by December 2011 (Chang, Luo, and Ren 2014; Deng and Gao 2018), when it became routine practice (Chang, Luo, and Ren 2014). However, short selling in China continues to require approval from the China Securities Regulatory Commission (CSRC) (Gu, Kang, and Xu 2018) and remains costly (Deng and Gao 2018). Most recently, the CRSC tightened the regulations on short-selling further and started to crack done investors with high turnover (Bloomberg News 2023; Ho-him 2024). Overall, short selling activities in China remained low, and instead of increasing liquidity as intended by the CSRC, lifting the short selling prohibition reduced it. The rationale behind this lies in the fact that typical Chinese investors anticipate only informed counterparts (e.g. insiders) to short the stocks and hence abstain from trading them altogether (Sharif, Anderson, and Marshall 2014).

China's dynamic short selling regulations might affect stock market efficiency. According to Bris et al. (2007), short selling generally improves market efficiency by rapidly pricing in negative information. Further, scholars often argue that a short selling ban results in overvaluation of stocks due to hindered arbitrage (Miller 1977; Boehme, Danielsen, and Sorescu 2006). Coherently, stocks ineligible for short selling are likely to continue experiencing pricing inefficiencies (Gu, Kang, and Xu 2018). When academically examining dynamics on the Chinese financial market, it therefore is crucial to consider impact of short selling bans and hindered arbitrage on market efficiency.

IPO Proceedings

Another mechanism to counter-steer market movements practiced by the CSRC is the regulation of IPOs. While IPO proceedings are generally loosened in bull markets, regulations are tightened in bear periods and have been suspended nine times between 1994 and 2020 (Brunnermeier, Sockin, and Xiong 2020; Cong and Howell 2021). Generally, the number of listings per year is tied to an annual quota per industry and province (Carpenter and Whitelaw 2017). The proceedings for an IPO are long and costly and when considering an IPO, the amount of fees to be paid has been one of the main disadvantages stated by listing candidates (Allen, Qian, and Qian 2005). The long approval period and subsequent lag were found to be a reason for severe IPO underpricing (Mok and Hui 1998; Chan, Wang, and Wei 2004). A company must fulfil several requirements before being approved by the CSRC, such as the need to show proof of consistent profitability over a period of two to three years, resulting in numerous companies bolstering their performance before the clearance (Chen et al. 2011; Allen et al. 2024). After the IPO, the returns of companies' stocks listed in mainland China tend to spike with high average first-day returns (Tian 2011). However, these returns often decline more steeply post-listing compared to other markets, and large-cap stocks generally underperform more significantly than their smaller counterparts (Allen et al. 2024). Following consecutive annual losses, businesses receive the designation of *special treatment* (ST) without undergoing (forced) delisting Han and Shi 2022; Allen et al. 2024). Despite a lack of underlying value, these companies persist in the stock market and remain actively traded. Instead of possessing intrinsic value, they maintain potential value for acquisition through a reverse merger, essentially functioning as shell firms (Liu, Stambaugh, and Yuan 2019). Hence, literature lists mispricing though speculations on shell value as a key influencer on the explainability of stock returns in the Chinese stock market (Liu, Stambaugh, and Yuan 2019; Li and Rao 2022; Liu, Zhou, and Zhu 2024).

2.1.3 The Chinese Investor

Unlike the U.S. stock market, the Chinese stock market is characterized by a high number of non-institutional investors. While over 99 percent of share accounts are owned by individuals (SZSE 2019; SSE 2023), they hold just under 25% of the value on the SSE (SSE 2023) but generate 80% of trading turnover Han and Li 2017; Carpenter and Whitelaw 2017; Hu, Pan, and Wang 2021; Ren et al. 2021; Hou, Qiao, and Zhang 2023). The change in the number of individual investors often co-moves with bull markets: in the 2000s, the rise of the Chinese stock market facilitated a promising investment alternative to Chinese private investors as interest payments on savings accounts were low, the bond market was undeveloped and investments in foreign markets restricted (Fernald and Rogers 2002; Siebert 2007; Hasan, Wachtel, and Zhou 2009). The fear of missing out (*envy*), the chance for quick returns (*greed*) and *speculation* among inexperienced investors drove individuals to act irrationally, e.g. the trading of out-of-the-money put options in 2005 to 2008 (Yao and Luo 2009; Xiong and Yu 2011). During the boom of 2015, again, a vast number of inexperienced and optimistic investors entered the Chinese stock market, encouraged by government reforms and relaxed capital market politics (Yi Huang, Miao, and Wang 2019; Chen and Haga 2021). Their investments were characterized by sizeable leverage to multiply short-term returns, resulting in a mass margin call upon the downturn of the market (Brunnermeier, Sockin, and Xiong 2020).

In 2019, Chinese investors were bullish, and while the COVID-19 pandemic generally hit financial markets hard, the initial response of Chinese retail investors was an increase in stock holdings (Sha, Zhang, and Lu 2022). Coherently, the number of individual investor accounts increased during 2018 and 2021, where it peaked before turning downward (Gao, Li, and Lu 2023).

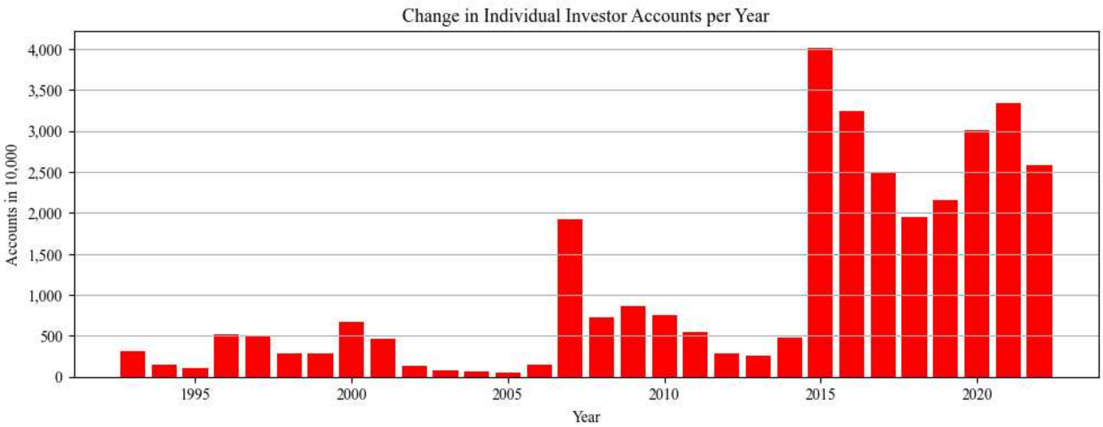

Figure 2 Number of New Individual Investor Accounts on Shanghai Stock Exchange per Year (SSE 2023).

This dominance influences investment decision dynamics in the Chinese stock market: Chinese individual investors are often young and deemed less (financially) educated or sophisticated and more restricted by liquidity constraints (SSE 2016; Han and Li 2017; Hu, Zhong, and Cai 2019; Brunnermeier, Sockin, and Xiong 2020; Wang et al. 2022). When translating these characteristics into investment behaviour, inexperience generally leads to mispricing (Greenwood and Nagel 2009), speculation (Xiong and Yu 2011; Carpenter and Whitelaw 2017; Brunnermeier, Sockin, and Xiong 2020), and biased behaviour, such as the local bias (Liao et al. 2012), conformation bias (Wang et al. 2022) and overconfidence (Han, Li, and Li 2020). Chinese retail portfolios are generally smaller (SZSE 2019; SSE 2023) and less diversified (Xie et al. 2024). Chinese retail investors exhibit gambling behaviour by buying lottery-like stocks and trading on perceived knowledge (Liu et al. 2022). Coherently, they tend to deprioritize proper research and rather buy-in and trade fast in bullish periods in a way that He and Fang (2019) describe as *fast in & fast out* (快进，快出). In bearish markets, however, investors rather wait as they hope for another market upturn (Li and Zhang 2008; Chiang and Zheng 2010). While the possibility of bubbles and bubble bursts is acknowledged, individuals anticipate losses on others rather than themselves (Kahneman and Tversky 1979; X. Li and Zhang 2008). Although the literature is still divided on whether Chinese retail investors are subject to herding (Chiang and Zheng 2010; Chiang, Li, and Tan 2010; Hu, Zhong, and Cai 2019; Zhang, Yuan, and Wu 2020; Liu et al. 2023) or not (Demirer and Kutan 2006; Yao, Ma, and He 2014), it is evident that they are prone to trade on noise (Han, Li, and Li 2020; Brunnermeier, Sockin, and Xiong 2020; Chen and Haga 2021). This could result in a less efficient yet more volatile market, which can also be observed in China (Brunnermeier, Sockin, and Xiong 2020; Blitz, Hanauer, and van Vliet 2021). Coherently, understanding the Chinese investors and their behaviour should help when explaining stock returns.

2.2 Empirical Asset Pricing Models

The evolution of academia in incorporating sentiment into asset pricing models represents a shift from traditional finance theories that focused primarily on rational expectations and efficient markets to recognizing the impact of investor sentiment on asset prices. The following chapter introduces traditional asset pricing models before exploring how to measure sentiment and incorporate it as an additional factor.

2.2.1 Traditional Asset Pricing Models

Both, institutional and private investors consider how to evaluate and price assets to make informed investment decisions. Hereby, asset pricing models are critical as they provide frameworks for understanding how returns are determined. Following Markowitz's (Markowitz 1952) Modern Portfolio Theory (MPT), investors were enabled to select and allocate assets to construct a diversified portfolio that maximizes expected return in accordance with individual risk aversion. However, while the MPT helps to evaluate the suitable asset allocation to reduce idiosyncratic risk, it does not directly price individual assets. Therefore, models like Capital Asset Pricing Model (CAPM) (Sharpe 1964;

Lintner 1965), Fama-French Three (Fama and French 1993) and Five Factor (Fama and French 2015) Models and subsequent models have been introduced.

Capital Asset Pricing Model (CAPM)

The Capital Asset Pricing Model (CAPM), introduced by William Sharpe (1964) and John Lintner (1965), builds on Harry Markowitz' portfolio theory: in its core, the CAPM calls for the investors to be compensated for time money value and additional risk. In coherence, expected return R_i is an equation of the risk-free rate R_f, which an investor could get by investing in a low-risk asset such as a government bond of a country with low default probability, plus the stock's return sensitivity to market returns β_i, multiplied by the risk premium $(R_m - R_f)$. The CAPM therefore creates a testable prediction about the correlation of systematic risk and expected return.

$$R_i = R_f + \beta_i(R_m - R_f) \qquad (1)$$

Although the CAPM is praised for its simplicity and predictive power, it is often criticized for its limitation to one period. Research hence continued to improve asset pricing models to span across time and include two or more factors.

Fama-French Three-Factor Model

Fama and French (1993) extended the CAPM by adding two additional factors: size as measured by market capitalization (SMB, small minus big) and value as measured by book-to-market ratio (HML, high minus low), resulting in the Fama-French Three-Factor Model (FF3):

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{MKT,i}(R_{m,t} - R_{f,t}) + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \varepsilon_{i,t} \qquad (2)$$

This model better captures the cross-section of expected stock returns, addressing anomalies that CAPM could not explain, namely the size (Banz 1981) and the value effect (Stattman 1980; Rosenberg, Reid, and Lanstein 1985).

2.2.2 Asset Pricing Models in China

Neither the Shanghai nor Shenzhen stock market follows a random walk and hence, they should exhibit a higher chance for predictability (Darrat and Zhong 2000). Yet, the traditional APMs are not sufficient in predicting asset prices in the Chinese market, which is often attributed to the market's regulatory environment, different economic conditions, and the speculative nature of Chinese investors (Cheung, Hoguet, and Ng 2014).

Fama-French in China

Scholars have been using the FF3 model to investigate the Chinese market with varying results. While most studies find a significant size effect, the value effect validated by some scholars (Cheung, Hoguet, and Ng 2014; Cakici, Tang, and Yan 2016; Carpenter, Lu, and Whitelaw 2021) and not by others (Wang and Xu 2004; Hilliard and Zhang 2015; Hu et al. 2019). Therefore, modifications are necessary.

CH-3 Model

One significant adaptation is the CH-3 model, developed by Liu, Stambaugh, and Yuan (2019), which is specifically tailored to the Chinese market. The CH-3 model adapts the Fama-French Three-Factor Model to better suit the Chinese market. Firstly, to address the speculative nature of small stocks, which are rather valued by their *shell value* instead of their fundamental value (see Chapter 2.1.2), the smallest 30% of firms by market capitalization are excluded. Secondly, the value factor uses the earnings-per-price (EP) ratio instead of the book-to-market ratio. The authors run a horserace among potential factors and find that the EP ratio better captures value effects in the Chinese context. The model is similar to FF3:

$$R_{i,t} = \alpha_i + \beta_{MKT,i} MKT_t + \beta_{SMB,i} SMB_t + \beta_{VMG,i} VMG_t + \varepsilon_{i,t} \quad (3),$$

but they define $R_{i,t}$ as the excess return of a portfolio or individual stock over the risk-free rate. MKT represents the market factor, which is the excess return of the market portfolio minus the risk-free rate. SMB remains the same, while VMG (Value minus

Growth) is the value factor based on EP.

CH-4 Model

Liu et al. (2019) also validate the thought that investor sentiment plays a significant role in the retail-investor-dominated Chinese market. Their CH-4 model adds a turnover factor to capture sentiment-driven trading and has shown promise in explaining asset returns in China (Liu, Stambaugh, and Yuan 2019; Li and Rao 2022).

$$R_{i,t} = \alpha_i + \beta_{MKT,i} MKT_t + \beta_{SMB,i} SMB_t + \beta_{VMG,i} VMG_t + \beta_{PMO,i} PMO_t + \varepsilon_{i,t} \quad (4)$$

2.3 Investor Sentiment

The emergence of behavioural finance prompted academia to thoroughly test, expand, and question the models introduced in the preceding chapter, allowing financial literature to enter a new era (Lee, Jiang, and Indro 2002; Brown and Cliff 2004; Ge et al. 2020). Therefore, when laying the foundation for including sentiment as a factor in asset pricing models, it is necessary to acknowledge seminal works that question the traditional assumptions of market efficiency and attribute stock market movements to alternative factors.

2.3.1 Emerge of Behavioural Finance

The stock market is not a rational system but rather a grid scheme of human interactions (Nofsinger 2005). Coherently, investors often deviate from rational decision-making due to limited information processing abilities (Simon 1955), market-wide information asymmetry and coherent costs for obtaining and processing information (Grossman and Stiglitz 1980) or cognitive biases (Kahneman and Tversky 1979; Shefrin and Statman 2000). While some investor expectations are true and therefore should be priced in, others are considered *noise*, which may be traded on as if it were information (Black 1986). Hence, it is argued that random and unpredictable factors can influence investment decisions without an underlying rationale. This trading behaviour is often attributed to individual (vs. institutional) investors (Bu and Pi 2014).

The question remains as to why arbitrageurs refrain from intervening and rectifying this market inefficiency as mispricing should be traded against by rational investors (Fama 1965; Black 1986). Delving into this, De Long et al. (1990) (DSSW) focus on the limitations of arbitrage in exploiting misperceptions caused by noise traders, noting time and risk constraints. Their model on closed-ended fund discounts considers rational and irrational (noise) traders, assuming short horizons for the former and stochastic, unpredictable sentiment for the latter. They argue that this unpredictability is undiversifiable and adds to market transaction risk, impeding arbitrage effectiveness and hence continuously deviating prices from intrinsic values. Abreu and Brunnermeier (2002) add that even rational investors buy into the bullish behaviour and delay arbitrage to earn returns based on the market movement, leading to short-term synchronization and amplification of mispricing and, ultimately, bubbles. Additionally, Abreu and Brunnermeier (2003) name the impact of a lack of coordination among arbitrageurs as a short-term dilator. Only in the long run, those bubbles burst, and pricing is rectified (Abreu and Brunnermeier 2002; 2003; Berger and Turtle 2015).

Lee, Shleifer, and Thaler (1991) amplify DSSW's model from closed-end fund specifics to market-wide investor sentiment influences by deriving coherences between the types of investors and assets impacted by investor sentiment. They highlight that close-ended funds are predominantly held and traded by individuals and conclude that if noise impacts the prices and returns of assets favoured by noise traders, it may also impact other asset classes owned by the same group of individual investors, such as small capitalization stocks. This coherence is backed by some scholars (Neal and Wheatley 1998; Kumar and Lee 2006; Barber and Odean 2000; Barber, Odean, and Zhu 2009) while challenged by others (Elton, Gruber, and Busse 1998; Fisher and Statman 2000; Lee, Jiang, and Indro 2002; Brown and Cliff 2004).

The decision-making of private investors as a single individual, however, is not likely to determine stock market prices. Generally, for investor sentiment to influence the stock market to a degree that is reliably quantifiable, either a few large transactions or a large number of small transactions have to take place (Shleifer 2000). Focusing on individual investors, the latter may be explained by herding. Humans seldomly make decisions in isolation, but patterns of reciprocal influence cause herding (Ge et al. 2020), meaning an individual rather follows the actions of others than relying on their own information (Shiller 1995).

2.3.2 Investor Sentiment in Chinese Stock Markets

As outlined in Chapter 2.1.3 of this thesis, the Chinese market is dominated by individual investors (Liu, Stambaugh, and Yuan 2019; SSE 2023). Hence, the impact of investor behaviour, especially sentiment, should be not only impacting the Chinese stock market prices and, coherently, returns, but also be reliably measurable.

Baker and Wurgler (2006, 2007) explore the influence of investor sentiment on stock returns, investigating its impact through two channels: *difficult-to-value* and *limits-to-arbitrage*. Both result in profound implications for the Chinese stock market: Firstly, they propose that stocks whose valuations are more subjective are less robust against speculation and sentiment-induced trading, using the example of a "young, unprofitable, extreme growth stock" (Baker and Wurgler 2007). These stocks generally have less reliable records of assets, earnings and management decisions and are hence subject to the investor's judgment, both positive and negative. As outlined in Chapter 2.1.2, stocks without a two or more consecutive years of profits are generally restricted from IPOs in China. Consequently, the likelihood of reverse mergers is priced into listed small- cap firms, disregarding their actual track record and consequently deviating from their fundamental value (Liu, Stambaugh, and Yuan 2019). For newly listed small companies, however, Allen et al. (2024) find that smaller firms perform better than larger companies after IPOs.

Secondly, Baker and Wurgler (2007) examine stocks that are more challenging to arbitrage and find that these stocks are more susceptible to sentiment-driven price movements. Short-selling restrictions on the Chinese stock market cause a lack of arbitrage opportunities to correct sentiment-induced stock prices. In conclusion, Chinese small cap stocks are likely to be influenced by investors sentiment.

2.3.3 Indirect Measures of Investor Sentiment

Generally, sentiment measures can be distinguished into indirect and direct measures, the former being mainly economic variables that can be obtained from market data. The latter refers to surveys or text-based analysis of the news, financial newsletters, search engine volume, social media and other sources (Zhou 2018).

As mentioned above, early papers use the *shift in closed-end fund discounts* as a sentiment indicator (De Long et al. 1990; Elton, Gruber, and Busse 1998; C. Lee, Shleifer, and Thaler 1991). Baker and Wurgler (2006, 2007) explore the influence of investor sentiment on stock returns and test their hypothesis on the U.S. stock market. Due to a lack of a universal sentiment measure, they form their composite index (*BW index*), utilizing six underlying proxies: "*the closed-ended fund discount, the NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium*" (Baker and Wurgler 2006).

However, the BW measure is insufficient to meet Chinese market particulars, due to the limited number of closed-end funds and the country's IPO legislation (Yang and Zhang 2014). Several researchers have been contributing to the literature by adapting the BW index accordingly, but with no particular success (Yang and Zhang 2014; Yang and Gao 2014; He, Zhu, and Gu 2017; Xu and Zhou 2018).[1] For one out of three sentiment measures taken into consideration, Allen et al. (2024) follow Baker, Wurgler, and Yuan (2012). They find that the positive sentiment of less informed private investors significantly inflates stock prices – they find that the more positive the sentiment, the lower the stock returns of A-shares listed in mainland China. Others do not per se remodel the BW index but rather take some of the six BW measures to recombine any of them with newly introduced economic variables, with a popular proxy being the growth rate of investor accounts (Zhang and Yang 2009; Bu and Pi 2014; Chen, Chong, and She 2014; Han and Li 2017; Chu, Wu, and Qiu 2016; Song, Peng, and Huang 2020; Cheema, Man, and Szulczyk 2020; Han and Shi 2022). The China Stock Market & Accounting

[1] Only (Xu and Zhou 2018) of the cited papers is published in a Q1-ranked journal.

Research Database (CSMAR) does provide raw BW Index component datasets in their Baker-Wurgler Investor Sentiment Index section, alongside two China-specific investor sentiment indices. The China Investor Composite Sentiment Index (CICSI) serves as a BW index equivalent in China (Han and Shi 2022) and is constructed from closed-fund discount, the number of IPOs, the average first-day returns on IPOs, new investor accounts, consumer confidence, and trading volume. The second index given, the Investor Sentiment Index (ISI), replaces trading volume with turnover. However, both sentiment indices are not capable to distinguish between institutional and individual investor sentiment (Sun et al. 2021). Zhang, Li and Malor (2004) examine and dismiss the applicability of closed-end fund discounts as a sentiment measure for the Chinese market. Further, market-based sentiment metrics suffer from the limitation of being the culmination of various economic factors beyond investor sentiment, thus being unable to purely represent the investors' view (Qiu and Welch 2006; Da, Engelberg, and Gao 2015; Sun, Najand, and Shen 2016). Altogether, the BW index, or coherent adaptions, are therefore excluded as a relevant metric in this thesis.

2.3.4 Direct Measures of Investor Sentiment

Another way to measure the impact of retail investor sentiment is through direct measurement, which means observing or gathering investors' perceptions of future market movements first-hand (Zhou 2018). An intuitive approach is through *investor surveys*, which are widely available for the U.S. market (Fisher and Statman 2000; Lee, Jiang, and Indro 2002; Brown and Cliff 2004; 2005; Lemmon and Portniaguina 2006; Qiu and Welch 2006; Ho and Hung 2009; Schmeling 2009; Greenwood and Shleifer 2014; Liu 2015; Wang, Su, and Duxbury 2021). Those are not an option for China as they focus on American investors and are not expected to mirror the Chinese private investors' sentiment. However, a general Consumer Confidence Index, comparable to the Index of Consumer Confidence for the U.S., is available for China (Wang, Su, and Duxbury 2021). Notwithstanding the availability, surveys are limited by factors such as low frequencies of conducting the polls, a potentially small number of participants, and subjectivity of the answers to the truthfulness of the surveyed as well as the design of the survey and the respondents' interpretation thereof (Zhou 2018; Da, Engelberg, and Gao 2015).

As discussed before (Chapter 2.3.1) investors are themselves influenced by information available to them. An investor's information intake may stem from newspapers, newsletters, and other texts, which is why *text- and media-based* sentiment (Zhou 2018; L. Sun, Najand, and Shen 2016) should be considered. Lu et al. (2023), construct a news coincidence index including the People's Daily, Guangming Daily and Hexun texts, but rather focusing on economic prosperity than stock markets. Using news articles sourced from CSMAR, Huang (2018) finds significant predictive power for market returns and volatility within his sample period from 2007 to 2014, especially for small firms. Shen (2022) confirms this effect when analysing international digital news in China.

With digitalization pushing more information, conversations, and – coherently – information acquisition and opinion formation online, the influence of texts and media on sentiment analysis becomes twofold: investors are not only influenced by information available to them but also express their opinions online, making them accessible to others (Sul, Dennis, and Yuan 2017). Thus, *online search engines*, as well as *social media and online forums*, facilitate further research on direct examination of investor sentiment (Nofsinger 2005; Preis, Moat, and Stanley 2013; Nardo, Petracco-Giudici, and Naltsidis 2016; Gan et al. 2020), as it allows investors to directly broadcast their thoughts, emotions and expectations towards all, including the financial market promptly manner which enables fast, continuous, inexpensive observation thereof (Oliveira, Cortez, and Areal 2017). These social media proxies tend to outperform surveys (Ge et al. 2020; Bollen, Mao, and Zeng 2011) and traditional new papers (Gan et al. 2020), both in term of direct measurement (rather than relying on conformity between survey answers or author's opinion and investor sentiment) and frequency. In China, the transcendence of social media sentiment indices may even be amplified, as 74.2 % of the population uses social media, totalling 1.06 billion users (We are Social and Meltwater 2024a), compared to U.S.'s 70.1 % (239 million) (We are Social and Meltwater 2024b). However, China proves to be a special case as media is often state-controlled (THuang 2018), and Western search engines and social media platforms used for research in other countries, such as

Google (Da, Engelberg, and Gao 2015; Brochado 2020; Joseph, Wintoki, and Zhang 2011) and Wikipedia (Moat et al. 2013; Behrendt, Peter, and Zimmermann 2020) are not available (Endeshaw 2004). While some researchers have been using Google search results regardless of the limited service (Gao, Ren, and Zhang 2020; Szczygielski et al. 2024), most turn to the Chinese-native alternative search engine Baidu (百度) (Fang et al. 2020; Huang, Chen, and Wu 2023).

