



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/243578/>

Version: Accepted Version

Article:

JIANG, ZHAOHUI, ANDERSON, KEITH PHILIP and STAFYLAS, DIMITRIOS (2026) A PLS value factor and investor sentiment: evidence from China. Global Finance Journal. 101306. ISSN: 1044-0283

<https://doi.org/10.1016/j.gfj.2026.101306>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A PLS value factor and investor sentiment: evidence from China

Zhaohui Jiang^{ab}, Keith Anderson^a, Dimitrios Stafylas^a

^aSchool for Business and Society, University of York, Church Lane, YO10 5ZF, England

^bCurrent affiliation: School of Economics, Shandong University, Jinan, Shandong, China

Abstract

We explore the value premium in China's Shanghai-A and Shenzhen-A stock markets, and how to exploit it using a new investor sentiment index. Using data from 2000 to 2021, we evaluate book to market (BM), earnings to price (EP) and sales to price (SP) factors. Consistent with evidence from other countries, BM generates a stronger value premium among small-cap stocks, whereas EP generates a larger value premium among large-cap stocks. Using partial least squares (PLS), we construct a novel value factor that outperforms all other value factors in terms of higher portfolio returns. We also explore the relationship between the value premium and investor sentiment. In China, unlike most developed countries, value stocks outperform growth stocks in bull markets. Our results suggest that long-term investing in value stocks is more (less) profitable when market sentiment is lower (higher).

Keywords: Value premium; Investor sentiment; Chinese stock market; Partial least squares

1. Introduction

Prior academic research finds that value stocks, which are stocks with low prices in relation to book value of equity (Fama & French, 1992, 1993; Chan et al., 1995), debt (Bhandari, 1988), earnings (Basu, 1977; Jaffe et al., 1989) or dividends (Lakonishok et al., 1994), produce higher long-term returns compared to growth stocks i.e. stocks with high prices in relation to these fundamental values' measurements. However, when the literature examines the value premium, the use of these "traditional" value factors such as book to market (BM), earnings to price (EP) and sales to price (SP), suffers from multicollinearity. As Sun (2024) notes, when two factors are strongly correlated, it can be hard to say which one to keep. It is important to use proper methods for reducing the number of variables or selecting the right ones, especially when many variables explain the outcome in similar ways (ibid.). In addition, the common use of multiple regression analysis has been criticized due to problems estimating models with latent variables, reciprocal causality, and interdependence between variables (see Ramli et al., 2018, among others).

In this article, we apply the Partial Least Squares (PLS) methodology to solve multicollinearity and create a novel valuation ratio in the Chinese market that outperforms BM, EP and SP in terms of return generation to investors. We further explore when the value premium is higher assisting investors to have higher returns, considering market sentiment as well. In an influential paper, Baker and Wurgler (2006) provide an investor sentiment index in the US. Their findings support the idea that future returns of the value premium are high when sentiment is high. Baker et al. (2012) verify the US results and extend this theory to five developed countries: Canada, France, Germany, Japan and the United Kingdom. In this article, we similarly create a market sentiment index for the Chinese market, and examine future returns of value stocks and sentiment, complementing the previous studies.¹ Our analysis is informed by the investor habitat theory of Kumar and Lee (2006), and we suggest that similar

¹ China has two of the world's five largest exchanges by turnover (Shanghai Stock Exchange, 2023) and is the largest developing market. China's stock market has grown substantially from 1990 into 2022 with a total market capitalization of 78.8 trillion RMB (People's Bank of China, 2023). Also, it has a large number of individual investors, and it currently attracting great attention from the research community (Cheema & Nartea, 2014; and Cornell et al., 2020). In the Shanghai Stock Exchange, individual investors hold 46 million accounts (accounts with shares), which is 99.72% of all accounts (Shanghai Stock Exchange Yearbook, 2022). Furthermore, individual investors hold stock valued at 10.6 trillion RMB (24 per cent of the total market value) which is more than institutional holdings of 8.3 trillion RMB). The largest stock owners are general corporations, which own 23.1 trillion RMB (53.11% of the market), though these are less liquid than shares held by individual and institutional investors. Lastly, foreign investors hold 1.4 trillion RMB of stock (3.28%).

retail-investor habitats exist in the Chinese market. In addition, Chinese investor sentiment is largely retail-driven. This is due to the limited possibilities for short selling in China, which results in multi-year bear markets when value stocks are unpopular.

A limitation of Fama and French (1992) is that their framework attributes all value premiums to risk, rather than recognising that value signals also capture the extent to which market prices deviate from intrinsic values. Earnings-to-price (EP), cash-flow-to-price (CFP), book-to-market (BM), and sales-to-price (SP) ratios therefore reflect not only a firm's underlying growth prospects and profitability capacity but also the degree of mispricing relative to intrinsic value (Shiller, 1984; Shleifer and Summers, 1990). Empirically, value premiums tend to be stronger in developing markets than in developed ones (Pätäri & Leivo, 2017), consistent with the notion that mispricing plays a more prominent role in emerging markets. Griffin (2002) argues that country-specific three-factor models outperform global or international specifications in explaining average stock returns. In this context, several studies have discussed the performance of value factors in the Chinese stock market (Chen et al., 2010; Cheung et al., 2015; Cakici et al., 2017; Liu et al., 2019). One of the most influential such papers is from Liu et al. (2019); they examine EP, BM and cash flow to price, and find EP can best capture the value effect in China. Expanding on this strand of literature, we shed further light on two important questions of how to apply value investing in China. The first question is: *how best can Chinese investors exploit the value premium to maximize returns?*

A familiar approach that uses multiple value indicators that work together to select value stocks is choosing stocks with consistently the highest values for EP, SP and BM (e.g., Dhatt et al., 1999) but this method has several drawbacks.² A better approach is provided by Van der Hart et al. (2003), who normalize variables within the range of plus-minus three by subtracting the cross-sectional median and dividing by the median absolute deviation. The sum of the pertinent normalized variables constitutes the score for the strategy. This approach solves the problem of too few stocks in the portfolio, but it does not capture the impact of changes in market capitalization on the value premium, as the BM ratio performs better for small stocks

² First, after the number of stocks is divided multiple times, there will be fewer and fewer stocks left in the portfolio. Investors are unable to go further using other indicators such as momentum and screening stocks. For example, when screening the top 1/5th of stocks at a time using the value indicator, after three screens of BM, EP and SP, there are only 1/125th of the total stocks in the pool. Second, when stocks are divided by several value factors into too many equal parts, a stock which has consistently the highest values for all EP, SP and BM might not exist. When the investor chooses a stock that meets all value indicators and is relatively profitable, this method cannot compare the value of two stocks with the same SP ratio, because one stock has a relatively higher EP, and the other has a relatively higher BM.

than large stocks (Fama & French, 2008). In this article we consider the above issues, and we improve the value factor using Partial Least Squares (PLS).

Although recent literature increasingly adopts machine learning models, PLS remains competitive. For example, Gu et al. (2020) show that the cumulative log returns of long-only portfolios sorted on out-of-sample predicted returns place PLS close to random forests and slightly below neural networks. Likewise, Jiang et al. (2024) find that in forecasting earnings direction, PLS delivers overall return levels similar to random forests and marginally underperforms neural networks. The strength of PLS lies in its interpretability—regression coefficients directly reflect variable importance—making it more accessible to both researchers and practitioners. We employ a Convolutional Neural Network (CNN) model to learn the size and valuation ratios and compare its performance with the PLS value factor. The results show that the CNN delivers a lower Sharpe ratio in the test set than the PLS value factor, largely due to overfitting. We also provide the full set of PLS loadings in Table 9 (Chinese stock market) and Appendix V (U.S. stock market), enabling investors to replicate the PLS value factor using only basic linear operations.

While numerous studies discuss the value premium extensively, very few of them examine when value stocks perform better than growth stocks. Only a portion of the literature in the field of behavioural finance addresses the relationship between investor sentiment and the value premium (Baker & Wurgler, 2006; Du, 2011; Baker et al., 2012; Huang et al., 2015). They suggest that sentiment may cause systematic patterns of growth stock mispricing (Baker & Wurgler, 2006; Huang et al., 2015).

Baker and Wurgler (2006) suggest investor sentiment has a more significant effect on securities whose valuations are highly subjective and difficult to arbitrage in the US. When sentiment is low, the subsequent returns for small, young, high volatility, extreme growth, unprofitable, non-dividend-paying, and distressed stocks tend to be high. Returns on growth companies are (on average) lower when beginning-of-period sentiment is high, representing the correction of a bubble in growth stocks. Thus, the value premium should vary depending on sentiment, largely influenced by growth stocks. Du (2011) confirmed this using US data and find that the cumulative return of high minus low-value portfolios in the current year is directly proportional to the change in investor sentiment in the same year. Furthermore, the current year's investor sentiment index position is inversely proportional to the cumulative

return of high minus low value portfolios for the coming year.³

Although there is no officially recognized stock market sentiment index in China, several prior works have shed some light on the impact of sentiment on China's stock market. Chen et al. (2014) combine economic indicators and market-based factors to construct a market-based Chinese sentiment index. Zhu and Niu (2016) provide an investor sentiment in the Chinese stock market considering expected earnings growth and the required rate of return. Han and Li (2017) explain investor sentiment as irrational investors who push the rational investors out of the market, because of short selling restrictions. They provide evidence that the level of investor sentiment is very likely to forecast positive stock returns in the short run, while still predicting stock returns negatively over longer horizons. So far, there has been no investigation into the relationship between the value premium and investor sentiment in the Chinese market. Our study examines this void considering investors' high returns as well. Thus, our second question is: *how best can investors use market sentiment for their investments in value stocks?*

Regarding the first research question (value stocks), we start by examining simple one-variable classifications of growth and value stocks. As shell resources guarantee a minimum value for the stock (Liu et al., 2019), small stocks are safer than large stocks in the Chinese stock market. Therefore, unlike developed countries, value stocks have on average a larger market cap than growth stocks in China (Liu et al., 2019). Furthermore, this single variable classification method is easily influenced by the size effect. After controlling for the size factor, the value premium becomes larger than previously. Finally, because of the latent variables and interdependence between value factors, we use the PLS method to build a combined value factor in each size decile.

Regarding the second research question (investor sentiment), following Baker and Wurgler (2006), we use principal component analysis (PCA) to build a monthly market sentiment proxy in the Chinese market, with five investor sentiment proxies. In addition, unlike previous studies (e.g., Chen et al., 2014; Zhu & Niu, 2016; Han & Li, 2017) our study is the first that adjusts the calculation methodology based on the Chinese market's characteristics. For instance, to determine RIPO⁴ for a stock listed on the Shanghai or Shenzhen Stock

³ Baker et al. (2012) expanded this theory worldwide; they constructed sentiment indices in Canada, France, Germany, Japan and the United Kingdom, and they find that future returns of growth stocks are low when sentiment is high.

⁴ RIPO is defined as the average first-day returns on IPOs in Baker and Wurgler (2006).

Exchange, we survey 3,838 IPOs and begin with the day of its initial public offering (IPO). Since the Shenzhen and Shanghai stock exchanges have rules governing the rise and fall of stocks on the opening day, if the stock reaches the daily trading limit on the first day, we use data from two days prior to the IPO. If it continues to reach the limit on the second day, we use data from the third day, and so on. If the stock experiences consecutive limit hits, we use data from the day following the last limit-hit day to calculate RIPO.⁵ Our sentiment proxy is consistent with historical market practices, and we believe that accurately reflects monthly market sentiment. Baker and Wurgler (2006) and Baker et al. (2012) consider monthly returns when they explore sentiment and its relationship with the value premium.

Our findings show that, first, consistent with other countries, BM generates more of a value premium in small cap stocks, but EP generates more of a value premium in large cap stocks. Second, our PLS value factor surpasses all other value factors tested when it comes to generating portfolio returns in the Chinese stock market in the hold back period (January 2012 to January 2021). It summarises the strengths of different value indicators, so the PLS value factor performs well in almost all market capitalizations. It performs on average better than any other single value factor (BM, EP and SP) over the short, medium and long term in both the training and hold-back periods. Moreover, the SP is superior to both EP and BM. Third, when sentiment is high, the value stocks identified by BM, EP, SP and our combined value factor provide higher short-term future returns than growth stocks which is contrary to Baker and Wurgler's (2006) study for US and Baker et al. (2012) study for five other developed countries. During periods of high sentiment, value stocks provide lower long-term returns compared to growth stocks.

Our study provides several contributions. First, we create the PLS value factor; to our knowledge, we are the first to construct composite value factors for stocks using partial least squares, which addresses several of the problems when using OLS regressions.⁶ Empirically, the PLS value factor enhances the explanatory power of multifactor models and helps investors identify higher-premium value portfolios. While Fama and French (1992) propose a linear relation between factors and returns in the three-factor model, our findings extend the evidence of Fama and French (2006, 2012, 2015), who acknowledge that the book-to-market premium

⁵ For more details on how we generated the sentiment proxies, see Appendix II.

⁶ The parameters of the PLS value factor, including the mean and standard deviation required for standardization, as well as loading weights, can be obtained directly from the paper for both U.S. (Appendix V) and Chinese stock market data (Table 9).

is more pronounced among small-capitalisation stocks than among large-capitalisation stocks—a persistent challenge to the linear framework. In Fama and French (2012), across four regions (North America, Europe, Japan, and Asia-Pacific), value premiums in average stock returns generally decline with firm size, and the global three-factor model underestimates the value–growth spread for microcaps while overestimating it for megacaps. Our work extends these findings by recognising that value premiums are size-dependent, rather than uniform across the cross-section. Building on this insight, we construct the PLS value factor, which explicitly captures how valuation ratios vary with firm size. Empirically, the PLS value factor improves the explanatory power of multifactor models and delivers higher premiums than traditional single-ratio value factors. It is also more interpretable and considerably less prone to overfitting than CNN-based approaches, which achieve good in-sample fits when learning low-dimensional inputs—valuation ratios and size—but deteriorate in the test set. In contrast, the PLS value factor provides more stable performance and higher realised premiums in the validation sample.

Second, this study provides further insights from the ongoing debate between efficient market hypothesis and behavioural finance theory, providing empirical evidence consistent with the “investor habitat” perspective. Under the Efficient Market Hypothesis, as proposed by Fama and French (1992), stock prices are assumed to reflect the present value of expected future cash flows, and correlations in returns merely capture correlations in fundamentals, with sentiment shocks fully arbitrated away. In contrast, behavioural theories (Shiller, 1984; Shleifer & Summers, 1990) posit that prices are jointly determined by the interaction between noise traders and rational arbitrageurs. Barberis et al. (2005) show that the influence of retail sentiment is stronger where retail participation is concentrated, while Kumar and Lee (2006) provide evidence of systematic co-movements in retail investors’ trades. They identify “retail investor habitats”—market segments dominated by small-cap, low-priced, high book-to-market, and low institutional ownership stocks—where sentiment-driven mispricing tends to persist. Our results extend this view by offering new empirical evidence consistent with the existence of such investor habitats. Complementing Baker et al.’s (2012) study, we are the first to examine the relationship between the value premium and investor sentiment in the Chinese context.⁷

⁷ In addition, the dividend premium proposed by Baker & Wurgler (2006) is questionable in constructing investor sentiment measures for the Chinese equity market because in most developed markets, rising sentiment is

Third, our findings provide an alternative explanation for the U-shaped relationship between earnings-to-price (EP) ratios and subsequent returns in the Chinese market.⁸ Specifically, we show that the U-shaped EP-return pattern is primarily driven by the extraordinary returns of firms involved in reverse takeovers. These are typically small cap, loss-making firms that, following the reverse merger, deliver exceptional post-event performance. This mechanism, unique to the institutional and regulatory environment of China, offers a novel contextual insight into the pricing dynamics of low-EP stocks in emerging markets. This finding fills a critical gap in the existing literature by clarifying a misinterpreted aspect of EP-related return patterns in the Chinese market. Prior studies, such as Chen et al. (2010), argue that the EP premium is weak in China.

Fourth, while several studies have compared the performance of various value indicators in the Chinese market (e.g., Chen et al., 2010; Cheung et al., 2015; Cakici et al., 2017; Liu et al., 2019), none has explicitly examined the predictive power of the sales-to-price (SP) factor. This study fills that gap by providing evidence that the SP factor consistently outperforms other single valuation metrics in explaining the cross-section of returns for both large-cap and small-cap stocks in China.⁹

Finally, the practical implications of our study are that we assist investors to exploit short-term and long-term investment strategies, for instance, considering the high sentiment which can lead to higher (lower) short-term (long-term) returns on value stocks compared to growth stocks. Also, they can match value signals to market cap (e.g. EP for large caps, BM for small caps). Fund managers can adopt PLS for custom factor construction and adjust their portfolio allocations and exposures based on sentiment and implement time-variant investment strategies when exploiting value premiums. Since investor behaviour differs structurally from that in developed markets, policy makers can use this to understand the unique dynamics and

accompanied by the overvaluation of growth stocks (Baker et al., 2012) and Baker & Wurgler's (2006) conclusion on the dividend premium is no longer applicable.

⁸ The U-shaped relationship between earnings-to-price (EP) ratios and subsequent returns has been documented in various markets, including China (Chen et al. 2010), Japan (Chan et al. 1991), and the United States (Jaffe et al., 1989; Fama & French, 1992). In most developed markets, this phenomenon is typically attributed to survivorship bias (Jaffe et al, 1989), where extremely low-EP stocks—often distressed—only appear to outperform due to the removal of failed firms from the sample.

⁹ The role of firm size in moderating the predictive power of different value indicators in the Chinese market has received limited attention in prior research, and it is explicitly examined in our study. Consistent with evidence from developed markets (Chan et al., 1991; Jaffe et al., 1989; Fama & French, 1992), we find that EP is more effective in identifying undervalued large-cap stocks, whereas BM is better suited for detecting value opportunities among small-cap firms in China.

design regulations that stabilize sentiment or to promote greater transparency and investor education to mitigate overreactions during high-sentiment periods. Two concrete recommendations we can make are, first, that policymakers could use our investor sentiment measure to help them accurately gauge prevailing market sentiment. At times of extremely high sentiment, regulators could increase the number of IPO approvals. This would reduce the gap between offer prices and prices in the secondary market and prevent stocks deviating excessively from their IPO valuations. Second, government pension funds could adopt a counter-cyclical strategy by selling small-capitalisation and value stocks when retail optimism is high and repurchasing them when sentiment is low. The pension system could then earn arbitrage profits at the same time as reducing the mispricing caused by a large influx of retail investors into their preferred investment habitat.

The remainder of this article is structured as follows. Section 2 provides a literature review of various aspects of the value premium and sentiment indices. Section 3 covers our methodology including our use of Partial Least Squares. Section 4 is our empirical analysis and discussion. Section 5 provides robustness tests including a comparison using this methodology in the US market. Section 6 provides a short illustration of the success of this strategy on the Chinese market and Section 7 concludes.

2. Literature review

In this section we briefly introduce the literature on value investing in the US and on global markets, then look in more detail at value premium research in China. The U-shaped returns chart when sorting companies by EP is a particular observation in China which has until now remained unexplained. We then survey the performance of different value factors such as BP, EP and SP. Finally, we briefly survey the literature on investor sentiment.

2.1. The history of value investing research

Value investing is traced back to Graham's investing strategies (Graham & Dodd, 1934). Academic studies have discussed the value premium since the 1960s (Nicholson, 1960; McWilliams, 1966; Breen, 1968; Nicholson, 1968). Value stocks are defined as those in which the market price is relatively low in relation to earnings per share (Basu, 1975), cash flow per share (Lakonishok, et al., 1994), book value per share (Fama & French, 1992), or sales per share (Barbee et al., 1996). Numerous studies have discussed which value factor provides the highest and most stable returns (e.g., Senchack & Martin, 1987; Fama & French, 1998). Van

Der Hart et al. (2003) explore various stock selection strategies in emerging markets in 32 countries. Their findings indicate that internationally diversified portfolios with high EP and BM stocks outperform growth stocks. Multi-factor models have no constant criterion for the value factor used, though the book value-to-market ratio (BM) is the most popular. Different value indicators perform differently for different market capitalization stocks. As Fama and French (2008) state, the BM ratio performs better for small stocks than large stocks.

2.2. Value investing research around the world and in China

After value investing was shown to perform well in the US market (Graham & Dodd, 1934; Basu, 1977), researchers collected evidence from global markets. Results are slightly different in different countries, at different periods, and using different databases (see Pätäri & Leivo, 2017, for a systematic literature review). The conflicting results can mostly be explained by differences in the samples, time periods, and methodologies employed. Additionally, as value factors perform differently in different countries, Griffin (2002) argues that country-specific three-factor models are more useful in explaining average stock returns than are world and international versions. Therefore, a number of researchers survey the performance of value factors in China (Chen et al., 2010; Cheung et al., 2015; Cakici et al., 2017; Liu et al., 2019).

