Finding Anomalies in China

Kewei Hou, Fang Qiao, and Xiaoyan Zhang*

First Draft: July 2019
This Draft: August 2024

Abstract

Using firm-level trading and accounting signals, we construct 454 univariate strategies in the Chinese stock market over the past 20 years. With the conventional single-testing *t*-statistic cutoff of 1.96, 104 and 22 strategies have significant value-weighted raw returns and alphas, respectively. To avoid false discoveries, we calibrate the multiple-testing *t*-statistic cutoff to 2.85. The raw returns of 38 strategies survive the higher hurdle rate, while none remain significant after risk adjustments. Composite strategies with multiple signals and machine learning techniques perform much better than univariate strategies. Interestingly, these anomaly returns comove significantly with market development, accounting quality, liquidity, and regulations.

Keywords: China, anomalies, multiple tests, univariate strategies, composite strategies.
JEL Classification: G12, G14, G15.

---

* We thank Lu Zhang, Xintong Zhan, Wenxin Jiang, Jinfei Sheng, Jianan Liu, Bing Han, Dacheng Xiu, and participants at the 2019 Summer Institute of Finance conference, the 2021 China Fintech Research Conference, the 2021 China Financial Research Conference, the 2021 China International Conference in Finance, the 2024 FMA Asia/Pacific Conference, Renmin University of China for helpful discussions and comments. Hou is from Fisher college of Business, The Ohio State University, 820 Fisher Hall, 2100 Neil Avenue, Columbus, Ohio 43210, US; Email: hou.28@osu.edu; Qiao is from China School of Banking and Finance, University of International Business and Economics, 10 Huixindongjie, Chaoyang District Beijing, 100029, P.R. China; Email: qiaofang@uibe.edu.cn. Zhang is from PBC School of Finance, Tsinghua University, 43 Chengfu Road, Haidian District Beijing, 100083, P.R. China; Email: zhangxiaoyan@pbcsf.tsinghua.edu.cn. Fang Qiao acknowledges the financial support from the Ministry of Education Project of Humanities and Social Sciences [Grant No. 23YJC790106]. All remaining errors are our own.

1. Introduction

After 30 years of rapid growth, the Chinese stock market is currently the second-largest stock market in the world, with 4,873 stocks listed and a total market capitalization of 11.7 trillion dollars by the end of August 2022. Meanwhile, its correlation with the rest of global markets stays relatively low, which makes it a natural choice for international diversification. Due to its enormous size, fast growth, and diversification benefits, the Chinese stock market has attracted considerable attention from investors, regulators, and academia from all over the world, all trying to understand the opportunities and risks therein. Unlike more advanced stock markets, the Chinese stock market has certain unique features, such as a large number of retail investors and an opaque information environment, which make it interesting yet challenging to understand the return dynamics of Chinese stocks.

Anomalies are important channels for understanding the cross-section of stock returns. Classic asset pricing theories state that returns are compensations for risks, and many asset pricing models have been proposed to explain cross-sectional return patterns, such as the CAPM of Sharpe (1964), the three- and five-factor models of Fama and French (1993, 2015), and the *q*-factor models of Hou, Xue, and Zhang (2015) and Hou, Mo, Xue, and Zhang (2021). Under the joint hypothesis of market efficiency and risk adjustments using the correct model, all assets or managed portfolios should not have significant abnormal returns. Hence, the unexplained return patterns are typically called "anomalies." Recent studies on U.S. anomalies, such as McLean and Pontiff (2016), Green, Hand, and Zhang (2017), and Hou, Xue, and Zhang (2020), have substantially influenced our understanding of return dynamics, market efficiency, and compensation for risks.

In this study, we aim to understand the cross-sectional return patterns in the Chinese stock market through the lens of anomalies. In contrast to existing studies, which mostly focus on a small number of anomalies, we collect 208 signals to construct 454 strategies for the Chinese market, which makes our study the most comprehensive one on Chinese anomalies. One might expect the anomaly patterns in China to be similar to those in the U.S. However, previous research, such as Titman, Wei, and Xie (2013) and Jacobs and Muller (2020), shows that anomaly patterns in the U.S. cannot be easily replicated in other markets. Given the distinctive features of the Chinese and U.S. markets, the anomaly patterns in China could be substantially different from those in the U.S., which would shed light on the sources of anomalies and provide insights on market efficiency as well as asset pricing models.

Among the 208 firm-level signals we collect, 73 are based on trading data, such as price, volume, and volatility, and 135 are based on accounting data, such as value, profitability, and investment. To examine whether they contain information about the cross-section of Chinese stock returns, we first construct univariate strategies to examine the signals individually. For each signal, we sort the largest 70% of Chinese stocks (excluding microcaps) into ten groups and long stocks in the top decile with the highest signal values and short stocks in the bottom decile with the lowest signal values. With different portfolio holding periods, we construct a total of 454 univariate strategies, and compute value- and equal-weighted high-minus-low raw and risk-adjusted returns for each of these strategies. If the high-minus-low raw or risk-adjusted return (alpha) on a strategy is significantly different from zero, then it is deemed a significant anomaly. The significant anomalies can be used for investment purposes; they can also provide information on cross-sectional return patterns, inferences on market efficiency, and the adequacy of asset pricing models.