Regarding Microblogs, Antweiler and Frank (2004) pioneered in the field by analysing messages from chat rooms of Yahoo Finance and Raging Bull and categorizing them into buy, sell, or hold recommendations. General posts (Bollen, Mao, and Zeng 2011), or firm-specific posts (Sul, Dennis, and Yuan 2017; Oliveira, Cortez, and Areal 2017) from Twitter (now: X) have been used in literature to significantly predict U.S. stock market returns(Checkley, Higón, and Alles 2017; Gu and Kurov 2020). Chinese alternatives include Weibo (微博, Twitter equivalent) (Li et al. 2014; Xu et al. 2017; Ge et al. 2020; Dong and Gil-Bazo 2020; Ren et al. 2021; Yingying Xu and Zhao 2022), Xueqiu (雪球, finance-focused online forum) (Zhang et al. 2018; Guo, Sun, and Qian 2017) or Guba (股吧) by Eastmoney (东方财富网, finance-focused online forum including a finance-focused Twitter equivalent) (Li et al. 2014; Li et al. 2020; Sun et al. 2021; Wang et al. 2022; Niu, Pan, and Xu 2023). These services operate in Chinese, limiting the sampling bias error from international investors to Chinese speakers. At the same time, it generates difficulties for people unfamiliar with the Chinese internet sphere to understand the websites' exact functionalities, features, and differences thereof, as well as their significance to Chinese investors. To put it in context, similiarweb's global *Top Websites Ranking* lists Eastmoney, Xueqiu and Sina Finance (Weibo's finance-focused discussion forum) the 16th, 33rd, and 41st, respectively, in the category *Finance, Investing* in February 2024 (similiarweb 2024).[2] This is in coherence with previous Chinese research stating that Guba is consistently ranked high or the highest when it comes to online forums on

[2] Websites known and used in the western world, such as Economic Times, Yahoo! Finance and Nasdaq.com rank 9th, 12th and 47th, respectively.

investing throughout past years (Yuqin Huang, Qiu, and Wu 2016; Li et al. 2020; Wang et al. 2022; Dong et al. 2022). Furthermore, the academic publications either do not disclose their data source or the data source was no longer available or accessible for all but Guba Senti, which is Eastmoney.com's native investor sentiment index available through the Chinese Research Data Services Platform (CNRDS).

III. Research Questions and Hypothesis

3.1 Research Objective

The primary objective of this thesis is to evaluate the effectiveness of an asset pricing model that incorporates individual investor sentiment as a fourth factor, in explaining stock returns in the Chinese market, compared to established factor models tailored to the Chinese market.

3.2 Quantitative Analysis and Research Hypothesis

Based on the gaps identified in the literature on multifactor asset pricing models and the role of investor sentiment in explaining stock returns, this thesis formulates a series of hypotheses to be empirically tested using Chinese stock market data. These hypotheses are designed to assess the effectiveness of an enhanced asset pricing model incorporating sentiment and to evaluate its robustness relative to traditional models.

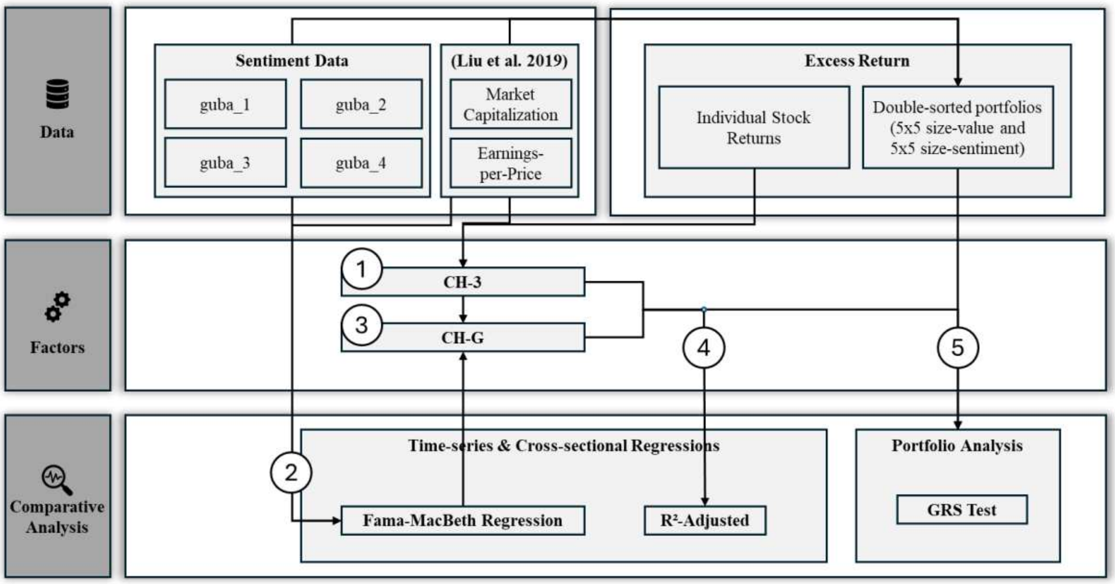

Figure 3 Methodology Overview

The quantitative methodology consecutively follows the analysis steps outlined in Figure

3. To ensure the validity of the factors and model, including the correctness of sampled data and code functionality, the first step is to replicate previous scholars' results. Therefore, Liu et al. (2019) are followed to remodel CH-3 factors (Step 1 in Figure 3) before adding a sentiment factor. In order to choose the most adequate Guba sentiment factor, a horse race among the candidates introduced in Chapter 5.3.1 is realized through Fama-MacBeth Regressions (Step 2) before constructing the factors *mkt*, *size*, *value* and *sentiment* which together form the CH-G factor model (Step 3, Chapter 5.3.2). This model is then thoroughly tested against the replicated CH-3* model through multiple linear regressions (Step 4). Next, size-neutral portfolios on value and sentiment are used as for GRS testing and conclusive comparison (Step 5, Chapter 5.4).

3.2.1 Fama-MacBeth Regressions

Fama-MacBeth regressions (Fama and Macbeth 1973), are a method for estimating parameters in asset pricing models and are implemented in this thesis as a two-step process: first, regressions are run cross-sectionally for each period to obtain a series of parameter estimates; second, these estimates are averaged over time to produce final coefficient estimates. The standard errors are adjusted to account for the time-series variation in the estimates, providing more reliable inference. Moreover, the Fama-MacBeth method mitigates the effects of heteroskedasticity and serial correlation by averaging across periods, thus offering more robust estimates compared to Ordinary Least Square methods (Cochrane 2009). This approach is used in the context of Chinese stock returns to run horse race comparisons among explanatory variables such as value (Liu, Stambaugh, and Yuan 2019), reversal and turnover (Lin et al. 2020) as well as order flow (Jones et al. 2021). Similarly, it is employed in this thesis to determine the leading sentiment factor among those introduced in Chapter 4.2.6.

H_0: $\bar{\beta}_{Sentiment_i} = 0$

The average coefficient of $\beta_{Sentiment_i}$ across all periods is zero.

H_1: $\bar{\beta}_{Sentiment_i} \neq 0$

The average coefficient of $\beta_{Sentiment_i}$ across all periods is not zero.

3.2.2 Average R-square and Average Adjusted R-square

The R-squares are a statistical measure that represents the proportion of the variance in the dependent variable (stock returns) that is explained by the independent variables (movement with the market, size, value, sentiment). It provides an indication of the goodness-of-fit of the model (Wooldridge 2015). It is remarkable that Liu et al. (2019) use average R-squares for their analysis rather than adjusted R-squares. The latter measure, however, seems to be the best-fit test going forward, as it incorporates that an additional factor is added to the total number of factors used to explain the dependent variable. The regressions are run over a rolling window to enhance the stability and robustness of the factor model over different periods.

3.2.3 Portfolio Analysis

Portfolio analysis using 5x5 sorting on size and value is employed to capture the nuanced effects of coherent factors on stock returns. Sorting stocks into portfolios based on size and value follows Fama and French (1993), and examines how the intersection of these characteristics impacts expected returns. This approach is then transferred to the analysis of the additional sentiment factor with a size-neutral sorting for sentiment. As sentiment is hypothesized to be a key factor in explaining returns in the Chinese market, especially in small stocks, constructing portfolios double-sorted by size and sentiment allows for direct testing of the influence of sentiment on returns. This dual approach provides a robust framework for comprehensively comparing the explanatory power of the established and enhanced factor models across different portfolios. All portfolios are rebalanced annually in accordance with Chinese annual report filing deadlines at the end of April (Li and Rao 2022) to account for changes in size, value, and sentiment scores.

3.2.4 GRS Test

The Gibbons, Ross, and Shanken (GRS) test evaluates the overall performance of multifactor models. It assesses whether the intercepts (alphas) in a set of time-series regressions are jointly equal to zero – as they should, if the model perfectly explains the returns and hence no excess return is left unaccounted for (Gibbons, Ross, and Shanken 1989). If the GRS test rejects the null hypothesis, it indicates that the model fails to fully capture the expected returns, suggesting that some systematic risk factors might be missing from the model. Following Li and Rao (2022) and Da et al. (2022), this test is performed on the 5x5 portfolios introduced in Chapter 3.2.3.

H_0: The intercepts (alphas) are jointly equal to zero across portfolios.

$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$, where α_i represent the intercept (alpha) for each of the N portfolios

H_1: At least one intercept (alpha) is not equal to zero.

$H_1: \alpha_i \neq 0$, where α_i represent the intercept (alpha) for each of the N portfolios

IV. Data Sourcing and Factor Model Construction

The subsequent chapters describe the data sources, sampling and cleaning processes as well as the calculation of factor models for comparison in their ability to explain stock returns on the Chinese market.

4.1 Stock and Sentiment Databases

The data is sampled from January 2000 to April 2024. The starting month of January 2000 is chosen due to the market specifics present in China: firstly, with Shanghai and Shenzhen stock exchanges opened in the 1990s and hence, the number of stocks listed after applying filtering criteria in accordance with Liu et al. (2019), is not sufficient for portfolio construction before 2000. Further, the implementation of laws and regulations concerning accounting standards and trading behaviour at the turn of the millennium not only smoothed the extreme volatility in the stock market but also resulted in more comparable and reliable company data for the construction of market, size, and value factors (Liu, Stambaugh, and Yuan 2019; Hou, Qiao, and Zhang 2023).

The stock data is retrieved from GuoTaiAn's China Stock Market Trading Research (CSMAR) database due to data access, data availability and quality (Lin 2017; Carpenter, Lu, and Whitelaw 2021; Han and Shi 2022). Further, the database includes data on delisted stocks, whereas e.g. the WIND Excel add-in does not. However, CSMAR seems to miss some data on the disclosure dates of financial statements, which is critical regarding the calculation of the Earnings-per-Price ratio. To compensate for the incompleteness, the EP dataset is complemented by the financial disclosure publishing dates provided by the Chinese Research Data Services Platform (CNRDS). CNRDS also provides sentiment datasets, described in more detail in Chapter 4.2.6. The exact data fields downloaded are listed in Appendix 1.

4.2 Data Preparation

After downloading the raw data, the data is being restructured into input variables for CH-3 by the help of Python code. With Liu et al. (2019) taking the data directly from WIND underlying database, the first step is to attempt to replicate and thereby validate their results. The CH-3 factor includes the market, size and value factors.

4.2.1 General Preparation

To simplify further steps in data preparation, merging, and other processing operations, all data files are read in a *pandas data frame* and units are standardized to 1 unit meaning that the market capitalization is multiplied by 1,000 and the risk-free rate is divided by 100. All columns are given standardized and identifiable headings across all datasets, e.g. *code* for *Stkcd* and *股票代码*. Further, a *date* column is added to unify the format of time, replacing e.g. *Trdmnt*, *Listdt*, *Annodt*, *统计日期*, *实际披露时间*. These two columns (*code* and *date*) also act as key variables in merging operations. As CSMAR provides data on both A- and B-shares, only stocks with *Markettype* other than 2 (SSE B-share market) and 8 (SZSE B-share market) are kept ensuring that only A-shares are included in the analysis. Lastly, columns with missing values are dropped.

4.2.2 Preparation of Stock Return Data

In addition to the universal data manipulation described above, the monthly stock returns, along with the end-of-period risk-free rate, are shifted by one period to ensure they are matched with the previous periods' firm characteristics, such as market value or price-to-earnings data. Specifically, firm characteristics from time t are aligned with returns from time t+1 (the following month). This alignment ensures that the analysis captures how characteristics like size and value influence subsequent returns, rather than being influenced by returns that occurred earlier. This approach not only prevents look-ahead bias—where future information might accidentally inform past data—but also avoids potential reverse causality where returns influence the characteristics.

4.2.3 Calculation of Earnings-per-Price Ratio

The earnings-per-price ratio is calculated as follows:

$$EP = \frac{Net\ profit\ excluding\ non\text{-}recurring\ gains\ and\ losses}{Market\ Value} \tag{5}$$

Liu et al. (2019) is followed closely for constructing all of the factors. However, EP is the most challenging to recreate as data is only available semi-annually or quarterly and the authors do not thoroughly describe the construction of their EP factor. First, *Net Profit Attributable to Shareholders of Listed Company after Deducting Non-recurring Gains and Losses* (as used by Liu et al. (2019)) available on CSMAR exhibit missing values. In these cases, the *net profit* excluding *minority interests* is used. Given that both are available only cumulative and semi-annually before 2002 and quarterly after, the first step is to transform the data to have quarterly reported net profits. This involves using the most recent incremental net profit data disclosed, meaning that the quarterly report is preferred over semi-annual or annual reports disclosed in the same month as well as decimating net profits of the second, third, and fourth quarter. However, investors only gain access to this information after the statements are officially disclosed. To address this, the data is prepared to identify the actual disclosure dates of the financial statements. The primary source for disclosure data is the CSMAR announcement date dataset. However, since the CSMAR database occasionally has missing values, additional publication dates from the CNDRS are incorporated to ensure completeness and accuracy. For values without net profits announced, the most recently available value is inserted, backdating a maximum of twelve months. This leads to highly autocorrelated data, which needs to be addressed during the data analysis.
Conclusively, the net profits per stock code are merged with the market value (size) per stock. EP is calculated in accordance with Formula 5 before excluding missing values, e.g., if no reports have been published for over a year.

4.2.4 Preparation of IPO Date Data

For CH-3 model replication, a later step includes filtering for companies that have been listed for less than six months (Liu, Stambaugh, and Yuan 2019). Accordingly, the data preparation for IPO data includes adding a column containing the date six months after the initial public offering for intelligible calculation after merging.

4.2.5 CH-3 Data Merging and Filtering

With all files readily prepared, the market value, EP ratio, IPO date, as well as returns and risk-free rate are matched in one file using the key variables *date* and *code*. Initially, the dataset includes over 5,500 unique stock codes, translating into 695,410 monthly observations before introducing the filter criteria following Liu et al. (2019). Appendix 2 illustrates the impact of the filtering criteria on the number of unique stock codes and the total number of observations after dropping:

1) Market: stock codes other than SSE main board, SZSE main board or ChiNext,
2) IPO: initial public offering (IPO) date less than six months ago,
3) Trading: less than 120 (15) trading days in the past year (month),
4) Size: bottom 30% by market value.

Firstly, neither the Shanghai Stock Exchange STAR market nor the Beijing Stock Exchange are included in their analysis, as did not come into existence until 2019 and 2021, respectively (SSE 2024b; BSE 2024). Secondly, the condition of the IPO date increases data reliability by skipping the first, highly volatile months after listing (Hou, Qiao, and Zhang 2023). The third ensures to exclude stocks with long trading suspensions (Liu, Stambaugh, and Yuan 2019). Lastly, dropping the smallest 30% by market capitalization is the most invasive filter criteria, coherently reducing the sample size by one-third. This is justified by the mispricing of microcap stocks based on potential shell value in reverse mergers rather than their intrinsic value, as discussed in Chapter 2.1.2. The applicability of this extreme filter will be thoroughly tested throughout the thesis.

Furthermore, considerations were made regarding negative EP ratios. However, Liu et al. (2019) do not mention their exclusion when calculating CH-3 factors. Also, financial firms are typically excluded (e.g., Fama and French 1993), but no differences were found including or excluding financial firms (Liu, Stambaugh, and Yuan 2019). Hence, these values are kept in sample. The resulting dataset includes the firm characteristics of 4,135 companies and is used as the input for replicating the CH-3 model.

4.2.6 Sentiment Data

The daily raw number of online posts on A-shares, published on Guba by Eastmoney is available via CNRDS from January 2008 to December 2023, with Appendix 3 listing the exact data fields downloaded from CNRDS. These posts are pre-processed by the data provider and labelled with positive, neutral, or negative sentiment. As Figure 4 depicts, the number in the year 2008 are neglectable and are therefore excluded.

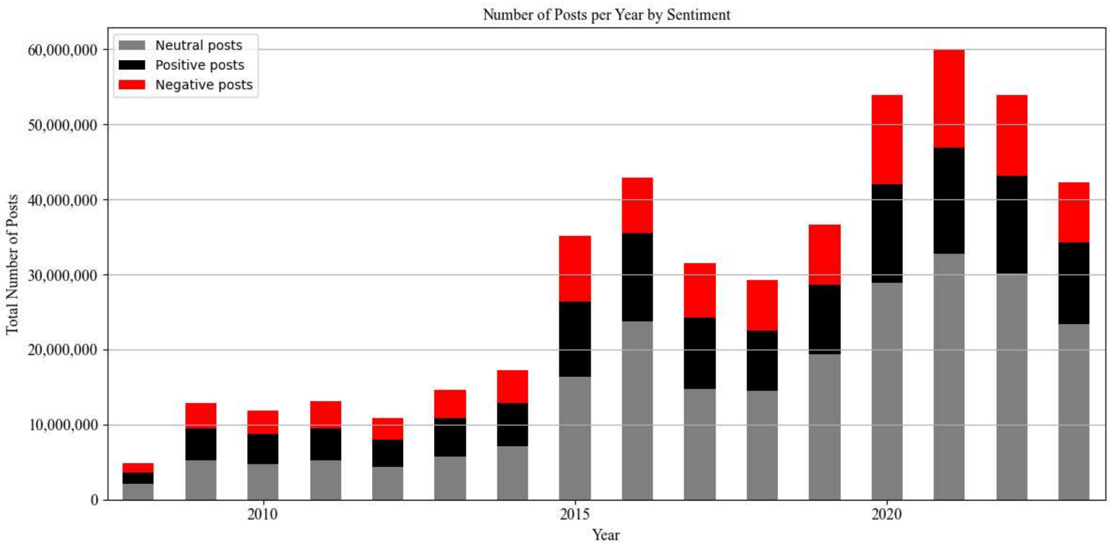

Figure 4 Number of Posts Per Year Per Sentiment (2008-2023)

Appendix 4 provides summary statistics about the number of posts across time. After retrieving the raw numbers on positive and negative posts per stock per day, these values are aggregated to a monthly level to match the periodicity of the monthly returns. Next, the monthly numbers of positive and negative stock posts are transformed to reflect a uniform sentiment factor by following the approach of Antweiler and Frank (2004). Although they categorize *Yahoo! Finance and Raging Bull* posts into *buy*, *hold* and *sell* indications before aggregating them into a *bullishness signal*, their approach is transferable to *positive*, *neutral* and *negative* sentiment (Checkley, Higón, and Alles 2017; Li et al. 2020; Liang et al. 2020). Accordingly, the sentiment factor GUBA at time t for company i is represented by the following equation:

$$GUBA\_1_{i,t} = \frac{M_{i,t}^{pos} - M_{i,t}^{neg}}{M_{i,t}^{pos} + M_{i,t}^{neg}} \tag{6}$$

with $M_{i,t}^{c} = \sum_{j \in D(t)} w_{i,j} * x_{i,j}^{c}$, where

$M_{i,t}^{c}$	weighted sum of posts of sentiment c ∈ {pos, neg},
$w_{i,j}$	weight of the message, is 1 if equal-weighted,
$x_{i,j}^{c}$	dichotomous variable, is 1 if post j is of type c ∈ {pos, neg}, 0 otherwise.

This measure is a normalized measure with homogeneity of degree zero, and hence is independent of the number of messages posted and only relies on the relative proportions of positive and negative posts.

An alternative scale-variant measure incorporates the total number of posts. This measure depends on both the relative and absolute magnitudes of the number of positive and negative posts. The logarithmic transformation helps to dampen the influence of excessively high values (Antweiler and Frank 2004; Y. Sun et al. 2021).

$$GUBA\_2_{i,t} = ln\left[\frac{1 + M_{i,t}^{pos}}{1 + M_{i,t}^{neg}}\right] \tag{7}$$

Further, the level of agreement among investors is measured by the level of consensus in sentiment among messages related to a particular company. When the messages are either all positive or all negative, the agreement at time t reaches a maximum value of 1. Conversely, if sentiment is sharply divided and, hence, the number of positive and negative messages is evenly split, the agreement drops to 0, weakening the overall predictive power of the sentiment measure (Antweiler and Frank 2004; Checkley, Higón, and Alles 2017). This third Guba measure is calculated as presented in Formula 8.

$$GUBA\_3_{i,t} = 1 - \sqrt{1 - \left(\frac{M_{i,t}^{pos} - M_{i,t}^{neg}}{M_{i,t}^{pos} + M_{i,t}^{neg}}\right)^2} \tag{8}$$

Before, the number of neutral (hold) sentiment posts are excluded as irrelevant (Antweiler and Frank 2004). Formula 9 includes them; however, this inclusion is more of an attention measure, recognizing that neutral messages contribute to overall investor attention rather than directly reflecting bullish or bearish sentiment (Li et al. 2020).

$$GUBA\_4_{i,t} = GUBA\_1_{i,t} \times \ln\left(1 + M_{i,t}^{pos} + M_{i,t}^{neu} + M_{i,t}^{neg}\right) \quad (9)$$

Next, the Guba values are merged with the data from Chapter 4.2.5 using *code* and *date* as key variables. Although daily sentiment is often incorporated as quickly as on the same day (Sul, Dennis, and Yuan 2017; Wang et al. 2022), to avoid both look-ahead bias and the potential for reverse causality, the Guba sentiment values are merged similar to size and EP. Specifically, these firm characteristics for a given month are matched with the return and risk-free rate data of the following month. This approach ensures that the analysis accurately reflects how sentiment influences future returns, rather than inadvertently capturing sentiment that could have been affected by earlier movements in returns.

To construct the CH-G dataset, the first three steps of filtering are repeated. Next, the stocks without any Guba posts are excluded and data before January 2009 is eliminated due to the low number of Guba posts available, resulting in 4,412 unique stock codes. Lastly, the lowest 30% by market capitalization is also excluded in this sample.

4.2.7 Factor Model Construction

To incorporate the sentiment factor into a factor model explaining stock returns, the CH-3 model (Formula 3) is extended to

$$R_{i,t} = \alpha_i + \beta_{MKT,i} MKT_t + \beta_{SMB,i} SMB_t + \beta_{VMG,i} VMG_t + \beta_{SENTI,i} SENTI_t + \varepsilon_{i,t} \quad (10)$$

Construction of the Market Factor (MKT)

The market factor is constructed by multiplying each stock's market capitalization by its return, summing these values across all stocks, and then dividing by the total market capitalization before subtracting the risk-free rate of each month, thereby generating a time series of excess market returns. Following traditional practices (Fama and French 1993; Liu, Stambaugh, and Yuan 2019), this method effectively captures the general movement of the entire stock market, which is crucial for explaining the returns of individual portfolios when combined with other factors such as size and value.