Chen et al. (2010) examine the cross-sectional predictive power of three fundamental valuation ratios—book-to-price (BP), earnings-to-price (EP), and cash-flow-to-price (CFP)—in the Chinese equity market. Their findings indicate that BP dominates both EP and CFP in explaining stock returns. Notably, the return pattern associated with EP is U-shaped: while returns increase from the third to the tenth EP decile portfolios, the lowest two EP deciles deliver unusually high returns. In particular, the bottom decile (lowest EP stocks) outperforms the top decile (highest EP stocks) by 0.01%, suggesting that low-EP stocks—traditionally viewed as growth-oriented—can yield unexpectedly strong performance in the Chinese context.

A cross-sectional return predictability comparison of three commonly used valuation ratios—book-to-price (BP), cash-flow-to-price (CFP), and earnings-to-price (EP), was examined by Cakici et al. (2017). Their empirical analysis reveals that BP and CFP consistently outperform EP in explaining cross-sectional variation in stock returns in China. Hu et al. (2019) investigate the explanatory power of size and value factors in the cross-section of stock returns in the Chinese A-share market. Using a standard double-sorting approach, they construct six portfolios by intersecting two size-sorted groups with three book-to-market (B/M) sorted groups. Their results reveal no robust value premium, suggesting that the value factor does not

significantly explain cross-sectional variation in Chinese stock returns during the sample period.

Liu et al. (2019) compare the predictive power of four valuation ratios—cash-flow-to-price (CFP), earnings-to-price (EP), book-to-market (BM), and assets-to-market (AM)—in explaining cross-sectional stock returns. Using Fama–MacBeth regressions that include all four ratios simultaneously, they find that EP exhibits the strongest explanatory power, outperforming the other value measures. Chen et al. (2010) provide yearly decile returns portfolios for different value factors in the Chinese stock market. They test 18 factors, including value, size, momentum and liquidity. According to their panel data, the PE returns are a U-shape. The bottom 30% of stocks sorted by PE and PCF ratio perform better, and for the remaining 70%, the stock returns increase as the value increases (*ibid.*). Moreover, the result is similar to the results in Japan found by Chan et al. (1991) and in the US stock market found by Jaffe et al. (1989) and Fama and French (1992). The U-shaped return severely weakens the return premium typically attributed to the EP factor. However, Chen et al. (2010) do not offer a theoretical explanation for this anomaly.

Despite the contributions of these studies, several important questions remain unanswered. First, the underlying economic rationale for the U-shaped return pattern associated with the earnings-to-price (EP) ratio observed in Chen et al. (2010) has not been fully explained. Second, the role of firm size in moderating the predictive power of different value indicators in the Chinese market warrants further investigation. Third, the performance and explanatory power of the sale-to-price (SP) factor remain underexplored. Finally, it is still unclear under what market conditions or over what time periods the value premium tends to be stronger.

Overall, two hypotheses have been proposed in the literature to explain the U-shaped return pattern observed in value stocks in countries other than China. First, the U-shaped return comes from investors over-extrapolating the performance of value firms with negative earnings in the future. Investors over extrapolate the performance of value firms with negative earnings in the future, compared to companies with less negative EP. Once the investors' expectational errors are corrected, this results in higher future returns (Gharghori et al., 2013). Second, the return of companies with negative earnings is susceptible to survivor bias (Jaffe et al., 1989). This explanation may have made sense in some other countries' markets in the past, but it is not applicable to the Chinese market.

Contrary to the previous literature, we provide another alternative explanation in the Chinese stock market. The U-shaped return comes from the returns of reverse takeovers.¹⁰ Companies subject to reverse takeovers will deliver extremely high subsequent performance. Investors often perceive these near-bankrupt companies as low-value and not worth purchasing. However, after undergoing a reverse merger, these companies can become new entities with strong earnings quality and attractive fundamentals. Even if a reverse merger has not yet occurred, the potential for such a transaction means that their market value will not fall below the value of their shell source—that is, the value associated with bypassing the time and regulatory costs of the traditional IPO process. We find the U-shaped return is completely absent for BM and SP, which supports this explanation. According to the Implementation Measures for the Suspension and Termination of Listing of Loss-making Listed Companies issued by the SFC in 2001, the stock exchange shall suspend the trading of the shares of a listed company for the third consecutive year of losses. Therefore, small size stocks with a low EP are more likely to be targeted for reverse takeovers than those with a low BM or SP.

2.3. Which value factor to use?

The effectiveness of different value factors varies across market capitalization segments. Value premiums, particularly those associated with the BP ratio, tend to be more pronounced in stocks with smaller market capitalizations. Fama and French (2006) find that in the U.S. markets, significant BP-based value premiums (return difference between the top two versus the bottom two BP quintiles) exist for all five size groups except for the largest capitalization group. According to Fama and French (2012), the BM value premium decreases when moving from small cap samples toward the larger cap ones in the North America, Europe, and Asia Pacific. They also suggest that larger BM value premiums for small stocks are typical.

Cheung et al. (2015) test the performance of value, size, momentum, dividend yield and volatility. They claim that the value premium is more sensitive among large size stocks in China. However, they only consider BM as a value factor in their study. These findings

¹⁰ A reverse merger in the A-share market typically involves a non-listed firm acquiring a controlling stake in a publicly listed company that serves as a shell. Following the acquisition, the listed entity undergoes a significant asset restructuring process, whereby its original business operations are divested and the core assets of the acquirer are injected into the vehicle. As a result, corporate control, business focus, and often the firm's name is transferred to the formerly unlisted entity. Effectively, the private firm “puts on the shell,” achieving public listing status through a non-IPO route.

contradict the results of Fama and French (2006). Moreover, their study does not test how the performance of other value factors varies across different size segments.

2.4. The sales-to-price ratio

The SP ratio is defined as the ratio of sales to price (Sales/Market Capitalization) and measures a firm's revenue generation relative to its market valuation, serving as an indicator of valuation efficiency and potential mispricing. Fisher (1984) and subsequent studies use the SP ratio; in particular, Senchack and Martin (1987) examine the efficacy of SP for value portfolio selection. SP becomes the single value factor that provides the highest value premium in other studies examining the US (Dhatt et al., 1999; Dhatt et al., 2004; Barbee et al., 2008), the UK (Leledakis & Davidson, 2001) and Australia (Gharghori et al., 2013). The SP premium is on average higher than the BM and EP premiums in the medium and long term. A potential justification is that SP is a little-used value measure, compared to BM and EP. The information in the BM and EP may be more effectively embedded into stock returns than the information in SP because investors may have a tendency to focus more on EP and BP than on SP (Barbee et al., 2008). Despite extensive investigation into value and other market factors in the Chinese context, previous studies (such as Chen et al., 2010; Cheung et al., 2015; Cakici et al., 2017; Liu et al., 2019) have not evaluated the performance of the SP factor.

2.5. Investor sentiment indices

Baker and Wurgler (2006) provide an investor sentiment index in the US. Their finding supports the idea that future returns of the value premium are high when sentiment is high. Baker, et al. (2012) verify the US results and extend this theory to five developed countries: Canada, France, Germany, Japan and the United Kingdom. Baker and Wurgler's (2006) market sentiment indicator was composed of several variables: the closed-end fund discount CEFD, turnover, the average first-day returns RIPO, the share of equity issues in total equity and debt issues, the share of equity issues in total equity and debt issues, and the dividend premium $p^D - ND$.

The turnover ratio is based on the ratio of reported share volume to average shares listed in the market (Rouwenhorst, 1999; Baker & Stein, 2004). Karpoff (1987) suggests that the market turnover should be high in a bull market and low in a bear market. Moreover, Baker & Stein (2004) suggest that turnover can work as a sentiment index. In a market with short-sales

constraints, irrational investors can only provide liquidity when they are excited. Therefore, it is a sign of overvaluation.

High first-day returns on IPOs are a leading indicator in Baker & Wurgler's indicators of investor sentiment (Baker & Wurgler, 2006). Moreover, Ritter (1991) documents long-term reversals in returns on IPO stocks. They infer that the evidence is consistent with positive sentiment that significantly affects the prices of young growth stocks. Low returns on IPOs are often analysed as a sign of market timing (ibid). The number of IPOs is a lagging indicator in Baker and Wurgler (2006). In addition, Ibbotson and Jaffe (1975), Lowry and Schwert (2002), and Benveniste et al. (2003) find that IPO volume lags the first-day returns on IPOs.

The dividend premium P^{D-ND} is the difference between the logs of the dividend-paying companies' and non-dividends-paying companies' average market-to-book ratios. If nonpayers' average market-to-book ratios are much higher than dividend payers' market-to-book ratios, the market sentiment is high (Baker & Wurgler, 2006). Baker & Wurgler (2004) give two justifications for dividend premiums, which are investor sentiment theory and catering theory. In investor sentiment theory, when the dividend premium is high, investors seek firms that show safety features, including dividend payment; when it is low, investors prefer firms with the characteristics of maximum capital preference, which means no dividends. Further evidence pointing to sentiment is the positive correlation between the dividend premium and the closed-end fund discount (Baker & Wurgler, 2006). The nature of catering theory is that managers deliver to investors what they currently like. In the case of dividends, catering means that managers tend to provide dividends when investors put a relatively high stock price on dividend payers and tend to skip dividends when investors prefer non-dividend-paying companies.

In China, an increasing number of new accounts shows a more significant demand for stocks from investors (Li & Zhang, 2008; Chen et al., 2014). When there is a bullish (bearish) sentiment, the number of new accounts will certainly increase (decrease) (ibid.). Moreover, "VIX is the world's leading measure of investor sentiment and market volatility," according to the CBOE. When the VIX is high, investor sentiment is presumed to be low since investors are assumed to be risk averse (Bandopadhyaya & Jones, 2008).

Overall, value investing has been widely examined, originally in the US starting in 1960s, but more recently around the world and in China. The U-shaped returns when sorting Chinese stocks by PE is a particularly unusual fact which has not until now been explained. Different statistics show different power to identify value stocks in different countries: BP is

the most popular but SP is particularly powerful. Much research has also been done on investor sentiment since the seminal paper of Baker and Wurgler (2006). We provide a rigorous methodology for determining an investor sentiment index suitable for the Chinese context below.

3. Methodology

Traditional multiple regression analysis suffers from multicollinearity issues (see appendix I and II). Therefore, we use the PLS to improve the value factor. Unlike ordinary least squares (OLS), which estimates coefficients directly from the original predictors, PLS extracts latent components that maximise the covariance between valuation ratios and future stock returns. This enables the model to retain the predictive information shared by BM, EP and SP while substantially reducing redundant information arising from multicollinearity. Because our objective is to construct a predictive value factor rather than merely reduce the dimensionality of the predictors, PLS is more appropriate than unsupervised approaches such as principal component analysis (PCA). A PLS model is trained by finding the multidimensional direction in the X space (BM, EP and SP) that best explains the maximum multidimensional variance direction in the Y space (one-year ahead excess return), therefore avoiding redundant information among independent variables. The PLS method was pioneered by Wold (1975) and extended by Kelly and Pruitt (2015). It has also been applied in finance by Kelly and Pruitt (2013), Huang et al. (2015) and Mselmi et al. (2017). Additionally, it has been widely applied in other fields, including strategic management (Birkinshaw et al., 1995; Hulland, 1999), information technology management (Gefen & Straub, 1997; Yi & Davis, 2003; Venkatesh & Agarwal, 2006) and marketing (Reinartz et al., 2004; Singh et al., 2006).

3.1. Partial Least Squares regression (PLS)

PLS and PCA methods can both handle multicollinearity (Kelly & Pruitt, 2013). The primary distinction between them in how they handle multicollinearity lies in their dimension reduction techniques. PCA condenses variables based on their covariance among predictors and is an unsupervised machine learning method. PLS consolidates variables based on their covariance with the target variable, aiming to identify the best linear combination of predictors for forecasting purposes (ibid.). More specifically, a PLS model finds the multidimensional direction in the X space (BM, EP and SP) that best explains the maximum multidimensional variance direction in the Y space (one-year ahead excess return), making a supervised machine

learning method (Wold, 1975). The purpose of the PLS procedure is therefore not only statistical dimension reduction but also economic aggregation. BM, EP and SP capture different aspects of relative firm valuation. Individually, each ratio contains useful but incomplete information regarding future expected returns. By combining these valuation measures according to their predictive contribution to subsequent excess returns, the resulting latent factor is expected to provide a more comprehensive representation of firm value than any single valuation ratio.

$$Cov(X, Y) = \sqrt{Var(X)Var(Y)}r(X, Y) \rightarrow \max \quad (1)$$

Specifically in this case, a stock possessing high EP, BM and SP shares common characteristics of value stocks not only in terms of relative profitability but also possibly due to the market's low expectations for growth (Cai & Keasey, 2022). When there are multiple features in explanatory variables, unsupervised machine learning methods tend to extract the common features of all explanatory variables, making it difficult to distinguish the specific features intended to be extracted from others. Therefore, we use the PLS approach to construct a new value factor from BM, EP and SP. For each stock, we define a latent variable V , which is jointly reflected by BM, EP and SP. Economically, this latent variable represents the common information contained in multiple valuation ratios about the firm's relative profitability and market valuation. Firms with simultaneously high BM, EP and SP are typically characterised by strong fundamentals relative to market prices and are therefore more likely to exhibit value characteristics. In a PLS model, latent variables are explicitly defined as linear combinations of their observed indicators. We assume that the one year ahead excess stock return, R , is explained by the relative profitability of stocks, as captured by the value factors, and follows the standard linear relation:

$$R = \alpha + \theta V + E \quad (2)$$

where R is the $NT \times 1$ vector of one year ahead excess stock return, V is a $NT \times 3$ matrix of stacked value factors, BM, EP and SP, and E is a $NT \times 1$ vector of residuals.

Previous studies suggest that the predictive ability of valuation ratios varies systematically across firm size (Fama & French, 2008). Consequently, imposing a single set of loading coefficients across all firms would implicitly assume that BM, EP and SP contribute equally to expected returns regardless of firm size. This assumption is inconsistent with the existing empirical evidence. To accommodate this cross-sectional heterogeneity, securities are first sorted into ten size deciles at the end of each month. Within each size decile, all firm-month observations from the training period (January 2000 to December 2010) are pooled to

form a training sample, and a separate PLS model is estimated for each size group.

We condense the set of predictors from dimension P ($P=3$) to a smaller number of K linear combinations of predictors. Prior to model estimation, all explanatory variables are standardised to have zero mean and unit variance to ensure that the estimated loading weights are not driven by differences in measurement scales across the valuation ratios. In line with Gu et al. (2020) and Leippold et al. (2022), we represent the PLS regression in its matrix form:

$$R = (ZW_K)\theta_K + \tilde{E} \quad (3)$$

where R is the $NT \times 1$ vector of one year ahead excess stock return, Z is the $NT \times P$ matrix of stacked value factors, BM, EP and SP, W_K is a $P \times K$ transformation matrix, θ_K is the $K \times 1$ vector of model parameters.

Next, we find the transformation matrix that optimizes the correlation between the transformed variables and the forecast goal that is the fundamental concept behind PLS regression. In this context, the columns of w_k , represented as $w_1, w_2, \dots, w_k$, are determined by solving a series of optimization problems, i.e.,

$$w_j = \arg \max_w Cov^2(R, Zw), \text{ s.t. } w'w = 1, Cov(Zw, Zw_l) = 0, l = 1, 2, \dots, j-1 \quad (4)$$

We adopt the algorithm developed by Venturini and Mehmetoglu (2019) to calculate of w_k . The PLS model is estimated using the full training sample, which consists of monthly firm-level observations from January 2000 to December 2010. The estimated loading weights are then fixed and applied to the hold-back sample from January 2011 to December 2021 to construct the PLS value factor. This procedure ensures that all out-of-sample analyses are based solely on information available during the training period and avoids look-ahead bias. Finally, the PLS combined value factor, V , is estimated by multiplying the explanatory variables BM, EP, and SP by their respective loading weights:

$$V = w_{BM}BM + w_{EP}EP + w_{SP}SP \quad (5)$$

where the loading weights w_{BM} , w_{EP} , and w_{SP} are coefficients that represent the contribution of each valuation ratio to the overall PLS combined value factor. The loading weights differ across market size deciles, as described later in the PLS value factor section of the empirical analysis. The estimated loading weights are subsequently applied, without further re-estimation, to the hold-back sample (January 2011 to December 2021) to construct the PLS value factor.

3.2. *PCA and investor sentiment*

Investor sentiment is inherently unobservable and therefore must be inferred indirectly from multiple market-based indicators. Individual sentiment proxies capture different dimensions of

investor optimism and pessimism, while also containing idiosyncratic noise. Prior studies propose several sentiment proxies as time-series conditioning variables, and combining multiple proxies provides better measures of investor sentiment than relying on a single proxy (Baker & Wurgler, 2006). Sentiment proxies are highly correlated with each other. Following Baker and Wurgler (2006), we use PCA to build up investor sentiment by reducing dimensional sentiment indicators.¹¹ PCA extracts the common variation shared across these indicators and therefore provides a more reliable measure of aggregate investor sentiment than any individual proxy. The independent variables are all standardised by subtracting the mean and dividing by the variance.

The first main component is to construct a new dimension by linearly combining several of the original dimensions in such a way that the variance of the projections of all the original samples on the new dimension is minimised. The second component is found using the eigenvector associated with the largest eigenvalues to the first component, and it should be orthogonal to the first component, and so on for the next components.

$$c_i = J_{1i}proxy_1 + J_{2i}proxy_2 + \dots + J_{qi}proxy_q \quad (6)$$

We only use the first principal component to construct the sentiment indicator and drop other components, because our aim is to find the parts of them that are co-linear.

4. Empirical analysis

In this section, we first describe our sample selection and data, then look at how valuation ratios vary widely in China for different firm sizes, and how we can ‘neutralise’ this effect. We next construct our PLS value indicator. In the final two sub-sections we construct our sentiment index then consider how it articulates with the value premium through time.

4.1. Data and sample selection

4.1.1. Sample

We employ a comprehensive dataset for the Shanghai-A and Shenzhen-A stock markets based on both the CSMAR and Choice databases. Four stock exchanges, the Shanghai Stock

¹¹ As we mentioned previously, PLS is also a method of dimensionality reduction. There are two reasons for choosing PCA rather than PLS in constructing the sentiment index. First, we cannot be sure that all investor sentiment proxies we have determined can predict the value premium or the market index in China. Second, if we apply PLS to fit investor sentiment by both future return and future value premium, and get different loading weights, there is a subjective appraisal of which one is closer to the real investor sentiment.

Exchange, the Shenzhen Stock Exchange, the Second Board market and STAR, exist in China. We remove stocks on STAR and the Second Board where some stocks lack liquidity. Our sample, Shanghai-A and Shenzhen-A, hold most of the large companies in the Chinese market, and the IPO review is more stringent. The sample period spans from January 2000 to December 2021; the data for future one-month ahead return ends in November 2021, one-year ahead returns end in January 2020, and five-year ahead returns end in January 2016. The number of stocks has grown from 777 in January 2000 to 3,127 in December 2021. In line with Liu et al. (2019), we only use data after 2000 because a package of Chinese securities laws was passed in December 1998 and implemented in July 1999. It regulates financial reporting and guarantees consistency in accounting information.

When calculating the time series of the value premium, we drop all data listed for less than one month to eliminate meaningless returns created by the IPO effect. The Chinese stock market's lengthy and strict IPO review process provides companies with an additional value called shell value (see Liu et al., 2019). The market price of a stock is often much higher than its IPO price. Therefore, after the IPO, the daily limit is usually reached for a few days or weeks because of the daily up and down limits.¹² These stocks contain value stocks, but investors have little chance to participate in the trade because the stockholders usually sell their stocks after the price reaches the maximum market price. We define the return for delisted stocks that will not be relisted as -1.

4.1.2. Variables

We examine the earnings to price ratio (EP), the cash flow to price ratio (CFP), the book to market ratio (BM) and the sales to price ratio (SP). In addition, we test eight investor sentiment proxies: turnover ratio (TURN) (Rouwenhorst, 1999; Baker & Stein, 2004), first-day return of IPO (RIPO) (Baker & Wurgler, 2006), the number of IPOs (NIPO) (Baker & Wurgler, 2006), the dividend payout ratio, new investor accounts (NIA) (Li & Zhang, 2008; Chen, et al., 2014) the dividend premium (Baker & Wurgler, 2004), the Volatility Index (VIX) (Bandopadhyaya & Jones, 2008), and market index movement (Hong & Stein, 1999). Turnover ratio, RIPO, NIPO, dividend payout ratio and NIA are finally selected as components of our sentiment index. The other three proxies are dropped because their correlation with other sentiment

¹² The Shanghai Stock Exchange and the Shenzhen Stock Exchange set regulations on the rise and fall of stocks on the first day. Since January 2014 the highest (lowest) price increase (decrease) of the new stocks listed on the Shanghai A and Shenzhen A on the first day has been 44% (36%) of the issue price and a 10% increase and decrease limit on subsequent days. We survey stocks that reached the daily limit after their IPO from 2014 to 2021 in the Choice database. There are 269 stocks that hit the daily limit for more than ten days.

proxies is very low.

Size

Market capitalisation data are collected from the Choice database. It represents the total market value of all of a company's outstanding shares.