To establish statistical significance, we start with the conventional single-testing t-statistic cutoff of 1.96 at the 5% significance level. For the value-weighted high-minus-low strategies, 104 out of 454 have significant raw returns. Among them, 38 are based on trading signals and liquidity measures in particular, and the other 66 are based on accounting signals, especially profitability measures. For the equal-weighted strategies, 189 out of 454 have significant raw returns, with 95 strategies based on trading signals and 94 based on accounting signals. Liquidity and profitability measures are again the biggest contributors. These results suggest that liquidity and profitability are important drivers of cross-sectional returns in China.

The significant high-minus-low raw returns may be compensation for exposures to systematic risk factors in China. Therefore, we compute risk-adjusted returns or alphas, using the CAPM and the Chinese three- and four-factor models (CH3 and CH4) of Liu, Stambaugh, and Yuan (2019). The CAPM considers only the market factor; the CH3 model incorporates the market, size, and value factors in China; and the CH4 model further adds a liquidity risk factor. While the CAPM cannot explain most of the significant high-minus-low raw returns, the CH3 and CH4 models can. For example, out of the 104 significant value-weighted strategies, 103 have significant CAPM alphas; 37 have significant CH3 alphas; and only 22 have significant CH4 alphas.

In assessing the evidence of significance in the factor pricing and anomaly literature, recent studies, such as Harvey, Liu, and Zhu (2016), Chordia, Goyal, and Saretto (2020), Hou, Xue, and Zhang (2020), Chen (2021), and Giglio, Liao, and Xiu (2021), highlight the possibility of false discoveries and offer various solutions. In this paper, we adopt the methodology of Harvey and Liu (2020) and use their multiple-testing framework and double bootstrapping to recalibrate the *t*-statistic cutoff to 2.85. The higher hurdle rate for *t*-statistics greatly reduces the number of significant anomalies in China. For the value-weighted strategies, only 38 remain significant in raw returns. After CH3 and CH4 risk adjustments, the number drops to four and zero, respectively. For the equal-weighted strategies, 108 remain significant in raw returns, 37 in CH3 alphas, and 21 in CH4 alphas.

To obtain economic intuitions on our results, we compare the anomaly patterns in China to those in the U.S., in terms of significance rate and information content. For the former, Hou, Xue, and Zhang (2020) document a significance rate of 35% in the U.S. for value-weighted raw returns under single testing and 18% under multiple testing. Our results show a significance rate of 23% in China for value-weighted raw returns under single testing and 8% under multiple testing. The patterns are similar for equal-weighted returns. Therefore, the significance rate in China is considerably lower than that in the U.S. In our view, the short time series combined with the high level of uncertainty and opaque information environment are responsible for the smaller number of significant anomalies in China.

What is perhaps more interesting is the information content of the anomalies. Our results show that trading signals contribute more to the significant anomalies in China than in the U.S., while the opposite is true for accounting signals. The relative importance of trading signals in China is likely driven by the fact that the Chinese stock market is still a young and under-developed market, with more than half of the daily trading volume coming from retail investors, whose excessive trading generates predictable return patterns that are picked up by trading signals. At the same time, due to the relatively lax financial reporting standards and disclosure practices and the generally opaque information environment, many accounting signals in China are too noisy to significantly predict returns during our sample period.

To further improve our understanding of the anomaly patterns in China, we study two additional important questions. First, do aggregate economic conditions affect anomaly returns in China? To answer this question, we relate anomaly returns to aggregate market-level variables, including trading frictions, financial market development, accounting data quality, investor sentiment, market liquidity, and government regulations. We find that anomaly returns comove significantly with financial market development, accounting quality, market liquidity, and government regulations. In particular, we find that periods of higher market liquidity are associated with larger anomaly returns, which is consistent with the retail trading interpretation.

Second, given the short history, noisy signals, and volatile returns of the Chinese stock market, can investors enhance the performance of their trading strategies by combining multiple signals into composite strategies? To answer this question, we consider both traditional signal aggregation methods, such as composite score and multiple regression, as well as advanced machine learning methods, such as Lasso and random forest. The composite strategies, especially those based on machine learning methods, deliver higher raw returns and alphas than univariate strategies, indicating that investors can potentially improve the profitability of their trading strategies by combining firm-level signals.