Construction of Size and Value Factors for CH-3 (SMB and VMG)

Liu et al. (2019) and Fama and French (1993) are followed to construct the size and value factors. Hereby, the stocks are split into two groups, *big* (B) and *small* (S) at the median of the market capitalization and three groups for EP, indicated as *growth* (G) for the lowest 30%, *middle* (M) for the next 40% and *value* (V) for the remaining 30% of stocks. This procedure is done monthly. Next, the size – EP – intersections are used to construct six value-weighted portfolios (*S/V*, *S/M*, *S/G* and *B/V*, *B/M* and *B/G*) before calculating the factors:

$$SMB_{Value} = \frac{1}{3}(S/V + S/M + S/G) - \frac{1}{3}(B/V + B/M + B/G) \tag{11}$$

$$VMG = \frac{1}{2}(S/V + B/V) - \frac{1}{2}(S/G + B/G) \tag{12}$$

Construction of Size, Value and Sentiment Factor for CH-G

The merging of the size, ep, return, and rf with the Guba dataset results in a change in the number of observations (Appendix 2) as not every stock is posted about daily. Hence, the construction of size and value factors is repeated. Hereby, the size factor is formed by using a simple average of the size factor in Formula 11 (SMB_{Value}) and the similarly constructed sentiment-neutral SMB (SMB_{Guba}). Further, the Guba factor is formed in coherence with constructing the value factor, with the lowest and highest 30% representing *pessimism* (P) and *optimism* (O) on stocks, respectively, while the other stocks are *neutral* (N):

$$SMB_{Guba} = \frac{1}{3}(S/O + S/N + S/P) - \frac{1}{3}(B/O + B/N + B/P) \tag{13}$$

$$SMB = \frac{1}{2}(SMB_{Value} + SMB_{Guba}) \tag{14}$$

$$Guba = \frac{1}{2}(S/O + B/O) - \frac{1}{2}(S/P + B/P) \tag{15}$$

V. Research Results

This chapter presents the results from the replication of the CH-3 factors, to ensure correct data input, computation, and coding. The data sampling is expanded to April 2024, to cover the longest period possible. Also, the model is run from 2009 to 2023, which is the main period of analysis of this thesis due to the data availability of the Guba dataset.

5.1 Descriptive Statistics of A-Shares on the Chinese Stock Market

Table 1 provides a comparison of firm characteristics and returns across different subsets of Chinese stocks.

Table 1

Descriptive statistics of the firm characteristics (size, value) and excess returns.

	Size	Value	Excess Returns
Panel A: All individual stocks in China			
Mean	13,897,420,000	0.0046	0.0103
St. Dev.	56,569,230,000	0.0138	0.1374
Min	95,562,110	-0.0777	-0.8918
Max	2,786,247,000,000	0.0585	3.6172
Panel B: All individual stocks but the smallest 30%			
Mean	18,857,630,000	0.0068	0.0076
St. Dev.	67,005,840,000	0.0132	0.1346
Min	629,602,100	-0.0777	-0.8690
Max	2,786,247,000,000	0.0585	2.9224
Panel C: Only the smallest 30%			
Mean	2,330,170,000	-0.0003	0.0166
St. Dev.	1,142,570,000	0.0137	0.1436
Min	95,562,110	-0.0777	-0.8918
Max	8,084,738,000	0.0585	3.6172

This table provides the descriptive statistics for market capitalization (size), EP (value) and stock returns across three panels. Panel A includes all individual stocks in China, Panel B excludes the smallest 30% of stocks, and Panel C focuses solely on the smallest 30%. The EP ratio is winsorized at 5% to reduce the impact of extreme values. All other values are not winsorized, following Liu et al. (2019). The sample period is January 2000 – April 2024.

Notably, the smallest 30% of stocks (Panel C) exhibit higher average returns (1.66%) compared to Panels A and B. This suggests that smaller stocks may offer higher returns, albeit with greater variability, as indicated by the higher standard deviation of returns in

Panel C (14.36%). These smaller firms show lower average EP values, with a mean of -0.03%, compared to 0.46% in Panel A and 0.68% in Panel B, indicating higher value for bigger companies.

5.2 Replication of the three-factor model (CH-3) by Liu et al. (2019)

This chapter compares the research results of Liu, Stambaugh, and Yuan's (2019) CH-3 with the self-sampled CH-3* model. Specifically, the summary statistics and average R-square values will be remodelled to ensure the robustness of the base factors (market, size, value) before adding the sentiment factor. For comparison, the data published by the authors themselves (Stambaugh 2024) is utilized for direct juxtaposition.[3] The comparative period is January 2000 to April 2024, exceeding the original paper's period by over eight years.

5.2.1 Comparative analysis of the summary statistics

Despite the difference in data sources (WIND vs. CSMAR) and the uncertainty surrounding the exact calculation methodology for EP, the results presented in Table 2 exhibit close alignment. The most notable discrepancy is observed in the mean of the MKT and MKT* factor, which differs by 0.13% per month. Generally, the correlation matrix aligns with the findings of Liu et al. (2019), who attribute the relatively high negative correlation between SMB and VMG to the tendency of smaller stocks in China to be growth stocks. The moderate positive correlation between MKT (MKT*) and SMB (SMB*) suggests that smaller firms' performance is somewhat aligned with overall market trends. Conversely, the negative correlation between MKT (or MKT*) and VMG (or VMG*) factors suggests that value stocks tend to move inversely. The last column of the correlation matrix reports the correlation of the market, size, and value factors of CH-3 and CH-3* with one another. MKT and MKT* as well as SMB and SMB* factors have a near-perfect correlation (0.98), indicating a high degree of similarity between the two models. Although there is some divergence with the VMG and VMG* factors, the

[3] Monthly CH-3 factors are regularly updated, daily CH-3 factors are updated until 2021.

correlation is still strong (0.96). This suggests that the replication was highly successful, with the slight deviations most likely resulting from the differences in databases or EP calculation methods. Going forward, the self-constructed CH-3* model will be the basis for further analysis.

Table 2

Summary statistics of CH-3 (Liu, Stambaugh, and Yuan 2019) and CH-3* replicated.

	CH-3, factors published by Liu et al. (2019)			CH-3*, replicated by following Liu et al. (2019)			
	MKT	SMB	VMG	MKT*	SMB*	VMG*	
Descriptive Statistics							
Mean	0.47	0.65	1.13	0.60	0.61	1.19	
St. Dev.	7.19	4.31	3.72	7.39	4.15	3.71	
t-stat	0.95	2.50	6.03	1.14	2.50	6.15	
Correlation Matrix							
MKT | MKT*	1.00	0.14	-0.29	1.00	0.14	-0.28	0.98
SMB | SMB*		1.00	-0.54		1.00	-0.54	0.98
VMG | VMG*			1.00			1.00	0.96

This table reports summary statistics for the CH-3 and the replicated CH-3* factors market (MKT), size (SMB) and value (VMG). The first block reports the mean returns and standard deviation (both in percent per month) and Newey-West (1987) adjusted t-statistics calculated with 4 lags. The second block provides the correlations matrix, with the last column reporting the correlation between CH-3 and CH-3* factors. The factors are constructed excluding shell stocks (following Liu et al. (2019)), and the sample period is January 2000 to April 2024.

5.2.2 Explanatory power of CH-3 over time

Liu et al. (2019) eliminate the smallest 30% of the stocks by market capitalization when constructing the CH-3 factors to avoid shell value contamination. They then use these factors to explain the return variance through R-squared values averaged over time and across stocks. Hereby, they split the Chinese stock returns into two groups: while Panel A includes all individual stocks in China, Panel B composes all but the smallest 30% of stocks in China. In the period January 2000 to December 2016 reported in their paper, Panel B exhibits increased average R-square values (0.562) over Panel A (0.536). This effect cannot be replicated, neither by constructing the factors from self-sampled data nor when using the factors provided via the author’s website, not in the original nor any other period. This is notable, as it is a divergence from the original findings of Liu et al. (2019), albeit not unusual: several scholars find similar results to their analysis including or excluding the microcap stocks (Lin et al. 2020; Carpenter, Lu, and Whitelaw 2021; Dong et al. 2022; Hou, Qiao, and Zhang 2023). Li and Rao (2022) argue that regulatory improvements and the declining number of reverse mergers refute Liu et al.'s (2019) arguments. Advancements in the regulatory environment in China (as discussed in Chapter 2.1.2) could indeed have reduced irregularities, especially in later years. Thus, the inclusion or exclusion of the smallest 30% by market cap should be analysed more thoroughly going forward.

Despite the difference in results per Panel, the replication of CH-3 (CH-3*) exhibits rather similar average R-squares, especially for Panel B. Extending the analysis to the period January 2000 to December 2024 results in a lower average R-square regardless of the panel choice (0.442 for Panel A, 0.443 for Panel B). The coherent values are provided in Appendix 5 and depict the robustness of both, the original CH-3 model and the replicated CH-3* within the original paper's timeframe. Clearly, the proximity of the results of the original and the replication the summary statistics and the average R-squares confirm the successful reconstruction of CH-3 and underscore the reliability of the sampled data and constructed factors. However, the observed deterioration in the R-squared values over time suggests that while the CH-3 model remains powerful, its predictive power has begun to wane as the market evolves. Consequently, the CH-3 model's factors may no longer capture all relevant aspects of asset pricing in the Chinese market.

5.3 Empirical Analysis of Adding a Sentiment Factor to the CH-3 Model

The preceding replication results attest to the need for further adaptation and refinement of the model to account for changes in market dynamics, regulatory adjustments, and evolving investor behaviour. At the same time, it may be relevant and feasible to include the smallest 30% of stocks, as their simple exclusion proves to be undertheorized: Carpenter, Lu, and Whitelaw (2021) find that excluding the smallest 30% of stocks has negligible impact on their analysis. Other scholars use the elimination of the lowest 30% as a robustness test in their research and find similar results including or excluding the microcap stocks (Lin et al. 2020; D. Dong et al. 2022; Hou, Qiao, and Zhang 2023). Especially in the context of adding a sentiment model, which may provide an additional explanation regarding the speculation on small companies' value (Baker and Wurgler 2006), the role of microcaps should be analysed. To cater to that, factors will be constructed on different panels and the CH-3* and CH-G model will be rigorously tested for their ability to explain the returns of various subsamples.

5.3.1 Sentiment Factor Decision through Fama-MacBeth Regressions

To test which of the individual investor sentiment factors introduced in Chapter 4.2.6 is most suitable to explain the stock returns in the Chinese market, Fama-MacBeth Regressions are conducted. This helps to evaluate and compare the factors across stocks and time. As a consequence of adding the Guba data to the analysis, the data period is shortened from January 2000 – April 2024 to January 2009 – December 2023. For comparison, the Fama-MacBeth Regressions results of Liu et al. (2019) and their replication spanning different periods are presented in Appendix 6.

Table 3

Fama-MacBeth regressions of stock returns on beta, size, value and sentiment.

Model	CH-3*			CH-G		
Intercept	0.1002	0.1005	0.1007	0.0992	0.1025	0.0924
	(3.07)	(3.10)	(3.10)	(3.04)	(3.18)	(2.90)
β	0.0018	0.0026	0.0027	0.0027	0.0025	0.0030
	(0.56)	(0.83)	(0.84)	(0.86)	(0.78)	(0.99)
logME	-0.0041	-0.0043	-0.0043	-0.0042	-0.0043	-0.0040
	(-3.14)	(-3.28)	(-3.29)	(-3.20)	(-3.37)	(-3.12)
EP+	0.4468	0.4357	0.4369	0.4421	0.4362	0.4349
	(4.19)	(4.11)	(4.13)	(4.17)	(4.12)	(4.13)
D(EP<0)	-0.0042	-0.0039	-0.0039	-0.0039	-0.0039	-0.0036
	(-2.48)	(-2.36)	(-2.36)	(-2.35)	(-2.40)	(-2.24)
Guba_1		0.0115				0.1003
		(4.38)				(1.84)
Guba_2			0.0052			-0.0418
			(4.26)			(-1.71)
Guba_3				0.0424		0.0431
				(3.99)		(2.52)
Guba_4					0.0017	-0.0017
					(3.97)	(-0.60)
Adjusted R^2	0.0514	0.0548	0.0547	0.0536	0.0548	0.0590

The table presents the average slope coefficients obtained from month-on-month Fama–MacBeth regressions, which cross-sectionally regress next month's individual stock returns on this month's individual firm characteristics. Each column represents a different regression model, with nonempty rows indicating the variables included in the analysis. The explanatory variables comprise of the monthlyized daily beta value, which quantifies the stock's sensitivity to movements within its bourse; the logarithm of month-end market capitalization (logME); EP+, representing positive earnings-to-price values, with zero assigned otherwise; and D(EP < 0), which takes the value of one if earnings are negative, and zero otherwise. The subsequent rows add one Guba factor at a time, with the last column displaying all factors at once. The final row displays the average adjusted R-squared for each regression specification. T-statistics, calculated using Newey-West (1987) standard errors with 4 lags, are shown in parentheses. The sample size is all individual stocks but the smallest 30% by market capitalization and the sample period is January 2009 – December 2023.

The average adjusted R^2 in Fama-MacBeth regressions reflects the mean explanatory power of the independent variables across the multiple cross-sectional regressions conducted for each period. Table 3 demonstrates that adding the Guba factor increases the average adjusted R-square of the Fama-MacBeth Regression, regardless of the choice of Guba calculation. Nevertheless, Guba_1 and Guba_4 increase the value the most, with Guba_1 having both the highest average coefficient and t-statistics. Additionally, the factor remains significant at a 10% level in the last column, which simultaneously includes all sentiment indicators in the regression. The addition of neutral stocks to the sentiment (Guba_4) to include an attention feature seems to rather dilute the expressiveness of the sentiment factor. Hence, going forward, Guba_1 will be added to CH-3* to construct the four-factor model CH-G.

The results are similar for other sample sizes (as reported in Appendix 7 and 8), with three distinctions: first, in none of the other Panels negative EP is statistically significant on a 5% level. Second, the intercepts of Panel C are much higher than those of Panel A and B, indicating that there is a higher baseline level of returns for the stocks in Panel C, independent of the explanatory variables included in the model. Third, Guba_3 exhibits both, a higher coefficient and a slightly higher t-stat for Panel C. However, as the adjusted R^2 of the regressions including Guba_3 remains lower than those of Guba_1 and both are insignificant when including all Guba measures, Guba_1 will be chosen for all Panels.

5.3.2 Descriptive statistics of CH-3 in comparison to CH-G

Table 4

Summary statistics of CH-3* and CH-G factor models.

	CH-3*			CH-G			
	MKT	SMB	VMG	MKT	SMB	VMG	GUBA
Summary Statistics							
Mean	0.45	0.69	0.92	0.45	0.60	0.92	0.56
Standard deviation	6.45	4.02	3.62	6.45	4.31	3.62	2.30
t-statistics	0.93	2.28	4.07	0.93	1.93	4.07	3.50
Sharpe Ratio	0.07	0.17	0.25	0.07	0.14	0.25	0.25
Skewness	-0.15	-0.16	0.06	-0.15	-0.22	0.06	0.02
Excess Kurtosis	1.72	2.86	1.35	1.72	3.11	1.35	3.53
Correlation Matrix							
MKT	1.00	0.11	-0.34	1.00	0.15	-0.34	-0.20
SMB		1.00	-0.52		1.00	-0.61	-0.45
VMG			1.00			1.00	0.45
GUBA							1.00

This table reports summary statistics for the CH-3 and CH-G factors market (MKT), size (SMB) and value (VMG) and CH-G sentiment (GUBA). The first block reports the mean returns, standard deviation, Newey-West (1987) adjusted t-statistics calculated with 4 lags as well as the monthly Sharpe Ratio, Skewness and Excess Kurtosis for each factor. The second block provides the correlations among factors. The sample size is all individual stocks but the smallest 30% by market capitalization and the sample period is January 2009 – December 2023.

Table 4 presents summary statistics for factors of the CH-3* and CH-G factor models in the sample period January 2009 to December 2024, with the factors MKT, SMB, VMG, and GUBA formed based on the sample of firm characteristics including all but the smallest 30% individual stocks by market capitalization. Coherent descriptive statistics of the firm characteristics used for the construction are presented in Appendix 9. The GUBA factor, introduced under the CH-G model, exhibits a mean of 0.56 and a standard deviation of 2.30, the latter being the lowest among all factors. The t-statistic for GUBA is 3.50, indicating statistical significance at the 5% level and well above the usual cutoff of 1.96. Hence, sentiment appears to carry significant premiums in China during the sample period, highlighting the importance of Guba_1 as a relevant factor in the CH-G. The Sharpe Ratios across both, CH-3 and CH-G, are moderately good, with an annualized Sharpe Ratio of 1.21 for both VMG and GUBA. This indicates that the excess returns relative to the risk taken are better than the risk-adjusted benchmark. Skewness values are close to zero, suggesting that the return distributions of the factors are symmetrical, with only slight (mostly negative) deviations from a normal distribution. Both, Sharpe Ratios and skewness, are consistent with recent research on the Chinese market (Liu, Zhou, and Zhu 2024). Positive excess kurtosis is present across CH-3* and CH-G, indicating that the return distributions exhibit heavier tails than a normal distribution, which implies a higher likelihood of extreme events. This is not unusual for equity returns, however presents challenges to the reliability of future analysis. One standard implemented in the course of this thesis to deal with both, heteroskedasticity and excess kurtosis, is to use Newey-West standard errors (Newey and West 1987) when testing time-series data for statistical significance.

Generally, Table 5 follows the methodology of Liu et al. (2019) by using only the largest 70% by market capitalization to form the factors and group the dependent variable into panels: Panel A includes all individual stocks, and Panel B excludes the smallest 30% by market size. Furthermore, Panel C is added and includes only the smallest 30%. Unlike Liu et al. (2019), this table reports average adjusted R-squared values instead of average R-squared values, correcting for the number of factors in the models to provide a more accurate measure of explanatory power.

The regressions are run over a 3-year (36 months) window to reduce the impact of any single period's abnormal behaviour before averaging the adjusted R-square values over time. Generally, Panels A and B are fairly similar and confirm that across all panels, the CH-G factors (MKT, SMB, VMG, GUBA), report the highest average adjusted R-square in the sample period. The increase in average adjusted R-squared when adding GUBA to the MKT factor is notably smaller than the increases observed when adding SMB or VMG individually as factors, suggesting that GUBA contributes less to explaining the variance in returns compared to the other factors.

Table 5

Average adjusted R-squares for individual stocks in China.

	Panel A (Returns)	Panel B (Returns)	Panel C (Returns)
MKT	0.245	0.261	0.235
MKT, SMB	0.350	0.346	0.390
MKT, VMG	0.302	0.319	0.297
MKT, GUBA	0.267	0.282	0.267
MKT, SMB, VMG	0.370	0.370	0.399
MKT, SMB, GUBA	0.356	0.352	0.395
MKT, VMG, GUBA	0.311	0.328	0.309
MKT, SMB, VMG, GUBA	0.378	0.377	0.406

The table presents average adjusted R-squares on regressions of monthly individual stocks' returns on factors in China's stock market. The factors are formed on all but the smallest 30% of stocks. Regressions are estimated on eight models: the first includes excess market returns (MKT) only, before successively adding size (SMB), value (VMG) and sentiment (GUBA), forming all combinations possible. All regressions are run over a rolling three-year (36 months) window. The mean of the adjusted R-squares is first calculated over time for each individual stock and then averaged across stocks. The factor model is used to explain three distinct Panels of returns: those of all individual stocks (Panel A), those of all but the smallest 30 % by market capitalization (Panel B) and those of only the smallest 30% by market capitalization (Panel C). The sample period is January 2009 – December 2023.

However, when combining the insights from Tables 4 and 5, it becomes clear that the lack of statistical significance for MKT and SMB in Table 4 suggests that the traditional market and size factors may not be the primary drivers of stock returns in this sample, and instead, factors like VMG and even GUBA, despite its relatively lower contribution, are more relevant in explaining the return variations in the Chinese stock market during this sample period. The summary statistics as well as adjusted R-squares for the factors formed on all individual stocks (Panel A) and only the smallest 30% of stocks (Panel C) are reported in Appendix 10 to 12.

5.4 Portfolio Analysis

To evaluate the pricing ability of both, CH-3* and CH-G, it is crucial to establish a consistent basis for comparison. Following standard methodology (Fama and French 1993; Li and Rao 2022), 25 portfolios are constructed by sorting stocks into quintiles according to their size and value (EP) at the end of April. April is chosen as this is the deadline for filing annual reports for listed firms in China (Li and Rao 2022). These portfolios are then rebalanced annually. In Chapter 5.4.2, sorting by size and sentiment allows for a comprehensive assessment of how the sentiment factor influences the pricing ability of the models. The portfolio returns are formed on Panel A, resulting in the lowest 30% of stocks by market cap to be included (Li and Rao 2022).

5.4.1 5x5 portfolios sorted for Size and Value

Table 6 presents summary statistics for the 5x5 Fama–French portfolios, which are sorted by firm size (from Small to Big) and value, measured by the Earnings-to-Price ratio (from Growth to Value). The stocks are categorized into quintiles of size and value and their intersection creates 25 portfolios. The mean returns indicate that smaller firms tend to have higher returns across all value categories, with the highest mean return of 1.71% per month observed in the small firm portfolio with medium value. The standard deviation values, reflecting the risk associated with each portfolio, generally decrease as firm size increases, indicating that larger firms exhibit less volatility. T-statistics suggest the statistical significance of these returns, with smaller portfolios generally showing higher significance levels, particularly within the small firm and high-value categories. This pattern underscores that the higher risk-return trade-off associated with small-cap, value-oriented portfolios compared to their large-cap, growth-oriented counterparts is also existent in the Chinese stock market.

Table 6
Summary statistics of the Fama–French portfolios sorted for size and value.

	Growth	EP 2	EP 3	EP 4	Value
Mean					
Small	1.62	1.70	1.71	1.62	1.48
Size 2	0.97	0.96	1.10	1.25	1.25
Size 3	0.46	0.56	0.70	0.98	0.92
Size 4	0.16	0.26	0.49	0.57	0.92
Big	0.13	0.08	0.23	0.23	0.55
Standard Error					
Small	0.69	0.70	0.70	0.68	0.63
Size 2	0.66	0.67	0.68	0.64	0.58
Size 3	0.66	0.69	0.65	0.61	0.55
Size 4	0.66	0.67	0.63	0.59	0.54
Big	0.65	0.59	0.57	0.54	0.52
T-statistics					
Small	2.34	2.42	2.45	2.36	2.34
Size 2	1.47	1.42	1.60	1.95	2.16
Size 3	0.69	0.81	1.09	1.61	1.67
Size 4	0.24	0.40	0.78	0.97	1.70
Big	0.20	0.14	0.41	0.42	1.06

This table presents the mean, standard deviation and t-statistics of 5x5 Fama–French portfolios sorted by size and value. Size and Value are sorted from smallest (lowest) to biggest (highest) with increasing number indicators (1-5). All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980).

5.4.2 5x5 portfolios sorted for Size and Sentiment

Although double-sorting by size and sentiment is unconventional, this method is justified when comparing CH-3* and CH-G, as this sorting method provides insights into the potential added value of incorporating individual investor sentiment.

Table 7 reveals important insights into the relationship between firm size and sentiment, such as that smaller firms again demonstrate higher and more statistically robust mean returns across all sentiment categories. The largest mean return with the highest significance above the 5% level is observed in the smallest size (Small) and highest sentiment (Optimistic) category. However, smaller companies also tend to be more volatile, as demonstrated by the higher standard deviations.

Table 7

Summary statistics of the Fama–French portfolios sorted for size and sentiment.

	Pessimistic	Sentiment 2	Sentiment 3	Sentiment 4	Optimistic
Mean Return					
Small	1.51	1.45	1.63	1.59	2.02
Size 2	0.96	0.89	1.05	1.14	1.41
Size 3	0.62	0.61	0.71	0.76	0.97
Size 4	0.28	0.45	0.51	0.51	0.96
Big	0.04	0.18	0.36	0.52	0.54
Standard Error					
Small	0.68	0.68	0.69	0.66	0.72
Size 2	0.65	0.65	0.66	0.65	0.66
Size 3	0.65	0.63	0.63	0.62	0.62
Size 4	0.61	0.62	0.60	0.61	0.60
Big	0.57	0.55	0.53	0.55	0.50
T-statistics					
Small	2.23	2.12	2.35	2.39	2.79
Size 2	1.47	1.38	1.59	1.76	2.13
Size 3	0.96	0.97	1.12	1.22	1.56
Size 4	0.45	0.72	0.85	0.84	1.65
Big	0.07	0.32	0.68	0.96	1.08

This table presents the mean, standard deviation and t-statistics of 5x5 Fama–French portfolios sorted by size and sentiment. Size and Sentiment are sorted from smallest (pessimistic) to biggest (optimistic) with increasing number indicators (1-5). All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

5.4.3 Factor-based Portfolio Performance Analysis

The factor-based portfolio performance analysis reveals how each model performs under varying market conditions and portfolio characteristics. Firstly, alpha (α) provides insights into the portfolio's performance beyond what is explained by the included factors. Secondly, the coefficients (β_{MKT}, β_{SMB}, β_{VMG}, β_{SENTI}) shed light on the portfolio's sensitivities to the model factors. The statistical significance of the results determines the reliability of the observed relationships.