Valuation ratios

We examine the earnings to price ratio (EP), the cash flow to price ratio (CFP), the book to market ratio (BM) and the sales to price ratio (SP). All except BM are trailing 12-month figures. The EP, CFP and SP are calculated by the sum of the last four quarterly earnings, free cash flow or sales divided by the current share price. We obtain BM by dividing the book value of the latest financial results by market capitalization. EP, CFP, BM, SP are available in the Choice database. To limit the influence of extreme observations, we winsorise each variable at the 5th and 95th percentiles.

Market turnover (TURN)

TURN measures the average ratio of monthly transaction value to market capitalisation outstanding in the A-share market. The turnover ratio is based on the ratio of reported share volume to average shares listed in the market (Rouwenhorst, 1999; Baker & Stein, 2004). Karpoff (1987) argues that the market turnover should be high in the bull market and low in the bear market. Moreover, Baker and Stein (2004) suggest that turnover can work as a sentiment index. In a market with short-sales constraints, irrational investors can only provide liquidity when they are enthusiastic. Therefore, it is a sign of overvaluation. The turnover ratios are obtained from the CSMAR database for the period January 2003 to December 2021.

Average first-day return on IPOs (RIPO)

RIPO is defined as the closing price divided by the IPO issue price minus one. Following Baker and Wurgler (2006), this variable is widely recognised as a leading indicator of investor sentiment. However, the calculation is adjusted to reflect the unique regulatory framework of the Chinese market, where daily price limits restrict price movements to +44% on the first day and +10% thereafter. To correct for these limits, if a stock hits the daily limit for n consecutive days after its IPO, the closing price on day $n+1$ is used in the RIPO calculation. Moreover, the “shell value” of small IPOs, as documented by Liu, Stambaugh, and Yuan (2019), induces abnormally high RIPO values among micro-cap stocks. To mitigate this distortion, the smallest decile (market cap below 227 million RMB) is excluded. Across 3,838 IPOs from January 2000 to December 2021, smaller firms show average IPO premiums of 214%, compared to 67% for large firms. Our RIPO specification therefore differs from previous Chinese studies by

incorporating trading-limit adjustments, providing a more suitable measure of market liquidity and investor sentiment under China's unique institutional constraints. Zhu and Niu (2016) employ a similar RIPO measure but do not adjust for the trading-limit constraint that frequently affects stock liquidity in the Chinese market. Other studies, such as Chen et al. (2014) and Han and Li (2017), do not include RIPO as a sentiment or liquidity proxy. The IPO sample, including listing dates, daily returns over the following 30 trading days, total shares, and issue prices, is obtained from the CSMAR database for the period January 2000 to December 2021. Firm size is calculated as the product of total shares and the issue price.

Number of IPOs (NIPO)

NIPO captures the monthly number of IPOs and follows Baker and Wurgler (2006) as a lagging indicator of market sentiment. Previous studies (Ibbotson & Jaffe, 1975; Lowry & Schwert, 2002; Benveniste et al., 2003) suggest that IPO volume tends to lag first-day IPO returns. In China, the IPO process is heavily influenced by government regulation. The China Securities Regulatory Commission (CSRC) has suspended IPOs on nine separate occasions between 1990 and 2022, typically during market downturns. These suspension periods are treated as null values in our dataset, consistent with low market sentiment. The IPO sample is obtained from the CSMAR database from January 2000 to December 2021.

Dividend premium

Following Baker and Wurgler (2004, 2006), the dividend premium is defined as the log difference between the average market-to-book ratios of dividend-paying and non-paying firms. We define dividend-paying stocks as those that have paid dividends in the past 12 months, that is, stocks with a dividend payout ratio greater than zero, where the dividend payout ratio is calculated as the sum of dividends over the past 12 months divided by the market capitalisation at the end of the month. A high dividend premium indicates investor preference for safety and stability, while a low premium reflects a speculative appetite for high-growth, non-dividend-paying firms. The measure captures the interplay between investor sentiment and managerial catering behaviour, in that managers tend to pay dividends when investors favour them. Market-to-book ratio and dividend payout ratio are obtained from the Choice database for January 2000 to December 2021.

Dividend payout ratio

The dividend payout ratio equals total dividends over the past 12 months divided by the stock price. It serves as a complementary sentiment proxy, assumed to be negatively correlated with investor optimism. When sentiment is low, more firms initiate dividends to attract conservative

investors; when sentiment is high, firms withhold dividends as investors chase capital gains (Baker & Wurgler, 2006). Dividend payout ratio is obtained from the Choice database for January 2000–December 2021.

New investor accounts (NIA)

NIA represents the number of newly opened investor accounts each month and proxies retail participation in the market. A rising number of new accounts signals bullish investor sentiment, while a decline indicates pessimism. This measure is particularly relevant in China, where retail investors dominate trading activity. The variable follows Li and Zhang (2008) and Chen and She (2014). New investor accounts in each month are from the CSMAR database for January 2003–December 2021.

Volatility index (VIX)

The VIX is proxied by the CBOE China ETF Volatility Index, as the domestic Chinese volatility index (SSE 50ETF VIX) was only available between February 2015 and February 2018. Following Gong, Zhang, Wang, and Wang (2022), who document a correlation of 0.85 between the two indices, we employ the CBOE China ETF volatility as a consistent time series measure of market uncertainty. Higher VIX values indicate greater fear and lower sentiment (Bandopadhyaya & Jones, 2008). VIX data are sourced from the Chicago Board Options Exchange for March 2011–December 2021.

Market movement (MM)

MM is measured as the return on the Shanghai A Index in the current month, calculated as the index value at the end of the current month divided by that at the end of the previous month minus one. It captures cyclical investor behaviour and the interaction between momentum and value effects. During market downturns, momentum traders often sell indiscriminately, while value investors hold or accumulate undervalued stocks (Hong & Stein, 1999; Lakonishok, Shleifer, & Vishny, 1994). Conversely, in rising markets, optimistic expectations suppress the value premium. Shanghai A Index data is drawn from the CSMAR database for January 2000–December 2021.

4.1.3. Descriptive statistics

Table 1 reports the descriptive statistics of size and four valuation ratios: book-to-market ratio (BM), cash flow-to-price ratio (CFP), earnings-to-price ratio (EP), and sales-to-price ratio (SP). The standard deviations show that BM and SP vary more widely across firms, while CFP and

EP are more concentrated. Most ratios are positively skewed, except EP, which is negatively skewed due to firms with low or negative earnings.

Insert Table 1 around here

Table 2 presents the pairwise correlations among firm size, BM, CFP, EP, and SP. All correlations are positive and statistically significant at the 1% level, implying that these ratios tend to move together. In particular, BM shows relatively high correlations with SP (0.525) and CFP (0.336), suggesting that these measures may capture overlapping information about firm value. The strong correlations across the valuation ratios further indicate the presence of multicollinearity.

Unlike many developed countries, the valuation ratios (EP, BM, CFP and SP) of large stocks are on average higher than small stocks in the Chinese stock market. In developed markets generally the valuation ratio of stocks is negatively correlated with the size of stocks (see e.g. Bali et al., 2016 for BM; Jaffe et al., 1989 for EP; Leledakis & Davidson, 2001 for SP). The main reason is that small stocks are, on average, riskier than large stocks, giving investors relatively higher expectations, which makes them have relatively higher BM, CFP, EP and SP. A very low valuation ratio means that the market expects the company may not survive as its price has fallen a lot. Large companies with more financial resources, different divisions, and political connections are more likely to recover from such a crisis than small companies. A large company has lower risk than a small company, so its value premium is relatively stable. According to Chan and Chen (1991), from 1956 to 1985, 51% of the companies in the top size quintile had been listed on the NYSE for more than ten years. However, just 20% of enterprises have remained in the bottom quintile for more than 10 years, and approximately 66% of the firms have moved from higher to lower quintiles. Additionally, the bottom quintile has more highly leveraged enterprises.

Insert Table 2 around here

In China, however, as the market size increases, the value indicator of the stock also tends to increase. Table 3 shows the average values of size (in million RMB), BM, CFP, EP and SP, plus the average one-month-ahead, one-year-ahead and five years ahead returns for each size decile portfolio. The average BM of the largest decile is 43.43% which is much larger than the smallest decile, 29.69%. The average SP of the largest decile is 69.90%, which is also much larger than the smallest decile at 48.32%. BM and SP rapidly increase in the smallest 20% of market size and approach large size stocks' average BM and SP by the third decile. The

correlation is even more pronounced for EP and CFP. The average value of EP and CFP increases with market capitalization in both small and large sizes. As the market capitalization increases, the average EP of stocks in each portfolio increases from -4.01% to 4.85%, and it is negative in decile portfolios one and two. The average CFP of stocks in each portfolio increases from -0.68% to 2.29%. The reason is that a listed company on the verge of bankruptcy has shell resources in China, so it will probably be a target for acquisition. Due to the lengthy listing process, a company wishing to go public will often take a shell listing of a listed company that is on the verge of bankruptcy. In addition, the number of companies aspiring to be listed far exceeds the number of shell companies. Losing the risk of delisting, small size stocks are even safer than large-cap stocks in China, as their stock prices have fallen to the bottom and cannot fall any more.

Insert Table 3 around here

Table 4 presents the descriptive statistics of our investor sentiment proxies. The distributions of the sentiment indicators are different. NIA, dividend payout ratio, NIPO, RIPO, turnover ratio, and VIX have positively skewed distributions. Only index return and dividend premium are slightly negatively skewed. Table 5 provides the correlations among the explanatory variables. Overall, there is a moderate correlation for the majority of sentiment variables. The most significant correlation is between the turnover ratio and NIA (0.642). These two proxies are also the most direct proxies of market sentiment in emerging markets. The dividend premium is not significantly correlated with any other investor sentiment except the dividend payout ratio. Contrary to Baker and Wurgler's (2006) study which finds that the dividend premium is negatively correlated with the turnover ratio, we find that the dividend premium is slightly positively correlated with turnover ratio and NIA (0.02 and 0.06). In most developed markets, rising sentiment is accompanied by the overvaluation of growth stocks (Baker et al., 2012). On this premise, nonpayers' average market-to-book ratios are much higher than dividend payers' market ratio when market sentiment is high, so it should be negatively correlated. However, in the Chinese stock market, with rising investor sentiment, investors are not chasing growth stocks, but value stocks.

Insert Tables 4 & 5 around here

Table 1: Descriptive statistics for the firm-month sample

Variable	Mean	SD	Skewness	Min	p25	p50	p75	Max	Observations
size	15400	76300	18.77491	104	2390	4440	9530	5670000	474101
BM	39.59%	25.85%	116.25%	3.59%	20.19%	0.3340596	0.5268296	1.402156	474101
CFPTTM	3.89%	7.12%	95.90%	-17.94%	0.13%	0.0259054	0.0658877	0.3416324	474101
EPTTM	2.20%	3.84%	-142.54%	-31.19%	0.77%	0.0213779	0.0392648	0.1562009	474101
SPTTM	55.48%	58.40%	201.05%	2.33%	16.64%	0.347523	0.7042601	3.923022	474101

Note: The table presents the descriptive statistics for the firm-month sample. The reported statistics include the mean, standard deviation (SD), skewness, minimum (Min), first quartile (p25), median (p50), third quartile (p75), maximum (Max) and the number of valid observations for size (in million RMB), BM, CFP, EP, SP. The sample covers the period from January 2000 to December 2021 and is constructed using a comprehensive dataset for the Shanghai A-share and Shenzhen A-share stock markets obtained from the CSMAR and Choice databases. The valuation ratios (BM, CFP, EP, and SP) are winsorised at the 5th and 95th percentiles to mitigate the influence of extreme observations.

Table 2: Correlation matrix for the firm-month sample

Variables	Size	BM	CFPTTM	EPTTM	SPTTM
Size	1				
BM	0.062***	1			
CFPTTM	0.104***	0.336***	1		
EPTTM	0.125***	0.272***	0.244***	1	
SPTTM	0.011***	0.525***	0.321***	0.190***	1

Note: The table reports the pairwise Pearson correlation matrix for the main variables, including firm size (in million RMB), BM, CFP, EP, and SP, measured in levels. The sample comprises firm-month observations from January 2000 to December 2021 for Shanghai A-share and Shenzhen A-share firms obtained from the CSMAR and Choice databases. The valuation ratios (BM, CFP, EP, and SP) are winsorised at the 5th and 95th percentiles. “***”, “**”, and “*” denote statistical significance at the 1%, 5% and 10% levels, respectively (two-tailed P-value).

Table 3: Size-sorted portfolio characteristics

Size	1(Small)	2	3	4	5	6	7	8	9	10(Large)
Average size	1529.20	2097.57	2595.63	3160.43	3849.23	4781.07	6203.65	8636.19	14022.20	85728.25
Average BM (%)	29.69	38.02	40.43	41.73	41.79	41.45	40.72	40.62	40.07	43.43
Average CFP (%)	-0.68	0.34	0.52	0.63	0.71	0.79	0.93	1.21	1.52	2.29
Average EP (%)	-4.01	-0.27	0.56	1.25	1.76	2.16	2.59	3.15	3.48	4.85
Average SP (%)	48.32	57.41	62.49	63.79	61.51	65.51	69.40	71.82	68.64	69.90
Average monthly return (%)	2.43	1.68	1.37	1.22	0.99	1.00	0.89	0.93	0.89	0.80
Average annual return (%)	26.28	20.88	18.01	17.07	15.03	15.07	14.04	14.43	14.57	13.58
Average 5 years return (%)	147.74	114.22	101.90	92.29	84.41	79.34	78.83	78.52	74.26	69.21

Note: The table reports the average values of firm size (in million RMB), BM, CFP, EP, SP, and the average one-month-ahead, one-year-ahead, and five-year-ahead returns for each size decile portfolio. At the end of each month, all stocks are sorted into 10 decile portfolios based on firm size. Portfolio returns are equal-weighted. The reported values are time-series averages over the sample period from January 2000 to December 2021. The sample comprises Shanghai A-share and Shenzhen A-share firms obtained from the CSMAR and Choice databases. The monthly return ends in December 2021. The annual return ends in January 2020. The five-year return ends in January 2016.

Table 4: Descriptive statistics for investor sentiment proxies

Variable	Mean	SD	Skewness	Min	p25	p50	p75	Max	Observations
NIA	120.95	129.72	3.03	6.89	37.38	97.13	151.35	892.42	228
Dividend premium	-11.93%	4.93%	-5.26%	-23.44%	-15.49%	-12.19%	-8.11%	-1.62%	264
dividend payout ratio	77.46%	29.29%	32.38%	27.97%	55.91%	72.29%	100.56%	149.35%	264
NIPO	14.54	14.02	1.13	0.00	4.00	10.00	23.50	67.00	264
RIPO	0.99	0.83	1.42	-0.05	0.34	0.91	1.48	6.27	264
Turnover ratio	0.24	0.12	1.38	0.08	0.16	0.21	0.30	0.69	228
VIX	25.85	6.51	1.72	15.09	21.49	24.66	28.93	60.25	130
c1	0.00	1.51	1.28	-2.52	-1.14	-0.17	0.77	6.44	228

Note: The table presents the descriptive statistics for the investor sentiment proxies, comprising NIA, Dividend Premium, Dividend Payout Ratio, NIPO, RIPO, Turnover Ratio, VIX, and the first principal component (c1). The reported statistics include the mean, standard deviation (SD), skewness, minimum, first quartile (p25), median (p50), third quartile (p75), maximum, and the number of monthly observations for each variable. The sample period spans from January 2000 to December 2021. Dividend Premium, Dividend Payout Ratio, NIPO, and RIPO are available from January 2000 to December 2021. NIA, Turnover Ratio, and the first principal component (c1) are available from January 2003 to December 2021. VIX is available from March 2011 to December 2021. None of the investor sentiment proxies reported in this table are winsorised or trimmed.

Table 5: Correlations of investor sentiment proxies

Variables	Index return	NIA	Dividend premium	Dividend payout ratio	NIPO	RIPO	Turnover ratio	VIX	c1
Index return	1								
NIA	0.222***	1							
Dividend premium	0.05	0.02	1						
Dividend payout ratio	-0.069	-0.446***	0.400***	1					
NIPO	0.001	0.255***	0.004	-0.093	1				
RIPO	0.062	0.460***	-0.062	-0.204***	0.270***	1			
Turnover ratio	0.351***	0.642***	0.06	-0.267***	0.068	0.261***	1		
VIX	-0.260***	0.093	0.052	0.003	-0.058	-0.093	0.223**	1	
c1	0.230***	0.889***	-0.108*	-0.592***	0.448***	0.645***	0.724***	0.052	1

Note: The table reports the pairwise Pearson correlation matrix for investor sentiment proxies, comprising NIA, Dividend Premium, Dividend Payout Ratio, NIPO, RIPO, Turnover Ratio, VIX, and the first principal component (c1). The sample comprises monthly time-series observations from January 2000 to December 2021. Pairwise Pearson correlations are computed using all non-missing observations for each variable pair. “***”, “**”, and “*” denote statistical significance at the 1%, 5% and 10% levels respectively (two-tailed P-value).

4.2. Comparing valuation ratios' profitability for different firm sizes

4.2.1. Limitations of single factor value portfolios in China

In China, single-factor value portfolios may represent two very different groups of stocks at different times. Sometimes they are large stocks and at other times small stocks. From January 2000 to December 2021, the average market value of the highest EP sorted portfolios fluctuated repeatedly, e.g. between July 2006 to August 2009 there was a low of only 0.74 of the market average at the end of October 2007; it reached a high in August 2009, which was 4.98 times the market average. The relative size is calculated by dividing the highest EP group's average value by the market's average value each month. Therefore, the main results of this article use the value premium after controlling for market size.

4.2.2. The value premium for different sizes of firm

Table 6 provides the average value portfolio return in the Chinese stock market from January 2000 to January 2021 in ten different size portfolios. In line with Chou et al. (2011), the securities are first assigned to ten deciles by size. Then, within each size decile, securities are assigned to ten deciles based on BM, EP and SP. The value premium is calculated by the equally weighted return of the top 10% of BM, EP or SP stocks minus the equally weighted return of the bottom 10%. Our holding periods are one month, twelve months and five years. BM, EP and SP provide a positive premium except for the smallest size decile.¹³

¹³ We also tested the performance of CFP (not reported). CFP cannot be used to identify future stock returns in China. In our analysis, when considering only one-month-ahead returns, the CFP hedge portfolio exhibited positive returns. However, when extending our analysis to one-year-ahead and five-year-ahead returns, the performance of CFP becomes notably more volatile irrespective of whether we control for market capitalization or remove stocks with negative earnings. CFP performs well in many developed countries, while its poor performance in this research in China may come from the varying definitions of cash flow. The definition of cash flow is inconsistent in how cash flow is defined and measured across different studies. In Fama & French (1998), Chan et al. (2006) and Barbee et al., (2008), cash flow is defined to be the sum of earnings and depreciation. In Israel & Moskowitz (2013) and Hou et al. (2015), cash flow is income before extraordinary items plus equity's share of depreciation plus deferred taxes. In a paper much closer to the present study about the Chinese market, Liu et al. define cash flow as "the net change in cash or cash equivalents between the two most recent cash flow statements." The cash flow to price ratio we employed in this study is the same as Liu et al. (2019). Liu et al. (2019) showed the CFP is second to EP as a value indicator in the Chinese stock market, and they did not evaluate the SP. They conducted cross-sectional regressions of monthly stock returns on individual stock characteristics, covering the period from January 1, 2000, to December 31, 2016.

As presented in Table 6, valuation ratios behave differently at different sizes. If the group with the smallest market value is excluded, in which the performance value factors can be heavily influenced by reverse mergers, EP is better able to identify large-size value stocks. Conversely, the BM indicator performs better in small cap stocks, which agrees with the findings of Fama and French in 2008 in the US. BM distinguishes value from growth stocks better than EP-sorted portfolios in sizes 2, 3, 4, 5, 7 and 8 in both monthly and annual portfolios on average. It also performs better than EP-sorted portfolios in sizes 2, 3, 4, 6, 7 in the 5 years ahead return. The higher premium mainly comes from the short portfolio, and the low BM stocks provide on average lower returns than low EP stocks. Taking annual data as an example, the growth BM portfolio provides 2.34% on average lower returns than the EP-sorted portfolio in the least profitable decile across all market capitalisation groups.

SP becomes the value factor which provides the highest value premium. The results are in alignment with other studies in the US (Dhatt et al., 1999; Dhatt et al., 2004; Barbee et al., 2008), the UK (Leledakis & Davidson, 2001) and Australia (Gharghori et al., 2013). The SP premium is on average higher than the BM and EP premiums in the medium and long term. Similar to EP, our results show that SP performs better for large cap stocks compared to BM. A potential justification is that SP is a little-used value measure. Investors focus more on BM and EP, so the information in BM and EP may be more effectively absorbed into stock returns than the information in SP (Barbee et al., 2008). In the literature there are far fewer papers discussing SP than BM and EP.