Our study is related to three different strands of the anomaly literature. First, it is related to the literature on U.S. anomalies. Recent studies, such as McLean and Pontiff (2016), Green, Hand, and Zhang (2017), Linnainmaa and Roberts (2018), Hou, Xue, and Zhang (2020), and Chen and Zimmermann (2022), document hundreds of firm-level characteristics that can predict the cross-section of stock returns; Jensen, Kelly, and Pedersen (2023) develop Bayesian anomaly replication models to show that anomalies work in the U.S. and global markets; Harvey, Liu, and Zhu (2016), Chordia, Goyal, and Saretto (2020), Harvey and Liu (2020), Chen (2021), and Giglio, Liao, and Xiu (2021) focus on the testing procedures for the significance of anomalies; and Gu, Kelly, and Xiu (2020) adopt machine learning techniques to combine anomaly signals. Second, our study of anomalies in China is related to previous papers on anomalies in international markets. For example, Ang, Hodrick, Xing, and Zhang (2009) examine the idiosyncratic volatility anomaly in 23 developed markets; Hou, Karolyi, and Kho (2011) investigate the size, dividend yield, earnings yield, cash flow-to-price, book-to-market equity, leverage, and momentum effects in 49 developed and emerging markets; Titman, Wei, and Xie (2013) and Watanabe, Xu, Yao, and Yu (2013) study the investment anomaly in more than 40 countries; Jacobs and Muller (2020) study 241 anomalies in 39 international markets. Some of these papers find that certain anomalies in the U.S. may not replicate in other markets, while others show that anomalies tend to be stronger in more developed markets. Finally, our work naturally connects to previous studies on Chinese anomalies, such as Chen, Kim, Yao, and Yu (2010), Hsu, Viswanathan, Wang, and Wool (2018), Jansen, Swinkels, and Zhou (2021), Wu, Wei, and Zhang (2021), Leippold, Wang, and Zhou (2022), and Li, Liu, Liu, and Wei (2023).

In comparison with previous research on international and Chinese anomalies, our study contributes to the literature in four different ways. First, we provide the most comprehensive study to date on Chinese stock market anomalies and obtain substantial new findings to improve our understanding of cross-sectional return patterns in China. Second, to address the concern of false discoveries, we calibrate the multiple-testing threshold for the Chinese market, which could be useful for future studies on anomalies in China. Third, we document significant differences in anomaly patterns between China and the U.S. to complement the international anomaly literature. In particular, we show that many trading-based signals, especially those related to liquidity, are significant in China. Finally, we link anomaly returns to aggregate market conditions and find that financial market development and accounting quality significantly attenuate anomaly performance, while market liquidity and government regulations significantly accentuate anomaly performance.

In particular, three studies—Liu, Stambaugh, and Yuan (2019), Leippold, Wang, and Zhou (2022), and Li, Liu, Liu, and Wei (2023)—investigate cross-sectional returns in the context of China and are closely related to our paper. However, our research perspectives are different. For instance, Liu, Stambaugh, and Yuan (2019) propose risk adjustment benchmark factor models. Leippold, Wang, and Zhou (2022) use machine learning methods to construct composite strategies with multiple signals. In contrast, we analyze cross-sectional return patterns through a comprehensive set of individual anomalies, covering both single and multiple testing, information content, and sources of anomalies. Li, Liu, Liu, and Wei (2023) compare the performance of various factor models based on replicated anomalies following Hou, Xue, and Zhang (2020). Conversely, we construct comprehensive anomalies and aim to address three important issues in the anomaly literature: false discoveries, economic insights, and driving forces. Therefore, based on different research perspectives, our study provides interesting new findings that differ from those in the three aforementioned papers.

The remainder of this paper is organized as follows. Sections 2 and 3 introduce the data and methodology used in this paper, respectively. We study univariate strategies and relate them to aggregate market conditions in Section 4, and examine composite strategies in Section 5. Section 6 presents robustness results with respect to microcaps, and Section 7 concludes.

2. Data

2.1 Data Sources

The Chinese stock market data are available from 1990 onwards, and our sample period is from July 2000 to December 2020. We start our analysis from 2000 for three reasons. First, to construct anomaly strategies, we need a large cross-section of stocks, but the sample size before 2000 cannot ensure a sufficient number of stocks for portfolio construction. Second, many regulations on financial reporting were implemented around 2000 in China, and accounting data became more comparable across firms afterwards. Third, several trading rules, such as the daily price limits, which greatly reduced noise and extreme movements in stock prices, also became effective around 2000.

We obtain daily stock trading data and quarterly accounting data from WIND Information Inc., institutional ownership data from CSMAR, and analyst earnings forecast data from SUNTIME. We also obtain social media coverage from GUBA, web search index from WSVI, and news data from CFND via the Chinese Research Data Services (CNRDS) platform. The WIND and CSMAR data are available starting from 2000. Other datasets start at later times, with SUNTIME starting from 2006, GUBA from 2008, WSVI from 2011, and CFND from 2001.

We merge the above datasets using unique stock identification codes. Following previous studies, we apply several filters to our sample. First, we drop the first six months of data after a firm goes public to avoid potentially extreme volatility and illiquidity following the IPO. Second, to guarantee minimum stock liquidity and data quality and to filter out firms with long trading suspensions, we exclude firms with less than 75% of daily trading records within a signal construction window. Third, to avoid the shell contamination effect documented in Liu, Stambaugh, and Yuan (2019), we drop the smallest 30% of firms based on market capitalization. Lastly, due to their different trading rules, we exclude firms listed on the Science and Technology Innovation Board. More details on the data are presented in Internet Appendix A.