Tables 8 and 9 present the analysis of the factors formed on all but the smallest 30% of stocks by market capitalization, and the 5x5 size-value sorted portfolios from January 2009 to December 2023.

Tables 8 and 9 reveal that alphas are consistently significant for smaller-sized companies across all value categories, suggesting that these stocks tend to outperform their expected returns which can be explained by the included factors – notwithstanding the choice of

the model. In contrast, larger companies show less significant alphas, indicating that their performance is better explained by the factor models. This is consistent with the lower volatility in larger companies, as shown in Tables 6 and 7. Another potential explanation could stem from the short-selling restrictions discussed in Chapter 2.12., which, combined with the dominance of retail investors, foster a market environment where prices can significantly deviate from their intrinsic values based on firm characteristics.
The market beta (β_{MKT}) is highly significant and almost similar across all portfolios, implying strong sensitivity to market-wide movements. The size beta (β_{SMB},) is most pronounced in smaller companies, while larger portfolios, especially those in the highest value category, show much lower or even negative size betas. The value beta (β_{VMG},) is significantly negative in smaller, growth portfolios, but becomes positive in larger, value-oriented portfolios, reflecting greater exposure to value stocks that tend to perform better in the Chinese market. The sentiment beta (β_{SENTI}) is significantly negative for smaller, growth-oriented portfolios, suggesting these stocks are more vulnerable to negative investor sentiment, while larger, value-oriented portfolios are less affected.
In comparing the CH-G and CH-3* models across the 5x5 size-value sorted portfolios, several key differences emerge that highlight the impact of including sentiment as a factor in the model. The CH-G model generally shows stronger explanatory power, particularly in the smaller size categories and in portfolios with extreme value or growth characteristics. For instance, the alpha values in the CH-G model tend to be slightly lower (with exemption of the big and the growth portfolios) compared to the CH-3* model, suggesting that the inclusion of sentiment helps to capture more of the unexplained variation in returns, leading to a reduction in alpha.
In terms of market (β_{MKT}), size (β_{SMB}), and value (β_{VMG}) coefficients, both models show similar patterns, with strong and significant loadings across most portfolios. However, the sentiment coefficient (β_{SENTI}) in the CH-G model shows significant values; this is something the CH-3* model, which lacks the sentiment factor, fails to capture, partly resulting in higher unexplained returns (alpha). On the other hand, in larger portfolios or those more aligned with value, the sentiment factor tends to be less significant,

suggesting that its influence diminishes as portfolios become larger and more value-oriented. Concluding, adding sentiment as a factor in the CH-G model appears to improve the model's ability to explain returns in smaller and growth-oriented portfolios, where sentiment-driven behaviours are more likely to impact prices. However, its impact is less pronounced in larger, value-oriented portfolios. Therefore, the CH-G model may be considered superior in certain market segments, particularly when sentiment is expected to play a significant role, while the CH-3* model remains robust in more traditional, value-driven contexts. Results for factor models formed on all or microcap stocks are presented in Appendix 15 to 18. Hereby, the change in significance levels of the sentiment coefficient is remarkable. In Panel A, β_{SENTI} is significant at a 5% or 1% level on all but two (small, high-value) portfolios. This suggests that sentiment-driven behaviours are present across stock categories, reflecting the retail investor-driven dynamics of the Chinese market. In Panel C the observation also that growth stocks are exhibiting stronger sentiment beta values statistically significant on a 1% level is confirmed. However, when forming the factors on microcaps only, the resulting β_{SENTI} seems to have a higher and statistically significant (on at least a 5% level) impact on returns of all but the smallest edge portfolios.

The analysis of the size-sentiment portfolios, as presented in Appendix 13, reveals results that are generally consistent with the findings from the size-value portfolios. It is noteworthy that the correlation coefficient of VMG and GUBA is 0.45, meaning that the factors should exhibit similarities. Notably, alphas remain significant for smaller portfolios, particularly those with optimistic sentiment, suggesting continued outperformance beyond what the factors can explain. This is consistent with the findings on individual investor behaviour in markets like China, where excessive optimism often leads to overvaluation (Han, Li, and Li 2020). Again, CH-G dominates in explaining all but the growth portfolios.

Also, the market and size betas remain consistently significant across all portfolios, reaffirming their strong influence on returns. However, differences can be observed in the value (β_{VMG}) and sentiment (β_{SENTI}) coefficients. While β_{VMG}, again, shows negative and significant coefficients for most smaller portfolios, this effect diminishes as size becomes bigger. Additionally, β_{SENTI} is more negative and significant in portfolios with sentiment level 2, indicating that these stocks are more impacted by negative sentiment. These variations suggest that sentiment plays a critical role in influencing returns in the Chinese market when portfolios are sorted by sentiment, underscoring the importance of considering investor sentiment in portfolio construction.

Table 8

CH-3* factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Growth	EP 2	EP 3	EP 4	Value
Alpha					
Small	0.93***	1.07***	0.98***	0.80***	0.66***
Size 2	0.41***	0.42***	0.58***	0.59***	0.31**
Size 3	-0.11	0.11	0.14	0.45***	0.02
Size 4	-0.21	-0.04	0.28**	0.19	0.03
Big	0.13	0.22	0.32*	-0.03	-0.28**
β_{MKT}					
Small	0.97***	0.99***	0.97***	0.98***	0.91***
Size 2	1.02***	0.99***	0.97***	0.97***	0.96***
Size 3	1.06***	1.03***	1.00***	0.97***	0.96***
Size 4	1.10***	1.05***	1.00***	0.97***	1.02***
Big	1.16***	1.02***	0.97***	1.01***	1.10***
β_{SMB}					
Small	1.16***	1.15***	1.22***	1.21***	1.08***
Size 2	0.94***	1.03***	1.08***	1.03***	0.97***
Size 3	0.90***	0.98***	0.94***	0.86***	0.82***
Size 4	0.63***	0.79***	0.68***	0.68***	0.65***
Big	0.05	0.18***	0.23***	0.25***	0.14***
β_{VMG}					
Small	-0.33***	-0.39***	-0.33***	-0.23***	-0.14**
Size 2	-0.33***	-0.41***	-0.46***	-0.28***	0.05
Size 3	-0.31***	-0.47***	-0.33***	-0.29***	0.11*
Size 4	-0.35***	-0.51***	-0.52***	-0.33***	0.18***
Big	-0.36**	-0.55***	-0.51***	-0.19***	0.43***

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-value sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB) and value (β_VMG). The portfolios are categorized into five value levels, ranging from Growth to Value, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all but the smallest 30 % by market capitalization and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Table 9

CH-G factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Growth	EP 2	EP 3	EP 4	Value
Alpha					
Small	1.00***	1.00***	0.87***	0.72***	0.61***
Size 2	0.46***	0.37***	0.51***	0.54***	0.34**
Size 3	-0.06	0.11	0.08	0.41***	0.10
Size 4	-0.03	0.00	0.25**	0.17	0.05
Big	0.39	0.24*	0.36*	0.02	-0.24*
β_{MKT}					
Small	0.96***	0.99***	0.97***	0.97***	0.91***
Size 2	1.02***	0.99***	0.97***	0.97***	0.95***
Size 3	1.06***	1.02***	1.00***	0.97***	0.95***
Size 4	1.09***	1.04***	1.00***	0.96***	1.02***
Big	1.15***	1.02***	0.96***	1.01***	1.10***
β_{SMB}					
Small	1.13***	1.17***	1.25***	1.23***	1.09***
Size 2	0.92***	1.04***	1.10***	1.04***	0.95***
Size 3	0.88***	0.97***	0.96***	0.86***	0.79***
Size 4	0.57***	0.77***	0.69***	0.68***	0.65***
Big	-0.03	0.18***	0.22***	0.24***	0.13***
β_{VMG}					
Small	-0.10	-0.20***	-0.14***	-0.04	0.04
Size 2	-0.15***	-0.24***	-0.29***	-0.11**	0.23***
Size 3	-0.14**	-0.31***	-0.18***	-0.15***	0.28***
Size 4	-0.20**	-0.36***	-0.40***	-0.21***	0.31***
Big	-0.28*	-0.50***	-0.46***	-0.13***	0.47***
β_{SENTI}					
Small	-0.27***	-0.01	0.07	0.01	-0.01
Size 2	-0.20***	-0.03	0.01	-0.01	-0.16*
Size 3	-0.21***	-0.10*	0.00	-0.03	-0.25***
Size 4	-0.38***	-0.15**	-0.02	-0.06	-0.11*
Big	-0.49***	-0.08	-0.11	-0.13**	-0.09

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-value sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB), value (β_VMG), and sentiment (β_SENTI). The portfolios are categorized into five value levels, ranging from Growth to Value, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all but the smallest 30 % by market capitalization and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

5.4.4 Model comparison by GRS Test

GRS tests are conducted to provide insights into the CH-G multifactor model's pricing ability, as this test is analysing whether the intercepts (alphas) of the regression equations for a set of portfolios are jointly equal to zero (H0) and low P-values ($P < 0.05$) lead to the rejection of the null hypothesis. The results are provided in Table 10.

Table 10
Model comparison by GRS test.

	CH-3*		CH-G	
	F-stat	p-value	F-stat	p-value
Panel A: Factors formed on all individual stocks				
5x5 size-value	2.50	0.0004	2.40	0.0006
5x5 size-sentiment	1.74	0.0231	1.56	0.0545
Panel B: Factors formed on all but the smallest 30% by market capitalization				
5x5 size-value	3.52	0.0000	3.90	0.0000
5x5 size-sentiment	3.23	0.0000	3.45	0.0000
Panel C: Factors formed on the smallest 30% by market capitalization only				
5x5 size-value	1.98	0.0068	2.00	0.0061
5x5 size-sentiment	1.39	0.1157	1.46	0.0861

The table presents the results of the GRS test, specifically the F-statistics and the P-value for each factor model and each set of 25 portfolios, once sorted for size and value and once for size and sentiment. The Panels distinguish between the input for factors. Sample period is January 2009 – December 2023.

The GRS test results exhibit evidence that neither of the factor models with factors formed on all but the smallest 30% can fully capture the returns of either portfolio. This is consistent with previous research regarding CH-3 (Liu, Stambaugh, and Yuan 2019; Li and Rao 2022). Contrasting, the p-values for both models exceed 0.05 for 5x5 size-sentiment portfolios within Panel C (factors formed on the smallest 30% by market capitalization). Hence, when deviating from the previous method to omit the microcaps and instead use only them to form the CH-3* and CH-G factors, these better explain the size-sentiment returns, with alphas not significantly different from zero. This outcome may be attributed to the greater susceptibility of smaller stocks to sentiment-driven effects. While both p-values fail to reject the null hypothesis on a 5% significance level, with CH-3* exhibiting a stronger p-value, the F-statistic of the CH-G model is slightly higher, indicating marginally higher pricing ability. When considering all individual stocks to form the CH-3* and CH-G factors (Panel A), the CH-3* model significantly rejects the null hypothesis at a 5% level, whereas the CH-G model marginally fails to do so, which suggests that these factors can explain the 5x5 size-sentiment portfolio returns.

5.5 Robustness checks

5.5.1 Horse Race among direct sentiment measures

Although the literature review (Chapter 2.3.4) led to the conclusion that the Chinese individual investor's sentiment is particularly present in and measurable through the sentiment of Guba postings, as these are published by the investors directly, other sentiment factors could also be relevant. Potential candidates could be the sentiment extracted from online or traditional newspapers as individual investors might use those as a source of information before making investment decisions. The CNDRS provides two datasets similar to the Guba dataset, which lists the number of positive and negative sentiments of online and traditional news articles. The online news includes more than 400 internet media sources, while the latter subsumes 600 sources (Yongan Xu et al. 2023; CNRDS 2024). The coherent factors have been calculated in accordance with Antweiler and Frank (2004) and hence, comparable to the Guba_1 sentiment factor constructed in Chapter 4.2.6. In the following, the Fama MacBeth regressions have been repeated to run a horse race among Guba, online, and traditional news sentiment.
First, the data sets need to be prepared. Following the methodology described in Chapter 4.2.6., the daily counts of positive and negative articles were summed to a monthly level and the *Online* and *Trad* factors were calculated. Next, the filtered dataset excluding shell companies (Chapter 4.2.7), the *Guba* dataset, the *Trad* dataset, and the *Online* dataset were merged into one to exclude combinations of stock code and dates without values for all factors tested and hence ensure comparability among the factors. Although data for traditional newspapers is available between October 2000 and December 2023, the merger with Guba and the decision to exclude the Guba values for 2008 shifts the start month to January 2009. Further, the online newspaper dataset is only available until December 2022. As a result, the period taken into consideration is shortened to January 2009 to December 2022 (168 months), with 3,537 unique stock codes and 162,327 observations (excluding the smallest 30% of stocks by market cap).
Table 11 replicates the procedure of the Fama-MacBeth regressions run in Chapter 5.3.1

over the period from January 2009 to December 2022 and adds the regression results for online and traditional news sentiment.

Table 11

Fama-MacBeth regressions of stock returns on beta, size, and value and sentiment factors.

Model	CH-3*	CH-G	CH-O	CH-T	All
Intercept	0.0892	0.0903	0.0890	0.0888	0.0900
	(2.61)	(2.67)	(2.61)	(2.61)	(2.66)
β	0.0027	0.0033	0.0027	0.0026	0.0033
	(0.60)	(0.76)	(0.60)	(0.59)	(0.75)
logME	-0.0037	-0.0039	-0.0037	-0.0037	-0.0039
	(-2.74)	(-2.89)	(-2.73)	(-2.74)	(-2.89)
EP+	0.3676	0.3530	0.3724	0.3665	0.3566
	(3.01)	(2.89)	(3.04)	(3.00)	(2.92)
D(EP<0)	-0.0046	-0.0039	-0.0046	-0.0045	-0.0043
	(-2.53)	(-2.36)	(-2.60)	(-2.52)	(-2.51)
Guba_1		0.0104			0.015
		(3.08)			(3.15)
Online			-0.0004		-0.0007
			(-0.71)		(-1.24)
Traditional				-0.0006	0.0005
				(1.09)	(1.04)
Adjusted R^2	0.0571	0.0610	0.0580	0.0581	0.0625

The table presents the average slope coefficients obtained from month-on-month Fama–MacBeth regressions, which cross-sectionally regress next month's individual stock returns on this month's individual firm characteristics. Each column represents a different regression model, with nonempty rows indicating the variables included in the analysis. The explanatory variables comprise of the monthlyized daily beta value, which quantifies the stock's sensitivity to movements within its bourse; the logarithm of month-end market capitalization (logME); EP+, representing positive earnings-to-price values, with zero assigned otherwise; and D(EP < 0), which takes the value of one if earnings are negative, and zero otherwise. The subsequent rows add one sentiment factor at a time, starting with Guba_1, followed by Online and Traditional news sentiment calculated similarly, with the last column displaying all factors at once. The final row displays the average adjusted R-squared for each regression specification. T-statistics, calculated using Newey-West (1987) standard errors with 4 lags, are shown in parentheses. The sample size is all individual stocks but the smallest 30% by market capitalization and the sample period is January 2009 – December 2022.

Among the sentiment factors tested, Guba_1 emerges as the most effective singular factor, significantly influencing stock returns with a coefficient of 0.0104 and a t-statistic of 3.08, which is significant at the 5% level. This is reflected in the relatively high adjusted R^2 of 0.0610.

In contrast, the online and traditional factors do not exhibit statistical significance at the

5% level, with t-statistics of -0.71 and 1.09, respectively, suggesting that these sentiment measures have no explanatory power when considered individually. The underperformance of traditional news sentiment is consistent with Xu et al. (2023). Further, news- and media-based research is subject to criticism for portraying the sentiment of the author rather than reflecting that of the investors (X. Zhang et al. 2018), with online news sentiment being attributed to the rational behaviour of institutional investors during recession periods (Yongan Xu et al. 2023) - both of which are groups which do not dominate the Chinese market (see Chapter 2.1.3).

Despite their lack of significance, when all sentiment factors are combined in the final regression model, there is a modest increase in the adjusted R^2 to 0.0625, indicating that while individually these factors may not be significant, their collective inclusion still marginally enhances the model's explanatory power. This suggests while the Guba_1 sentiment measure remains the most reliable predictor of stock returns among the sentiment factors tested a combined approach might still capture additional nuances in market sentiment.

5.5.2 Turnover as a sentiment factor

As highlighted in Chapter 2.1.3., turnover does measure the individual investors sentiment, contrary to traditional and online news sentiment. The turnover factor is introduced as the fourth factor of the CH-4 model and yielded significant test statistics for explaining anomalies in the Chinese stock market (Liu, Stambaugh, and Yuan 2019). Liu et al. (2019) is followed in constructing the factor PMO (Pessimistic minus Optimistic), which is based on abnormal turnover, defined as the previous month's turnover divided by the previous year's turnover calculated on an individual stock level. Next, the stocks get sorted for size-neutralized abnormal turnover with the high turnover (optimistic) stocks being subtracted from the lower (pessimistic) ones. The SMB factor is the average of SMB_{Value} and $SMB_{Turnover}$, similar to the EP-neutralized and Guba-neutralized SMBs in Chapter 4.2.7. Market and value factor construction remains unchanged and together, they build the reconstructed four-factor model CH-4*.

The summary statistics exhibit values similar to other scholars (Liu, Stambaugh, and Yuan 2019; Li and Rao 2022), with turnover exhibiting a much higher excess kurtosis and a more negative skew than the sentiment measure (Appendix 23), making coherent analysis results more difficult to interpret, even when using adapted t-statistics such as Newey-West (1987). Average adjusted R-square values are slightly lower than those of CH-G (Appendix 26). Next, the GRS test is conducted to analyse the CH-4* model's ability to explain the stock returns in comparison to CH-G within the analysis period and for the size-neutral value- and sentiment-sorted portfolios.

Table 12 demonstrates that CH-4* is not able to explain the cross-sectional returns of the different portfolio sorts with factors constructed on Panels A and B and is marginally failing to reject the null hypothesis on a 5% level when sorting for size and sentiment and forming the factors on microcap. This result contrasts the original authors (Liu, Stambaugh, and Yuan 2019) while supporting subsequent research (Z Li and Rao 2022) and substantiating CH-G as a powerful model.

Table 12

CH-4* model GRS test results.

	CH-4*	
	F-stat	p-value
Panel A: Factors formed on all individual stocks		
5x5 size-value	2.23	0.0017
5x5 size-sentiment	1.63	0.0390
Panel B: Factors formed on all but the smallest 30% by market capitalization		
5x5 size-value	3.08	0.0000
5x5 size-sentiment	2.90	0.0000
Panel C: Factors formed on the smallest 30% by market capitalization only		
5x5 size-value	1.93	0.0088
5x5 size-sentiment	1.56	0.0560

The table presents the results of the GRS test, specifically the F-statistics and the P-value for the reconstructed CH-4* factor model and each set of 25 portfolios, once sorted for size and value and once for size and sentiment. The Panels distinguish between the input for factors. Sample period is January 2009 – December 2023.

5.5.3 Daily Sentiment Measure

Daily data is more volatile, which results in difficulties in interpreting statistical test results. Hence, the thesis has been focusing on monthly data. However, it is theorized that social-media-based sentiment proxies impact stock returns as fast as on the same day (Sul, Dennis, and Yuan 2017). In the following, the periodically is reduced from one month to one day while the factor construction remains similar, meaning that the firm characteristics (including sentiment) of time t will be used to analyse the excess returns of t+1. EP is backdated up to 240 days if no data is available due to the periodic nature of semi-annually or quarterly report publications. The difference in monthly and daily unique stock code numbers in Appendix 2 stems from dissimilar data availability in the source databases (e.g. if returns are available daily but not monthly) and slight differences in filtering. The latter is most evident in Filter 3 (Trading Status), as the original paper eliminates stocks with less than 120 trading records a year or 15 trading days per month to prevent the inclusion of returns following long trading suspensions (Liu, Stambaugh, and Yuan 2019). Hereby, the latter should lead to an irregularity: during the Chinese New Year period in spring, the Chinese stock market is prone to experience less than 15 trading days per month in February due to the annual nation-wide holidays and the relatively short month. Coherently, this will impact all stocks in these months and, consequently, there should be no factor values calculatable. However, this is not the case in the factor data published by Liu et al. (2019) on the corresponding website (Stambaugh 2024). Although the thought process of the original authors cannot be replicated regarding this matter, the filtering criteria are kept for monthly data. However, the daily records are included if the trading volume on that day is larger than zero and the trading status of the stock is *normal* (e.g. no *special treatment* designation).

The GRS test is utilized for determining if the CH-G's success in Panel A and Panel C can persist across different periods periodically. Due to limitations of memory capacity on the private laptop of the author, the data had to be split in different periods to be able to perform the data analysis. Therefore, 3-year and 5-year periods have been chosen.

Similar to the monthly results (Table 10), the p-values in Table 13 are very low for the 5x5 size-value portfolios, indicating that the model fails to fully capture the factors that affect daily returns in these sample portfolios. However, the 5x5 size-sentiment sort exhibits a more mixed result: During the 2012 – 2014, 2015 – 2017 2021 – 2023 and 2019 – 2023 periods, the p-values are above 0.05, indicating that for these years, the CH-

G model sufficiently explains the daily stock returns. In other terms, there is no strong evidence during these periods to suggest that key risk factors are missing from the model. It is intriguing that these periods are coinciding with highly irregular events in the Chinese market, namely the Asian Financial Crisis 2015 and the COVID-19 pandemic, indicating that those events could be influenced by individual investor sentiment. For subpanels B and C, the GRS test does yield p-values above 0.05 across all the periods and factor sorts. Hence, this suggests that the CH-G model might not be capturing all the necessary risk factors impacting the returns, if the factors are created using these subpanels.

Table 13
GRS test results for daily CH-G.

	3-year periods					5-year period		
	2009 –2011	2012 – 2014	2015 –2017	2018 –2020	2021 –2023	2009 –2013	2014 –2018	2019 –2023
Panel A: Factors formed on all individual stocks								
5x5 size-value	1.97	2.77	2.06	2.31	2.20	2.88	3.00	2.94
	(0.0033)	(0.0000)	(0.0018)	(0.0003)	(0.0007)	(0.0000)	(0.0000)	(0.0000)
5x5 size-	1.80	1.48	1.42	2.29	0.98	2.92	1.71	1.39
sentiment	(0.0101)	(0.0639)	(0.0833)	(0.0004)	(0.4877)	(0.0000)	(0.0166)	(0.0984)

The table presents the results of the GRS test on daily data for Panel A (factors formed on all individual stocks). The F-statistics are reported for each set of 25 portfolios, once sorted for size and value and once for size and sentiment. P-values are presented in brackets. The Panels distinguish between the input for factors. Each sample period starts on the first day (01.01) and ends on the last day (31.12.)

VI. Limitations and Further Research

6.1 Limitations

6.1.1 China's Regulatory Environment

One limitation of this thesis arises from the unique characteristics of the Chinese stock market, which introduces challenges in applying traditional asset pricing models. Regulatory interventions, such as the requirements for and restrictions on IPOs, can distort market behaviour. Given that short-selling can significantly influence stock prices and market efficiency, the disregard for said restrictions in the construction of the factor models might lead to incomplete or biased interpretations of the Chinese market dynamics.

6.1.2 Data and Data Sampling

Another limitation lies in the data and data sampling methodologies employed. While this thesis tests both, the inclusion and exclusion of microcaps for factor construction, other filters have not been adapted and tested.
First, the analysis is confined to a the 2009 to 2023 period, which may not fully capture the long-term trends or the effects of significant economic and market developments. Second, the thesis focuses on selected stock markets and A-shares, while excluding the Beijing or STAR stock market and all B-shares. This exclusion could result in a loss of insights regarding young and regulatorily enhanced stock markets (Allen et al. 2024) as well as the behaviour of foreign investors, who predominantly trade in B-shares (Carpenter and Whitelaw 2017). Third, the sentiment data used, particularly from Guba, also presents challenges: while Guba is the leading platform for online stock discussions (Yuqin Huang, Qiu, and Wu 2016; Y. Li et al. 2020; X. Wang et al. 2022; D. Dong et al. 2022), other blogs may provide exclusive insights. Furthermore, sentiment data for non-trading days may be lost, as the merging operation for firm characteristics and sentiment

on date and stock code is done after excluding stocks with long trading suspensions (Appendix 2, Filter 3). This potentially skews the results and underestimates the influence of sentiment during critical market periods.