Insert Table 6 around here

4.2.3. Short-term, medium-term, and long-term value premium comparisons

In line with Fama & French (2012), we find the BM premium provides relatively poor performance in the longer term and a lower premium than EP and SP in the two deciles with the largest firms. A possible reason is that the BM indicators cannot reflect profitability information promptly. Even if the company loses competitive advantage, the book value of large-cap stocks usually only changes slowly. In contrast to book value, earnings and sales can change significantly with each quarterly report.

In China, the compounded monthly value premium is slightly larger than the

annual value premium, and much larger than the five year value premium.¹⁴ This is different from the results that the value premium is concentrated in the first two years after the portfolio formation in Singapore (see Yen et al., 2004) 2-3 years in Europe (see Bird & Casavecchia, 2007) and its effects continue for at least five years in the UK (see Anderson & Brooks, 2006).

Two aspects may lead to higher monthly value premiums than longer-term value premiums in China. First, individual investors and unintelligent momentum traders may cause prices to deviate from market value in the short term. China's stock market has been established a relatively short time, with a history of only about 30 years. As mentioned, there is a large number of individual investors in the Chinese stock market. Also, some momentum investors in the Chinese stock market may not be trading smart enough. For example, Hong and Stein (1999) suggest that momentum traders sell stock portfolios, but they do not justify the value of each stock. The irrational behaviour of individual investors and momentum traders causes prices to deviate from rational values, and rational investors need to spend more money to push prices back to rational values.

Second, although irrational investment behaviour is caught by diversified arbitrageurs who eliminate their influence on prices (Shleifer, 2000), not enough arbitrage is taking place in the Chinese stock market. One reason is that the friction cost of China's stock market is higher than that of other countries. Provided that the transaction fees are low enough, the monthly value premium can be arbitrated and narrowed by higher frequency transactions, such as day and intraday trades. However, the stamp tax has been 0.1 per cent since 2008 and the commission fee is usually 2-3 bps for retail investors in China (Leippold et al., 2022), which is much higher than the transaction costs (\$8.00 per million dollars) in the US (U.S. Securities and Exchange Commission, 2023). As Shleifer and Vishny (1997) state, idiosyncratic risk, which might be fundamental or noise trader idiosyncratic risk, deters arbitrageurs. High stamp tax makes traders tend to adopt low-frequency trading strategies. The other reason is that the lack of a short-selling mechanism also reduces arbitrage methods. The history of arbitrage in the Chinese market is only a decade old. In 2010, 90 constituent stocks from the Shanghai Stock Exchange (SSE) were first allowed for margin trading and

¹⁴ In calculating the compounded return, we assume a 0.3% transaction cost in each transaction.

short selling. This was increased to 900 in 2013. In 2023, a total of 2200 stocks, 1000 on the Shanghai Stock Exchange and 1200 on the Shenzhen Stock Exchange, are available for short sale.

4.2.4. The U-shaped EP return

Table 6 shows high annual returns (28%) for the smallest deciles of stocks. However, much of this is due to reverse mergers, as presented in Table 7. Panel A shows EP value portfolio returns from January 2012 to December 2021. In the smallest group, growth stocks (low EP) offer the highest returns. In the second and third smallest group, there is a U-shaped EP value stock performance. The portfolio for EP1 performs on average 4.35% better than EP2 in the smallest size group. As the size increases the U-shaped performance decreases. Comparing the portfolios of EP1 and EP2 to the portfolios of EP3, the U-shaped performance is significant in the smallest three size deciles. We also test the cases when the holding period is one month and five years, and they both present similar results. We provide two possible explanations for this phenomenon of U-shaped value stock performance in different sizes in China. First, the U-shaped return comes from investors under-estimating the future performance of value firms with negative earnings. Investors under-estimate the performance of value firms with negative earnings in the future, compared to companies with less negative EP. Once the investors' expectational errors are corrected, this results in higher future returns (Gharghori et al., 2013). Smaller companies are more likely to have extremely negative profits which is disliked by investors. Therefore, the U-shaped EP value stock performance is clear in the small stocks.

Secondly, the U-shaped return comes from the returns of reverse takeovers. Companies subject to reverse takeovers will deliver extremely high subsequent performance. There is no U-shaped return for BM and SP, so this supports that explanation. The Implementation Measures for the Suspension and Termination of Listing of Loss-making Listed Companies as issued by the SFC in 2001 stated that the stock exchange must suspend trading in a share if the company has suffered three consecutive years of losses. Small size stocks with a low EP are therefore more likely to be targets for reverse takeovers, compared to using BM or SP.

To further investigate whether the U-shaped value premium is primarily driven by reverse-merger activity or by market overreaction, we examine and remove stocks

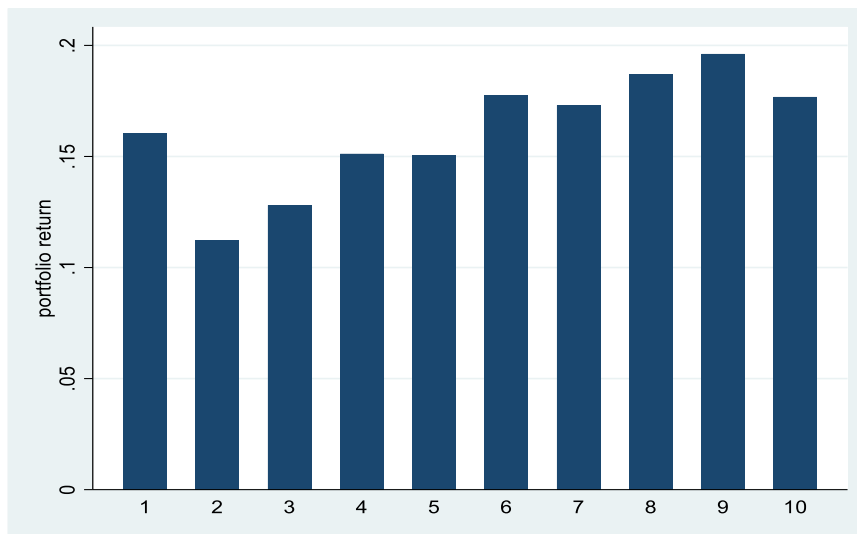
that have undergone reverse mergers. Based on information from the Choice database, we identify 148 reverse-merger transactions between January 2007 and December 2021. Table 8 summarises the post-announcement performance of reverse-merger stocks. The results indicate substantial positive returns following the announcement of a reverse-merger event. In the announcement month, the average return is 6.21%. When extending the window to include both the announcement month and the subsequent month, the cumulative return rises to 17.01%. The cumulative gains continue to build over longer horizons, reaching 56.38% over twelve months and 67.80% over thirty-six months. Although volatility increases over time, the large and persistent cumulative returns suggest that shell-resource stocks experience significant market revaluation upon reverse-merger announcements, consistent with strong investor expectations regarding the value-creation opportunities associated with obtaining a listing through this channel.

Therefore, we remove reverse merged stocks and provide the results in panel B. After removing the reverse takeover, in the smallest three market caps the returns of growth stocks fell by 4.35%, 0.86% and 1.19% respectively and consequently the value premium widened by 6.29%, 0.85% and 2.19%. Figures 1 and 2 present the average portfolio returns for each EP decile within the third-smallest market capitalisation decile, before and after excluding firms that had undergone reverse takeovers. The returns of the least valuable stocks fell by 1.19%. In market capitalisation groups 3.4.5.9.10, the least valuable stock returns will still be slightly higher than the second least valuable stock. The most notable of these is group 4, where the least valuable stock returns (13.92%) are 2.6% higher than the second least valuable stock (11.32%). This error is acceptable as we have removed all stocks where reverse takeovers occurred, which also creates survivorship bias. We conclude that the U-shaped return mainly comes from the returns of reverse takeovers in China.

Insert Tables 7 & 8 around here

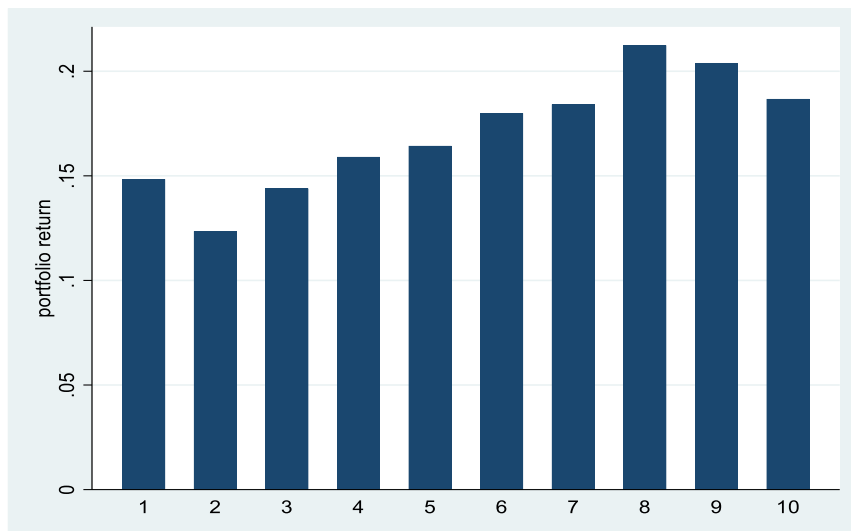
Insert Figure 1 & 2 around here

Figure 1: Average Returns by EP Decile in the Third-Smallest Size Group (Before Excluding Reverse Takeovers)



Note: This figure reports the average monthly returns of portfolios formed by sorting stocks in the third-smallest market capitalisation decile into EP deciles, prior to excluding firms that experienced reverse takeovers.

Figure 2: Average Returns by EP Decile in the Third-Smallest Size Group (After Excluding Reverse Takeovers)



Note: This figure presents the average monthly returns of portfolios formed by sorting stocks in the third-smallest market capitalisation decile into EP deciles, after removing firms that have undergone reverse takeovers.

Table 6: Value portfolio return for different size deciles

	Size	1(Small)	2	3	4	5	6	7	8	9	10(Large)
Monthly BM	Growth	2.17%	0.90%	0.52%	0.60%	0.03%	0.67%	-0.04%	0.39%	0.50%	0.41%
	Value	2.59%	2.27%	1.84%	1.76%	1.52%	1.49%	1.21%	1.52%	1.23%	1.29%
	premium	0.42%	1.37%	1.32%	1.16%	1.49%	0.82%	1.24%	1.14%	0.73%	0.87%
Monthly EP	Growth	2.79%	1.87%	1.33%	1.12%	0.64%	0.56%	0.54%	0.63%	0.33%	0.26%
	Value	2.42%	2.40%	1.95%	1.90%	1.49%	1.71%	1.45%	1.36%	1.46%	1.47%
	premium	-0.37%	0.52%	0.62%	0.78%	0.85%	1.15%	0.91%	0.73%	1.12%	1.20%
Monthly SP	Growth	1.81%	1.27%	0.78%	0.70%	0.51%	0.67%	0.39%	0.57%	0.39%	0.25%
	Value	2.57%	2.39%	1.89%	1.94%	1.63%	1.78%	1.28%	1.58%	1.06%	1.29%
	premium	0.76%	1.12%	1.12%	1.24%	1.13%	1.11%	0.90%	1.02%	0.67%	1.04%
Annual BM	Growth	28.32%	15.90%	13.02%	14.24%	9.75%	11.43%	9.56%	10.21%	12.78%	10.83%
	Value	28.56%	27.78%	22.91%	22.93%	18.43%	19.04%	18.53%	20.40%	18.70%	20.43%
	premium	0.24%	11.89%	9.89%	8.69%	8.69%	7.61%	8.97%	10.19%	5.91%	9.60%
Annual EP	Growth	28.33%	21.52%	18.04%	16.82%	13.27%	13.50%	12.62%	13.88%	10.48%	10.94%
	Value	25.73%	24.49%	23.18%	22.10%	19.10%	21.58%	20.16%	22.00%	21.92%	22.19%
	premium	-2.60%	2.97%	5.15%	5.27%	5.84%	8.08%	7.54%	8.13%	11.44%	11.24%
Annual SP	Growth	22.28%	18.91%	17.81%	14.60%	12.15%	12.78%	11.62%	9.01%	10.04%	7.64%
	Value	26.18%	26.63%	22.81%	23.88%	19.90%	21.85%	20.30%	21.92%	18.25%	23.84%
	premium	3.91%	7.72%	5.00%	9.27%	7.75%	9.06%	8.68%	12.91%	8.21%	16.20%
5 years BM	Growth	170.21%	87.60%	84.57%	70.21%	66.69%	58.74%	57.66%	58.61%	57.00%	53.66%
	Value	142.94%	141.11%	118.40%	111.80%	104.62%	102.92%	96.93%	87.99%	72.09%	95.82%
	premium	-27.27%	53.51%	33.83%	41.59%	37.93%	44.18%	39.27%	29.37%	15.09%	42.16%
5 years EP	Growth	174.27%	115.45%	109.42%	87.34%	74.39%	77.90%	63.86%	55.07%	43.39%	26.23%
	Value	164.02%	143.79%	121.51%	117.04%	121.79%	95.30%	92.44%	111.47%	112.79%	95.27%
	premium	-10.25%	28.34%	12.09%	29.70%	47.40%	17.40%	28.58%	56.40%	69.39%	69.03%
5 years SP	Growth	143.40%	109.82%	89.46%	81.06%	70.70%	59.61%	64.98%	61.78%	55.77%	38.82%
	Value	164.68%	135.26%	133.45%	118.67%	103.98%	119.59%	106.09%	95.21%	91.14%	108.00%
	premium	21.28%	25.44%	44.00%	37.61%	33.28%	59.98%	41.11%	33.43%	35.37%	69.18%

Note: The table provides the average value portfolio return in the Chinese stock market from January 2000 to January 2021 in ten different sizes. The securities are firstly assigned to ten deciles by size. Then, in each size, securities are assigned to ten deciles based on BM, EP and SP, and the value premium is calculated by the equal-weighted return of the top 10% BM, EP and SP stocks minus the equal-weighted return of the bottom 10% BM, EP and SP stocks. The holding period includes one-month, one-year and five-year period. BM, EP and SP provide a positive premium except for the smallest size decile.

Table 7: EP Value portfolio return in different size deciles from January 2012 to December 2021

Panel A: Annual EP value portfolio return										
Size	1(Small)	2	3	4	5	6	7	8	9	10(Large)
1(Growth)	31.00%	19.20%	16.03%	10.82%	7.03%	7.03%	7.74%	7.85%	13.43%	11.52%
2	26.17%	19.44%	11.22%	10.19%	8.14%	9.63%	12.71%	8.03%	10.03%	9.00%
3	28.36%	15.61%	12.80%	12.19%	10.78%	9.66%	9.63%	12.24%	16.50%	9.23%
4	23.65%	17.20%	15.12%	15.84%	12.81%	11.01%	11.75%	11.85%	12.37%	9.23%
5	26.18%	18.73%	15.06%	14.87%	13.84%	13.45%	12.07%	13.43%	8.71%	12.38%
6	26.39%	23.07%	17.75%	14.23%	15.55%	17.99%	11.59%	11.48%	12.99%	14.55%
7	25.94%	21.82%	17.30%	16.09%	15.66%	12.74%	10.81%	10.72%	11.11%	11.38%
8	28.34%	19.06%	18.72%	16.56%	16.02%	15.66%	15.08%	12.03%	10.52%	12.25%
9	26.43%	21.03%	19.64%	15.86%	14.15%	12.99%	13.39%	14.19%	11.05%	14.31%
10(Value)	24.09%	21.76%	17.67%	16.38%	13.77%	12.46%	16.19%	15.72%	11.98%	14.82%
(10-1)	-6.91%	2.56%	1.64%	5.56%	6.74%	5.43%	8.45%	7.86%	-1.45%	3.30%
Panel B: Annual EP value portfolio return (after removing the reverse merged stock)										
Size	1(Small)	2	3	4	5	6	7	8	9	10(Large)
1(Growth)	26.65%	18.34%	14.84%	13.92%	8.71%	7.28%	6.50%	5.52%	8.29%	7.16%
2	26.82%	20.86%	12.35%	11.32%	9.78%	10.52%	8.36%	5.35%	6.02%	5.19%
3	30.35%	17.55%	14.39%	12.10%	11.42%	9.77%	7.62%	8.54%	10.78%	5.24%
4	25.66%	18.76%	15.90%	16.23%	12.92%	9.14%	10.01%	9.03%	9.62%	4.98%
5	27.82%	19.95%	16.43%	14.91%	13.48%	13.03%	10.42%	10.07%	7.01%	8.01%
6	26.19%	22.82%	17.99%	15.15%	15.14%	16.87%	10.60%	10.07%	11.08%	10.32%
7	27.31%	23.16%	18.41%	17.14%	14.52%	14.28%	12.08%	11.05%	9.53%	8.66%
8	30.02%	22.96%	21.24%	17.17%	17.97%	15.17%	14.75%	11.28%	9.07%	10.05%
9	28.54%	23.88%	20.40%	18.97%	16.77%	13.35%	15.25%	13.21%	10.29%	11.82%
10(Value)	26.03%	21.75%	18.68%	17.75%	14.36%	13.85%	13.46%	15.73%	12.10%	11.52%
(10-1)	-0.61%	3.41%	3.84%	3.83%	5.65%	6.57%	6.96%	10.21%	3.81%	4.35%

Note: The table provides EP Value portfolio return in different sizes from January 2012 to December 2021. In panel A, the securities are first assigned to ten deciles by size. Then, in each size, securities are assigned to ten deciles based on EP. The holding period is one year. In Panel B, we found a total of 148 reverse-merger deals from January 2007 to December 2021 on the Choice database. The stocks that have undergone reverse mergers are removed. Then we repeat the process for panel A.

Table 8: Post-Announcement Performance of Reverse-Merger Stocks

Return	m1r	fm1r	fm3r	fm6r	fm12r	fm18r	fm24r	fm36r
Mean	6.21%	17.01%	18.22%	25.93%	56.38%	51.43%	58.94%	67.80%
SD	17.77%	42.31%	42.41%	58.76%	162.48%	117.20%	133.35%	186.01%

Note: The table reports the average returns and standard deviations of 148 reverse-merger stocks in the announcement month and the subsequent 1- to 36-month horizons. m1r denotes the announcement-month return; fm1r, fm3r, fm6r, fm12r, fm18r, fm24r, and fm36r denote cumulative returns over the announcement month and the subsequent 1, 3, 6, 12, 18, 24, and 36 months, respectively.

4.2.5. The PLS value factor

As Domingos (2012) states, “the fundamental goal of machine learning is to generalize beyond the examples in the training set.” We design a training and hold-back period to test our results, ensuring we do not overfit the data. The training period is from January 2000 to December 2010 and the results for the hold back period are from January 2011 to December 2021.

Using the procedures for building up the PLS value factor in the Section Partial Least Squares regression (PLS) in the Methodology, we obtain the PLS value factors in each size. Table 9 provides each size's standardised loadings¹⁵ of BM, EP and SP. Ignoring the smallest market size group, as market size increases, the weighting of EP gradually increases. Conversely, the loading weights of BM fall as the market capitalization rises.

The varying PLS loadings across market-capitalisation groups indicate that these valuation measures capture mispricing differently for small and large firms, reflecting heterogeneous investor expectations and behavioural biases across size segments. It is known that the intrinsic value of a stock is determined by the present value of its expected future dividends or free cash flows (with earnings being the starting point), and in some cases, by the synergies resulting from mergers and acquisitions. Under market efficiency, the EP ratio framework suggests that the product of expected earnings growth and the current earnings–price ratio should be comparable across all stocks. Hence, a firm with higher expected earnings growth would exhibit a higher price given similar current earnings and thus a lower current EP ratio. However, in less efficient or developing markets, a low EP ratio may arise not from high growth potential but from irrational investor optimism, which pushes up prices and leads to overvaluation. Consequently, portfolios formed from high-EP stocks tend to outperform those based on low-EP stocks, capturing this mispricing effect. Similarly, in the SP premium, higher current sales often signal stronger potential for future earnings, thereby offering a complementary perspective to valuation. In the smallest market size group, returns of stocks depends strongly on reverse takeovers, as we

¹⁵ "standardised loading" refers to the loadings of variables on latent factors that have been standardised to have a mean of zero and a standard deviation of one. Sales and book value will usually be several times the earnings. PE, BM and PS have very different values, so there is a need to standardised them.

explained previously. Valuation ratios such as earnings-to-price (EP), cash-flow-to-price (CFP), book-to-market (BM), and sales-to-price (SP) reflect both the firm's underlying growth potential and profitability capacity and the degree to which a stock's market price deviates from its intrinsic value.

Insert Table 9 around here

Table 10 provides the one-month ahead, one-year ahead, and five-year ahead average return of the PLS value factor sorted portfolio in the Chinese stock market from January 2012 to January 2021 in ten different sizes. It delivers an average monthly long-short return of 0.85%, an average annual long-short return of 7.71%, and an average year's long-short return of 38.86% over all size deciles. The combined value factor can provide a similar premium across market sizes, except for the smallest size. In the ninth size decile, the annual premium was negative; however, the monthly value premium and five-year premium were both positive, which seems slightly confusing. Growth stock portfolios in the ninth size decile have high one year ahead returns because high volatility causes an increase in numerical returns.¹⁶ The long-term and short-term value premiums remain positive because short-term returns are numerically less volatile where the frequency of one year ahead return and one month ahead return are monthly, and the long-term value premium is high enough to ignore this issue as it allows annual returns to average out.