2.2 Signals

We construct 199 signals following Hou, Xue, and Zhang (2020), one of the most comprehensive studies on anomalies. In addition, we construct nine signals that are specific to the Chinese stock market, including a state ownership (SOE) indicator, a margin trading indicator, and signals based on social media comments, coverage, and web search volume. Altogether, we have a total of 208 signals. Based on the information content of the signals, we separate them into two broad groups: 73 trading-based signals and 135 accounting-based signals. The trading signals are further divided into three subgroups: 35 liquidity measures, such as volume and turnover; 18 risk proxies, such as beta and volatility; and 20 past return signals, including momentum, reversal, and seasonality. The accounting signals are also divided into four subgroups: 35 profitability measures, such as return on equity and return on assets; 24 value proxies, including book-to-market ratio and earnings-to-price ratio; 37 investment measures, such as asset growth; and 39 other fundamental variables, including R&D expense ratio and analyst forecast dispersion. We list the 208 signals in Table 1 and provide their detailed definitions in Internet Appendix B. Given that Chinese trading rules and accounting standards are sometimes different from those in the U.S., we make adjustments to signal construction when necessary.

[Insert Table 1 Here]

2.3 Summary Statistics

To have a basic idea of the overall features of the Chinese stock market, we provide summary statistics of the Chinese market in comparison to the U.S. over the sample period 2000 – 2020 in Table 2. In the first two rows, we report the time-series averages of the number of listed firms and total market capitalization for China and the U.S. Over the past 20 years, on average there are more than 2,000 listed firms in China, whereas there are more than 4,000 listed firms in the U.S. The average aggregate market cap is 3.81 trillion dollars for China, compared to 18.54 trillion dollars for the U.S. The next four rows of Table 2 report the average pooled firm-level statistics for each market. The average firm size is 1.81 billion dollars in China, less than half of the U.S. average firm size of 4.32 billion dollars. The average book-to-market ratio in China is 0.40, which is lower than that of 0.84 in the U.S., indicating higher market valuations in China. We compute share turnover in China as the total number of shares traded in a year divided by the number of free-float A-shares outstanding at the prior year-end. The average annual share turnover in China is 6.53, which almost triples the average U.S. share turnover of 2.38, suggesting that Chinese firms are much more liquid than U.S. firms. In the last row, the average annual return for Chinese firms is 18% with a volatility of 0.70, compared to the average U.S. firm annual return of 12% and volatility of 0.66. Overall, Chinese firms have smaller size, higher valuations, and higher liquidity.

[Insert Table 2 Here]

We also present the time series of several important market characteristics in Figure 1. Panels A and B plot the number of listed firms and total market cap over time, respectively. The number of Chinese listed firms rose from 973 in 2000 to 3,736 in 2020, and the total market cap rose from 0.2 trillion dollars in 2000 to 11 trillion dollars in 2020. By comparison, over the same period, the number of listed firms in the U.S. decreased from 6,358 in 2000 to 3,744 in 2020, while the total market cap grew from 14 trillion dollars in 2000 to 39 trillion dollars in 2020.

[Insert Figure 1 Here]

As mentioned in the introduction, the Chinese stock market differs from the U.S. market in two major aspects: trading and information environment. Panel C of Figure 1 plots the average firm-level share turnover by year. Over the 20 years in our sample, the average share turnover is always higher in China than in the U.S. We also observe substantial time-series variation in turnover for Chinese firms. It exhibits large spikes in 2007, 2009, and 2015, likely driven by excessive trading during market boom and bust periods. In terms of information environment, Boone and White (2015) use institutional ownership as a proxy for information quality, arguing that institutional investors are more sophisticated, and their participation improves the information environment of a firm. We compute firm-level institutional ownership in China as the shares held by mutual funds, brokers, insurance companies, security funds, entrusts, etc. divided by the number of free-float A-shares outstanding. We plot the average firm-level institutional ownership in Panel

D of Figure 1. The average institutional ownership in China varies between 3% and 22% over our sample period, which is much lower than the 33–65% range for the U.S. over the same period, suggesting that information quality is worse in China than in the U.S.

3. Methodologies

3.1 Univariate Strategies

The univariate strategies are each based on one firm-level signal. They are constructed following the [k, m, n] timing convention, with k, m, and n referring to the length of the signal estimation period, waiting period, and holding period, respectively. At the beginning of each month t, we first collect firm-level information over months [t-k-m, t-m-1] to construct a signal. Then we sort the largest 70% of A-share stocks into decile portfolios based on their signal values. Decile 1 contains the 10% of stocks with the lowest signal values, and decile 10 contains the 10% of stocks with the highest signal values. We hold these decile portfolios for n months over months [t, t+n-1] and compute their value-weighted and equal-weighted raw returns over the holding period. If n is greater than one, for a given decile each month there exist n sub-deciles, each of which is based on a signal estimation window ending in each of the n-month period prior to month t-m. We take the simple average of the sub-decile returns as the return of the decile.[1] Altogether, we construct 454 univariate strategies. If a signal contains valuable information about the cross-section of stock returns in China, we expect to see a significant average high-minus-low return (the return difference between decile 10 and decile 1).