6.1.3 Methodology and Model Construction

Another limitation of this thesis is related to the methodology and construction of factor models. Notably, the thesis does not adequately control for market volatility and turnover, both of which are significant factors in asset pricing, especially in a market as dynamic as China's. Cohesively, other variations such as the CH-4 (Liu, Stambaugh, and Yuan 2019), or the CH-4-R model Li and Rao 2022), which include a turnover factor were not utilized. Further, the method employed to calculate the Guba sentiment factor, based on Antweiler and Frank's (2004) approach, oversimplifies sentiment by only accounting for the number of positive and negative posts. Although the thesis' focus was on sentiment rather than attention, this method fails to capture the nuanced differences in sentiment across posts and does not adequately reflect the impact of post volume on the sentiment index (Y. Sun et al. 2021; Wang et al. 2022). Additionally, while this thesis attempts to control for heteroscedasticity using Newey-West (1987) and White (1980) standard errors, the presence of heteroscedasticity in the data remains a potential issue that could impact the robustness of the results.

6.1.4 Additional Control Variables

A further limitation of this thesis is the omission of several critical control variables that could enhance the robustness and accuracy of the findings. This thesis has not incorporated external macroeconomic factors, such as changes in interest rates, exchange rates, or global economic conditions, which can have significant effects on asset prices. Additionally, firm fixed effects or other micro-level variables are also omitted. The exclusion of these variables means that the findings may not fully account for the broader economic environment that influences stock returns.

6.2 Future Research

Future research could improve upon these limitations by exploring the impact of regulatory changes, such as short-selling restrictions and IPO reforms, on asset pricing. Incorporating these elements into the analysis could provide a more accurate depiction of market dynamics and help in developing models that are more tailored to the Chinese financial landscape. Expanding the dataset to include a broader range of market segments, such as B-shares, the Beijing stock market, and new stock exchanges such as the STAR market, as well as extending the period, could provide a more comprehensive view. Moreover, exploring alternative factor models could help to identify more specific anomalies and improve the explanatory power of the models in the Chinese context. Concurrently, integrating the missing control variables, particularly market volatility, turnover, and external macroeconomic factors should be considered. In terms of factor construction, the methodology could be adapted in accordance with Liu, Zhou, and Zhu (2024), who triple-sort the factors Market, Size, Value, and *Trend*.

Further, future research could also either incorporate sentiment data from other online platforms or consider attention measures alongside sentiment, similar to other scholars who either disregarded Chinese factor model research, social media sentiment measures in China, or the most recent years of data available (Da, Engelberg, and Gao 2011; Sun, Fang, and Wang 2018; Zhang and Tao 2019; Dong et al. 2022; Chen et al. 2022). Another research focus could be the impact of sentiment across time, especially in tumultuous periods as daily sentiment showed promising results during the subsamples 2015 – 2017 and 2020 – 2021.

The adoption of more sophisticated sentiment analysis techniques, such as machine learning-based methods, could help to capture the nuances in investor sentiment and attention more accurately. Here, the use of Large Language Models trained in both, Chinese language and Chinese financial idioms, is inevitable Li et al. 2014; Y. Li et al. 2020). Future research including advanced models may rely on existing papers on the Chinese market (Tan, Yan, and Zhu 2019; Luo 2021; Leippold, Wang, and Zhou 2022) or adapt new techniques tested on other markets (Renault 2017; Shapiro, Sudhof, and

Wilson 2022). These advanced techniques would also allow for a shorter periodicity, e.g. intra-day sentiment and trading, as researched on other markets (Sun, Najand, and Shen 2016; Renault 2017).

VII. Conclusion

In recent years, sentiment has entered the spotlight of multifactor models for explaining stock returns. With China exhibiting a large number of individual investors, the interlink between China-specific asset pricing models and sentiment factors is worth exploring. This thesis has focused on the explanation of stock returns in the Chinese financial market from 2009 to 2023. By concentrating on China-specific multifactor models, this research sought to enhance the understanding of stock returns in a market that is rapidly evolving but remains insufficiently explored in academic research.

Firstly, an extensive literature review was conducted to gain deeper insight into the dynamics unique to the Chinese stock market. Regulatory frameworks, such as short-selling restrictions, were examined for their potential impact on pricing efficiency, which can lead to overvaluation of stocks due to constrained arbitrage opportunities. Additionally, the study reviewed IPO regulations to understand Liu et al.'s (2019) decision to exclude the smallest 30% of stocks by market capitalization. This exclusion was premised on the argument that small stocks are contaminated by shell value through reverse mergers, avoiding the more stringent official IPO process. While Liu et al. (2019) observed improved explanatory power by excluding microcaps for factor construction, this thesis yielded opposing results and, hence, concurs with other scholars (Lin et al. 2020; Carpenter, Lu, and Whitelaw 2021; Dong et al. 2022; Hou, Qiao, and Zhang 2023) Rather, the results of this thesis contends that including all stocks for factor model construction may provide a more comprehensive understanding and argues in favour of the inclusion of all stocks in future research.

To represent the influence of a dominating group of individual investors particular to the Chinese market, the key contribution of this thesis is the integration of direct sentiment measurement derived from individual investor posts on the social media platform Guba. This addition led to the development of a sentiment-inclusive four-factor model, named CH-G. The CH-G model was rigorously tested against the established CH-3 model to assess the added value of incorporating sentiment to be able to explain stock returns. Throughout these tests, the exclusion of microcaps, as suggested by Liu et al. (2019), was carefully scrutinized to evaluate its impact on the model's performance.

The results of the analysis demonstrate that incorporating a sentiment factor significantly enhances the explanatory power of asset pricing models in the Chinese market. This improvement is evident in both the adjusted R-squared values and the results of portfolio analysis, which indicate that sentiment plays a crucial role in influencing stock returns in the Chinese market. For instance, the 5x5 portfolio analysis reveals that small stocks are more prone to overvaluation, and the CH-G model exhibits greater explanatory power across all but growth stock categories. This indicates that while growth stocks might be driven by different factors, sentiment plays a vital role in explaining the returns of other stock categories.

Furthermore, the research highlights that sentiment extracted from social media blog posts of individual investors is more effective in explaining stock returns than sentiment derived from traditional or online news sources. Additionally, CH-G beats a fourth-factor model including turnover, which was previously introduced to measure individual investor sentiment. The results of both robustness tests underscore the importance of considering and directly measuring the sentiment of retail investors in the Chinese stock market. Hence, these findings have significant implications for both theory and practice: integrating sentiment factors into asset pricing models can enhance their accuracy and relevance in markets dominated by retail investors, and providing valuable insights for both academics and practitioners seeking to understand or predict stock market behavior in China.

While this thesis provides new insights into the Chinese stock market, it also acknowledges the limitations posed by the market's relative youth and the frequent – and often severe – regulatory changes. These factors contribute to the complexity and unpredictability of the market and as the market continues to grow and mature, the need for robust, adaptable models will remain critical. Future research could benefit from incorporating more sophisticated sentiment analysis techniques, such as machine learning-based methods that account for nuanced investor sentiment and attention, especially during volatile periods, to better capture the dynamic nature of the Chinese financial landscape.

In conclusion, this thesis not only contributes to the existing literature on asset pricing in emerging markets but also highlights the significant impact of investor sentiment in the Chinese context. The findings advocate for an inclusion of microcap stocks in the course of future multifactor model research focused on explaining stock returns on the Chinese financial market. Further, it aims to show that promising adaptions to asset pricing models can be discovered through studying the main influencing factors of the Chinese market.

Bibliography

Abreu, Dilip, and Markus K Brunnermeier. 2002. "Synchronization Risk and Delayed Arbitrage." *Journal of Financial Economics* 66 (2): 341–60. https://doi.org/https://doi.org/10.1016/S0304-405X(02)00227-1.

Abreu, Dilip, and Markus K. Brunnermeier. 2003. "Bubbles and Crashes." *Econometrica* 71 (1): 173–204. https://doi.org/10.1111/1468-0262.00393.

Allen, Franklin, Jun Qian, and Meijun Qian. 2005. "Law, Finance, and Economic Growth in China." *Journal of Financial Economics* 77 (1): 57–116. https://doi.org/10.1016/J.JFINECO.2004.06.010.

Allen, Franklin, Jun Qian, Chenyu Shan, and Julie Lei Zhu. 2024. "Dissecting the Long-Term Performance of the Chinese Stock Market." *Journal of Finance* 79 (2): 993–1054. https://doi.org/10.1111/jofi.13312.

Antweiler, Werner, and Murray Z Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance* 59 (3): 1259–94. https://doi.org/https://doi.org/10.1111/j.1540-6261.2004.00662.x.

Baker, Malcolm, and Jeffrey Wurgler. 2007. "Investor Sentiment in the Stock Market." *Journal of Economic Perspectives* 21 (2): 129–51. https://doi.org/10.1257/jep.21.2.129.

Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan. 2012. "Global, Local, and Contagious Investor Sentiment." *Journal of Financial Economics* 104 (2): 272–87. https://doi.org/10.1016/J.JFINECO.2011.11.002.

Baker, Malcom, and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *The Journal of Finance* 61 (4): 1645–80. https://doi.org/10.1111/j.1540-6261.2006.00885.x.

Banz, Rolf W. 1981. "The Relationship between Return and Market Value of Common Stocks."

Barber, Brad M, and Terrance Odean. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *The Journal of Finance* 55 (2): 773–806. https://api.semanticscholar.org/CorpusID:154973050.

Barber, Brad M, Terrance Odean, and Ning Zhu. 2009. "Do Retail Trades Move Markets?" *The Review of Financial Studies* 22 (1): 151–86. https://doi.org/10.1093/rfs/hhn035.

Behrendt, Simon, Franziska J Peter, and David J Zimmermann. 2020. "An Encyclopedia for Stock Markets? Wikipedia Searches and Stock Returns." *International Review* *of Financial Analysis* 72:101563. https://doi.org/https://doi.org/10.1016/j.irfa.2020.101563.

Berger, David, and Harry J. Turtle. 2015. "Sentiment Bubbles." *Journal of Financial Markets* 23:59–74. https://doi.org/https://doi.org/10.1016/j.finmar.2015.01.002.

Black, Fischer. 1986. "Noise." *Journal of Finance* 41 (3): 529–43. https://doi.org/10.1111/j.1540-6261.1986.tb04513.x.

Blitz, David, Matthias X. Hanauer, and Pim van Vliet. 2021. "The Volatility Effect in China." *Journal of Asset Management* 22 (5): 338–49. https://doi.org/10.1057/S41260-021-00218-0/TABLES/6.

Bloomberg News. 2023. "China Tightens Rules on Short Selling in a Bid to Boost Stocks." October 16, 2023. https://www.bloomberg.com/news/articles/2023-10-16/china-tightens-rules-on-short-selling-in-a-bid-to-boost-stocks.

Boehme, Rodney D, Bartley R Danielsen, and Sorin M Sorescu. 2006. "Short-Sale Constraints, Differences of Opinion, and Overvaluation." *Source: The Journal of Financial and Quantitative Analysis*. Vol. 41.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2 (1): 1–8. https://doi.org/https://doi.org/10.1016/j.jocs.2010.12.007.

Bris, Arturo, William N. Goetzmann, and Ning Zhu. 2007. "Efficiency and the Bear: Short Sales and Markets Around the World." *Journal of Finance* 62 (3): 1029–79. https://doi.org/10.1111/J.1540-6261.2007.01230.X.

Brochado, Ana. 2020. "Google Search Based Sentiment Indexes." *IIMB Management Review* 32 (3): 325–35. https://doi.org/https://doi.org/10.1016/j.iimb.2019.10.015.

Brown, Gregory W., and Michael T. Cliff. 2004. "Investor Sentiment and the Near-Term Stock Market." *Journal of Empirical Finance* 11 (1): 1–27. https://doi.org/10.1016/j.jempfin.2002.12.001.

Brown, Gregory W., and Michael T. Cliff. 2005. "Investor Sentiment and Asset Valuation." *Journal of Business* 78 (2): 405–39. https://doi.org/10.1086/427633.

Brunnermeier, Markus K., Michael Sockin, and Wei Xiong. 2020. "China's Model of Managing the Financial System." 27171. NBER Working Paper Series.

BSE. 2024. "Introduction." 2024. https://www.bse.cn/company/introduce.html.

Bu, Hui, and Li Pi. 2014. "Does Investor Sentiment Predict Stock Returns? The Evidence from Chinese Stock Market." *Journal of Systems Science and Complexity* 27 (1): 130–43. https://doi.org/10.1007/S11424-013-3291-Y/METRICS.

Cakici, Nusret, Yi Tang, and An Yan. 2016. "Do the Size, Value, and Momentum Factors

Drive Stock Returns in Emerging Markets?” *Journal of International Money and Finance* 69:179–204. https://doi.org/https://doi.org/10.1016/j.jimonfin.2016.06.001.

Carpenter, Jennifer N., Fangzhou Lu, and Robert F. Whitelaw. 2021. “The Real Value of China’s Stock Market.” *Journal of Financial Economics* 139 (3): 679–96. https://doi.org/10.1016/J.JFINECO.2020.08.012.

Carpenter, Jennifer N., and Robert F. Whitelaw. 2017. “The Development of China’s Stock Market and Stakes for the Global Economy.” *Annual Review of Financial Economics* 9 (1): 233–57. https://doi.org/10.1146/annurev-financial-110716-032333.

Chan, Kalok, Junbo Wang, and K.C.John Wei. 2004. “Underpricing and Long-Term Performance of IPOs in China.” *Journal of Corporate Finance* 10 (3): 409–30. https://doi.org/https://doi.org/10.1016/S0929-1199(03)00023-3.

Chan, Kam C, Hung-Gay Fung, and Samanta Thapa. 2007. “China Financial Research: A Review and Synthesis.” *International Review of Economics & Finance* 16 (3): 416–28. https://doi.org/https://doi.org/10.1016/j.iref.2005.09.004.

Chang, Eric C, Yan Luo, and Jinjuan Ren. 2014. “Short-Selling, Margin-Trading, and Price Efficiency: Evidence from the Chinese Market.” *Journal of Banking & Finance* 48:411–24. https://doi.org/https://doi.org/10.1016/j.jbankfin.2013.10.002.

Checkley, M S, D Añón Higón, and H Alles. 2017. “The Hasty Wisdom of the Mob: How Market Sentiment Predicts Stock Market Behavior.” *Expert Systems with Applications* 77:256–63. https://doi.org/https://doi.org/10.1016/j.eswa.2017.01.029.

Cheema, Muhammad A, Yimei Man, and Kenneth Szulczyk. 2020. “Does Investor Sentiment Predict the Near-Term Returns of the Chinese Stock Market?” *International Review of Finance* 20 (1): 225–33. https://EconPapers.repec.org/RePEc:bla:irvfin:v:20:y:2020:i:1:p:225-233.

Chen, Haiqiang, Terence Tai Leung Chong, and Yingni She. 2014. “A Principal Component Approach to Measuring Investor Sentiment in China.” *Quantitative Finance* 14 (4): 573–79. https://doi.org/10.1080/14697688.2013.869698.

Chen, Jean J. 2004. “Determinants of Capital Structure of Chinese-Listed Companies.” *Journal of Business Research* 57 (12 SPEC.ISS.): 1341–51. https://doi.org/10.1016/S0148-2963(03)00070-5.

Chen, Jian, Guohao Tang, Jiaquan Yao, and Guofu Zhou. 2022. “Investor Attention and Stock Returns.” *Journal of Financial and Quantitative Analysis* 57 (2): 455–84. https://doi.org/DOI: 10.1017/S0022109021000090.

Chen, Shimin, Zheng Sun, Song Tang, and Donghui Wu. 2011. "Government Intervention and Investment Efficiency: Evidence from China." *Journal of Corporate Finance* 17 (2): 259–71. https://doi.org/10.1016/j.jcorpfin.2010.08.004.

Chen, Sze Ting, and Kai Yin Allison Haga. 2021. "Using E-GARCH to Analyze the Impact of Investor Sentiment on Stock Returns Near Stock Market Crashes." *Frontiers in Psychology* 12 (July). https://doi.org/10.3389/fpsyg.2021.664849.

Cheung, C, George Hoguet, and S Ng. 2014. "Value, Size, Momentum, Dividend Yield, and Volatility in China's A-Share Market." *The Journal of Portfolio Management* 41 (2): 57–70. https://doi.org/10.3905/jpm.2015.41.5.057.

Chiang, Thomas C., Jiandong Li, and Lin Tan. 2010. "Empirical Investigation of Herding Behavior in Chinese Stock Markets: Evidence from Quantile Regression Analysis." *Global Finance Journal* 21 (1): 111–24. https://doi.org/10.1016/j.gfj.2010.03.005.

Chiang, Thomas C., and Dazhi Zheng. 2010. "An Empirical Analysis of Herd Behavior in Global Stock Markets." *Journal of Banking and Finance* 34 (8): 1911–21. https://doi.org/10.1016/j.jbankfin.2009.12.014.

Chu, Xiaojun, Chongfeng Wu, and Jianying Qiu. 2016. "A Nonlinear Granger Causality Test between Stock Returns and Investor Sentiment for Chinese Stock Market: A Wavelet-Based Approach." *Applied Economics* 48 (21): 1915–24. https://doi.org/10.1080/00036846.2015.1109048.

CNRDS. 2024. "Financial Report Disclosure Time Database."

Cochrane, John H. 2009. *Asset Pricing: (Revised)*. Princeton University Press.

Cong, Lin William, and Sabrina T Howell. 2021. "Policy Uncertainty and Innovation: Evidence from Initial Public Offering Interventions in China." *Management Science* 67 (11): 7238–61. https://doi.org/10.1287/mnsc.2020.3807.

Da, Jiaqi, Annuar Nassir, Mohd Padzil Hashim, and Wei Theng Lau. 2022. "The Impacts of The Systematic, Idiosyncratic Risks and Market Sentiment on China A-Shares Performances." *International Journal of Economics and Management* 16 (2): 147–62. https://doi.org/10.47836/ijeam.16.2.01.

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. "In Search of Attention." *The Journal of Finance* 66 (5): 1461–99. https://doi.org/10.1111/j.1540-6261.2011.01679.x.

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2015. "The Sum of All FEARS Investor Sentiment and Asset Prices." *The Review of Financial Studies* 28 (1): 1–32. https://doi.org/10.1093/rfs/hhu072.

Darrat, Ali F, and Maosen Zhong. 2000. "On Testing the Random-Walk Hypothesis: A Model-Comparison Approach." *Financial Review* 35 (3): 105–24.

https://doi.org/https://doi.org/10.1111/j.1540-6288.2000.tb01423.x.

Demirer, Riza, and Ali M. Kutan. 2006. “Does Herding Behavior Exist in Chinese Stock Markets?” *Journal of International Financial Markets, Institutions and Money* 16 (2): 123–42. https://doi.org/10.1016/J.INTFIN.2005.01.002.

Deng, Xiaohu, and Lei Gao. 2018. “The Monitoring of Short Selling: Evidence from China.” *Research in International Business and Finance* 43 (January):68–78. https://doi.org/10.1016/j.ribaf.2017.07.087.

Dong, Dayong, Keke Wu, Jianchun Fang, Giray Gozgor, and Cheng Yan. 2022. “Investor Attention Factors and Stock Returns: Evidence from China.” *Journal of International Financial Markets, Institutions and Money* 77:101499. https://doi.org/https://doi.org/10.1016/j.intfin.2021.101499.

Dong, Hang, and Javier Gil-Bazo. 2020. “Sentiment Stocks.” *International Review of Financial Analysis* 72:101573. https://doi.org/https://doi.org/10.1016/j.irfa.2020.101573.

Elton, Edwin J., Martin J. Gruber, and Jeffrey A. Busse. 1998. “Do Investors Care about Sentiment?” *Journal of Business* 71 (4): 477–500. https://doi.org/10.1086/209754.

Endeshaw, Assafa. 2004. “Internet Regulation in China: The Never-ending Cat and Mouse Game.” *Information & Communications Technology Law* 13 (1): 41–57. https://doi.org/10.1080/1360083042000190634.

Fama, Eugene F. 1965. “The Behavior of Stock-Market Prices.” *The Journal of Business* 38 (1): 34–105. http://www.jstor.org/stable/2350752.

Fama, Eugene F. 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work.” *The Journal of Finance* 25 (2): 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x.

Fama, Eugene F. 1991. “Efficient Capital Markets: II.” *The Journal of Finance* 46 (5): 1575–1617. https://doi.org/10.1111/j.1540-6261.1991.tb04636.x.

Fama, Eugene F., and Kenneth R. French. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics* 33 (1): 3–56. https://doi.org/10.1016/0304-405X(93)90023-5.

Fama, Eugene F., and Kenneth R. French. 2015. “A Five-Factor Asset Pricing Model.” *Journal of Financial Economics* 116 (1): 1–22. https://doi.org/10.1016/J.JFINECO.2014.10.010.

Fama, Eugene F, and James D Macbeth. 1973. “Risk, Return, and Equilibrium: Empirical Tests.” *Journal of Political Economy* 81 (3): 607–36.

Fan, Wenzhong. 2021. “Construction Methods for the Shanghai Stock Exchange Indexes:

1870-1940." New Haven. https://som.yale.edu/sites/default/files/2021-12/SSE-CC.pdf.

Fang, Jianchun, Giray Gozgor, Chi-Keung Marco Lau, and Zhou Lu. 2020. "The Impact of Baidu Index Sentiment on the Volatility of China's Stock Markets." *Finance Research Letters* 32 (January). https://doi.org/10.1016/j.frl.2019.01.011.

Fernald, John, and John H Rogers. 2002. "Puzzles in the Chinese Stock Market." *The Review of Economics and Statistics* 84 (3): 416–32.

Fisher, Kenneth L., and Meir Statman. 2000. "Investor Sentiment and Stock Returns." *Financial Analysts Journal* 56 (2): 16–23. https://doi.org/10.2469/faj.v56.n2.2340.

Gan, Baoqing, Vitali Alexeev, Ron Bird, and Danny Yeung. 2020. "Sensitivity to Sentiment: News vs Social Media." *International Review of Financial Analysis* 67 (January):101390. https://doi.org/10.1016/j.irfa.2019.101390.

Gao, Jianwei, Haiwei Li, and Zhou Lu. 2023. "Impact of COVID-19 on Investor Sentiment in China's Stock Markets." *Heliyon* 9 (10): e20801. https://doi.org/10.1016/j.heliyon.2023.e20801.

Gao, Zhenyu, Haohan Ren, and Bohui Zhang. 2020. "Googling Investor Sentiment around the World." *Journal of Financial and Quantitative Analysis* 55 (2): 549–80. https://doi.org/DOI: 10.1017/S0022109019000061.

Ge, Yidi, Jiangnan Qiu, Zhiyong Liu, Wenjing Gu, and Liwei Xu. 2020. "Beyond Negative and Positive: Exploring the Effects of Emotions in Social Media during the Stock Market Crash." *Information Processing & Management* 57 (4): 102218. https://doi.org/10.1016/J.IPM.2020.102218.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken. 1989. "A Test of the Efficiency of a given Portfolio." Vol. 57.

Greenwood, Robin, and Stefan Nagel. 2009. "Inexperienced Investors and Bubbles." *Journal of Financial Economics* 93 (2): 239–58. https://doi.org/10.1016/J.JFINECO.2008.08.004.

Greenwood, Robin, and Andrei Shleifer. 2014. "Expectations of Returns and Expected Returns." *Review of Financial Studies* 27 (3): 714–46.

Griffin, John M. 2002. "Are the Fama and French Factors Global or Country Specific?" *The Review of Financial Studies* 15 (3): 783–803. http://www.jstor.org/stable/2696721.

Grossman, Sanford J, and Joseph E Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *The American Economic Review* 70 (3): 393–408. http://www.jstor.org/stable/1805228.

Gu, Chen, and Alexander Kurov. 2020. "Informational Role of Social Media: Evidence from Twitter Sentiment." *Journal of Banking & Finance* 121:105969. https://doi.org/https://doi.org/10.1016/j.jbankfin.2020.105969.

Gu, Ming, Wenjin Kang, and Bu Xu. 2018. "Limits of Arbitrage and Idiosyncratic Volatility: Evidence from China Stock Market." *Journal of Banking & Finance* 86:240–58. https://doi.org/https://doi.org/10.1016/j.jbankfin.2015.08.016.

Guo, Dezhi, Yiyin Zheng, Weishen Wang, Preng-Nien Hu, Ziqi Yang, and Zejun Chen. 2022. "The Impact of Different Sentiment in Investment Decisions: Evidence from China's Stock Markets IPOs." *Economic Research-Ekonomska Istraživanja* 36 (August):1–21. https://doi.org/10.1080/1331677X.2022.2113739.

Guo, Kun, Yi Sun, and Xin Qian. 2017. "Can Investor Sentiment Be Used to Predict the Stock Price? Dynamic Analysis Based on China Stock Market." *Physica A: Statistical Mechanics and Its Applications* 469:390–96. https://doi.org/https://doi.org/10.1016/j.physa.2016.11.114.