Insert Table 10 around here

Table 11 compares the difference between the premium of the PLS value factor and the premiums of other single value factors from January 2012 to December 2021. In most portfolios, the combined value factor outperforms the single value indicator. These returns exceed those of BM (0.21%), EP (0.17%), and SP (0.15%) in one month ahead returns. In addition, it delivers long-short returns 1.11% higher than BM, 4.40% higher than EP, and 0.65% higher than SP for one-year ahead returns. Moreover, for long-term projections, it delivers five years of long-short returns, surpassing BM by

¹⁶ For example, financial information service provider Shanghai DZH had a market capitalisation of 13.58 billion RMB at the beginning of the bull market in 2014. After entering the bull market in April 2015, capitalisation grew to 66.55 billion RMB. After the bull market ended in April 2016, the market capitalisation returned to 18.37 billion RMB and remains at 15.4 billion RMB at time of writing (October 2023). The one year ahead returns from April 2014 and April 2015 were 3.9 and -0.72, respectively, resulting in an equally-weighted return of 1.59 times but the geometric average return for the year is 0.16.

0.26%, EP by 10.64%, and SP by 6.65%.

In addition, the performance of PLS can be pulled up or down by particularly good or particularly bad value indicators when an indicator works very well or very badly. For example, due to the number of reverse mergers, which we discussed above, EP cannot provide a positive premium for small stocks when the holding period is one year. Therefore, PLS performs better than EP and worse than SP and BM for small stocks. Similarly, among large-cap stocks, the PLS value premium outperforms BM but underperforms EP and SP when BM's performance declines.

As Stambaugh and Yuan (2017) suggested, the choice of breakpoints is not crucial to the relative performance of our value factor. We believe that the extremes of our value factor best identify value, so we use breakpoints of 10% and 90% rather than the 30% and 70% used by Fama & French (2015) and Hou et al. (2015). If we use 30% and 70%, our conclusions remain unchanged (results not reported here).

Insert Table 11 around here

Table 12 presents the Sharpe ratios and distributional characteristics of four monthly value premiums—BM, EP, SP, and PLS—over the period from January 2012 to December 2021. The mean monthly premiums range between 0.64% and 0.85%, indicating consistent positive excess returns across all value-based factors. Among them, the PLS value premium exhibits the highest average return (0.85%) and Sharpe ratio (0.182), suggesting relatively superior risk-adjusted performance.

Insert Table 12 around here

Table 13 shows the difference between the value premium in the training (January 2000 to December 2010) and hold-back periods (January 2012 to January 2021). In our two time periods, ignoring the smallest size, we observe a significant reduction in the BM, EP and SP value premiums in the short and medium term. This is consistent with the Chinese market becoming more efficient. It is also in line with the behavioural finance hypothesis of Lakonishok et al. (1994) that the value premium mainly comes from investor bias. The value premium will decrease as more arbitrageurs join in and investor overreaction decreases.

Insert Table 13 around here

4.2.6. Comparison Between ANN and PLS Value Strategies

In light of evidence that Convolutional Neural Networks (CNNs) achieve leading

performance in forecasting profitability and returns (Gu et al., 2020; Jiang et al., 2024), we compare the CNN-predicted scores with the PLS value factor. The parameters of CNN in this research are in line with Jiang et al. (2024). We apply the geometric pyramid rule (Masters, 1993) to determine the number of neurons in each layer and we use a CNN with five hidden layers. Tangent sigmoid is applied as an activation function.

Table 14 reports the performance comparison between the ANN-based value strategy and the PLS-based benchmark during the training and test periods. In the training period (2000–2010), the ANN model performs broadly similarly to PLS in terms of the value premium across size deciles. The average premium generated by the ANN strategy is 1.41%, - lower than the 1.47% achieved by PLS. The Sharpe ratios show a similar pattern, with ANN delivering comparable performance to PLS, albeit with marginally lower risk-adjusted returns in most deciles. This indicates that both models are able to identify value-related cross-sectional return patterns in-sample. However, the difference in performance doubles in the out-of-sample period (2012–2021). Although both models continue to produce positive value premiums on average, the performance of ANN deteriorates considerably relative to PLS. The average out-of-sample premium of ANN drops to 0.78% - less than the 0.90% observed for PLS. More notably, the ANN model exhibits weaker Sharpe ratios in several size deciles and even produces negative performance in the smallest-size portfolios, suggesting overfitting in the training stage. In contrast, PLS demonstrates more stable and consistent out-of-sample performance across size groups, reflecting superior generalisation ability.

Insert Table 14 around here

In summary, the ANN and PLS value factors deliver broadly comparable results in the training period, indicating that both models can identify cross-sectional return patterns in-sample. However, the ANN exhibits a modest decline in performance in the test period, suggesting reduced predictive power when applied to new data. This decline stems from overfitting to the training sample in a setting where the number of explanatory variables is limited, which constrains the network’s ability to generalise. By contrast, the PLS value factor maintains more stable out-of-sample performance and offers clearer interpretability by explicitly modelling the relationships among the predictors.

4.3. The investor sentiment index

Using the PCA procedures detailed in Section 2.2, we adopt NIA (New investor

accounts), Dividend payout ratio, Turnover, NIPO (the number of IPOs) and RIPO (The average first-day returns on IPOs) as our sentiment index.¹⁷ We obtain the aligned investor sentiment index c_1 , which is the first component, from the five individual sentiment proxies. The proportion of variation explained, c_1 , is 0.4564.

$$c1_t = 0.5886NIA_t - 0.3918Dividend\ Payout\ Ratio_t + 0.4791Turnover_t + 0.2968NIPO_t + 0.4272RIPO_t \quad (7)$$

We verify the validity of our sentiment indicators in two ways. First, the coefficients of the original explanatory variables of the principal component need to satisfy financial theory. The first principal component satisfies exactly the hypothesis that investor sentiment should be negatively related to dividend payout ratio and positively related to the other proxies.

¹⁷ In the Chinese stock market, IPO premiums will also be affected by the market capitalisation of the stock and the regulations stipulated by the China Securities Regulatory Commission. The return of an IPO is strongly affected by the size of the IPO, because IPO stocks have an additional shell value (Liu, et al. 2019). The value of the shell is not relevant to the size, therefore the smaller the size of the stock the higher the shell value, which would lead to an extremely high RIPO after the IPO process.

Table 9: Standardised loadings and statistics by size decile

	Size	1 (Small)	2	3	4	5	6	7	8	9	10 (Large)
Loadings	EP	0.72	0.30	0.43	0.34	0.40	0.37	0.36	0.53	0.81	0.75
	BM	0.96	0.97	0.93	0.73	0.85	0.77	0.79	0.79	0.66	0.75
	SP	0.25	0.49	0.36	0.88	0.71	0.86	0.85	0.83	0.73	0.79
Mean	EP	-0.06	0.00	0.00	0.01	0.01	0.02	0.02	0.02	0.03	0.04
	BM	0.22	0.33	0.35	0.35	0.36	0.36	0.35	0.34	0.33	0.33
	SP	0.41	0.51	0.55	0.56	0.56	0.56	0.55	0.58	0.59	0.54
Std. dev.	EP	0.32	0.09	0.08	0.07	0.05	0.05	0.04	0.04	0.04	0.04
	BM	0.46	0.26	0.25	0.25	0.26	0.25	0.24	0.23	0.21	0.21
	SP	0.63	0.73	0.73	0.78	0.84	0.77	0.79	0.83	0.83	0.84

Note: The table provides each size's standardised loadings, mean and standard deviation of BM, EP and SP. The training period spans from January 2000 to December 2010.

Table 10: PLS value portfolio return by size decile

	Size	1 (Small)	2	3	4	5	6	7	8	9	10 (Large)	average
Monthly	Growth	2.68%	1.62%	0.64%	0.40%	-0.20%	0.35%	0.41%	0.32%	0.80%	0.81%	0.78%
	Value	2.64%	1.99%	1.88%	1.77%	1.16%	1.65%	1.38%	1.41%	1.17%	1.31%	1.64%
	premium	-0.04%	0.37%	1.24%	1.37%	1.36%	1.30%	0.97%	1.09%	0.37%	0.50%	0.85%
Annual	Growth	27.67%	11.87%	11.18%	6.93%	2.93%	4.95%	6.25%	7.65%	15.99%	11.67%	10.71%
	Value	28.60%	21.07%	19.11%	17.58%	15.45%	17.10%	18.22%	15.50%	13.09%	18.51%	18.42%
	premium	0.92%	9.20%	7.94%	10.64%	12.52%	12.15%	11.96%	7.85%	-2.89%	6.84%	7.71%
5 years	Growth	75.83%	23.50%	16.58%	13.30%	14.79%	8.22%	17.27%	12.50%	16.14%	23.66%	22.18%
	Value	84.42%	75.16%	61.89%	55.87%	44.89%	57.21%	66.87%	57.07%	31.83%	75.22%	61.04%
	premium	8.59%	51.66%	45.30%	42.57%	30.10%	48.99%	49.60%	44.57%	15.69%	51.56%	38.86%

Note: The table provides the one-month ahead, one-year ahead, and five-year ahead average return of the PLS value factor sorted portfolio in the Chinese stock market from January 2012 to January 2021 in ten different sizes.

Table 11: Differences between combined value and BM, EP and SP premium by size decile

Size	1 (Small)	2	3	4	5	6	7	8	9	10 (Large)	average
Panel A: Difference between PLS value premium return and EP premium return											
1 month	0.47%	-0.38%	0.36%	0.58%	0.32%	0.66%	0.07%	0.14%	-0.35%	-0.17%	0.17%
1 year	7.83%	6.64%	6.29%	5.09%	5.78%	6.72%	3.52%	-0.01%	-1.44%	3.55%	4.40%
5 years	-7.34%	32.44%	41.75%	23.20%	-0.74%	26.47%	21.78%	1.25%	-12.52%	-19.85%	10.64%
Panel B: Difference between PLS value premium return and BM premium return											
1 month	-0.19%	-0.52%	-0.03%	0.58%	0.20%	0.67%	0.49%	0.18%	0.36%	0.35%	0.21%
1 year	-0.09%	-0.22%	-1.81%	1.09%	2.64%	5.67%	4.90%	2.96%	-2.43%	-1.58%	1.11%
5 years	2.02%	-5.56%	-8.29%	-7.63%	-0.89%	-3.08%	2.79%	6.51%	-0.45%	17.20%	0.26%
Panel C: Difference between PLS value premium return and SP premium return											
1 month	-0.48%	-0.22%	-0.05%	0.02%	0.51%	0.16%	0.19%	0.93%	0.59%	-0.11%	0.15%
1 year	-4.63%	2.02%	1.08%	-0.11%	4.90%	3.55%	1.99%	2.33%	0.06%	-4.66%	0.65%
5 years	10.23%	22.51%	2.85%	2.04%	10.77%	-2.29%	5.65%	10.38%	9.58%	-5.23%	6.65%

Note: The table compares the difference between the premium of the PLS value factor and other single factor value factors from January 2012 to December 2021. The securities are first assigned to ten deciles by size. Then, in each size, we obtained the results using the premium calculated by the PLS value factor minus the premium calculated by BM, EP and SP.

Table 12: Sharpe Ratios and Distributional Characteristics of Value Premiums

Variable	Mean	Standard Deviation	Skewness	Sharpe ratio
Monthly BM premium	0.0064	0.0492	0.7721	0.1308
Monthly EP premium	0.0068	0.0298	0.2497	0.1464
Monthly SP premium	0.0069	0.0385	0.8618	0.1818
Monthly PLS value premium	0.0085	0.0468	0.8706	0.18232

Note: This table reports the summary statistics of four monthly value premiums—BM, EP, SP, and PLS—over the sample period from January 2012 to December 2021. Reported statistics include the mean, standard deviation, skewness, and Sharpe ratio for each premium. The risk-free rate in China is assumed to be 3% on an annual basis when computing the Sharpe ratios.

Table 13: Differences between value premium in training and hold back periods by size decile

Size	1 (Small)	2	3	4	5	6	7	8	9	10 (Large)	average
Monthly BM premium	-0.55%	-0.94%	-0.06%	-0.80%	-0.66%	-0.37%	-1.39%	-0.47%	-1.27%	-1.39%	-0.79%
Annual BM premium	0.33%	-5.69%	-1.07%	0.93%	1.92%	-2.55%	-4.39%	-10.01%	-11.95%	-3.28%	-3.58%
Five years BM premium	49.69%	0.33%	25.94%	6.33%	-13.51%	6.71%	4.59%	10.94%	3.64%	-10.22%	8.44%
Monthly EP premium	-0.31%	0.45%	0.50%	0.05%	0.33%	-0.97%	-0.09%	0.45%	-0.68%	-0.82%	-0.11%
Annual EP premium	-8.14%	-0.68%	-6.40%	0.85%	1.74%	-5.11%	1.58%	-0.38%	-23.60%	-13.52%	-5.37%
Five years EP premium	37.32%	-13.61%	-12.74%	-17.03%	-26.47%	8.06%	-2.88%	-20.22%	-63.76%	3.67%	-10.77%
Monthly SP premium	-0.54%	-1.09%	0.31%	-0.04%	-0.42%	0.04%	-0.27%	-1.69%	-1.69%	-0.80%	-0.62%
Annual SP premium	2.41%	-1.98%	3.20%	1.81%	-0.87%	-1.99%	1.38%	-14.62%	-20.85%	-10.25%	-4.18%
Five years SP premium	-35.93%	1.79%	-5.78%	-0.50%	-24.28%	-19.99%	-4.51%	-4.20%	-46.62%	-24.07%	-16.41%

Note: The table shows the difference between BM, EP, and SP value premium in training and hold-back periods. The securities are first assigned to ten deciles by size. Then, in each size, we obtained the results using the premium calculated by BM, EP and SP in the hold back period (January 2012 to January 2021) minus the premium calculated by BM, EP and SP in the training period (January 2000 to December 2010).

Table 14: Performance Comparison Between ANN and PLS Value Strategies

	size	1	2	3	4	5	6	7	8	9	10	average
PLS in the training period	premium (10-1)	0.32%	1.80%	1.20%	1.34%	2.06%	1.19%	1.58%	1.34%	2.13%	1.72%	1.47%
	std premium	7.08%	4.56%	5.08%	5.07%	4.85%	5.44%	5.00%	6.30%	6.26%	7.36%	3.41%
	Sharpe ratio	1.08%	34.19%	18.79%	21.53%	37.38%	17.37%	26.67%	17.30%	30.09%	20.07%	35.81%
ANN in the training period	premium (10-1)	0.21%	1.15%	1.20%	1.41%	2.43%	1.08%	1.46%	1.33%	1.82%	2.02%	1.41%
	std premium	5.77%	4.99%	4.90%	5.04%	5.10%	5.24%	4.94%	6.19%	6.08%	6.36%	3.27%
	Sharpe ratio	-0.64%	18.04%	19.49%	22.98%	42.80%	15.82%	24.49%	17.44%	25.98%	27.91%	35.56%
PLS in the test period	premium (10-1)	0.06%	0.48%	1.47%	1.37%	1.36%	1.30%	0.97%	1.09%	0.37%	0.50%	0.90%
	std premium	4.96%	3.95%	5.32%	6.02%	5.64%	5.73%	6.36%	6.52%	8.05%	8.89%	4.70%
	Sharpe ratio	-3.67%	5.95%	23.02%	18.58%	19.79%	18.35%	11.32%	13.01%	1.48%	2.90%	13.86%
ANN in the test period	premium (10-1)	0.01%	0.59%	1.00%	1.63%	1.39%	1.39%	0.91%	0.86%	-0.25%	0.23%	0.78%
	std premium	4.02%	3.60%	5.24%	5.68%	5.80%	5.10%	5.71%	6.26%	7.16%	5.41%	3.90%
	Sharpe ratio	-5.92%	9.46%	14.37%	24.32%	19.65%	22.48%	11.64%	9.82%	-6.98%	-0.31%	13.55%

Note: The table compares the performance of ANN and PLS in forming value portfolios. In each month, stocks are first divided into ten size deciles and then into ten value deciles using scores generated by ANN and PLS, respectively. The value premium is calculated as the average one-month-ahead return of stocks in the highest-value decile minus that of the lowest-value decile. The training period runs from January 2000 to December 2010, and the test period runs from January 2012 to December 2021.

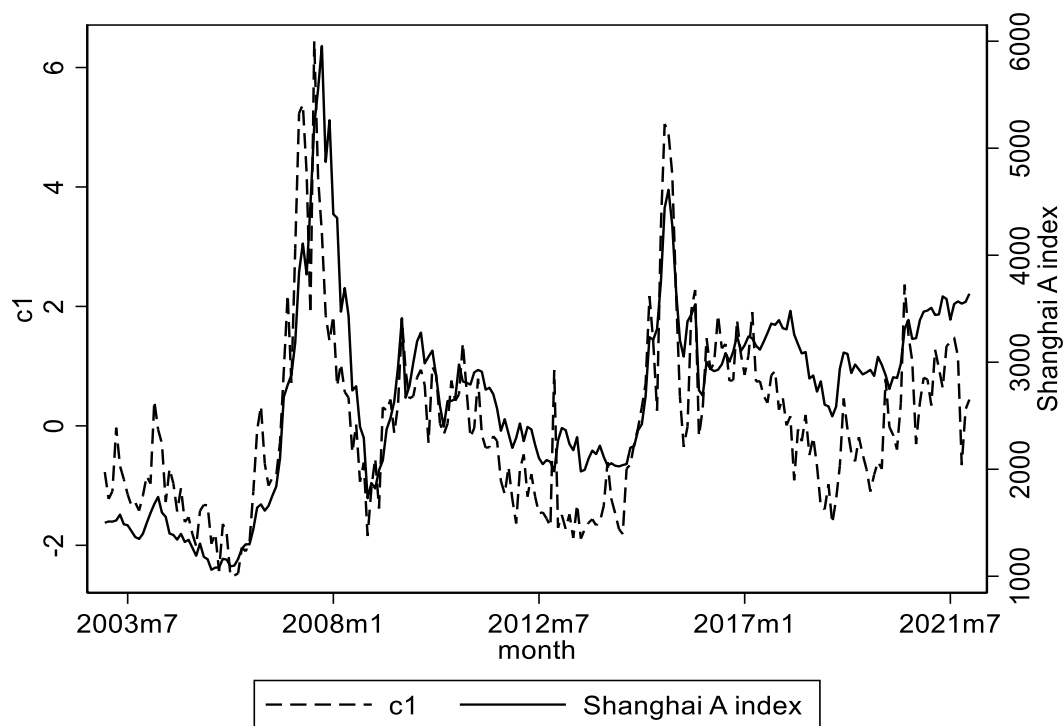
Second, following Baker and Wurgler (2006), we briefly summarise the Chinese stock market bubbles between 2003 and 2021, as a rough knowledge of speculative episodes in China could help us make preliminary judgments about the accuracy of our sentiment proxies. Figure 3 shows an overview of the Shanghai A index and c1 from January 2003 to December 2021. China's stock market experienced two bull markets in 2007 and 2015. The 2007 bull market started with the stock market reform in 2005. The stock reform in 2005 made the entire stock market expect to move towards total circulation as many SOEs started to move towards having tradable shares. The signal of the end of the bull market was rising stamp duty.¹⁸ High investor sentiment kept stock prices high for three months before finally falling. Following a prolonged fall in the stock market, the government enacted a series of favourable policies contributing to the bull market in 2014.¹⁹ The bull market attracted many new investors until the bubble burst in June 2015. It can be seen that our sentiment index is closely associated with the Shanghai A index. It led it in the 2006-8 boom and crash, and also during the long bear market to 2013.

Insert Figure 3 around here

¹⁸ In May 2007, the Chinese Ministry of Finance announced that stamp duty would be raised to 0.3%.

¹⁹ The Third Plenary Session of the 18th Central Committee of the People's Republic of China in 2013 illustrated the promotion of the reform of the registration system for issuing shares and the promotion of equity financing through multiple channels. In addition, the People's Bank of China decided to lower the benchmark interest rates for RMB loans and deposits of financial institutions with effect from 22 November 2014. Benchmark interest rate for one-year loans to financial institutions reduced by 0.4 percentage points.

Figure 3: c1 and market index



Note: The figure shows an overview of c1 and the Shanghai A index from January 2003 to December 2021.

4.3.1. Empirical analysis of the investor sentiment index and the value premium

In this section, we examine the investor sentiment indicator's ability to predict movement in the Shanghai A index, value premiums (BM, EP, SP, PLS) in the next month and in five years' time (we also tested the relationship between one year ahead value premium and investor sentiment. The results (not reported here) were between the one month and five-year results). In line with Han and Li (2017), we use predictive regression with a single predictor variable to examine the relationship between value premium and investor sentiment. If we use only the value factor to distinguish between value or growth stocks, the selected value (growth) stocks are mainly large-size (small-size) stocks in the Chinese stock market. To avoid the effect of the market capitalization factor, we use size-controlled value premiums, which are calculated by the average value premium in each decile size in each month. Table 15 reports the estimation results of the regression with a single regressor.