3.2 Risk Adjustments

To compute risk-adjusted returns, the first benchmark model we consider is the Chinese CAPM with a local market factor. We follow Liu, Stambaugh, and Yuan (2019) and construct the

[1] Details on portfolio construction (including exceptions to the above method) are provided in Internet Appendix B.

Chinese market factor as the difference between the return on the value-weighted portfolio of the largest 70% of A-share stocks and the one-year deposit rate (a proxy for the risk-free rate in China). For a strategy *l*, we estimate its CAPM alpha, α_l^{CAPM}, as:

$$R_{l,t}^{high} - R_{l,t}^{low} = \alpha_l^{CAPM} + \beta_{MKT,l}^{CAPM} MKT_t + e_{l,t}, \tag{1}$$

where $R_{l,t}^{high}$ is the return of decile 10 in month *t*, $R_{l,t}^{low}$ is the return of decile 1, MKT_t is the excess return on the market portfolio, and $\beta_{MKT,l}^{CAPM}$ is strategy *l*'s exposure to the market factor.

To control for the size and value effects in China, Liu, Stambaugh, and Yuan (2019) propose a three-factor model, CH3, which contains the Chinese market, size, and value factors. In addition, they introduce a four-factor model, CH4, which adds a liquidity factor to control for the liquidity effect in China. We follow Liu, Stambaugh, and Yuan (2019) to construct the size, value, and liquidity factors, and provide details on their construction in Internet Appendix C1. Following Equation (1), we compute the alphas for the CH3 and CH4 models, α_l^{CH3} and α_l^{CH4}, and their associated factor exposures as:

$$R_{l,t}^{high} - R_{l,t}^{low} = \alpha_l^{CH3} + \beta_{MKT,l}^{CH3} MKT_t + \beta_{SMB,l}^{CH3} SMB_t + \beta_{VMG,l}^{CH3} VMG_t + e_{l,t}, \tag{2}$$

$$R_{l,t}^{high} - R_{l,t}^{low} = \alpha_l^{CH4} + \beta_{MKT,l}^{CH4} MKT_t + \beta_{SMB,l}^{CH4} SMB_t + \beta_{VMG,l}^{CH4} VMG_t + \beta_{PMO,l}^{CH4} PMO_t + e_{l,t}, \tag{3}$$

where SMB_t, VMG_t, and PMO_t are the Chinese size, value, and liquidity factors, respectively.

3.3 Multiple Tests

When testing whether the average high-minus-low return or alpha of a strategy is significantly different from zero, the conventional statistical inference is based on a "single-testing" framework, which assumes that there is only one hypothesis being tested, and the *t*-statistic cutoff is set to 1.96 for the 5% significance level. However, in our setting, there are hundreds of anomalies being tested on the same dataset, and consequently using the single-testing hurdle rate could lead to false discoveries. Several recent studies, such as Harvey, Liu, and Zhu (2016) and Chordia, Goyal, and

Saretto (2020), advocate the alternative of “multiple testing” (simultaneous testing of more than one hypothesis) to address the issue of false discoveries in anomaly studies.

Prior literature (e.g., Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Barras, Scaillet, and Wermers, 2010; Harvey and Liu, 2020) proposes several different methods for computing multiple testing statistics. Here we adopt the method of Harvey and Liu (2020) for two reasons. First, Harvey and Liu (2020) consider the tradeoff between Type I and II errors, while most other methods only control for Type I errors.[2] Second, Harvey and Liu (2020) implement a double bootstrapping procedure to alleviate finite sample concerns, which is particularly relevant given the short history of the Chinese stock market. Details on applying the multiple-testing method of Harvey and Liu (2020) to Chinese data are provided in Internet Appendix C2.

Following Harvey and Liu (2020), we first choose the parameter p_0, the percentage of strategies that are true, to control for Type II errors. Given that the choice of p_0 is inherently subjective, Harvey and Liu (2020) recommend selecting p_0 to be less than the single-testing significance rate in the data to accommodate multiple-testing adjustments. Since 23% of the univariate strategies in our sample have absolute *t*-statistics greater than 1.96, we set p_0 to 15%, assuming that 8% out of the 23% significant strategies are false discoveries.

Next, we determine the *t*-statistic cutoff under multiple testing. Harvey and Liu (2020) argue that an appropriate hurdle rate would balance between Type I and II errors, resulting in a desirable odds ratio (the ratio of Type I to Type II errors). Panel A of Figure 2 plots the Type I and II error rates and the odds ratio (Y-axis) against the *t*-statistic hurdle rate (X-axis) based on applying the double bootstrapping procedure of Harvey and Liu (2020) to our sample. When the hurdle rate

[2] In Internet Appendix C2, we report multiple-testing results using the methods of Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001), and Barras, Scaillet, and Wermers (2010). Their *t*-statistic cutoffs are around 2.85. Given that these methods only control for Type I errors and ignore Type II errors, we focus on the Harvey and Liu’s (2020) method for our main analysis.

increases, the Type I error rate decreases, the Type II error rate increases, and the odds ratio decreases. When we fix the Type I error rate at 5%, the Type II error rate is at 6%, and the resulting odds ratio is at a reasonable 19% (close to 20% recommended by Harvey and Liu (2020)). The corresponding t-statistic cutoff is 2.85, which we use as our benchmark hurdle rate for multiple testing at the 5% significance level.