Han, Chunmao, and Yongdong Shi. 2022. "Chinese Stock Anomalies and Investor Sentiment." *Pacific-Basin Finance Journal* 73:101739. https://doi.org/https://doi.org/10.1016/j.pacfin.2022.101739.

Han, Xing, Kai Li, and Youwei Li. 2020. "Investor Overconfidence and the Security Market Line: New Evidence from China." *Journal of Economic Dynamics and Control* 117:103961. https://doi.org/https://doi.org/10.1016/j.jedc.2020.103961.

Han, Xing, and Youwei Li. 2017. "Can Investor Sentiment Be a Momentum Time-Series Predictor? Evidence from China." *Journal of Empirical Finance* 42 (June):212–39. https://doi.org/10.1016/j.jempfin.2017.04.001.

Hao, Jing, and Xiong Xiong. 2021. "Retail Investor Attention and Firms' Idiosyncratic Risk: Evidence from China." *International Review of Financial Analysis* 74:101675. https://doi.org/https://doi.org/10.1016/j.irfa.2021.101675.

Hasan, Iftekhar, Paul Wachtel, and Mingming Zhou. 2009. "Institutional Development, Financial Deepening and Economic Growth: Evidence from China." *Journal of Banking & Finance* 33 (1): 157–70. https://doi.org/https://doi.org/10.1016/j.jbankfin.2007.11.016.

He, Gang, Shuzhen Zhu, and Haifeng Gu. 2017. "On the Construction of Chinese Stock Market Investor Sentiment Index." Edited by David McMillan. *Cogent Economics & Finance* 5 (1): 1412230. https://doi.org/10.1080/23322039.2017.1412230.

He, Qing, and Cai Fang. 2019. "Regulatory Sanctions and Stock Pricing Efficiency: Evidence from the Chinese Stock Market." *Pacific Basin Finance Journal* 58

(December). https://doi.org/10.1016/j.pacfin.2019.101241.

Hess, Patrick. 2010. “Securities Clearing and Settlement in China: Markets, Infrastructures and Policy-Making.” Frankfurt am Main. http://ssrn.com/abstract_id=1632989.

Hilliard, Jitka, and Haoran Zhang. 2015. “Size and Price-to-Book Effects: Evidence from the Chinese Stock Markets.” *Pacific-Basin Finance Journal* 32:40–55. https://doi.org/https://doi.org/10.1016/j.pacfin.2015.02.003.

Ho, Chienwei, and Chi Hsiou Hung. 2009. “Investor Sentiment as Conditioning Information in Asset Pricing.” *Journal of Banking and Finance* 33 (5): 892–903. https://doi.org/10.1016/j.jbankfin.2008.10.004.

Ho-him, Chan. 2024. “Chinese Regulators Curb Short Selling as Market Downturn Deepens.” January 28, 2024. https://www.ft.com/content/1f9775a8-1dc9-48c1-aa74-c633ca95390f.

Hou, Kewei, Fang Qiao, and Xiaoyan Zhang. 2023. “Finding Anomalies in China.” https://EconPapers.repec.org/RePEc:ecl:ohidic:2023-02.

Hou, Kewei, Chen Xue, and Lu Zhang. 2015. “Digesting Anomalies: An Investment Approach.” *The Review of Financial Studies* 28 (3): 650–705. https://doi.org/10.1093/rfs/hhu068.

Hu, Grace Xing, Can Chen, Yuan Shao, and Jiang Wang. 2019. “Fama–French in China: Size and Value Factors in Chinese Stock Returns.” *International Review of Finance* 19 (1): 3–44. https://doi.org/10.1111/irfi.12177.

Hu, Shulan, Meiling Zhong, and Yanli Cai. 2019. “Impact of Investor Behavior and Stock Market Liquidity: Evidence from China.” *Entropy* 21 (11): 1111. https://doi.org/10.3390/e21111111.

Hu, Xing Grace, Jun Pan, and Jiang Wang. 2021. “Chinese Capital Market: An Empirical Overview.” *Critical Finance Review* 10 (2): 125–206. https://doi.org/10.1561/104.00000097.

Huang, Tzu-Lun. 2018. “The Puzzling Media Effect in the Chinese Stock Market.” *Pacific-Basin Finance Journal* 49:129–46. https://doi.org/https://doi.org/10.1016/j.pacfin.2018.04.005.

Huang, Yi, Jianjun Miao, and Pengfei Wang. 2019. “Saving China’s Stock Market?” *IMF Economic Review* 67 (2): 349–94. https://doi.org/10.1057/s41308-019-00079-z.

Huang, Yuqin, Huiyan Qiu, and Zhiguo Wu. 2016. “Local Bias in Investor Attention: Evidence from China’s Internet Stock Message Boards.” *Journal of Empirical Finance* 38:338–54. https://doi.org/https://doi.org/10.1016/j.jempfin.2016.07.007.

Jones, Charles M, Donghui Shi, Xiaoyan Zhang, and Xinran Zhang. 2021. "Understanding Retail Investors: Evidence from China."

Joseph, Kissan, M. Babajide Wintoki, and Zelin Zhang. 2011. "Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search." *International Journal of Forecasting* 27 (4): 1116–27. https://doi.org/https://doi.org/10.1016/j.ijforecast.2010.11.001.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–91. https://doi.org/10.2307/1914185.

Kumar, Alok, and Charles Lee. 2006. "Retail Investor Sentiment and Return Comovements." *Journal of Finance* 61 (February):2451–86. https://doi.org/10.2139/ssrn.502843.

Lee, Charles, Andrei Shleifer, and Richard Thaler. 1991. "Investor Sentiment and the Closed-End Fund Puzzle." *Journal of Finance* 46 (1): 75–109.

Lee, Wayne Y., Christine X. Jiang, and Daniel C. Indro. 2002. "Stock Market Volatility, Excess Returns, and the Role of Investor Sentiment." *Journal of Banking & Finance* 26 (12): 2277–99. https://doi.org/10.1016/S0378-4266(01)00202-3.

Leippold, Markus, Qian Wang, and Wenyu Zhou. 2022. "Machine Learning in the Chinese Stock Market." *Journal of Financial Economics* 145 (2): 64–82. https://doi.org/10.1016/J.JFINECO.2021.08.017.

Lemmon, Michael, and Evgenia Portniaguina. 2006. "Consumer Confidence and Asset Prices: Some Empirical Evidence." *The Review of Financial Studies* 19 (4): 1499–1529. https://doi.org/10.1093/rfs/hhj038.

Li, Qing, Tiejun Wang, Qixu Gong, Yuanzhu Chen, Zhangxi Lin, and Sa Kwang Song. 2014. "Media-Aware Quantitative Trading Based on Public Web Information." *Decision Support Systems* 61 (1): 93–105. https://doi.org/10.1016/j.dss.2014.01.013.

Li, Xindan, and Bing Zhang. 2008. "Stock Market Behavior and Investor Sentiment: Evidence from China." *Frontiers of Business Research in China* 2 (2): 277–82. https://doi.org/10.1007/s11782-008-0016-5.

Li, Yelin, Hui Bu, Jiahong Li, and Junjie Wu. 2020. "The Role of Text-Extracted Investor Sentiment in Chinese Stock Price Prediction with the Enhancement of Deep Learning." *International Journal of Forecasting* 36 (4): 1541–62. https://doi.org/10.1016/J.IJFORECAST.2020.05.001.

Li, Zhiyong, and Xiao Rao. 2022. "Evaluating Asset Pricing Models: A Revised Factor Model for China." *Economic Modelling* 116 (November). https://doi.org/10.1016/j.econmod.2022.106001.

Liang, Chao, Linchun Tang, Yan Li, and Yu Wei. 2020. “Which Sentiment Index Is More Informative to Forecast Stock Market Volatility? Evidence from China.” *International Review of Financial Analysis* 71:101552. https://doi.org/https://doi.org/10.1016/j.irfa.2020.101552.

Liang, Priscilla, and Thomas D Willett. 2015. “Chinese Stocks during 2000–2013: Bubbles and Busts or Fundamentals?” *The Chinese Economy* 48 (3): 199–214. https://doi.org/10.1080/10971475.2015.1031600.

Liao, Li, Zhisheng Li, Weiqiang Zhang, and Ning Zhu. 2012. “Does the Location of Stock Exchange Matter? A within-Country Analysis.” *Pacific-Basin Finance Journal* 20 (4): 561–82. https://doi.org/https://doi.org/10.1016/j.pacfin.2011.12.007.

Lin, Hung Wen, Jing Bo Huang, Kun Ben Lin, Joyce Zhang, and Shu Heng Chen. 2020. “Which Is the Better Fourth Factor in China? Reversal or Turnover?” *Pacific-Basin Finance Journal* 62:101347. https://doi.org/https://doi.org/10.1016/j.pacfin.2020.101347.

Lin, Qi. 2017. “Noisy Prices and the Fama–French Five-Factor Asset Pricing Model in China.” *Emerging Markets Review* 31 (June):141–63. https://doi.org/10.1016/J.EMEMAR.2017.04.002.

Lintner, John. 1965. “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets.” Vol. 47.

Liu, Hongqi, Cameron Peng, Wei A. Xiong, and Wei Xiong. 2022. “Taming the Bias Zoo.” *Journal of Financial Economics* 143 (2): 716–41. https://doi.org/10.1016/j.jfineco.2021.06.001.

Liu, Jianan, Robert F. Stambaugh, and Yu Yuan. 2019. “Size and Value in China.” *Journal of Financial Economics* 134 (1): 48–69. https://doi.org/10.1016/J.JFINECO.2019.03.008.

Liu, Shuming. 2015. “Investor Sentiment and Stock Market Liquidity.” *Journal of Behavioral Finance* 16 (1): 51–67. https://doi.org/10.1080/15427560.2015.1000334.

Liu, Tengdong, Dazhi Zheng, Suyan Zheng, and Yang Lu. 2023. “Herding in Chinese Stock Markets: Evidence from the Dual-Investor-Group.” *Pacific-Basin Finance Journal* 79 (June):101992. https://doi.org/10.1016/J.PACFIN.2023.101992.

Liu, Yang, Guofu Zhou, and Yingzi Zhu. 2024. “Trend Factor in China: The Role of Large Individual Trading.” *The Review of Asset Pricing Studies*, January, raae003. https://doi.org/10.1093/rapstu/raae003.

Long, J. Bradford De, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann.

1990. “Noise Trader Risk in Financial Markets.” *Journal of Political Economy* 98 (4): 703–38. https://doi.org/10.1086/261703.

LSEG. 2023. “Ground Rules Guide to Chinese Share Classes.” https://www.lseg.com/content/dam/ftse-russell/en_us/documents/policy-documents/guide-to-chinese-share-classes.pdf.

Lu, Wanbo, Yifu Wang, and Xingjian Zhang. 2023. “Which News Topics Drive Economic Prosperity in China?” *PLOS ONE* 18 (10): e0291862. https://doi.org/10.1371/journal.pone.0291862.

Luo, Yiyao. 2021. “Optimization of a Four-Factor Model for China Using Machine Learning Algorithm and Intelligent Computational Method.” *Proceedings of 2021 IEEE International Conference on Emergency Science and Information Technology, ICESIT 2021*, 606–11. https://doi.org/10.1109/ICESIT53460.2021.9696664.

Ma, Rui, Hamish D. Anderson, and Ben R. Marshall. 2018. “Stock Market Liquidity and Trading Activity: Is China Different?” *International Review of Financial Analysis* 56 (March):32–51. https://doi.org/10.1016/J.IRFA.2017.12.010.

Markowitz, Harry. 1952. “Portfolio Selection.” *The Journal of Finance*. Vol. 7.

Miller, Edward M. 1977. “Risk, Uncertainty, and Divergence of Opinion.” *The Journal of Finance* 32 (4): 1151–68. https://doi.org/10.1111/j.1540-6261.1977.tb03317.x.

Moat, Helen Susannah, Chester Curme, Adam Avakian, Dror Y Kenett, H Eugene Stanley, and Tobias Preis. 2013. “Quantifying Wikipedia Usage Patterns Before Stock Market Moves.” *Scientific Reports* 3 (1): 1801. https://doi.org/10.1038/srep01801.

Mok, Henry M K, and Y V Hui. 1998. “Underpricing and Aftermarket Performance of IPOs in Shanghai, China.” *Pacific-Basin Finance Journal* 6 (5): 453–74. https://doi.org/https://doi.org/10.1016/S0927-538X(98)00023-7.

Nardo, Michela, Marco Petracco-Giudici, and Minás Naltsidis. 2016. “Walking Down Wall Street with a Tablet: A Survey of Stock Market Predictions Using the Web.” *Journal of Economic Surveys* 30 (2): 356–69. https://doi.org/https://doi.org/10.1111/joes.12102.

Neal, Robert, and Simon M. Wheatley. 1998. “Do Measures of Investor Sentiment Predict Returns?” *Journal of Financial and Quantitative Analysis* 33 (4): 523–47. https://doi.org/10.2307/2331130.

Newey, Whitney K, and Kenneth D West. 1987. “Hypothesis Testing with Efficient Method of Moments Estimation.” *International Economic Review* 28 (3): 777–87. https://doi.org/10.2307/2526578.

Niu, Hongli, Qiaoying Pan, and Kunliang Xu. 2023. “Hybrid Deep Learning Models with Multi-Classification Investor Sentiment to Forecast the Prices of China's Leading Stocks." *PLOS ONE* 18 (11): e0294460. https://doi.org/10.1371/journal.pone.0294460.

Nofsinger, John R. 2005. "Social Mood and Financial Economics." *Journal of Behavioral Finance* 6 (3): 144–60. https://doi.org/10.1207/s15427579jpfm0603_4.

Oliveira, Nuno, Paulo Cortez, and Nelson Areal. 2017. "The Impact of Microblogging Data for Stock Market Prediction: Using Twitter to Predict Returns, Volatility, Trading Volume and Survey Sentiment Indices." *Expert Systems with Applications* 73:125–44. https://doi.org/https://doi.org/10.1016/j.eswa.2016.12.036.

O'Neill, Aaron. 2023. "Countries with the Largest Gross Domestic Product (GDP) 2022." November 9, 2023. https://www.statista.com/statistics/268173/countries-with-the-largest-gross-domestic-product-gdp/.

Preis, Tobias, Helen Susannah Moat, and H Eugene Stanley. 2013. "Quantifying Trading Behavior in Financial Markets Using Google Trends." *Scientific Reports* 3 (1): 1684. https://doi.org/10.1038/srep01684.

Qiu, Lily, and Ivo Welch. 2006. "Investor Sentiment Measures." 10794. Cambridge, MA. https://doi.org/10.3386/w10794.

Ren, Jie, Hang Dong, Balaji Padmanabhan, and Jeffrey V. Nickerson. 2021. "How Does Social Media Sentiment Impact Mass Media Sentiment? A Study of News in the Financial Markets." *Journal of the Association for Information Science and Technology* 72 (9): 1183–97. https://doi.org/10.1002/asi.24477.

Renault, Thomas. 2017. "Intraday Online Investor Sentiment and Return Patterns in the U.S. Stock Market." *Journal of Banking & Finance* 84:25–40. https://doi.org/https://doi.org/10.1016/j.jbankfin.2017.07.002.

Robert F. Stambaugh. 2024. "Robert Stambaugh's Home Page." Size and Value in China. 2024. https://finance.wharton.upenn.edu/~stambaug/.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein. 1985. "Persuasive Evidence of Market Inefficiency." *The Journal of Portfolio Management* 11 (3): 9–16. https://doi.org/10.3905/jpm.1985.409007.

Schmeling, Maik. 2009. "Investor Sentiment and Stock Returns: Some International Evidence." *Journal of Empirical Finance* 16 (3): 394–408. https://doi.org/10.1016/j.jempfin.2009.01.002.

Sha, Yezhou, Yong Zhang, and Xiaomeng Lu. 2022. "Household Investment Diversification amid Covid-19 Pandemic: Evidence from Chinese Investors." *Finance Research Letters* 47 (June):102820.

https://doi.org/10.1016/j.frl.2022.102820.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J Wilson. 2022. “Measuring News Sentiment.” *Journal of Econometrics* 228 (2): 221–43. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.07.053.

Sharif, Saqib, Hamish D. Anderson, and Ben R. Marshall. 2014. “Against the Tide: The Commencement of Short Selling and Margin Trading in Mainland China.” *Accounting and Finance* 54 (4): 1319–55. https://doi.org/10.1111/acfi.12032.

Sharpe, William F. 1964. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *The Journal of Finance* 19 (3): 425–42. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Shefrin, Hersh, and Meir Statman. 2000. “Behavioral Portfolio Theory.” *The Journal of Financial and Quantitative Analysis* 35 (2): 127–51. https://doi.org/10.2307/2676187.

Shen, Shulin, Le Xia, Yulin Shuai, and Da Gao. 2022. “Measuring News Media Sentiment Using Big Data for Chinese Stock Markets.” *Pacific-Basin Finance Journal* 74 (September):101810. https://doi.org/10.1016/J.PACFIN.2022.101810.

Shiller, Robert J. 1995. “Conversation, Information, and Herd Behavior.” *The American Economic Review* 85 (2): 181–85. http://www.jstor.org/stable/2117915.

Shleifer, Andrei. 2000. *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford University Press. https://doi.org/10.1093/0198292279.001.0001.

Siebert, Horst. 2007. “China: Coming to Grips with the New Global Player.” *World Economy* 30 (6): 893–922. https://doi.org/10.1111/j.1467-9701.2007.01037.x.

similiarweb. 2024. “Top Websites Ranking.” March 1, 2024. https://www.similarweb.com/top-websites/finance/investing/.

Simon, Herbert A. 1955. “A Behavioral Model of Rational Choice.” *The Quarterly Journal of Economics* 69 (1): 99–118. https://doi.org/10.2307/1884852.

Song, Huilin, Diyun Peng, and Xin Huang. 2020. “Incorporating Research Reports and Market Sentiment for Stock Excess Return Prediction: A Case of Mainland China.” *Scientific Programming*, May, 1–7. https://doi.org/10.1155/2020/8894757.

SSE. 2016. “Shanghai Stock Exchance Statistics Annual.” https://www.sse.com.cn/aboutus/publication/yearly/documents/c/tjnj_2016.pdf.

SSE. 2023. “Statistics Annual Yearbook 2023.” https://www.sse.com.cn/aboutus/publication/yearly/documents/c/5733825.pdf.

SSE. 2024a. “Milestone.” 2024. https://english.sse.com.cn/aboutsse/milestone/.

SSE. 2024b. “Overview - History and Development.” 2024.

http://english.sse.com.cn/aboutsse/overview/.

State Council of the People's Republic of China. 2023. "China to Lower Minimum Margin Ratio of Financing for Securities Purchases." August 28, 2023. https://english.www.gov.cn/news/202308/28/content_WS64ebdecfc6d0868f4e8dede8.html.

Stattman, D. 1980. "Book Values and Stock Returns." *The Chicago MBA: A Journal of Selected Papers* 4:25–45.

Sul, Hong Kee, Alan R. Dennis, and Lingyao (Ivy) Yuan. 2017. "Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns." *Decision Sciences* 48 (3): 454–88. https://doi.org/10.1111/deci.12229.

Sun, Licheng, Mohammad Najand, and Jiancheng Shen. 2016. "Stock Return Predictability and Investor Sentiment: A High-Frequency Perspective." *Journal of Banking & Finance* 73:147–64. https://doi.org/https://doi.org/10.1016/j.jbankfin.2016.09.010.

Sun, Yunchuan, Mengting Fang, and Xinyu Wang. 2018. "A Novel Stock Recommendation System Using Guba Sentiment Analysis." *Personal and Ubiquitous Computing* 22 (3): 575–87. https://doi.org/10.1007/s00779-018-1121-x.

Sun, Yunchuan, Xiaoping Zeng, Siyu Zhou, Han Zhao, Peter Thomas, and Haifeng Hu. 2021. "What Investors Say Is What the Market Says: Measuring China's Real Investor Sentiment." *Personal and Ubiquitous Computing* 25 (3): 587–99. https://doi.org/10.1007/s00779-021-01542-3.

Szczygielski, Jan Jakub, Ailie Charteris, Princess Rutendo Bwanya, and Janusz Brzeszczyński. 2024. "Google Search Trends and Stock Markets: Sentiment, Attention or Uncertainty?" *International Review of Financial Analysis* 91 (January):102549. https://doi.org/10.1016/J.IRFA.2023.102549.

SZSE. 2019. "Shenzhen Stock Exchange Fact Book 2019."

SZSE. 2024. "Overview." 2024. https://www.szse.cn/English/about/overview/index.html.

Tan, Zheng, Ziqin Yan, and Guangwei Zhu. 2019. "Stock Selection with Random Forest: An Exploitation of Excess Return in the Chinese Stock Market." *Heliyon* 5 (8): e02310. https://doi.org/10.1016/j.heliyon.2019.e02310.

Tang, Siyuan. 2023. "Price Limit Performance: New Evidence from a Quasi-Natural Experiment in China's ChiNext Market." *International Review of Financial Analysis* 89 (October):102747. https://doi.org/10.1016/J.IRFA.2023.102747.

The World Bank. 2022a. "Market Capitalization of Listed Domestic Companies (Current

US$).” 2022. https://data.worldbank.org/indicator/CM.MKT.LCAP.CD?most_recent_value_desc=true.

The World Bank. 2022b. “Stocks Traded, Total Value (% of GDP).” 2022. https://data.worldbank.org/indicator/CM.MKT.TRAD.GD.ZS?most_recent_value_desc=true.

Tian, Lihui. 2011. “Regulatory Underpricing: Determinants of Chinese Extreme IPO Returns.” *Journal of Empirical Finance* 18 (1): 78–90. https://doi.org/https://doi.org/10.1016/j.jempfin.2010.10.004.

Wang, Fenghua, and Yexiao Xu. 2004. “What Determines Chinese Stock Returns?” *Financial Analysts Journal* 60 (6): 65–77. https://doi.org/10.2469/faj.v60.n6.2674.

Wang, Wenzhao, Chen Su, and Darren Duxbury. 2021. “Investor Sentiment and Stock Returns: Global Evidence.” *Journal of Empirical Finance* 63 (September):365–91. https://doi.org/10.1016/J.JEMPFIN.2021.07.010.

Wang, Xinjie, Zhiqiang Xiang, Weike Xu, and Peixuan Yuan. 2022. “The Causal Relationship between Social Media Sentiment and Stock Return: Experimental Evidence from an Online Message Forum.” *Economics Letters* 216:110598. https://doi.org/https://doi.org/10.1016/j.econlet.2022.110598.

We are Social, and Meltwater. 2024a. “Digital 2024: China.” 2024. https://datareportal.com/reports/digital-2024-china.

We are Social, and Meltwater. 2024b. “Digital 2024: The United States of America.” 2024. https://datareportal.com/reports/digital-2024-united-states-of-america.

White, Halbert. 1980. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica* 48 (4): 817–38. https://doi.org/10.2307/1912934.

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach*. 6th ed. Cengage Learning.

Xie, Yuxin, Ruohua Tang, Athanasios A Pantelous, and Xiaomeng Lu. 2024. “Narrow Framing and Under-Diversification: Empirical Evidence from Chinese Households.” *China Economic Review* 83:102095. https://doi.org/https://doi.org/10.1016/j.chieco.2023.102095.

Xiong, Wei, and Jialin Yu. 2011. “The Chinese Warrants Bubble.” *American Economic Review* 101 (6): 2723–53. https://doi.org/10.1257/aer.101.6.2723.

Xu, Hai-Chuan, and Wei-Xing Zhou. 2018. “A Weekly Sentiment Index and the Cross-Section of Stock Returns.” *Finance Research Letters* 27:135–39.

https://doi.org/https://doi.org/10.1016/j.frl.2018.02.009.

Xu, Yingying, Zhixin Liu, Jichang Zhao, and Chiwei Su. 2017. “Weibo Sentiments and Stock Return: A Time-Frequency View.” *PLOS ONE* 12 (7): e0180723. https://doi.org/10.1371/journal.pone.0180723.

Xu, Yingying, and Jichang Zhao. 2022. “Can Sentiments on Macroeconomic News Explain Stock Returns? Evidence from Social Network Data.” *International Journal of Finance and Economics* 27 (2): 2073–88. https://doi.org/10.1002/ijfe.2260.

Xu, Yongan, Jianqiong Wang, Zhonglu Chen, and Chao Liang. 2023. “Sentiment Indices and Stock Returns: Evidence from China.” *International Journal of Finance & Economics* 28 (1): 1063–80. https://doi.org/10.1002/ijfe.2463.

Yang, Chunpeng, and Bin Gao. 2014. “The Term Structure of Sentiment Effect in Stock Index Futures Market.” *The North American Journal of Economics and Finance* 30:171–82. https://doi.org/https://doi.org/10.1016/j.najef.2014.09.001.