Market index return is negatively correlated with investor sentiment in the Chinese stock market, which is similar to developed countries (Baker & Stein, 2004; Brown & Cliff, 2005; Schmeling, 2009; Baker et al., 2012; Huang et al., 2015). Five-year ahead return is significantly correlated with the current $c1$, with a coefficient of -0.228. When sentiment is high at the beginning, the return of the market index would be low, representing the market correcting the mispricing bubble (Baker & Wurgler, 2006).

The sentiment indicator can predict the value premium in the short and long term, but the result is the opposite of Baker and Wurgler's (2006) experience in the US market. However, it is consistent with the conclusions of Kumar and Lee (2006). In the short term, the high sentiment indicator represents a positive return to the value premium. In the long term, we find significantly negative regression coefficients for the five-year value premium. In times of high (low) market sentiment, the value of the stock market is higher (lower), and it has a lower (higher) expected return in the long term with high significance. The regression coefficients for SP and BM are statistically significant at the 1% and 5% level, respectively. One possible explanation is that small-cap and low-priced stocks serve as the natural habitat of retail investors. Since investor sentiment in China largely reflects retail sentiment, periods of elevated sentiment are accompanied by a surge of retail investors entering the stock market, who tend to favour value stocks. Another possible explanation is that value stocks are more likely to be undervalued, and value investors typically maintain longer holding horizons than growth investors. Growth investments are more prone to bubble-like behaviour, where investors may rapidly sell once quarterly earnings fall short of expectations or following sharp

price increases. By contrast, expectations for value stocks are generally lower; even after price appreciation, value investors may still perceive them as undervalued relative to their benchmarks, thereby continuing to hold.

Due to short-selling restrictions, China's stock market takes longer to retrace after a bull market, when investors can ignore value stocks for a long time. In the long term, we find significantly negative regression coefficients for the five-year value premium. In times of high (low) market sentiment, the value of the stock market is higher (lower), and it has expected a lower (higher) return in the long term with high significance. In the short term, the high sentiment indicator represents a positive return to the value premium. The regression coefficient for SP is significant at the 1% level and for BM at 5%.

Insert Table 15 around here

Table 15: Single-factor Predictive Regression

		one month ahead return	one year ahead return	Five years ahead return
Index return	Coefficient	-0.0004	-0.115***	-0.228***
	Std. err.	(0.910)	(0.000)	(0.000)
	constant	0.0063	0.121***	0.332***
	Std. err.	(0.201)	(0.000)	(0.000)
	adj. R ²	-0.004	0.133	0.451
BM premium	Coefficient	0.0039*	0.0006	-0.0571**
	Std. err.	(0.044)	(0.932)	(0.007)
	constant	0.0097***	0.077***	0.227***
	Std. err.	(0.001)	(0.000)	(0.000)
	adj. R ²	0.014	-0.005	0.038
EP premium	Coefficient	0.0022	-0.0004	-0.0643**
	Std. err.	(0.153)	(0.954)	(0.005)
	constant	0.0086***	0.0668***	0.269***
	Std. err.	(0.000)	(0.000)	(0.000)
	adj. R ²	0.005	-0.005	0.042
SP premium	Coefficient	0.0046**	-0.0001	-0.0794***
	Std. err.	(0.002)	(0.982)	(0.000)
	constant	0.0099***	0.0884***	0.326***
	Std. err.	(0.000)	(0.000)	(0.000)
	adj. R ²	0.039	-0.005	0.072
PLS Value premium	Coefficient	0.0032	-0.0057	-0.0822***
	Std. err.	(0.079)	(0.376)	(0.001)
	constant	0.0116***	0.0951***	0.307***
	Std. err.	(0.000)	(0.000)	(0.000)
	adj. R ²	0.009	-0.001	0.062

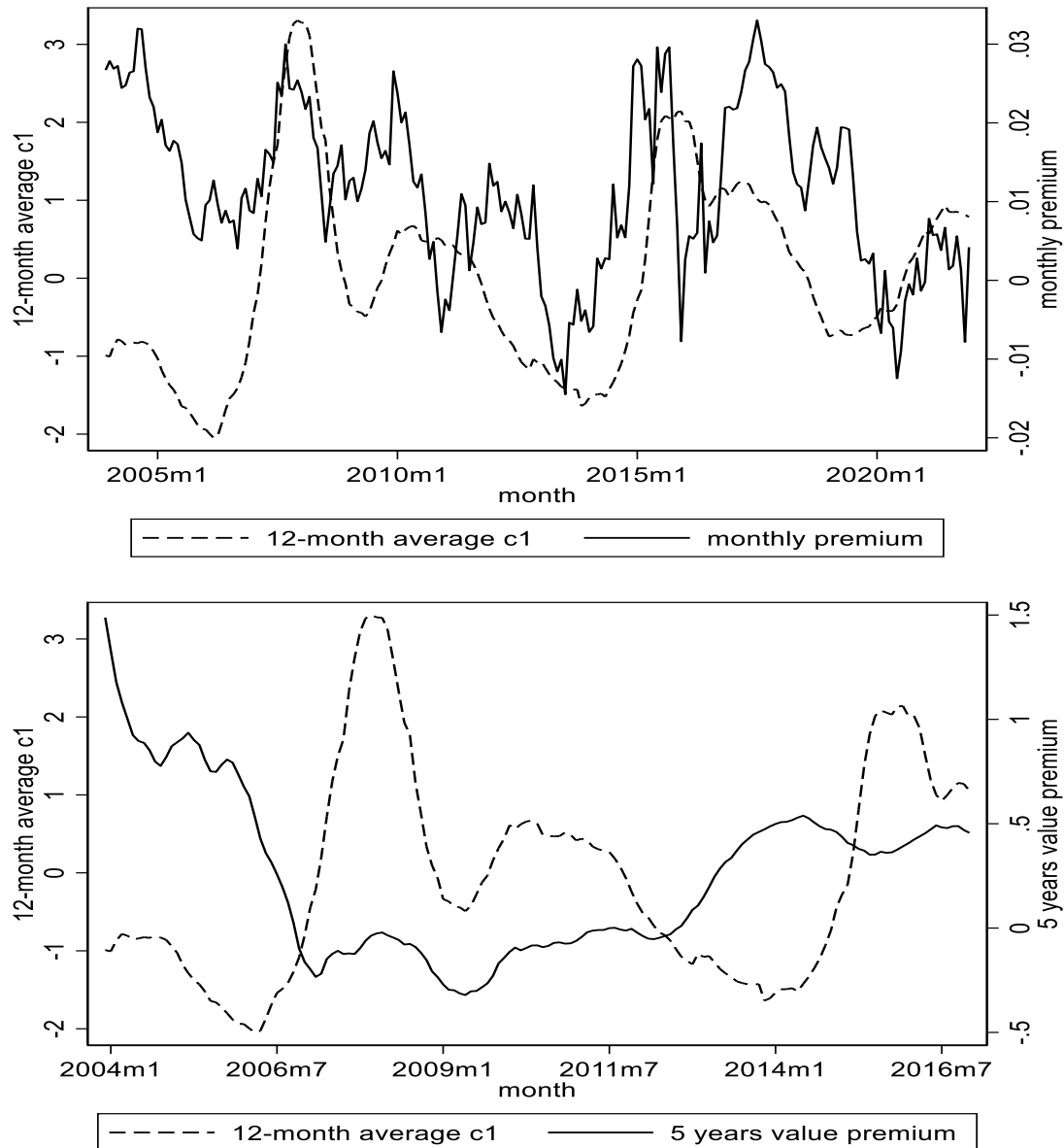
Note: The table presents the coefficient, standard errors, and adjusted R² of BM, EP, SP, and PLS value premiums for the future one-month, one-year and five-year periods, obtained through a simple ordinary least squares (OLS) linear regression with c1 as the independent variable. Future returns are regressed separately on each explanatory variable. Standard errors are conventional OLS standard errors. No fixed effects or clustering procedures are employed. “***”, “**”, and “*” denote statistical significance at the 0.1%, 1% and 5% levels, respectively (two-tailed P-value).

Figure 4 illustrates a twelve-month moving average of value premiums, which is calculated by hedging the returns of one month growth and value investment portfolios composed of stocks ranked by PLS value factor, and twelve months moving average c1. It confirms our results stated above. Two bull markets in China have been accompanied by the abnormal performance of value stocks. Investor sentiment is positively correlated with one month ahead value premium. Extremely high (low) market sentiment can lead to a positive (negative) value premium. In early 2006 and 2014, market sentiment hit the bottom and heralded the start of a bull market, with a value premium over the next year. During 2006-2007 and 2014-2015, market sentiment improved, and value stocks outperformed in a bull market. Despite the relationship, the PLS value premium in the Chinese equity market is positive most of the time. In the long term, the value premium negatively correlates with market sentiment. Holding value stocks for the long term when sentiment is low will give investors very high returns in the Chinese stock market. Conversely, when investor sentiment is high, it is high-risk to hold a long-term value position and sell growth stocks. Investor sentiment rose to a peak in 2007 and declined over the next five years, falling to very low levels in 2012. The value premium in 2007 for the subsequent five years was minimal.

Insert Figure 4 around here

The foregoing is due to a combination of the lack of a short-selling mechanism and changes in government policy. In the beginning, there are overvalued growth stocks and reasonably priced or relatively undervalued value stocks in the Chinese stock market. As Miller (1977) states, "A market with a large number of well-informed investors may not have any grossly undervalued securities, but if those investors are unwilling to sell short (as they often are) their presence is consistent with a few investments being overvalued impediments to short selling play a significant role in limiting the ability of rational traders to exploit over-pricing." Then, favourable information, such as the reduction of stamp duty by the government, is reported. It benefits all listed stocks, and a bull market emerges. Relatively undervalued value stocks are bought up, market sentiment rises, and a bubble emerges. When the bubble bursts, because of the lack of a short-selling mechanism the market needs more time to adjust overpriced stocks. This also confirms a phenomenon where, unlike markets in developed countries, Chinese stock markets have short bull and long bear markets. At this point, almost all stocks are over-priced, and value stocks enter a prolonged downturn. Eventually the market returns to the initial stage and the cycle repeats itself.

Figure 4: c1 and value premiums



Note: The figure illustrates a twelve-month moving average of value premiums and investor sentiments, denoted as c1. The monthly premiums and five-year value premiums are computed by hedging the returns of one-month and five-year growth and value investment portfolios, which consist of stocks ranked by the PLS value factor. The securities are initially categorized into ten deciles based on size. Within each size category, the securities are further assigned to ten deciles based on the PLS value factor. The value premiums are determined by the equal-weighted return of the top 10% PLS factor stocks minus the equal-weighted return of the bottom 10% PLS factor stocks. The holding periods are one month and five years, respectively. The sample period spans from January 2003 to January 2021. The 12-month average c1 represents the average c1 over the past 12 months. The data frequency is monthly.

4.3.2. Empirical analysis of the calendar effect on investor sentiment and the value premium

Table 16 provides evidence on the calendar effects in the Chinese stock market by examining the monthly performance of value premiums and their relationship with investor sentiment (denoted as $c1$) from January 2000 to December 2021. The table reports the average monthly value premiums based on four commonly used valuation measures: EP (earnings-to-price), BM (book-to-market), SP (sales-to-price), and the PLS value factor. For each month (from January to December), the premiums are computed by first sorting stocks into ten deciles based on market capitalization, then calculating the return spread between the highest and lowest value portfolios within each size decile, and finally averaging the ten spreads across all size groups. The "Average Value Premium" row presents the mean of the four individual factor premiums for each calendar month, serving as a general measure of the strength of value investing performance throughout the year. The last row ($c1$) reflects the corresponding average investor sentiment in each month, based on a standardized sentiment index.

The results suggest that value premiums exhibit significant calendar variation. Notably, November and December show the highest average value premiums (2.30% and 2.10%, respectively). March and April also coincide with the highest investor sentiment levels in the year, recorded at 0.34 and 0.35 respectively, reflecting increased market attention and trading activity during these months. These seasonal patterns in the Chinese stock market may reflect institutional and cultural influences on investor sentiment. The Lunar New Year typically falls between January 21 and February 20, suggesting that the post-holiday rebound in investor sentiment occurs in February or March. Additionally, the Chinese accounting year runs from January 1 to December 31, and the Shanghai Stock Exchange mandates that all listed firms must release their annual reports by the end of April. This regulatory requirement contributes to a surge in information flow during March and April, leading to increased investor attention and elevated sentiment. Consequently, March and April are often characterized by heightened investor optimism and trading activity in the Chinese equity market.

For the value premium, Haug and Hirschey (2006) document that it tends to be significantly larger in January in the U.S. equity market. In the context of China, Liang et al. (2022) find that the January effect is associated with the Lunar New Year rather than the Gregorian calendar year. If a January effect in value premiums were to exist in

China, it would likely manifest in either January (Gregorian) or February–March (Lunar). However, our findings show that value premiums are concentrated in November and December, inconsistent with the expected timing of a January effect. Therefore, we do not observe evidence of a January effect in value premiums in the Chinese equity market.

Insert Table 16 around here

Table 17 presents the average monthly returns of portfolios sorted by firm size across calendar months from January 2000 to December 2021. At the end of each calendar month, securities are ranked into ten deciles based on market capitalisation, and equal-weighted portfolios are subsequently formed. The reported figures represent the average return for each size decile (rows 1–10) during the corresponding calendar month (columns 1–12), averaged across all years in the sample period. The results reveal a clear calendar effect resembling the traditional January effect, albeit with a temporal shift to February and March, coinciding with the timing of the Lunar New Year in China, which typically occurs between late January and mid-February. The return spikes in these months are particularly evident among small-cap stocks, with the smallest decile (Decile 1) generating 7.45% in February and 6.79% in March, far exceeding their typical performance in other months. This pattern is consistent with the findings of Liang et al. (2022), who demonstrate that the January effect in China is driven by the Lunar New Year rather than the Gregorian calendar. The festive period tends to boost investor sentiment and trading activity, particularly among retail participants. Consequently, the heightened optimism in February and March appears to disproportionately benefit smaller firms, likely due to their greater sensitivity to sentiment-driven trading.

Insert Table 17 around here

Table 16: The Impact of Calendar Effects on Investor Sentiment and the Value Premium

Month	1	2	3	4	5	6	7	8	9	10	11	12
Monthly EP premium	-1.67%	0.27%	0.73%	-1.29%	2.54%	1.45%	-1.13%	0.79%	1.68%	0.46%	3.14%	2.10%
Monthly BM premium	1.18%	0.72%	1.26%	-0.53%	0.72%	2.15%	0.34%	0.90%	0.40%	1.96%	1.88%	1.75%
Monthly SP premium	1.36%	-0.39%	1.53%	-0.09%	0.79%	2.04%	0.58%	0.99%	0.60%	0.94%	1.76%	2.03%
Monthly PLS premium	0.73%	0.14%	1.62%	-0.72%	1.67%	1.98%	-0.04%	1.29%	0.97%	1.64%	2.44%	2.53%
Average Value premium	0.40%	0.18%	1.28%	-0.66%	1.43%	1.90%	-0.06%	0.99%	0.91%	1.25%	2.30%	2.10%
cl	-0.28	-0.28	0.34	0.35	0.22	0.21	0.10	0.14	-0.17	-0.56	0.00	-0.10

Note: The table compares the differences in value factor premiums across calendar months from January 2000 to December 2021. Securities are first sorted into ten deciles based on market capitalization. Within each size decile, value premiums are computed using four valuation measures: BM (book-to-market), EP (earnings-to-price), SP (sales-to-price), and the PLS value factor. Within each of the ten market capitalization deciles, we calculate the return spread between the top (most undervalued) and bottom (most overvalued) value portfolios, and then take the average of these ten spreads to obtain the overall value premium. The reported values represent the average monthly premiums across size deciles. Each column presents the average value premium and the corresponding investor sentiment level for a given calendar month. Note that (for example) the value "1" refers to portfolios formed at the end of January, reflecting stock performance throughout February.

Table 17: The Impact of Calendar Effects on stocks in different sizes

Month	1	2	3	4	5	6	7	8	9	10	11	12
1	7.45%	6.79%	1.00%	3.75%	-1.24%	2.60%	3.03%	1.08%	-0.16%	4.65%	1.38%	-1.15%
2	7.05%	5.49%	0.33%	3.16%	-2.10%	1.70%	1.45%	-0.06%	-0.80%	4.06%	0.93%	-1.02%
3	6.45%	5.06%	0.15%	2.93%	-2.86%	1.72%	0.95%	-0.23%	-0.77%	3.75%	0.52%	-1.09%
4	6.49%	4.25%	0.54%	2.32%	-2.96%	1.65%	0.70%	-0.42%	-0.88%	3.07%	0.79%	-0.96%
5	6.30%	3.75%	0.28%	2.59%	-3.12%	1.57%	0.34%	-1.09%	-0.84%	3.13%	0.09%	-1.03%
6	6.14%	3.43%	0.67%	2.09%	-2.79%	1.65%	-0.12%	-0.99%	-1.23%	3.11%	0.49%	-0.42%
7	5.91%	2.82%	0.79%	2.18%	-2.82%	1.79%	-0.11%	-0.91%	-1.22%	2.52%	0.58%	-0.71%
8	5.50%	2.36%	1.36%	1.54%	-2.56%	2.25%	-0.52%	-0.86%	-0.70%	1.94%	1.13%	-0.35%
9	4.79%	1.85%	1.61%	1.64%	-2.19%	2.04%	-0.66%	-0.26%	-0.80%	1.51%	1.41%	-0.29%
10	3.21%	0.51%	2.39%	0.25%	-1.17%	1.96%	-1.31%	-0.22%	-0.05%	1.07%	2.88%	0.03%
Average	5.93%	3.63%	0.91%	2.24%	-2.38%	1.89%	0.37%	-0.40%	-0.75%	2.88%	1.02%	-0.70%

Note: The table compares the returns of portfolios sorted by firm size across calendar months from January 2000 to December 2021. At the end of each month, securities are sorted into ten deciles based on market capitalization, and equal-weighted portfolios are constructed accordingly. The monthly return is then computed for each size portfolio. Each column presents the average return of the corresponding calendar month across all sample years. Note that (for example) the value "1" refers to portfolios formed at the end of January, reflecting stock performance throughout February.

5. Robustness tests

5.1. Drop the smallest 30% of stocks

At the smallest market capitalisations, reverse takeovers lead to extremely high future earnings, so extremely high future earnings come from the least valuable stocks. Some of these companies are on the verge of bankruptcy and more susceptible to a reverse takeover. In order to avoid the impact of the reverse mergers on our results, following Liu et al. (2019), we drop the smallest 30% of stocks in our sample and repeat the previous process. Appendix III compares the difference between premium of the PLS value factor and other single factor value factors from January 2012 to December 2021 after we drop the smallest 30% of stocks. In the hold-back period, we compare the PLS value premium with the EP, BM and SP premiums in the short, medium and long term. It is still on average higher than other premiums. In addition, the PLS combined value factor outperforms other single-factor models in monthly and annual terms for all five smallest deciles, in which quantitative investors are more interested.

5.2. Covid-19

Appendix IV reports the value premiums associated with EP, BM, SP, and the PLS value factor before and after the covid-19 outbreak. The sample period is divided into a pre-covid window (January 2000 to November 2019) and a post-covid window (December 2019 to December 2021). Securities are first sorted into ten deciles based on market capitalisation. Within each size decile, value premiums are calculated for each factor.

Before the onset of covid-19, the PLS value factor delivers the highest average monthly premium of 1.29 per cent, followed by BM at 1.19 per cent, SP at 1.04 per cent, and EP at 0.87 per cent. This pattern indicates that, under normal market conditions, the PLS factor consistently outperforms the individual valuation metrics by capturing a broader and more informative cross-section of firm fundamentals.

After the outbreak of covid-19, however, the performance of traditional value factors deteriorates notably. EP and BM record negative premiums of -0.47 per cent and -0.21 per cent, respectively. Although the PLS factor remains positive, its premium declines substantially to 0.14 per cent. In contrast, SP is the only factor that sustains a strong positive premium, at 0.68 per cent. These results suggest that while the PLS value factor performs best on average across varying market conditions, in unusual periods when a specific valuation metric aligns more closely with prevailing investor behaviour or market conditions—such as the strong post-

covid performance of SP—strategies based on that single factor may temporarily yield higher returns than the PLS-based approach.

5.3. Applying the PLS methodology to the US market

We also conducted robustness tests on the U.S. stock market. Consistent with Fama and French (2015), we calculated the monthly returns of the top 20% of value stocks minus the monthly returns of the bottom 20% of value stocks. The results indicate that, in the U.S. stock market, the PLS value factor outperforms the SP factor, which in turn outperforms the BM and EP factors. (See Appendix V for detailed results.) It is similar to our monthly results in China, where the PLS premium in China outperforms PS, EP and BM. The distinction is that in China the monthly value premium of PLS far surpasses that of the second-ranked PS. In contrast, in the United States, the value premium of PLS is only slightly higher than that of the second-ranked PS.