[Insert Figure 2 Here]

Since the choice of p_0 is somewhat subjective, we also consider alternative values of p_0, from 5% to 20%, and plot the corresponding t-statistic cutoff against p_0 in Panel B of Figure 2. The figure shows that the hurdle rate declines slightly to 2.70 when we increase p_0 from 15% to 20%, as we are less likely to miss a true anomaly in the latter scenario (when a larger fraction of strategies are assumed to be true anomalies). On the other hand, the hurdle rate increases slightly to 2.95 when we decrease p_0 to 10%. If we further decrease p_0 to 5% (and thus assume only 5% of strategies are true anomalies, whereas 18% are false discoveries), then the hurdle rate becomes 3.10. Given that the variation in the hurdle rate is relatively small for the range of p_0 we consider, we use the benchmark hurdle rate of 2.85 (under a 15% p_0) for our main analysis.

4. Empirical Results on Univariate Strategies

We summarize the significant univariate strategies using the single-testing hurdle rate in Section 4.1. Results using the multiple-testing hurdle rate are discussed in Section 4.2. Details on the significant strategies are provided in Section 4.3. We compare the anomaly patterns in China to those in the U.S. in Section 4.4. In Section 4.5, we examine the relations between anomaly returns and aggregate market conditions.

4.1 Univariate Strategies under Single Testing

We provide an overview of the univariate strategies with significant high-minus-low returns in Table 3. The statistical significance is based on the Newey and West (1987) standard errors with four lags, using the conventional single-testing t-statistic cutoff of 1.96.

[Insert Table 3 Here]

Panel A of Table 3 shows that 104 out of 454 univariate strategies (23% significance rate) have significant value-weighted high-minus-low raw returns under single testing. Of the 104 significant strategies, 25, 6, 7, 40, 2, 2, and 22 are from the liquidity, risk, past returns, profitability, value, investment, and others categories, respectively.

Given the possibility that high-minus-low raw returns could represent compensation for exposures to systematic risk factors, we also report risk-adjusted returns (alphas) using the CAPM, the CH3 model, and the CH4 model. We find that the CAPM cannot explain the returns of the significant strategies, as the CAPM alphas are significant for 103 strategies, almost identical to the number of significant strategies in raw returns (104). On the other hand, when we use the CH3 model to adjust returns, 37 strategies (8% significance rate) have significant alphas, indicating that the majority of the significant strategies in raw returns can be attributed to the size and value effects in China. Among the 37 significant CH3 alphas, 22 are based on trading signals (14 of which are from the liquidity category), and the other 15 are based on accounting signals (14 from the others category). When we further use the CH4 model for return adjustments, only 22 strategies (5% significance rate) have significant alphas, 10 of which are trading signals (3 from the liquidity category), and the other 12 are accounting signals (7 from the others category). Trading signals, especially liquidity measures, thus seem to be an important driver of the cross-section of stock returns in China not captured by existing factor models. Accounting signals, such as value and investment, on the other hand, are less effective in generating significant alphas, suggesting that they contain limited information about future returns after controlling for existing factor models.

When we examine equal-weighted returns in Panel B of Table 3, 189 out of 454 univariate strategies (42% significance rate) have significant high-minus-low raw returns under single testing, which almost doubles the number of significant strategies in value-weighted returns. This might not be surprising because equal-weighted returns are tilted towards smaller firms, which have been shown by prior literature to be associated with higher arbitrage costs and more significant anomalies in the U.S. and other international markets. Among the 189 significant strategies, 55, 23, 17, 45, 13, 9, and 27 are from the liquidity, risk, past returns, profitability, value, investment, and others categories, respectively.

When we apply the factor models to equal-weighted high-minus-low raw returns, 188 CAPM alphas are significant, again indicating that the market factor cannot explain the returns of the significant strategies. For the CH3 and CH4 models, 91 (20% significance rate) and 72 (16% significance rate) alphas are significant, respectively, with 39 and 27 significant alphas from the liquidity category alone. These equal-weighted results thus confirm the finding that liquidity is an important driver of the cross-section of Chinese stock returns.[3]

4.2 Univariate Strategies under Multiple Testing

When using the conventional single-testing *t*-statistic cutoff of 1.96, there is a concern that the significant strategies are actually false discoveries due to a low hurdle rate. To address this concern, we follow Harvey and Liu (2016) and apply a higher multiple-testing hurdle rate of 2.85 to the univariate strategies.

Table 4 reports the number of significant strategies under multiple testing, which declines substantially from its single-testing counterpart. For the value-weighted strategies in Panel A, 38

[3] Due to short-shelling constraints in China, we report the results for long-leg portfolios in Internet Appendix D Table D1. Only two strategies have significant value-weighted long-leg raw returns under single testing, both from the liquidity category. After risk adjustments, only one CH4 alpha is significant.