Yang, Chunpeng, and Rengui Zhang. 2014. “Does Mixed-Frequency Investor Sentiment Impact Stock Returns? Based on the Empirical Study of MIDAS Regression Model.” *Applied Economics* 46 (9): 966–72. https://doi.org/10.1080/00036846.2013.864041.

Yao, Juan, Chuanchan Ma, and William Peng He. 2014. “Investor Herding Behaviour of Chinese Stock Market.” *International Review of Economics and Finance* 29 (January):12–29. https://doi.org/10.1016/j.iref.2013.03.002.

Yao, Shujie, and Dan Luo. 2009. “The Economic Psychology of Stock Market Bubbles in China.” *The World Economy* 32 (5): 667–91. https://doi.org/https://doi.org/10.1111/j.1467-9701.2009.01176.x.

Zhang, Qiang, and Shu-e Yang. 2009. “Noise Trading, Investor Sentiment Volatility, and Stock Returns.” *Systems Engineering - Theory & Practice* 29 (3): 40–47. https://doi.org/https://doi.org/10.1016/S1874-8651(10)60010-5.

Zhang, Tao, Jian Li, and Phil Malone. 2004. “Closed-End Fund Discounts in Chinese Stock Markets.” *The Chinese Economy* 37 (3): 17–38. https://doi.org/10.1080/10971475.2004.11033495.

Zhang, Tonghui, Ying Yuan, and Xi Wu. 2020. “Is Microblogging Data Reflected in Stock Market Volatility? Evidence from Sina Weibo.” *Finance Research Letters* 32:101173. https://doi.org/https://doi.org/10.1016/j.frl.2019.04.030.

Zhang, Xi, Jiawei Shi, Di Wang, and Binxing Fang. 2018. “Exploiting Investors Social Network for Stock Prediction in China’s Market.” *Journal of Computational Science* 28:294–303. https://doi.org/https://doi.org/10.1016/j.jocs.2017.10.013.

Zhang, Yihao, and Lingfeng Tao. 2019. "Haze, Investor Attention and China's Stock Markets: Evidence from Internet Stock Forum." *Finance Research Letters* 31. https://doi.org/https://doi.org/10.1016/j.frl.2018.12.001.

Zhou, Guofu. 2018. "Measuring Investor Sentiment." *Annual Review of Financial Economics* 10 (1): 239–59. https://doi.org/10.1146/annurev-financial-110217-022725.

Appendix

Appendix 1: Raw data sourced across all databases for monthly values

Raw data sourced across all databases for monthly data.

Field Name	Field Content	Description
Stkcd [1, 2, 3, 4, 5, 6]	Stock Code	6-digit stock code
Trdmnt	Trading Month	YYYY-MM
Msmvttl	Monthly Total Market Capitalization	Number of shares issued multiplied by monthly closing price, in CNY 1,000
Ndaytrd	Monthly Trading Days	Actual trading days in the month
Mretwd[1]	Monthly Return with Cash Dividend Reinvested	$r_{n,t} = \frac{P_{n,t}}{P_{n,t-1}} - 1$, $P_{n,t}$ ($P_{n,t-1}$) = closing price of stock n (with cash dividend reinvested) on the last trading day of month t (t-1)
Markettype[1]	Stock Market	1 =SSE A-share market (excl. STAR), 2 = SSE B-share market, 4 = SZSE A-share market (excl. ChiNext), 8 = SZSE B-share market, 16 = ChiNext, 32 = STAR market, 64 = BSE A-share market
Listdt[2]	Listing Date	YYYY-MM-DD
Nrrmtdt[3]	Risk-free Interest Rate	In %, converted from annualized risk-free interest rate based on compound interest
Accper[4, 5, 6]	Ending Date of Fiscal Year	YYYY-MM-DD
F020102[4]	Net Profit Attributable to Shareholders after Deducting Non-recurring Gains & Losses	Available starting from 2007. If missing, net profit (excluding minority interests) is used
B002000000[5]	Net Profit	The net profit realized by an enterprise.
B002000201[5]	Minority Interest	The profit or loss attributable to minority stockholders based on their equity ratio
Typrep[5]	Statement Type	A=Consolidated Statements B=Parent Statements
Reptyp[6]	Report Type	1 = Q1 Report, 2 = Interim Report (cumulated), 3 = Q3 Report (cumulated), 4 = Annual Report (cumulated)
Annodt[6]	Announcement Date	YYYY-MM-DD
股票代码 [7]	Stock Code	Equivalent to Stkcd
统计日期 [7]	Ending Date of Fiscal Year	Equivalent to Accper
实际披露时间 [7]	Date of Disclosure	Equivalent to Annodt

[1]CSMAR – China Stock Market Trading Database: Monthly Stock Price & Returns, [2]CSMAR – China Stock Market Trading Database: Company Profile, [3]CSMAR – China Stock Market Trading Database: Risk-free Interest Rate, [4]CSMAR – China Listed Firms Research Series: Financial Indicators Database, [5]CSMAR – China Listed Firms Research Series: Financial Statements Database, [6]CSMAR – China Stock Market Financial Database – Statements Release Dates, [7]CNRDS – Financial Report Disclosure Time Database

Appendix 2: Filters and number of observations

The number of unique stock codes and observations per filtering or merging step.

	Individual Stock Codes		Total Observations (N)	
	Monthly	Daily	Monthly	Daily
CH-3 Filters				
Total Input	5,540	5,690	695,410	13,655,252
1) Market	4,747	4,761	673,793	12,679,900
2) IPO	4,727	4,742	655,649	12,432,360
3) Trading Status	4,547	4,742	596,536	12,170,070
4) Bottom 30%	4,135	4,478	417,504	8,517,571
CH-G Filters				
Total Input	5,540	5,690	695,410	13,655,252
1) Market	4,747	4,761	673,793	12,679,900
2) IPO	4,727	4,742	655,649	12,432,360
3) Trading Status	4,547	4,742	596,536	12,170,070
4) Guba Posts	4,413	4,605	484,062	9,379,387
5) Before 2008	4,412	4,604	471,338	9,218,156
6) Bottom 30%	3,996	4,278	329,891	6,451,807

This table describes the data filtering procedure. The first column lists the filters. The total input rows equal the merged firm characteristics (market capitalization, earnings-to-price ratio, IPO data, trading days), returns and risk-free data, expressed in the number of unique stock codes and the total number of observations between January 2000 and April 2024. CH-3 Filters follow Liu et al.(2019), whereas the trading status filter for monthly data excludes stocks with less than 20 trading days per month or 150 trading days per year as described in Chapter 4.2.5. For daily data, all stocks with a trading status other than *normal* were excluded. Until here, the data set mimics the CH-3 dataset and is referred to as such. The subsequent row represents the merger with Guba data, which excludes stocks without posts published in a specific month or day and reduces the sampling period to January 2008 to December 2023. Due to the low number of posts on Guba in this year, the data of 2008 is excluded. Lastly, the lowest 30% of stocks by market cap are excluded.

Appendix 3: Raw data fields for the calculation of the GUBA factor

Raw data fields for the calculation of the GUBA factor.

Field Name	Field Content	Description
股票代码	Stock code	Equivalent to Stkcd
统计日期	Date of posting	YYYY-MM-DD
正面帖子量	Number of positive posts	Count of positive posts per stock per day
负面帖子量	Number of negative posts	Count of negative posts per stock per day
中性帖子量	Number of neutral posts	Count of neutral posts per stock per day

All data fields have been downloaded from CNRDS – GUBA Stock Forum Comments Database.

Appendix 4: Number of posts per year by sentiment

Posts per year by sentiment.

	Total	Positive	Negative
2008	3,975,763	1,234,750	1,010,522
2009	11,406,447	3,796,616	3,037,682
2010	9,562,543	3,234,133	2,563,268
2011	10,258,579	3,311,226	2,846,660
2012	8,721,868	2,830,930	2,402,708
2013	13,141,560	4,593,699	3,472,932
2014	15,131,680	4,973,263	3,928,920
2015	28,933,576	8,159,688	7,370,248
2016	33,531,612	9,191,945	5,871,375
2017	23,960,981	7,307,023	5,578,091
2018	22,745,840	6,282,994	5,236,405
2019	32,359,089	8,279,878	7,011,536
2020	47,794,521	11,631,082	10,580,259
2021	51,471,122	12,032,655	11,189,468
2022	47,680,408	11,349,830	9,609,981
2023	37,626,645	9,557,572	7,230,833
2008 – 2023	398,302,234	107,767,284	88,940,888
2009 – 2023	394,326,471	106,532,534	87,930,366

This table present the total number of posts per sentiment and total across years. This only includes stocks which are included into the construction of the CH-3* and CH-G model.

Appendix 5: Overview on reported and replicated average R-squares for individual stocks in China

Average R-squares for individual stocks in China.

Model	CH- 3, as reported in paper (Liu, Stambaugh, and Yuan 2019)	CH- 3*, factors from website (Stambaugh 2024)	CH-3* replicated, factors based on self-sourced data, constructed by following Liu et al. (2019)	
Period	Jan 2000 – Dec 2016	Jan 2000 – Dec 2016	Jan 2000 – Dec 2016	Jan 2000 – April 2024
Panel A: All individual stocks in China				
MKT	0.385	0.396	0.400	0.283
MKT, SMB	0.507	0.547	0.537	0.404
MKT, VMG	0.471	0.510	0.513	0.361
MKT, SMB, VMG	0.536	0564	0.564	0.442
Panel B: All but the smallest 30% of stocks in China				
MKT	0.417	0.409	0.414	0.302
MKT, SMB	0.528	0.524	0.525	0.402
MKT, VMG	0.501	0.505	0.509	0.379
MKT, SMB, VMG	0.562	0.553	0.553	0.443
Panel C: Only the smallest 30% of stocks in China				
MKT	-	0.400	0.406	0.283
MKT, SMB	-	0.576	0.579	0.454
MKT, VMG	-	0.524	0.528	0.370
MKT, SMB, VMG	-	0.599	0.601	0.481

This table compares the average R-squares on regressions of monthly individual stocks' returns on factors in China's stock market reported by Liu et al. (2019) with the average R-square calculated using the factors published (Liu, Stambaugh, and Yuan 2019; Stambaugh 2024) and the average R-square calculated based on own data sourcing, using the methodology described by Liu et al. (2019). Regressions are estimated on four models: the first includes excess market returns (MKT) only, before successively adding size (SMB) and value (VMG), forming all combinations possible. All regressions are run over a rolling three-year (36 months) window. The mean of the R-squares is first calculated over time for each individual stock and then averaged across stocks. The factor model is used to explain three distinct Panels of returns: those of all individual stocks (Panel A), those of all but the smallest 30 % by market capitalization (Panel B) and those of only the smallest 30% by market capitalization (Panel C).

Appendix 6: Fama-MacBeth regressions of stock returns on beta, size, and value over time

Fama-MacBeth regressions of stock returns on beta, size, and value over time.

Model	CH- 3, as reported in paper (Liu et al., 2019)	CH-3* replicated, factors based on self-sourced data, constructed by following Liu et al. (2019)		
Period	Jan 2000 – Dec 2016	Jan 2000 – Dec 2016	Jan 2000 – April 2024	Jan 2009 – Dec 2023
Intercept	0.0690	0.1118	0.0797	0.1002
	(4.03)	(3.02)	(2.77)	(3.07)
β	0.0002	0.0028	0.0025	0.0018
	(0.07)	(0.93)	(1.00)	(0.56)
logME	-0.0068	-0.0046	-0.0033	-0.0041
	(-4.34)	(-3.10)	(-2.86)	(-3.14)
EP+	0.9503	0.7265	0.6211	0.4468
	(4.88)	(4.95)	(5.52)	(4.19)
D(EP<0)	0.0006	-0.0035	-0.0042	-0.0042
	(0.31)	(-2.09)	(-3.047)	(-2.48)
Adjusted R^2	0.0615	0.0613	0.0581	0.0514

The table presents the average slope coefficients obtained from month-by-month Fama–MacBeth regressions, which cross-sectionally regress next month’s individual stock returns on this month’s stock characteristics. Each column represents a different regression model, with nonempty rows indicating the variables included in the analysis. The explanatory variables comprise of the monthlyized daily beta value specific to a sub-market, which quantifies the stock's sensitivity to movements within its particular exchange; the logarithm of month-end market capitalization (logME); EP+, representing positive earnings-to-price values, with zero assigned otherwise; and D(EP < 0), which takes the value of one if earnings are negative, and zero otherwise. The final row displays the average adjusted R-squared for each regression specification. T-statistics, calculated using Newey-West (1987) standard errors with 4 lags, are shown in parentheses.

Appendix 7: Fama-MacBeth regressions on all stocks (Panel A)

Fama-MacBeth regressions of stock returns on beta, size, and value and sentiment.

Model	CH-3*	CH-G				
Intercept	0.1718	0.1714	0.1717	0.1699	0.1736	0.1652
	(4.92)	(4.95)	(4.95)	(4.89)	(5.04)	(4.82)
β	0.0009	0.0016	0.0016	0.0017	0.0015	0.0019
	(0.33)	(0.56)	(0.57)	(0.61)	(0.51)	(0.70)
logME	-0.0072	-0.0073	-0.0073	-0.0023	-0.0074	-0.0070
	(-5.02)	(-5.15)	(-5.15)	(-5.06)	(-5.24)	(-5.06)
EP+	0.5193	0.5062	0.5073	0.5219	0.5061	0.5069
	(4.86)	(4.76)	(4.77)	(4.82)	(4.76)	(4.79)
D(EP<0)	-0.0025	-0.0022	-0.0022	-0.0022	-0.0023	-0.0020
	(-1.57)	(-1.43)	(-1.43)	(-1.43)	(-1.47)	(-1.33)
Guba_1		0.0114				0.0465
		(4.95)				(1.24)
Guba_2			0.0052			-0.0218
			(4.84)			(-1.35)
Guba_3				0.0456		0.0388
				(4.72)		(2.89)
Guba_4					0.0018	-0.0003
					(4.55)	(0.09)
Adjusted R^2	0.0488	0.0516	0.0516	0.0501	0.0517	0.0547

The table presents the average slope coefficients obtained from month-by-month Fama–MacBeth regressions, which cross-sectionally regress next month's individual stock returns on this month's stock characteristics. Each column represents a different regression model, with nonempty rows indicating the variables included in the analysis. The explanatory variables comprise of the monthlyized daily beta value specific to a sub-market, which quantifies the stock's sensitivity to movements within its particular exchange; the logarithm of month-end market capitalization (logME); EP+, representing positive earnings-to-price values, with zero assigned otherwise; and D(EP < 0), which takes the value of one if earnings are negative, and zero otherwise. The subsequent rows add one Guba factor at a time, with the last column displaying all factors at once. The final row displays the average adjusted R-squared for each regression specification. T-statistics, calculated using Newey-West (1987) standard errors with 4 lags, are shown in parentheses. The sample size is all individual stocks and the sample period is January 2009 – December 2023.

Appendix 8: Fama-MacBeth regressions on microcap stocks only (Panel C)

Fama-MacBeth regressions of stock returns on beta, size, and value and sentiment.

Model	CH-3*	CH-G				
Intercept	0.4808	0.4719	0.4723	0.4692	0.4752	0.4692
	(6.22)	(6.15)	(6.15)	(6.08)	(6.20)	(6.15)
β	0.0042	0.0044	0.0044	0.0044	0.0043	0.0046
	(1.43)	(1.50)	(1.51)	(1.52)	(1.50)	(1.60)
logME	-0.0215	-0.0212	-0.0212	-0.0210	-0.0213	-0.0211
	(-6.23)	(-6.19)	(-6.19)	(-6.12)	(-6.24)	(-6.21)
EP+	0.6733	0.6624	0.6637	0.6715	0.6615	0.6598
	(4.80)	(4.74)	(4.76)	(4.83)	(4.73)	(4.78)
D(EP<0)	-0.0027	-0.0024	-0.0024	-0.0024	-0.0025	-0.0023
	(-1.75)	(-1.61)	(-1.62)	(-1.59)	(-1.64)	(-1.55)
Guba_1		0.0102				0.0427
		(3.93)				(0.63)
Guba_2			0.0046			-0.0320
			(3.81)			(-0.95)
Guba_3				0.0465		0.0388
				(4.05)		(1.49)
Guba_4					0.0017	-0.0038
					(3.68)	(0.99)
Adjusted R^2	0.0233	0.0258	0.0258	0.0249	0.0259	0.0280

The table presents the average slope coefficients obtained from month-by-month Fama–MacBeth regressions, which cross-sectionally regress next month's individual stock returns on this month's stock characteristics. Each column represents a different regression model, with nonempty rows indicating the variables included in the analysis. The explanatory variables comprise of the monthlyized daily beta value specific to a sub-market, which quantifies the stock's sensitivity to movements within its particular exchange; the logarithm of month-end market capitalization (logME); EP+, representing positive earnings-to-price values, with zero assigned otherwise; and D(EP < 0), which takes the value of one if earnings are negative, and zero otherwise. The subsequent rows add one Guba factor at a time, with the last column displaying all factors at once. The final row displays the average adjusted R-squared for each regression specification. T-statistics, calculated using Newey-West (1987) standard errors with 4 lags, are shown in parentheses. The sample size is microcap stocks only and the sample period is January 2009 – December 2023.

Appendix 9: Descriptive statistics of the firm characteristics (size, value, sentiment) and excess return for different panels

Descriptive statistics of the firm characteristics (size, value, sentiment) and excess return.

	Size	Value (EP)	Sentiment (Guba_1)	Excess Return
Panel A: All individual stocks in China				
Mean	15,985,380,000	0.0047	0.1342	0.0100
St. Dev.	61,461,100,000	0.0139	0.2284	0.1350
Min	95,562,110	-0.0777	-1.0000	-0.8918
25%	3,091,189,000	0.0003	-0.0133	-0.0672
50%	5,446,775,000	0.0042	0.1168	-0.0015
75%	11,526,060,000	0.0097	0.2644	0.0718
Max	2,786,247,000,000	0.0585	1.0000	2.9860
Panel B: All individual stocks but the smallest 30%				
Mean	21,722,120,000	0.0069	0.1402	0.0066
St. Dev.	72,711,520,000	0.0136	0.2240	0.1325
Min	1,277,522,000	-0.0777	-1.0000	-0.8690
25%	4,973,904,000	0.0013	-0.0045	-0.0695
50%	8,092,141,000	0.0055	0.1236	-0.0047
75%	16,401,370,000	0.0115	0.2690	0.0682
Max	2,786,247,000,000	0.0585	1.0000	2.9224
Panel C: Only the smallest 30%				
Mean	2,605,803,000	-0.0002	0.1202	0.0179
St. Dev.	1,082,445,000	0.0135	0.2377	0.1405
Min	95,562,110	-0.0777	-1.0000	-0.8918
25%	1,842,370,000	-0.0024	-0.0329	-0.0617
50%	2,425,165,000	0.0016	0.1004	0.0065
75%	3,133,445,000	0.0057	0.2525	0.0797
Max	8,084,738,000	0.0585	1.0000	2.9860

This table provides the descriptive statistics for firm size by market capitalization, earnings-to-price ratio (EP), sentiment, and stock returns across three panels. Panel A includes all individual stocks in China, Panel B excludes the smallest 30% of stocks, and Panel C focuses solely on the smallest 30%. The EP ratio is winsorized at 5% to reduce the impact of extreme values. The sample period is January 2009 – December 2023.

Appendix 10: Summary statistics for CH-G factors formed on Panel A and C

Summary statistics of factors formed on panels varying in stock selection.

	MKT (CH-G)	SMB (CH-G)	VMG (CH-G)	GUBA (CH-G)
Panel A: Factors formed all individual stocks in China				
Mean	0.51	1.12	0.78	0.53
Standard deviation	6.50	4.46	3.16	2.23
t-statistics	1.05	3.43	3.84	3.32
Sharpe Ratio	0.08	0.25	0.25	0.24
Skewness	-0.17	-0.28	0.06	-0.07
Excess Kurtosis	1.72	2.89	2.16	3.81
Panel C: Factors formed on only the smallest 30% of stocks in China				
Mean	1.83	0.81	0.82	0.57
Standard deviation	8.91	1.43	2.41	1.82
t-statistics	2.84	7.14	5.48	4.43
Sharpe Ratio	0.21	0.56	0.34	0.31
Skewness	0.04	0.48	0.16	0.30
Excess Kurtosis	1.4	0.86	0.62	1.60

This table reports summary statistics for the CH-G factors market (MKT), size (SMB) and value (VMG) and CH-G sentiment (GUBA), formed on different sample sizes, varying in the stocks included. The monthly mean returns, standard deviation, Newey-West (1987) adjusted t-statistics calculated with 4 lags as well as the monthly Sharpe Ratio, Skewness and Excess Kurtosis for each factor is reported. The sample size is all stocks (Panel A) and the smallest 30 % by market capitalization (Panel C). Panel B is reported in Chapter 5.3.2. Sample period is January 2009 – December 2024.

Appendix 11: Correlation matrix for CH-3* and CH-G factors formed on Panel A and C

Correlation matrix of factors formed on panels varying in stock selection.

	CH-3*			CH-G			
	MKT	SMB	VMG	MKT	SMB	VMG	GUBA
Panel A: Factors formed all individual stocks in China							
MKT	1.00	0.13	-0.37	1.00	0.17	-0.37	-0.20
SMB		1.00	-0.55		1.00	-0.63	-0.42
VMG			1.00			1.00	0.50
GUBA							1.00
Panel C: Factors formed on only the smallest 30% of stocks in China							
MKT	1.00	0.20	-0.18	1.00	0.21	-0.18	-0.08
SMB		1.00	-0.08		1.00	-0.17	-0.04
VMG			1.00			1.00	0.49
GUBA							1.00

This table reports the correlations among factors formed on all individual stocks (Panel A) and on microcap stocks (Panel C). Panel B is reported in Chapter 5.3.2 The sample period is January 2009 – December 2023.

Appendix 12: Average adjusted R-squares for factors formed on Panel A and C

Average adjusted R-squares for individual stocks in China.

	Panel A (Returns)	Panel B (Returns)	Panel C (Returns)
Panel A: Factors formed all individual stocks in China			
MKT	0.256	0.270	0.252
MKT, SMB	0.349	0.339	0.400
MKT, VMG	0.305	0.318	0.312
MKT, GUBA	0.274	0.287	0.278
MKT, SMB, VMG	0.364	0.358	0.407
MKT, SMB, GUBA	0.354	0.345	0.405
MKT, VMG, GUBA	0.312	0.326	0.320
MKT, SMB, VMG, GUBA	0.372	0.366	0.414
Panel C: Factors formed on only the smallest 30% of stocks in China			
MKT	0.334	0.294	0.397
MKT, SMB	0.365	0.309	0.402
MKT, VMG	0.362	0.301	0.409
MKT, GUBA	0.362	0.302	0.407
MKT, SMB, VMG	0.392	0.318	0.414
MKT, SMB, GUBA	0.392	0.317	0.411
MKT, VMG, GUBA	0.386	0.306	0.414
MKT, SMB, VMG, GUBA	0.415	0.321	0.418

The table presents average adjusted R-squares on regressions of monthly individual stocks' returns on factors in China's stock market. The factors are formed on panels varying in stock selection, with the results for factors formed on Panel B being reported in Chapter 5.3.2. Regressions are estimated on eight models: the first includes excess market returns (MKT) only, before successively adding size (SMB), value (VMG) and sentiment (GUBA), forming all combinations possible. All regressions are run over a rolling three-year (36 months) window. The mean of the adjusted R-squares is first calculated over time for each individual stock and then averaged across stocks. The factor model is used to explain three distinct Panels of returns: those of all individual stocks (Panel A), those of all but the smallest 30 % by market capitalization (Panel B) and those of only the smallest 30% by market capitalization (Panel C). The sample period is January 2009 – December 2023.