Moreover, the Sharpe ratio of the equally-weighted portfolio constructed from the PLS value factor (7.75%) is lower than that of SP (9.09%). This difference primarily results from the lower volatility of SP premiums among large-cap stocks (average standard deviation 7.75% for SP vs. 9.47% for PLS), which reduces the denominator in the Sharpe ratio and yields a higher risk-adjusted return for SP despite its smaller raw premium.

5.4. Robustness analysis of the investor sentiment index

Appendix VI presents the results of univariate regressions in which monthly value premiums and investor sentiment are regressed on a survey-based Investor Confidence Index. The dependent variables include one-month-ahead and five-years-ahead value premiums based on EP, BM, SP, and the PLS value factor, as well as our constructed investor sentiment index. The data span from January 2003 to December 2021, with the Investor Confidence Index obtained from the CSMAR database at each month-end. For the five-year-ahead premiums, the sample ends in January 2017, allowing a full forward-looking horizon.

The Consumer Confidence Index is positively correlated with our Investor Sentiment Index, providing external validation for the reliability of our sentiment proxy. The results reveal that the one-month-ahead premiums of all four value factors are negatively related to investor confidence, though the relationships are not statistically significant. In contrast, all four five-years-ahead value premiums show strong and statistically significant negative associations with the Investor Confidence Index. Specifically, the EP, BM, SP, and PLS-based value factors

have coefficients of -0.0159 , -0.0132 , -0.0123 , and -0.0155 , respectively, all significant at the 0.1% level. These findings suggest that when investor sentiment is high, value stocks also tend to be relatively highly priced.

5.5. Robustness analysis across firm size subsamples

Appendix VII reports the subsample analysis by firm size. Overall, the results are consistent with the main findings and indicate that the predictive relationship between investor sentiment and future value premiums remains robust after controlling for firm size. Specifically, investor sentiment continues to exhibit a significantly negative association with long-horizon value premiums, while its relationship with one-month-ahead value premiums remains weak. The predictive effect is more pronounced for the large-firm subsample, particularly for the EP, SP and PLS value premiums, whereas the corresponding coefficients for the small-firm subsample are generally weaker and often statistically insignificant. These findings suggest that the documented relationship is not driven by firm size and remains robust across size-based subsamples.

5.6. Sensitivity Analysis

In the construction of the sentiment index, PCA is applied to a set of selected proxies, with the dividend premium excluded on the grounds that periods of rising investor sentiment in China are often characterised by capital flowing into value stocks rather than growth stocks. While this exclusion is theoretically justifiable, the original framework does not provide a sensitivity analysis to verify whether including or excluding specific proxies would materially affect the resulting components. To address this concern, this study incorporates the dividend premium into the PCA and examines its effect on the component structure. Table 18 reports the loading weights of each variable on the first two components, c_1 and c_2 , as derived from the principal component analysis. After including the dividend premium, the loading weights of the original five variables on c_1 remain largely unchanged. This indicates that the structure of c_1 is robust to the introduction of the dividend premium. Meanwhile, the dividend premium itself has a very small loading on c_1 (around -0.13), suggesting that it provides limited incremental explanatory power and plays only a marginal role in shaping the primary component.

Insert Table 18 around here

From Table 19, we observe that the inclusion of the dividend premium does not materially alter the predictive performance of the principal components. This is consistent with the very low loading weight of the dividend premium in the PCA results, indicating that it contributes only marginal information to the factor structure. Consequently, the regression

outcomes remain broadly stable across horizons. The only noticeable change is a relative small reduction in the long-run predictive significance of the BM and EP premiums, where the five-year coefficients change from 5% to 1% significance. This suggests that, while dividend-related information is not entirely irrelevant, its incremental explanatory power is limited and does not meaningfully reshape the overall factor's forecasting ability.

Insert Table 19 around here

Table 18: Component Loadings and Regression Results for c1 and c2

Variable	c1	c2	regression	
NIA	0.5886	0.5744	Coefficient	0.990***
Dividend payout ratio	-0.3918	-0.4203	Std. err.	(0.000)
NIPO	0.2968	0.2929	constant	7.51e-10
RIPO	0.4272	0.4197	Std. err.	(1.000)
Turnover ratio	0.4791	0.4619	adj. R ²	0.990
Dividend premium		-0.1349		

Note: This table reports the loading weights of each variable on the first two components, c1 and c2, as derived from the principal component analysis. The table also presents the regression results obtained by regressing c1 on c2, including the estimated coefficient, standard error, and constant term. The adjusted R² indicates the explanatory power of c2 for c1. “***”, “**”, and “*” denote statistical significance at the 0.1%, 1% and 5% levels respectively (two-tailed P-value).

Table 19: Single-factor Predictive Regression

		one month ahead return	five years ahead return
Index return	Coefficient	-0.0006	-0.231***
	Std. err.	(0.846)	(0.000)
	constant	0.0063	0.338***
	Std. err.	(0.201)	(0.000)
	adj. R ²	-0.004	0.470
BM premium	Coefficient	0.0038*	-0.0476*
	Std. err.	(0.045)	(0.024)
	constant	0.0097***	0.229***
	Std. err.	(0.001)	(0.000)
	adj. R ²	0.013	0.025
EP premium	Coefficient	0.0022	-0.0543*
	Std. err.	(0.134)	(0.016)
	constant	0.0086***	0.271***
	Std. err.	(0.000)	(0.000)
	adj. R ²	0.006	0.028
SP premium	Coefficient	0.0044**	-0.0801***
	Std. err.	(0.002)	(0.000)
	constant	0.0099***	0.328***
	Std. err.	(0.000)	(0.000)
	adj. R ²	0.036	0.074
PLS Value premium	Coefficient	0.0031	-0.0748**
	Std. err.	(0.080)	(0.002)
	constant	0.0116***	0.310***
	Std. err.	(0.000)	(0.000)
	adj. R ²	0.009	0.051

Note: The table presents the coefficient, standard error of coefficient, constant, standard error of constant, and adjusted R² of BM, EP, SP, and PLS value premiums for the future one month, one-year and five-year periods, obtained through a simple linear regression with c1 as the independent variable. “***”, “**”, and “*” denote statistical significance at the 0.1%, 1% and 5% levels respectively (two-tailed P-value).

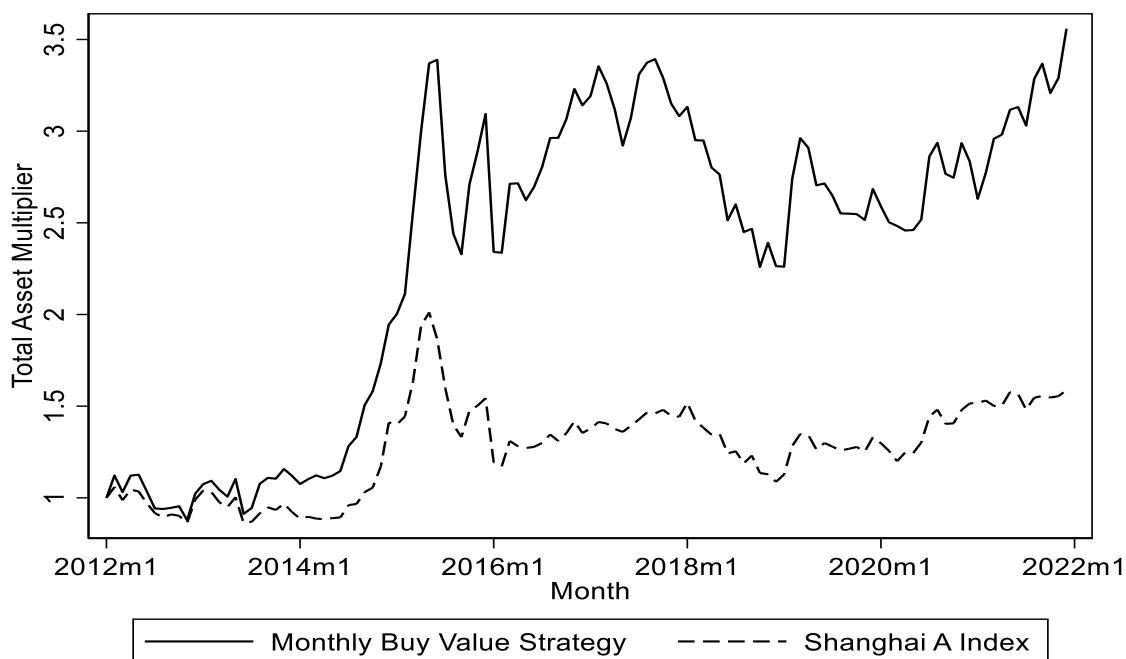
6. Illustrative investment strategy

As an illustration we offer a new PLS value investment strategy in China. Each month, the investor divides the stocks into deciles by market capitalization, then calculates the PLS value of each stock at each market capitalization according to the weighting calculated from January 2000 to December 2010. He or she then selects the most valuable 10% of stocks at each market capitalization to build an equal weighted portfolio, holds it for one month and sells it, then repeats the previous investment steps.

In line with Leippold et al. (2022), we consider commission (less than 3 bps), stamp tax (10 bps), and slippage (around 10 bps) as transaction fees. We consider trading costs of 30 bps for the effect of transaction costs on portfolio performance. Figure 5 shows what could be achieved by an investor who followed our new strategy from January 2012 to December 2021, along with the Shanghai A market index as a benchmark. If we consider transaction fees, the assets have grown by 2.56 times with an annualised return of 13.54%. The assets have grown by 4.07 times with an annualised return of 17.62% if we ignore transaction costs.

Insert Figure 5 around here

Figure 5: Investment Strategy



Note: Each month, the investor divides the stocks into deciles by market capitalization, then calculates the PLS value of each stock at each market capitalization. He or she then selects the most valuable 10% of stocks at each

market capitalization to build an equally weighted portfolio, holds it for one month and sells it, then repeats the previous investment steps. A fee of 30 bps is charged for each transaction.

7. Conclusion

We examine the performance of the traditional value factors (EP, BM and SP) in China, after controlling for the size effect. Due to multicollinearity issues between the value indicators, we use partial least squares, and we create a novel PLS value factor which outperforms traditional value metrics in terms of portfolio returns, for one-month, one-year and five-year holding periods. Following Baker and Wurgler (2006), we also construct a new market sentiment indicator.

We find that value stocks in the Chinese stock market perform better when market sentiment is high. Large stocks are generally safer than small stocks in developed markets (Chan & Chen, 1991). However, in our sample period, small stocks in China are likely to be safer because they have a shell value as a floor to their value. Therefore, different from many developed countries, the average value ratio (EP, BM and SP) of a stock, increases as market size increases in China. Thus, we suggest that the size of stocks should be considered when investors apply valuation ratios to judge the value of stocks. In line with developed markets (Chan et al., 1991; Jaffe et al., 1989; Fama & French, 1992), EP can better identify large stocks than small-cap stocks in China. However, we find some very high returns in the smallest stocks. This U-shaped EP value premium return mainly comes from reverse takeovers. In most of our sample period, investing in potential shell stocks could bring extremely high returns. However, investing in shell stocks cannot be recommended in the future, due to the reform of the company registration system in 2019. Our PLS combined value factor outperforms traditional value factor models in terms of returns when using short, medium and long term long-short portfolios.

This study sheds new light on Han & Li's (2017) statement that investor sentiment is a "size story" in China. Controlling for the size effect, the EP, BM, SP and the combined value factor premiums are all significantly correlated with market sentiment. In the short term, the value premium is positively correlated with positive investor sentiment, however there is a negative correlation between market sentiment and value premiums in the long term in China. This is contrary to the evidence in the US (Baker & Wurgler, 2006) and other developed countries (Baker et al., 2012). We explain this by proposing that most stocks are overvalued due to the market's lack of a short-selling mechanism. When the market falls, it takes longer to adjust to market bubbles. In a bear market, value stocks are unpopular because of their

chronically depressed prices. They only rise in conjunction with favourable government policies when investors become aware of undervalued value stocks and value stocks rise with market sentiment.

In practical terms, our findings help investors exploit both short- and long-term strategies. For example, high sentiment may boost short-term returns on value stocks but reduce long-term gains. Investors can also align value signals with firm size (e.g., EP for large caps, BM for small caps). Fund managers may use machine learning (like PLS) to build better value factors and adjust portfolios based on sentiment. Policymakers can recognize China's unique investor behavior and design measures to stabilize sentiment, improve transparency, and promote financial literacy. In particular, we propose that policymakers can increase the number of IPO approvals during times of high sentiment to reduce the gap between offer prices and the secondary market, and also government pension schemes could adopt a counter-cyclical policy, stabilising the market during extremes of sentiment while earning arbitrage profits.

This article has the limitation that while the PLS value factor outperforms traditional valuation factors on average over the long term, its performance may be adversely affected under extreme market conditions—particularly when one underlying component substantially underperforms the others. For example, in the two years following the covid-19 outbreak, the Chinese equity market exhibited negative premiums for book-to-market (BM) and earnings-to-price (EP) ratios. The performance of the PLS factor can be adversely affected by the weaker signals from the underperforming variables. In addition, while the results provide consistent evidence of a strong association between investor sentiment and value premium dynamics, we acknowledge that this relationship is mainly based on correlational evidence derived from regression analysis.

Avenues for future work can be the PLS value factor examination in other developing markets. Furthermore, the PLS value factor could be applied to further improve popular multi-factor models (e.g. the three-factor model of Fama & French, 1992; or the five-factor model of Fama & French, 2015).

References

- Anderson, K., & Brooks, C. (2006). The long-term price-earnings ratio. *Journal of Business Finance & Accounting*, 33(7-8), 1063-1086.
- Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets*, 7(3), 271-299.
- Baker, M., & Wurgler, J. (2004). A catering theory of dividends. *Journal of Finance*, 59(3), 1125-1165.
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4), 1645-1680.
- Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, 104(2), 272-287.
- Bali, T. G., Engle, R. F., & Murray, S. (2016). Empirical asset pricing: The cross section of stock returns. *John Wiley & Sons*.
- Bandopadhyaya, A., & Jones, A. L. (2008). Measures of investor sentiment: A comparative analysis put-call ratio vs. volatility index. *Journal of Business & Economics Research*, 6(8).
- Barbee, W.C., Jeong, J.G., & Mukherji, S. (2008) Relations between portfolio returns and market multiples. *Global Finance Journal* 19: 1–10.
- Barbee, W. C., Mukherji, S., & Raines, G. A. (1996). Do sales–price and debt–equity explain stock returns better than book–market and firm size?. *Financial Analysts Journal*, 52(2), 56-60.
- Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, 75(2), 283-317.
- Basu, S. (1975). The information content of price-earnings ratios. *Financial Management*, 53-64.
- Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis. *Journal of Finance*, 32(3), 663-682.
- Benveniste, L. M., Ljungqvist, A., Wilhelm Jr, W. J., & Yu, X. (2003). Evidence of information spillovers in the production of investment banking services. *Journal of Finance*, 58(2), 577-608.
- Bhandari, L.C. (1988). Debt/equity ratio and expected common stock returns: empirical evidence. *Journal of Finance*, 43(2), 507-528.

- Bird, R., & Casavecchia, L. (2007). Value enhancement using momentum indicators: the European experience. *International Journal of Managerial Finance*, 3(3), 229-262.
- Birkinshaw, J., Morrison, A., & Hulland, J. (1995). Structural and competitive determinants of a global integration strategy. *Strategic Management Journal*, 16(8), 637-655.
- Breen, W. (1968) Low price-earnings ratios and industry relatives. *Financial Analyst Journal*, 24: 125–127.
- Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *Journal of Business*, 78(2), 405-440.
- Cai, C. X., & Keasey, K. (2022). The Experts and the Evidence: A Practical Guide to Stock Investing. *John Wiley & Sons*.
- Cakici, N., Chan, K., & Topyan, K. (2017). Cross-sectional stock return predictability in China. *The European Journal of Finance*, 23(7-9), 581-605.
- Chan, L., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and stock returns in Japan, *Journal of Finance*, 46, 1739-1764.
- Chan, K., Chan, L.K.C., Jegadeesh, N., & Lakonishok, J. (2006) Earnings quality and stock returns. *Journal of Business*, 79: 1041–1082.
- Chan, K. C., & Chen, N. F. (1991). Structural and return characteristics of small and large firms. *The Journal of Finance*, 46(4), 1467-1484.
- Chan, L. K. C., Jegadeesh, N., & Lakonishok, J. (1995). Evaluating the performance of values versus glamour stocks: the impact of selection bias. *Journal of Financial Economics*, 38(3), 269-296.
- Cheema, M. A., & Nartea, G.V. (2014). Momentum returns and information uncertainty: evidence from China, *Basific-Basin Finance Journal*, 30, 173-188.
- Chen, H., Chong, T. T. L., & She, Y. (2014). A principal component approach to measuring investor sentiment in China. *Quantitative Finance*, 14(4), 573-579.
- Chen, X., Kim, K. A., Yao, T., & Yu, T. (2010). On the predictability of Chinese stock returns. *Pacific-Basin Finance Journal*, 18(4), 403-425.
- Cheung, C., Hoguet, G., & Ng, S. (2015). Value, size, momentum, dividend yield, and volatility in China's A-share market. *Journal of Portfolio Management*, 41(5), 57-70.
- Chou, J., Das, P. K., & Rao, S. U. (2011). The value premium and the January effect. *Managerial Finance*, 37(6), 517-536.
- Cornell, B., Hsu, J., Kiefer, P., & Wool, P. (2020). Assessing mutual fund performance in China, *Journal of Portfolio Management*, 46(5), 118-127.

- Dhatt, M. S., Kim, Y. H., & Mukherji, S. (1999). The value premium for small-capitalization stocks. *Financial Analysts Journal*, 55(5), 60-68.
- Dhatt, M. S., Kim, Y.H., & Mukherji, S. (2004) Can composite value measures enhance portfolio performance? *Journal of Investing* 13, 42–48.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87.
- Du, D. (2011). Value premium and investor sentiment. *Journal of the Academy of Behavioral Finance*, 1(2), 87-101
- Fama, E., & French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, 46, 427-466.
- Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56.
- Fama, E. F., & French, K. R. (1998). Value versus growth: The international evidence. *Journal of Finance*, 53(6), 1975-1999.
- Fama, E. F., & French, K. R. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, 82(3), 491-518.
- Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, 63(4), 1653-1678.
- Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105(3), 457-472.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22.
- Fisher, F. M. (1984). The misuse of accounting rates of return: Reply. *American Economic Review*, 74(3), 509-517.
- Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS quarterly*, 389-400.
- Gharghori, P., Strykowski, S., & Veeraraghavan, M. (2013). Value versus growth: Australian evidence. *Accounting and Finance*, 53: 393–417.
- Graham, B., & Dodd, D. (1934). *Security Analysis*. New York: McGraw-Hill.
- Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average US monthly stock returns. *Review of Financial Studies*, 30(12), 4389-4436.

- Griffin, J.M. (2002). Are the Fama and French Factors Global or Country Specific? *Review of Financial Studies*, 15(3), 783–803.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223-2273.
- Han, X., & Li, Y. (2017). Can investor sentiment be a momentum time-series predictor? Evidence from China. *Journal of Empirical Finance*, 42, 212-239.
- Haug, M., & Hirschey, M. (2006). The January effect. *Financial Analysts Journal*, 62(5), 78-88.
- Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54(6), 2143-2184.
- Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3), 650-705.
- Hu, G. X., Chen, C., Shao, Y., & Wang, J. (2019). Fama–French in China: size and value factors in Chinese stock returns. *International Review of Finance*, 19(1), 3-44.
- Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies*, 28(3), 791-837.
- Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal*, 20(2), 195-204.
- Ibbotson, R. G., & Jaffe, J. F. (1975). “Hot issue” markets. *Journal of Finance*, 30(4), 1027-1042.
- Israel, R., & Moskowitz, T. J. (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics*, 108(2), 275-301.
- Jaffe, J., Keim, D. B., & Westerfield, R. (1989). Earnings yields, market values, and stock returns. *Journal of Finance*, 44(1), 135-148.
- Jiang, Z., Anderson, K. P., & Stafylas, D. (2024). Predicting Future Earnings Changes: a Modest Integration of Artificial Neural Networks and Financial Factors. *Available at SSRN 4730021*.
- Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1), 109-126.
- Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *Journal of Finance*, 68(5), 1721-1756.
- Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2), 294-316.