(8% significance rate) have significant high-minus-low raw returns, compared with 104 under single testing (Table 3 Panel A). Out of the 38 significant strategies, 5, 20, and 9 are from the liquidity, profitability, and others categories, respectively, with the remaining 4 coming from the risk, past returns, and value categories. After factor model risk adjustments, the number of significant CAPM alphas remains unchanged at 38, indicating that the market factor cannot explain the significant strategies. When we control for the size and value factors using the CH3 model, the number of significant alphas drops to four, with two from the liquidity category, one each from the risk and past returns categories, and none from the profitability or other accounting categories. When we further control for the liquidity factor using the CH4 model, none of the strategies have significant CH4 alphas.

[Insert Table 4 Here]

For the equal-weighted strategies in Panel B, the higher 2.85 hurdle rate cuts the number of significant strategies by almost half from 189 under single testing (Table 3 Pane B) to 108 under multiple testing. As before, the CAPM cannot explain these significant strategies, but the CH3 and CH4 models explain most of the strategies based on accounting signals and some based on trading signals. The number of significant alphas is 37 for the CH3 model and 21 for the CH4 model, with more than half of the significant CH3 alphas (20) and more than one third of the significant CH4 alphas (8) coming from the liquidity category and the rest distributing about evenly across the risk, past returns, and profitability categories.

In sum, the higher *t*-statistic cutoff under multiple testing substantially reduces the significance rate of univariate strategies. Very few of the value-weighted strategies survive the higher hurdle rate after the CH3 and CH4 risk adjustments. Among the 30 or so equal-weighted strategies that do survive the higher hurdle rate and CH3 and CH4 risk adjustments, the majority are based on trading signals, liquidity measures in particular, while accounting signals largely fail to generate significant alphas.

4.3 Details on Significant Univariate Strategies

In this subsection, we provide detailed results on the significant univariate strategies. We focus on the 38 strategies with significant value-weighted high-minus-low raw returns under multiple testing, as indicated in Table 4. Table 5 reports their value-weighted and equal-weighted average raw returns and CH4 alphas as well as the associated *t*-statistics.[4]

[Insert Table 5 Here]

Among the 38 significant univariate strategies, five are liquidity measures, two are risk proxies, one is based on past returns, 20 are profitability measures, one is a value measure, and the remaining nine are from the others category. The significant liquidity measures are prior-month average daily share turnover, variation of daily share turnover, variation of daily RMB trading volume, turnover-adjusted number of zero daily trading volume, and GUBA media coverage. Using prior-month turnover-adjusted number of zero daily trading volume as an example, the average value-weighted high-minus-low raw return is 1.22% per month with a *t*-statistic of 3.12, and its CH4 alpha is much smaller at 0.23% per month with a *t*-statistic of 0.60. The average equal-weighted high-minus-low raw return is 1.46% per month with a *t*-statistic of 5.93, and its CH4 alpha is still substantial at 0.62% per month with a *t*-statistic of 2.30. For the other liquidity measures, the average value-weighted high-minus-low raw returns range from -1.17% to -1.56% per month with *t*-statistics from -2.87 to -3.71, and their CH4 alphas range from -0.16% to -0.69% with *t*-statistics from -0.46 to -2.22. The average equal-weighted high-minus-low raw returns range from -1.70% to -2.10% per month with *t*-statistics all above 5 in absolute value, and their CH4

[4] The complete results for all univariate strategies are reported in Internet Appendix D Table D2.

alphas range from -0.68% to -1.07% per month with *t*-statistics from -2.82 to -5.16. Overall, the evidence is consistent with a liquidity premium in the Chinese stock market, and the CH4 model with a liquidity factor can only partially explain it.

The two significant risk proxies are idiosyncratic skewness and total skewness. For idiosyncratic skewness, we compute it as the skewness of residuals from regressing a stock's daily excess returns on the CAPM model over the prior month. The average value-weighted and equal-weighted high-minus-low raw returns are -0.67% and -0.74% per month with *t*-statistics of -3.15 and -4.87, respectively. Therefore, Chinese firms with higher idiosyncratic skewness underperform firms with lower idiosyncratic skewness. After the CH4 risk adjustments, the value-weighted and equal-weighted alphas actually increase in magnitude to -0.90% and -1.01% per month with *t*-statistics of -2.78 and -5.29, respectively, indicating that the CH4 model exacerbates the anomaly. The results for prior-month total skewness are somewhat weaker. The average value-weighted and equal-weighted high-minus-low raw returns are -0.88% and -0.74% per month with *t*-statistics of -2.96 and -4.48, respectively, and their CH4 alphas are -0.52% and -0.73% per month with *t*-statistics of -1.46 and -3.82, respectively.