Appendix 13: CH-G portfolio performance analysis across 5x5 size-sentiment-sorted portfolios for factors formed on Panel B

CH-G factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Growth	EP 2	EP 3	EP 4	Value
Alpha					
Small	0.73***	0.79***	0.85***	0.90***	1.26***
Size 2	0.39***	0.24**	0.39***	0.48***	0.80***
Size 3	0.11	0.02	0.05	0.19**	0.39***
Size 4	-0.15	0.10	-0.01	-0.02	0.54***
Big	-0.21	-0.04	0.03	0.18	0.25**
β_{MKT}					
Small	0.98***	0.98***	0.99***	0.95***	0.94***
Size 2	0.96***	1.01***	1.00***	1.00***	0.95***
Size 3	1.01***	1.01***	1.01***	0.99***	0.97***
Size 4	0.98***	1.05***	1.04***	1.06***	0.95***
Big	1.07***	1.06***	1.06***	1.11***	0.97***
β_{SMB}					
Small	1.15***	1.11***	1.20***	1.12***	1.30***
Size 2	1.03***	0.98***	1.04***	0.99***	1.04***
Size 3	0.91***	0.87***	0.94***	0.86***	0.90***
Size 4	0.76***	0.64***	0.65***	0.68***	0.65***
Big	0.27***	0.15***	0.16***	0.05*	0.13***
β_{VMG}					
Small	-0.01	-0.11**	-0.10**	-0.15***	-0.23***
Size 2	-0.10**	-0.10***	-0.13***	-0.14***	-0.31***
Size 3	-0.15***	-0.07*	-0.05	-0.13***	-0.17***
Size 4	-0.11**	-0.14***	-0.09*	-0.11***	-0.21***
Big	-0.04	-0.08	0.10*	0.01	-0.02
β_{SENTI}					
Small	-0.18**	-0.18**	-0.07	-0.04	0.05
Size 2	-0.26***	-0.10*	-0.09*	-0.02	0.15**
Size 3	-0.20***	-0.15***	-0.13***	-0.07*	-0.01
Size 4	-0.25***	-0.25***	-0.07	-0.03	-0.00
Big	-0.27***	-0.14**	-0.25***	-0.00	-0.07

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-value sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB), value (β_VMG), and sentiment (β_SENTI). The portfolios are categorized into five sentiment levels, ranging from Pessimistic to Optimistic, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all but the smallest 30 % by market capitalization (Panel B) and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 14: CH-3 portfolio performance analysis across 5x5 size-sentiment-sorted portfolios for factors formed on Panel B

CH-3 factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Pessimistic	Sentiment 2	Sentiment 3	Sentiment 4	Optimistic
Alpha					
Small	0.71***	0.76***	0.88***	0.95***	1.36***
Size 2	0.31***	0.25**	0.40***	0.52***	0.95***
Size 3	0.06	-0.02	0.03	0.20**	0.44***
Size 4	-0.24*	0.00	-0.00	0.01	0.59***
Big	-0.34**	-0.11	-0.09	0.18	0.22*
β_{MKT}					
Small	0.99***	0.99***	0.99***	0.95***	0.94***
Size 2	0.97***	1.01***	1.00***	1.00***	0.95***
Size 3	1.01***	1.02***	1.01***	0.99***	0.97***
Size 4	0.99***	1.05***	1.04***	1.06***	0.96***
Big	1.08***	1.06***	1.06***	1.11***	0.97***
β_{SMB}					
Small	1.16***	1.13***	1.19***	1.11***	1.27***
Size 2	1.06***	0.98***	1.04***	0.98***	0.99***
Size 3	0.93***	0.89***	0.95***	0.85***	0.89***
Size 4	0.79***	0.67***	0.65***	0.67***	0.64***
Big	0.31***	0.17***	0.19***	0.05*	0.14***
β_{VMG}					
Small	-0.22***	-0.31***	-0.30***	-0.33***	-0.43***
Size 2	-0.30***	-0.27***	-0.31***	-0.30***	-0.45***
Size 3	-0.32***	-0.24***	-0.22***	-0.28***	-0.31***
Size 4	-0.27***	-0.29***	-0.20***	-0.23***	-0.32***
Big	-0.12**	-0.13***	0.03	-0.00	-0.05

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-sentiment sorted portfolios. The analysis considers market excess return (β_MKT), size (β_SMB) and value (β_VMG). The portfolios are categorized into five sentiment levels, ranging from Pessimistic to Optimistic, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all but the smallest 30 % by market capitalization (Panel B) and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 15: CH-G portfolio performance analysis across 5x5 size-value-sorted portfolios for factors formed on Panel A

CH-G factor-based portfolio performance analysis across 5x5 size-value-sorted portfolios.

	Growth	EP 2	EP 3	EP 4	Value
Alpha					
Small	0.55***	0.50***	0.26*	0.09	0.14
Size 2	0.13	-0.02	0.04	0.03	-0.01
Size 3	-0.32**	-0.22*	-0.31**	0.09	-0.13
Size 4	-0.14	-0.26	0.02	-0.09	-0.12
Big	0.54	0.20	0.27	-0.06	-0.21
β_{MKT}					
Small	0.95***	0.99***	0.99***	0.99***	0.92***
Size 2	1.01***	0.99***	0.98***	0.98***	0.96***
Size 3	1.05***	1.02***	1.01***	0.98***	0.96***
Size 4	1.08***	1.05***	1.01***	0.97***	1.02***
Big	1.13***	1.02***	0.97***	1.01***	1.09***
β_{SMB}					
Small	1.03***	1.06***	1.16***	1.15***	1.00***
Size 2	0.80***	0.90***	0.97***	0.94***	0.83***
Size 3	0.73***	0.81***	0.82***	0.73***	0.66***
Size 4	0.42***	0.61***	0.54***	0.56***	0.52***
Big	-0.15	0.07*	0.11*	0.14***	0.05
β_{VMG}					
Small	-0.30***	-0.33***	-0.16***	-0.02	0.04
Size 2	-0.31***	-0.36***	-0.36***	-0.10**	0.22***
Size 3	-0.31***	-0.44***	-0.22***	-0.19***	0.28***
Size 4	-0.38***	-0.46***	-0.47***	-0.23***	0.32***
Big	-0.51***	-0.61***	-0.50***	-0.14**	0.50***
β_{SENTI}					
Small	-0.30***	-0.33***	-0.16***	-0.02	0.04
Size 2	-0.31***	-0.36***	-0.36***	-0.10**	0.22***
Size 3	-0.31***	-0.44***	-0.22***	-0.19***	0.28***
Size 4	-0.38***	-0.46***	-0.47***	-0.23***	0.32***
Big	-0.51***	-0.61***	-0.50***	-0.14**	0.50***

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-value sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB), value (β_VMG), and sentiment (β_SENTI). The portfolios are categorized into five value levels, ranging from Growth to Value, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all stocks and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 16: CH-G portfolio performance analysis across 5x5 size-value-sorted portfolios for factors formed on Panel C

CH-G factor-based portfolio performance analysis across 5x5 size-value-sorted portfolios.

	Growth	EP 2	EP 3	EP 4	Value
Alpha					
Small	0.06	-0.22*	-0.49***	-0.57***	-0.32
Size 2	0.18	-0.26**	-0.43***	-0.14	0.20
Size 3	-0.14	-0.50***	-0.42***	0.06	0.23
Size 4	0.10	-0.43**	-0.25	-0.17	0.33
Big	0.53	0.02	-0.15	-0.04	0.72
β_{MKT}					
Small	0.99***	1.04***	1.05***	1.02***	0.91***
Size 2	0.98***	1.02***	1.04***	0.98***	0.86***
Size 3	0.99***	1.03***	0.98***	0.92***	0.81***
Size 4	0.95***	1.01***	0.95***	0.89***	0.77***
Big	0.81***	0.82***	0.80***	0.75***	0.61***
β_{SMB}					
Small	0.45***	0.27***	0.22***	0.07	0.18
Size 2	-0.56***	-0.46***	-0.41***	-0.52***	-0.46***
Size 3	-0.80***	-0.68***	-0.74***	-0.77***	-0.70***
Size 4	-1.15***	-1.03***	-1.00***	-0.90***	-0.74***
Big	-1.29***	-1.21***	-1.07***	-1.04***	-1.04***
β_{VMG}					
Small	0.45***	0.27***	0.22***	0.07	0.18
Size 2	-0.56***	-0.46***	-0.41***	-0.52***	-0.46***
Size 3	-0.80***	-0.68***	-0.74***	-0.77***	-0.70***
Size 4	-1.15***	-1.03***	-1.00***	-0.90***	-0.74***
Big	-1.29***	-1.21***	-1.07***	-1.04***	-1.04***
β_{SENTI}					
Small	0.05	0.11	0.16**	0.20**	-0.15
Size 2	-0.12	-0.03	0.14	0.05	-0.26**
Size 3	-0.24***	-0.01	0.00	-0.04	-0.38***
Size 4	-0.52***	-0.20*	-0.10	-0.19*	-0.49***
Big	-0.76***	-0.35**	-0.20	-0.44***	-0.77***

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-value sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB), value (β_VMG), and sentiment (β_SENTI). The portfolios are categorized into five value levels, ranging from Growth to Value, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on the smallest 30 % of stocks by market capitalization and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 17: CH-3 portfolio performance analysis across 5x5 size-value-sorted portfolios for factors formed on Panel A

CH-3 factor-based portfolio performance analysis across 5x5 size-value-sorted portfolios.

	Growth	EP 2	EP 3	EP 4	Value
Alpha					
Small	0.53***	0.58***	0.36**	0.16	0.15
Size 2	0.11	0.03	0.11	0.08	-0.06
Size 3	-0.35**	-0.22*	-0.26**	0.11	-0.25
Size 4	-0.27	-0.30	0.03	-0.10	-0.18
Big	0.38	0.20	0.23	-0.12	-0.27*
β_{MKT}					
Small	0.96***	0.99***	0.99***	1.00***	0.93***
Size 2	1.02***	1.00***	0.98***	0.99***	0.96***
Size 3	1.05***	1.03***	1.01***	0.98***	0.96***
Size 4	1.09***	1.05***	1.01***	0.98***	1.02***
Big	1.13***	1.01***	0.97***	1.02***	1.09***
β_{SMB}					
Small	1.04***	1.05***	1.14***	1.13***	1.00***
Size 2	0.81***	0.89***	0.95***	0.93***	0.85***
Size 3	0.74***	0.81***	0.81***	0.72***	0.69***
Size 4	0.45***	0.62***	0.53***	0.56***	0.53***
Big	-0.12	0.06	0.11*	0.15***	0.06
β_{VMG}					
Small	-0.52***	-0.51***	-0.36***	-0.23***	-0.16**
Size 2	-0.49***	-0.53***	-0.53***	-0.28***	0.01
Size 3	-0.48***	-0.61***	-0.38***	-0.34***	0.06
Size 4	-0.55***	-0.62***	-0.59***	-0.37***	0.17***
Big	-0.59***	-0.64***	-0.56***	-0.21***	0.45***

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-value sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB) and value (β_VMG). The portfolios are categorized into five value levels, ranging from Growth to Value, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all stocks and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 18: CH-3 portfolio performance analysis across 5x5 size-value-sorted portfolios for factors formed on Panel C

CH-3 factor-based portfolio performance analysis across 5x5 size-value-sorted portfolios.

	Growth	EP 2	EP 3	EP 4	Value
Alpha					
Small	0.07	-0.19	-0.46***	-0.55***	-0.34
Size 2	0.17	-0.27**	-0.42***	-0.14	0.17
Size 3	-0.17	-0.51***	-0.42***	0.05	0.17
Size 4	0.03	-0.47**	-0.27	-0.20	0.27
Big	0.45	-0.02	-0.19	-0.10	0.63
βMKT					
Small	0.99***	1.04***	1.05***	1.02***	0.91***
Size 2	0.98***	1.02***	1.04***	0.98***	0.86***
Size 3	0.99***	1.03***	0.98***	0.93***	0.81***
Size 4	0.95***	1.01***	0.95***	0.89***	0.77***
Big	0.82***	0.82***	0.80***	0.75***	0.61***
β_{SMB}					
Small	0.46***	0.27***	0.24***	0.11	0.17
Size 2	-0.58***	-0.46***	-0.39***	-0.51***	-0.50***
Size 3	-0.83***	-0.68***	-0.74***	-0.77***	-0.74***
Size 4	-1.22***	-1.05***	-1.01***	-0.92***	-0.80***
Big	-1.41***	-1.27***	-1.10***	-1.10***	-1.15***
β_{VMG}					
Small	-0.45***	0.04	0.39***	0.55***	0.28***
Size 2	-0.28***	0.03	0.27***	0.34***	0.17***
Size 3	-0.25***	0.03	0.25***	0.17***	0.14
Size 4	-0.37***	0.06	0.16**	0.20***	0.14
Big	-0.46**	-0.12	0.12	0.09	-0.07

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-value sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB) and value (β_VMG). The portfolios are categorized into five value levels, ranging from Growth to Value, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on the smallest 30 % of stocks by market capitalization and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 19: CH-G portfolio performance analysis across 5x5 size-sentiment-sorted portfolios for factors formed on Panel A

CH-G factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Pessimistic	Sentiment 2	Sentiment 3	Sentiment 4	Optimistic
Alpha					
Small	0.25**	0.34***	0.33***	0.36***	0.67***
Size 2	-0.02	-0.14*	-0.00	0.04	0.36***
Size 3	-0.22*	-0.27**	-0.27**	-0.12	0.08
Size 4	-0.39**	-0.07	-0.17	-0.25*	0.29*
Big	-0.26*	-0.06	0.06	0.18	0.20
β_{MKT}					
Small	0.98***	0.98***	0.99***	0.96***	0.95***
Size 2	0.96***	1.01***	1.00***	1.01***	0.96***
Size 3	1.01***	1.01***	1.01***	1.00***	0.98***
Size 4	0.99***	1.05***	1.04***	1.06***	0.96***
Big	1.07***	1.06***	1.05***	1.10***	0.97***
β_{SMB}					
Small	1.05***	1.00***	1.10***	1.04***	1.20***
Size 2	0.91***	0.86***	0.90***	0.88***	0.91***
Size 3	0.77***	0.73***	0.79***	0.72***	0.75***
Size 4	0.61***	0.49***	0.50***	0.54***	0.53***
Big	0.16***	0.06	0.05	-0.03	0.06*
β_{VMG}					
Small	-0.14***	-0.25***	-0.22***	-0.22***	-0.31***
Size 2	-0.20***	-0.18***	-0.23***	-0.19***	-0.38***
Size 3	-0.24***	-0.17***	-0.12**	-0.19***	-0.23***
Size 4	-0.17**	-0.24***	-0.17***	-0.15***	-0.21***
Big	-0.08	-0.15**	0.05	-0.01	0.01
β_{SENTI}					
Small	-0.12**	-0.13**	-0.00	0.03	0.06
Size 2	-0.24***	-0.09*	-0.08**	-0.00	0.14*
Size 3	-0.19***	-0.15***	-0.17**	-0.10*	-0.06
Size 4	-0.30***	-0.27***	-0.09	-0.08	-0.08
Big	-0.29***	-0.12*	-0.28***	0.00	-0.12*

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-sentiment sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB), value (β_VMG), and sentiment (β_SENTI). The portfolios are categorized into five sentiment levels, ranging from Pessimistic to Optimistic, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all stocks and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 20: CH-G portfolio performance analysis across 5x5 size-sentiment-sorted portfolios for factors formed on Panel C

CH-G factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Pessimistic	Sentiment 2	Sentiment 3	Sentiment 4	Optimistic
Alpha					
Small	-0.12	-0.09	-0.23**	-0.19*	-0.46**
Size 2	-0.12	-0.04	-0.22*	0.02	-0.25*
Size 3	-0.29**	-0.16	-0.17	-0.13	-0.06
Size 4	-0.31	0.02	0.04	-0.18	0.14
Big	0.05	0.18	0.49	0.59	0.36
β_{MKT}					
Small	0.99***	1.00***	1.03***	0.99***	1.06***
Size 2	0.98***	0.98***	1.00***	0.99***	0.99***
Size 3	0.98***	0.94***	0.96***	0.94***	0.94***
Size 4	0.91***	0.91***	0.89***	0.91***	0.86***
Big	0.79***	0.74***	0.70***	0.70***	0.66***
β_{SMB}					
Small	0.29***	0.14**	0.27***	0.15*	0.53***
Size 2	-0.45***	-0.63***	-0.41***	-0.66***	-0.25***
Size 3	-0.70***	-0.75***	-0.78***	-0.76***	-0.68***
Size 4	-0.95***	-0.98***	-1.04***	-0.91***	-0.81***
Big	-1.21***	-1.14***	-1.15***	-1.15***	-0.96***
β_{VMG}					
Small	-0.11**	-0.18***	-0.04	0.02	0.25***
Size 2	-0.03	0.01	0.10*	0.07	0.24***
Size 3	0.09	0.00	0.08	0.12**	0.16**
Size 4	0.14*	0.04	0.07	0.22***	0.25***
Big	0.08	0.03	0.11	0.13	0.22*
β_{SENTI}					
Small	-0.11**	-0.00	0.11	0.15**	0.30***
Size 2	-0.12	-0.15**	-0.08	0.08	0.16*
Size 3	-0.22**	-0.16**	-0.11	-0.12	-0.03
Size 4	-0.32***	-0.42***	-0.23**	-0.31***	-0.16
Big	-0.56***	-0.44***	-0.69***	-0.59***	-0.46***

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-sentiment sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB), value (β_VMG), and sentiment (β_SENTI). The portfolios are categorized into five sentiment levels, ranging from Pessimistic to Optimistic, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all but the smallest 30 % of stocks by market capitalization and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 21: CH-3 portfolio performance analysis across 5x5 size-sentiment-sorted portfolios for factors formed on Panel A

CH-3 factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Pessimistic	Sentiment 2	Sentiment 3	Sentiment 4	Optimistic
Alpha					
Small	0.25**	0.33***	0.38***	0.43***	0.76***
Size 2	-0.09	-0.13	0.01	0.09	0.49***
Size 3	-0.26**	-0.30**	-0.30**	-0.12	0.10
Size 4	-0.49***	-0.17	-0.17	-0.24*	0.30*
Big	-0.38***	-0.11	-0.05	0.19	0.15
β_{MKT}					
Small	0.99***	0.99***	1.00***	0.96***	0.95***
Size 2	0.97***	1.01***	1.00***	1.01***	0.96***
Size 3	1.01***	1.01***	1.02***	1.00***	0.98***
Size 4	0.99***	1.05***	1.04***	1.06***	0.97***
Big	1.07***	1.06***	1.05***	1.10***	0.98***
β_{SMB}					
Small	1.06***	1.01***	1.09***	1.02***	1.18***
Size 2	0.93***	0.86***	0.90***	0.87***	0.88***
Size 3	0.78***	0.74***	0.80***	0.72***	0.74***
Size 4	0.64***	0.52***	0.50***	0.54***	0.53***
Big	0.19***	0.07*	0.07	-0.03	0.07**
β_{VMG}					
Small	-0.36***	-0.46***	-0.42***	-0.41***	-0.51***
Size 2	-0.42***	-0.36***	-0.41***	-0.35***	-0.51***
Size 3	-0.43***	-0.35***	-0.31***	-0.35***	-0.38***
Size 4	-0.36***	-0.40***	-0.29***	-0.28***	-0.34***
Big	-0.19***	-0.19***	-0.03	-0.01	-0.03

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-sentiment sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB) and value (β_VMG). The portfolios are categorized into five sentiment levels, ranging from Pessimistic to Optimistic, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on all stocks and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 22: CH-3 portfolio performance analysis across 5x5 size-sentiment-sorted portfolios for factors formed on Panel C

CH-3 factor-based portfolio performance analysis across 5x5 size-sentiment-sorted portfolios.

	Pessimistic	Sentiment 2	Sentiment 3	Sentiment 4	Optimistic
Alpha					
Small	-0.13	-0.08	-0.21*	-0.17	-0.41**
Size 2	-0.15	-0.07	-0.23**	0.03	-0.23
Size 3	-0.32**	-0.19	-0.19	-0.16	-0.07
Size 4	-0.36*	-0.04	-0.00	-0.23	0.11
Big	-0.03	0.12	0.41	0.52	0.31
β_{MKT}					
Small	0.99***	1.00***	1.03***	0.99***	1.06***
Size 2	0.98***	0.98***	1.00***	0.99***	0.99***
Size 3	0.98***	0.94***	0.96***	0.94***	0.94***
Size 4	0.91***	0.92***	0.89***	0.91***	0.86***
Big	0.79***	0.74***	0.70***	0.70***	0.66***
β_{SMB}					
Small	0.27***	0.13**	0.28***	0.17*	0.56***
Size 2	-0.45***	-0.65***	-0.43***	-0.65***	-0.23**
Size 3	-0.73***	-0.77***	-0.79***	-0.77***	-0.69***
Size 4	-0.99***	-1.03***	-1.06***	-0.95***	-0.82***
Big	-1.28***	-1.20***	-1.24***	-1.24***	-1.02***
β_{VMG}					
Small	-0.17***	-0.19***	-0.01	0.08**	0.34***
Size 2	-0.05	-0.02	0.09*	0.14***	0.32***
Size 3	0.04	-0.02	0.08	0.11**	0.18***
Size 4	0.06	-0.08	0.03	0.14**	0.23***
Big	-0.08	-0.09	-0.11	-0.04	0.08

This table presents the estimated coefficients for the factor-based portfolio performance analysis using 5x5 size-sentiment sorted portfolios. The analysis considers the market excess return (β_MKT), size (β_SMB) and value (β_VMG). The portfolios are categorized into five sentiment levels, ranging from Pessimistic to Optimistic, and five size categories from Small to Big. The significance levels are indicated by asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Factors are built on the smallest 30 % of stocks by market capitalization and sample period is January 2009 to December 2023. All t-statistics are computed with the heteroskedasticity-consistent standard errors (White 1980)

Appendix 23: Summary statistics for CH-4* factors formed on different panels

Summary statistics of factors formed on panels varying in stock selection.

	MKT (CH-4*)	SMB (CH-4*)	VMG (CH-4*)	PMO (CH-4*)
Panel A: Factors formed all individual stocks in China				
Mean	0.51	1.05	0.77	0.83
Standard deviation	6.50	4.28	3.17	3.39
t-statistics	1.05	3.26	3.83	3.48
Sharpe Ratio	0.08	0.24	0.24	0.25
Skewness	-0.17	-0.17	0.06	-0.82
Excess Kurtosis	1.72	2.51	2.16	4.83
Panel B: Factors formed all but the smallest 30% of stocks in China				
Mean	0.45	0.54	0.92	0.79
Standard deviation	6.45	4.14	3.62	3.55
t-statistics	0.93	1.77	4.07	3.34
Sharpe Ratio	0.07	0.13	0.25	0.22
Skewness	-0.15	-0.06	0.06	-0.88
Excess Kurtosis	1.72	2.67	1.35	5.23
Panel C: Factors formed on only the smallest 30% of stocks in China				
Mean	1.83	0.79	0.82	0.87
Standard deviation	8.91	1.42	2.32	2.49
t-statistics	2.84	7.06	4.55	4.90
Sharpe Ratio	0.21	0.56	0.34	0.35
Skewness	0.04	0.52	0.15	-0.23
Excess Kurtosis	1.40	0.98	0.62	0.84

This table reports summary statistics for the replicated CH-4 factors market (MKT), size (SMB) and value (VMG) and turnover (PMO), formed on different sample sizes, varying in the stocks included. The monthly mean returns, standard deviation, Newey-West (1987) adjusted t-statistics calculated with 4 lags as well as the monthly Sharpe Ratio, Skewness and Excess Kurtosis for each factor is reported. Sample period is January 2009 – December 2024.

Appendix 24: Average adjusted R-squares for factors formed on different panels

Average adjusted R-squares for individual stocks in China.

	Panel A (Returns)	Panel B (Returns)	Panel C (Returns)
Panel A: CH-4* factors (MKT, SMB, VMG, PMO) formed all individual stocks in China	0.368	0.363	0.410
Panel B: CH-4* factors (MKT, SMB, VMG, PMO) formed all but the smallest 30% of stocks in China	0.374	0.373	0.402
Panel C: CH-4* factors (MKT, SMB, VMG, PMO) formed on only the smallest 30% of stocks in China	0.339	0.319	0.419

The table presents average adjusted R-squares on regressions of monthly individual stocks' returns on factors in China's stock market. The factors are formed on panels varying in stock selection. Regressions are estimated on the CH-4* model including excess market returns (MKT), size (SMB), value (VMG) and turnover (PMO). All regressions are run over a rolling three-year (36 months) window. The mean of the adjusted R-squares is first calculated over time for each individual stock and then averaged across stocks. The factor model is used to explain three distinct Panels of returns: those of all individual stocks (Panel A), those of all but the smallest 30 % by market capitalization (Panel B) and those of only the smallest 30% by market capitalization (Panel C). The sample period is January 2009 – December 2023.

Acknowledgements

I would like to acknowledge the support of my friends and family, without whom I would not have been able to finish my studies, let alone my thesis.

Thank you to my family for always cheering on me, celebrating the small and big successes alike. Thank you to my friends who have kept me sane. Special thanks to Yannick, Luma, Jeremy and Chris who have been discussing my thoughts with me. But mostly, thank you Markus, for all the small things you do and for supporting me in every aspect of life.

I want to dedicate this thesis to Opa, who had been my biggest supporter, and to Oma, whose values continue to guide me through life.