- Kumar, A., & Lee, C. M. (2006). Retail investor sentiment and return comovements. *Journal of Finance*, 61(5), 2451-2486.
- Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49(5), 1541-1578.
- Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, 145(2), 64-82.
- Leledakis, G., & Davidson, I. (2001). Are two factors enough? The UK evidence. *Financial Analysts Journal*, 57(6), 96-105.
- Li, X., & Zhang, B. (2008). Stock market behavior and investor sentiment: Evidence from China. *Frontiers of Business Research in China*, 2(2), 277-282.
- Liu, J., Stambaugh, R. F., & Yuan, Y. (2019). Size and value in China. *Journal of Financial Economics*, 134(1), 48-69.
- Liang, X., Liu, Q., & Zebedee, A. A. (2022). One Country, Two Calendars: Lunar January Effect in China's A-Share Stock Market. *Asia-Pacific Journal of Financial Studies*, 51(6), 859-895.
- Lowry, M., & Schwert, G. W. (2002). IPO market cycles: Bubbles or sequential learning?. *Journal of Finance*, 57(3), 1171-1200.
- McWilliams, J. D. (1966). Prices, Earnings and PE ratios. *Financial Analysts Journal*, 22(3), 137-142.
- Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *Journal of Finance*, 32(4), 1151-1168.
- Mselmi, N., Lahiani, A., & Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis*, 50, 67-80.
- Nicholson, S. F. (1960). Price-earnings ratios. *Financial Analysts Journal*, 16(4), 43-45.
- Nicholson, S. F. (1968). Price ratios in relation to investment results. *Financial Analysts Journal*, 24(1), 105-109.
- Pätäri, E., & Leivo, T. (2017). A closer look at value premium: Literature review and synthesis. *Journal of Economic Surveys*, 31(1), 79-168.
- People's Bank of China (2023). Financial Market Statistics. Available at <http://www.pbc.gov.cn/en/3688247/3688975/4505202/4505217/index.html>. Accessed 21st May 2025.

- Ramli, N. A., Latan, H., & Nartea, G. V. (2018). Why should PLS-SEM be used rather than regression? Evidence from the capital structure perspective. *Partial least squares structural equation modeling: Recent advances in banking and finance*, 171-209.
- Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of Marketing Research*, 41(3), 293-305.
- Ritter, J. R. (1991). The long-run performance of initial public offerings. *Journal of Finance*, 46(1), 3-27.
- Rouwenhorst, K. G. (1999). Local return factors and turnover in emerging stock markets. *Journal of Finance*, 54(4), 1439-1464.
- Schmeling, M. (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, 16(3), 394-408.
- Senchack, A.J., Jr., & Martin, J.D. (1987). The relative performance of the PSR and PER investment strategies. *Financial Analysts Journal* 43: 46–56.
- Shanghai Stock Exchange (2023). Overview. Available at <http://english.sse.com.cn/aboutsse/overview/>. Accessed 21st May 2025.
- Shanghai Stock Exchange Yearbook (2022). Statistical Yearbook. Available at <http://www.sse.com.cn/aboutus/publication/yearly/>. Accessed 21st May 2025.
- Shiller, R. J. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity* 2, 457–510.
- Shleifer, A. (2000). Inefficient markets: An introduction to behavioural finance. *Oup Oxford*.
- Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic Perspectives*, 4(2), 19-33.
- Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *Journal of Finance*, 52(2), 737-783.
- Singh, N., Fassott, G., Chao, M. C., & Hoffmann, J. A. (2006). Understanding international web site usage: A cross-national study of German, Brazilian, and Taiwanese online consumers. *International Marketing Review*, 23(1), 83-97.
- Stambaugh, R. F., & Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, 30(4), 1270-1315.
- Sun, C. (2024). Factor correlation and the cross section of asset returns: A correlation-robust machine learning approach. *Journal of Empirical Finance*, 77, 101497.

- U.S. Securities and Exchange Commission (2023). Overview. Available at <https://www.sec.gov/news/press-release/2023-15>. Accessed 9th September 2023.
- Van der Hart, J., Slagter, E., & Van Dijk, D. (2003). Stock selection strategies in emerging markets. *Journal of Empirical Finance*, 10(1-2), 105-132.
- Venkatesh, V., & Agarwal, R. (2006). Turning visitors into customers: A usability-centric perspective on purchase behavior in electronic channels. *Management Science*, 52(3), 367-382.
- Venturini, S., & Mehmetoglu, M. (2019). Plssem: a stata package for structural equation modelling with partial least squares. *Journal of Statistical Software*, 88, 1-35.
- Wold, H. (1975). Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability*, 12(S1), 117-142.
- Yen, J.Y., Sun, Q., & Yan, Y. (2004) Value versus growth stocks in Singapore. *Journal of Multinational Financial Management* 14: 19–34.
- Yi, M. Y., & Davis, F. D. (2003). Developing and validating an observational learning model of computer software training and skill acquisition. *Information Systems Research*, 14(2), 146-169.
- Zhu, B., & Niu, F. (2016). Investor sentiment, accounting information and stock price: Evidence from China. *Pacific-Basin Finance Journal*, 38, 125-134.

Appendix I: List of abbreviations

BM – Book to market ratio

CCI – Consumer confidential index

CFP – Price to cash flow ratio, which is calculated by the last four quarterly net cash flow divided by the current share price

CSMAR – China Stock Market & Accounting Research Database (One of the main Chinese databases)

CSRC – China Securities Regulatory Commission

EP – Earnings to price ratio, which is calculated by the last four quarterly earnings divided by the current share price.

NIA – New investor accounts

NIPO – The number of IPO

p^{D-ND} – Dividend premium

PLS – Partial Least Squares

RIPO – The return of average first-day IPO

RMB – Renminbi, which is the official currency of the People's Republic of China. At the time of writing (July 2023) \$1US $\approx$ 7.3 RMB

SOE – State-Owned Enterprise

SSE – Shanghai Stock Exchange

SP – Sales to price ratio, which is calculated by the last four quarterly sales divided by the current share price.

STAR – The Science and Technology Innovation Board

TURN – Market turnover ratio

VIX – The volatility index, which refer to CBOE China ETF volatility

Appendix II: Methodology

Multiple linear regression and multicollinearity

Traditional multiple regression analysis has been criticized due to problems estimating models with latent variables, reciprocal causality, and interdependence between variables (Ramli et al., 2018). As Sun (2024) points out, when two factors are highly correlated, it often becomes unclear which one should be retained and which should be dropped. This ambiguity underscores the importance of proper dimensionality reduction or variable selection techniques, especially when multiple explanatory variables exhibit similar explanatory power. In addition, while multicollinearity does not lead to bias in estimated slope coefficients, it does increase their standard errors, even in the context of Fama-MacBeth regressions (Green et al., 2017).

Traditional multifactor regression overfits when there is multicollinearity in the independent variables. Multicollinearity is a condition when two or more regressors in a set of regression data are redundant. ‘Redundant’ means what one explanatory variable explains about explained variable is exactly what the other variable explains. The following are detailed explanations:

$$Y = \beta_0 + \beta X + \varepsilon \quad (8)$$

Where Y is a $n \times q$ vector of observations of dependent variable. β_0 is a $n \times 1$ vector. X is a $n \times p$ matrix consisting of n observations on p independent variables. β is a $p \times 1$ vector. ε is a $n \times 1$ vector of residual value. Investors normalise or standardise EP, BM, SP when they use them together. If the variables included in the matrix X and the vector y are mean centred, we omit β_0 and get the equation:

$$Y = \beta X + \varepsilon \quad (9)$$

The ordinary least squares regression estimator β_{ols} can be obtained by minimizing the sum of squared residuals.

$$\varepsilon' \times \varepsilon = (Y - X\beta)' \times (Y - X\beta) \quad (10)$$

$$\beta_{ols} = (X'X)^{-1} \times (X'Y) \quad (11)$$

If the independent variables are highly correlated, the matrix $(X'X)$ is nearly singular or ill-conditioned, making the computation of its inverse

numerically unstable. For example, all valuation ratios contain information on the extent to which a stock is overvalued or undervalued by investors (Lakonishok et al., 1994). The multiple regression model cannot distinguish their individual effects on the dependent variable, so the coefficients associated with these redundant variables cannot be accurately estimated. This increase in variance implies that the estimates of the regression coefficients become less precise and more sensitive to small changes in the data.

Appendix II: Variable Inflation Factors

Variable	VIF	SQRT VIF	Tolerance	R-Squared
BM	1.24	1.11	0.81	0.195
EPTTM	1.08	1.04	0.93	0.072
CFPTTM	1.01	1.01	0.99	0.013
SPTTM	1.18	1.08	0.85	0.15
Mean VIF	1.13			
Dividendpremium	1.23	1.11	0.81	0.19
Dividendpremium	1.28	1.13	0.78	0.22
NIPO	1.13	1.06	0.89	0.11
RIPO	1.12	1.06	0.89	0.11
Mean VIF	1.19			

Note: The table provides the tests for the variable inflation factors. The VIF values indicate no major multicollinearity issues as they are below the threshold of 5.

Appendix III: Robustness test of the PLS value factor

Size	1(Small)	2	3	4	5	6	7	8	9	10(Large)	average
Panel A: Difference between PLS value premium return and EP premium return											
1 month	0.23%	0.91%	0.26%	0.56%	0.00%	0.26%	0.28%	0.07%	-0.42%	0.07%	0.22%
1 year	3.68%	8.10%	4.86%	6.47%	4.99%	1.65%	-1.54%	3.03%	-4.60%	10.05%	3.67%

5 years	30.03%	9.70%	-8.04%	28.85%	40.08%	-4.60%	-0.01%	9.40%	-13.86%	-14.57%	7.70%
Panel B: Difference between PLS value premium return and BM premium return											
1 month	0.65%	0.40%	0.31%	0.52%	0.91%	-0.01%	0.15%	0.29%	0.40%	0.14%	0.38%
1 year	0.60%	2.25%	4.78%	4.47%	8.01%	3.49%	0.85%	-4.05%	0.57%	-1.09%	1.99%
5 years	-5.72%	-8.72%	9.35%	1.30%	11.59%	-2.46%	10.13%	-1.33%	6.92%	22.71%	4.38%
Panel C: Difference between PLS value premium return and SP premium return											
1 month	0.43%	0.21%	0.46%	0.39%	0.08%	0.20%	0.88%	1.33%	-0.06%	-0.06%	0.39%
1 year	0.08%	2.99%	5.32%	4.23%	2.20%	-1.12%	2.79%	3.04%	-5.12%	-0.75%	1.37%
5 years	-1.53%	8.85%	10.74%	2.49%	4.40%	-2.58%	15.51%	2.88%	19.26%	-4.57%	1.16%

Note: The table provides the difference between combined value and BM, EP and SP premium by size decile after dropping the smallest 30% of stocks. After dropping the smallest 30% of stocks, we divide the remaining stocks into deciles by size again and train the PLS value factor to obtain loading weights, mean, and standard deviation from January 2000 to December 2010. Then we compare the performance between the premium of the PLS value factor and other single factor value factors from January 2012 to December 2021. The difference is calculated by PLS value premium minus EP/BM/SP value premium.

Appendix IV: Value premium before and after COVID-19

	Before covid	After covid
Monthly EP premium	0.87%	-0.47%
Monthly BM premium	1.19%	-0.21%
Monthly SP premium	1.04%	0.68%
Monthly PLS premium	1.29%	0.14%

Note: The table reports the average monthly value premiums before and after the COVID-19 outbreak. The pre-COVID period spans January 2000 to November 2019, while the post-COVID period spans December 2019 to December 2021. At the end of each month, stocks are first sorted into ten size deciles. Within each size decile, value premiums are calculated as the return on the highest-value portfolio minus the return on the lowest-value portfolio based on the EP, BM, SP, and PLS value factors. The reported monthly premiums are the equal-weighted averages across the ten size deciles and then averaged over each sample period.

Appendix V: Robustness test of the PLS value factor in the US

Panel A: Standardised loadings and statistics by size decile in the US						
	Size	1(Small)	2	3	4	5(Large)
Loadings	EP	0.18	0.05	0.31	0.55	0.59
	BM	0.93	0.89	0.88	0.8	0.79
	SP	0.79	0.91	0.89	0.82	0.79
Mean	EP	-0.07	0	0.03	0.05	0.06
	BM	1.06	0.84	0.73	0.65	0.6
	SP	2.99	2.14	1.79	1.48	1.25
Std. dev.	EP	0.21	0.16	0.12	0.1	0.08
	BM	0.64	0.56	0.5	0.45	0.42
	SP	2.57	2.19	1.9	1.59	1.28
Panel B: BM, EP and SP premium by size decile in the US						
Size	1(Small)	2	3	4	5(Large)	average
BM	2.98%	0.50%	0.12%	-0.08%	-0.66%	0.57%
PE	2.00%	0.09%	0.07%	0.08%	0.21%	0.49%
SP	2.63%	1.67%	0.90%	0.29%	-0.31%	1.04%
PLS	4.00%	0.95%	0.48%	0.23%	-0.30%	1.07%
Panel C: Standard deviations of BM, EP and SP premium by size decile in the US						
Size	1(Small)	2	3	4	5(Large)	average
BM	9.94%	10.35%	9.67%	8.87%	6.00%	5.87%
PE	10.94%	9.46%	8.13%	7.29%	6.07%	7.04%
SP	12.99%	10.40%	8.86%	8.62%	6.66%	7.75%
PLS	11.20%	12.30%	11.02%	10.33%	8.54%	9.47%
Panel D: Sharpe Ratios of BM, EP and SP premium by size decile in the US						
Size	1(Small)	2	3	4	5(Large)	average
BM	26.61%	1.59%	-2.23%	-4.68%	-16.59%	4.00%
PE	15.21%	-2.59%	-3.27%	-3.50%	-2.07%	2.20%
SP	17.66%	12.83%	6.37%	-0.53%	-9.69%	9.09%
PLS	32.72%	5.00%	1.31%	-1.02%	-7.44%	7.75%

Note: The table presents the parameters and performance of the PLS value factor in the United States, comparing it using BM, PE, and PS. Panel A provides each size's standardised loadings, mean and standard deviation of BM, EP and SP. The training period spans from January 1970 to December 2010. In line with Fama & French (2015), panel B provides the average value portfolio return in the US stock market from January 2012 to December 2021 in five different sizes. The securities are firstly assigned to five quintiles by size. Then, in each size, securities are assigned to five quintiles based on BM, EP, SP and PLS value factor, and the value premium is calculated by the equally-weighted one-month return of the top 20% BM, EP and SP stocks minus the equally-weighted return of the bottom 20% BM, EP and SP stocks. Panel C presents the standard deviations of the PLS value factor and the BM, EP, and SP premiums across different size quintiles. Panel D presents the Sharpe ratios of the PLS value factor and the BM, EP, and SP premiums across different size quintiles. The risk-free rate is assumed to be 4.1% per annum. The data for the United States are sourced from Wharton Research Data Services. According to Wharton Research Data Services, Market Capitalization and Price are acquired from both Center for Research in Security Prices and Compustat. Center for Research in Security Prices is designated as the primary source for pricing data. Earnings-related information is obtained from the Institutional Brokers' Estimate System database.

Appendix VI: Value Premiums and Investor Sentiment Regressed on the Investor Confidence Index

	One month ahead EP premium	One month ahead BM premium	One month ahead SP premium	One month ahead PLS value premium	Five years ahead EP premium	Five years ahead BM premium	Five years ahead SP premium	Five years ahead PLS value premium	c1
Coefficient	-0.0003	-0.0005	-0.0002	-0.0004	-0.0159***	-0.0132***	-0.0123***	-0.0155***	0.0209*
Std. err.	(0.141)	(0.082)	(0.284)	(0.110)	(0.000)	(0.000)	(0.000)	(0.000)	(0.026)
constant	0.0101***	0.0120***	0.0110***	0.0136***	0.239***	0.157***	0.190***	0.213***	-0.103
Std. err.	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.346)
N	227	227	227	227	192	192	192	192	228
adj. R ²	0.005	0.009	0.001	0.007	0.134	0.151	0.153	0.163	0.018

Note: The table reports the results of univariate OLS regressions in which one-month-ahead and five-year-ahead value premiums (based on EP, BM, SP, and the PLS value factor), together with the investor sentiment index (c1), are separately regressed on the Investor Confidence Index. The sample period spans from January 2000 to December 2021. Because the dependent variables are constructed using future returns, the sample for the one-month-ahead value premiums ends in November 2021, the sample for the one-year-ahead value premiums ends in January 2020, and the sample for the five-year-ahead value premiums ends in January 2016. The reported coefficients are estimated using separate regressions with a single explanatory variable. Adjusted R² is reported for each regression. “***”, “**”, and “*” denote statistical significance at the 0.1%, 1% and 5% levels, respectively.

Appendix VII: Single-factor Predictive Regression across Firm Size Subsamples

		Small		Large	
		one month ahead return	Five years ahead return	one month ahead return	Five years ahead return
BM premium	Coefficient	0.0023	-0.0743**	0.0054*	-0.0400
	Std. err.	(0.132)	(0.002)	(0.042)	(0.100)
	constant	0.0111***	0.230***	0.0084*	0.223***
	Std. err.	(0.000)	(0.000)	(0.038)	(0.000)
	adj. R ²	0.006	0.052	0.014	0.010
EP premium	Coefficient	0.0001	-0.0512	0.0042*	-0.0775**
	Std. err.	(0.936)	(0.053)	(0.027)	(0.002)
	constant	0.0056*	0.149***	0.0116***	0.390***
	Std. err.	(0.017)	(0.001)	(0.000)	(0.000)
	adj. R ²	-0.004	0.016	0.017	0.051
SP premium	Coefficient	0.0031*	-0.0433*	0.0061**	-0.115***
	Std. err.	(0.018)	(0.031)	(0.002)	(0.000)
	constant	0.0109***	0.257***	0.0091**	0.395***
	Std. err.	(0.000)	(0.000)	(0.002)	(0.000)
	adj. R ²	0.020	0.022	0.037	0.096
PLS Value premium	Coefficient	0.0021	-0.0663*	0.0043	-0.0980***
	Std. err.	(0.199)	(0.011)	(0.080)	(0.000)
	constant	0.0109***	0.207***	0.0123***	0.407***
	Std. err.	(0.000)	(0.000)	(0.001)	(0.000)
	adj. R ²	0.003	0.032	0.009	0.081

Note: The table reports the results of the subsample analysis based on firm size. At the end of each month, stocks are first sorted into ten size deciles by market capitalization and, independently, into ten value deciles based on BM, EP, SP, and the PLS value factor. For each size, the value premium is calculated as the equal-weighted return of the highest value decile minus the equal-weighted return of the lowest value decile. The small subsample consists of stocks in the bottom 50% of the size distribution, while the large subsample consists of stocks in the top 50%. The reported small (large) value premium is computed as the equal-weighted average of the five value premiums corresponding to Size Deciles 1–5 (6–10), respectively. Holding periods include one month and five years. The reported coefficients are obtained from separate OLS regressions of future value premiums on the investor sentiment index (C1). The table reports the estimated coefficients, conventional OLS standard errors, intercepts, standard errors of the intercepts, and adjusted R² values. The sample period spans from January 2000 to January 2021. No fixed effects or clustering procedures are employed

Appendix VIII: Variable Definitions and Data Sources

Proxy	Definition	Sample period and sources
size	The total market value of all of a company's outstanding shares at the end of each month.	January 2000 to December 2021. Choice database
Book to market ratio	The ratio of the most recently available quarterly book value of equity to the firm's market capitalisation at the end of each month.	January 2000 to December 2021. Choice database
Price to cash flow ratio	The ratio of net cash flow over the most recent four quarters to the firm's market capitalisation at the end of each month.	January 2000 to December 2021. Choice database
Earnings-to-price ratio	The ratio of net income over the most recent four quarters to the firm's market capitalisation at the end of each month.	January 2000 to December 2021. Choice database
Sales-to-price ratio	The ratio of sales over the most recent four quarters to the firm's market capitalisation at the end of each month.	January 2000 to December 2021. Choice database
One-month-ahead return	The stock return over the month following portfolio formation, including cash dividends.	January 2000 to December 2021. Choice database
One-year-ahead return	The cumulative stock return over the 12 months following portfolio formation, including cash dividends.	January 2000 to December 2021. Choice database
Five-year-ahead return	The cumulative stock return over the 60 months following portfolio formation, including cash dividends.	January 2000 to December 2021. Choice database
Market turnover	Average ratio of monthly transaction value to monthly market capitalization outstanding.	January 2000 to December 2021. CSMAR database
The average first-day returns on IPOs (RIPO)	The ratio of the IPO closing price to the IPO issue price minus one, excluding firms in the smallest market capitalisation decile (market cap lower than 227 million RMB). For IPOs that reach the daily price limit for n consecutive trading days, the closing price on trading day n+1 is used instead of the first-day closing price.	January 2000 to December 2021. CSMAR database
The number of IPOs (NIPO)	The number of initial public offerings recorded during each month.	January 2000 to December 2021. CSMAR database
Dividend premium	The dividend premium is the difference between the logs of the dividend-paying companies' and non-dividends-paying companies' average market-to-book ratios	January 2000 to December 2021. Choice database
Dividend payout ratio	The ratio of total dividends paid over the most recent 12 months to the firm's share price at the end of each month.	January 2000 to December 2021. Choice database
New investor accounts (NIA)	The number of newly opened investor accounts at the end of each month.	January 2003 to December 2021. CSMAR database
The Volatility Index (VIX)	The CBOE China ETF volatility index.	March 2011 to December 2021. Chicago Board Options Exchange
Market movement (MM)	The monthly return on the Shanghai A Index, measured as the percentage change in the index value from the end of the previous month to the end of the current month.	January 2000 to December 2021. CSMAR database