There is only one significant strategy based on past returns, which is the seasonality return between year *t*-2 and *t*-5 (the average return across months *t*-24, *t*-36, *t*-48, and *t*-60). The average value-weighted and equal-weighted high-minus-low raw returns are 0.76% to 0.49% per month with *t*-statistics of 2.93 and 3.13, respectively. Thus, the seasonality anomaly of Heston and Sadka (2008) also exists in the Chinese market. Again, the CH4 model cannot explain the anomaly. The value-weighted and equal-weighted CH4 alphas are 0.89% and 0.49% per month with *t*-statistics of 2.52 and 2.56, respectively.

The 20 significant profitability measures include return on equity, change in return on equity, return on assets, change in return on assets, standard unexpected earnings, return on net operating assets, assets turnover, gross profits-to-lagged assets, operating profits-to-lagged book equity, operating profits-to-lagged assets, and sales growth. The average value-weighted high-minus-low raw returns are very large ranging from 0.49% to 1.26% per month with a minimum *t*-statistic of 2.90. However, controlling for the CH4 factors reduces the alphas substantially to 0.20–0.65% per month, all but two of which have *t*-statistics that are less than 2. The average equal-weighted high-minus-low raw returns range from 0.44% to 1.24% per month with a minimum *t*-statistic of 3.09. Similar to the value-weighted results, the CH4 alphas are considerably lower ranging from 0.16% to 0.78% per month with *t*-statistics from 0.84 to 4.54.

The only significant value measure is quarterly earnings-to-price ratio. The average value-weighted high-minus-low raw return is 1.23% per month with a *t*-statistic of 3.43. The CH4 model which contains a value factor can explain it, leaving a tiny alpha of 0.08% with a *t*-statistic of 0.38. The average equal-weighted high-minus-low raw return and CH4 alpha are higher than their value-weighted counterparts at 1.35% and 0.39% per month with *t*-statistics of 5.04 and 2.51, respectively.

The remaining nine significant strategies are from the others category, including tangibility, asset liquidity, and tax expense surprises. The average value-weighted high-minus-low raw returns range from 0.46% to 1.02% per month with a minimum *t*-statistic of 3.12. The CH4 model can largely explain them, reducing alphas to 0.34–0.92% per month with all but two of the *t*-statistics below 2. The average equal-weighted high-minus-low raw returns range from 0.33% to 0.54% per month with *t*-statistics from 1.50 to 3.69. The CH4 model reduces the alphas to 0.20–0.47% per month with a maximum *t*-statistic of 1.88.

To heuristically compare the magnitude of high-minus-low raw returns and CH4 alphas across the 38 significant strategies, we plot them in Figure 3. The blue columns represent the absolute values of the value-weighted or equal-weighted high-minus-low raw returns, and the red columns represent the absolute values of their CH4 alphas. The strategies are ordered, from the largest to the smallest, based on their absolute high-minus-low raw returns. Panel A shows that the absolute value-weighted high-minus-low raw returns vary from 0.46% per month for tax expense surprises to 1.56% per month for prior-month variation of daily share turnover, and the absolute CH4 alphas vary from 0.08% per month for quarterly earnings-to-price ratio to 0.92% per month for asset liquidity. Panel B shows that the absolute equal-weighted high-minus-low raw returns vary from 0.33% per month for tangibility to 2.10% per month for GUBA media coverage, and the absolute CH4 alphas vary from 0.16% per month for change in return on equity to 1.07% per month for GUBA media coverage. Overall, strategies based on trading signals appear to have larger absolute high-minus-low raw returns and CH4 alphas than strategies based on accounting signals.

[Insert Figure 3 Here]

4.4 Comparison with U.S. Anomaly Patterns

To gain perspective on the anomaly patterns in China, we compare them with the patterns in the U.S., focusing on the significance rate and information content. Hou, Xue, and Zhang (2020) and Hou, Mo, Xue, and Zhang (2021) present comprehensive evidence on anomaly patterns in the U.S. We tabulate their results and report them next to our results from China in Table 6. We consider significant univariate strategies under both single testing and multiple testing. For multiple testing, the *t*-statistic cutoff is set to 2.85 for Chinese strategies and 2.78 for U.S. strategies following Hou, Xue, and Zhang (2020). For risk adjustments, we use the CH4 model for Chinese strategies and the Hou, Mo, Xue, and Zhang (2021) q^5 model for U.S. strategies.[5]

[Insert Table 6 Here]

Panel A of Table 6 presents comparison results on value-weighted strategies. For high-minus-low raw returns, the significance rate across all categories is 23% for Chinese strategies under single testing vs. 35% for U.S. strategies. After risk adjustments, the significance rate drops to 5% in China vs. 5% in the U.S. Using the higher multiple-testing t-statistic cutoff, the significance rate for high-minus-low raw returns shrinks considerably to 8% in China vs. 18% in the U.S., and it drops further to 0% in China vs. 1% in the U.S. after risk adjustments. For the equal-weighted strategies reported in Panel B, the pattern is similar. For example, the significance rate for high-minus-low raw returns is 42% in China under single testing vs. 56% in the U.S., and it drops to 24% in China under multiple testing vs. 47% in the U.S.[6] Overall, the significance rate is lower, especially in raw returns, for the univariate strategies in China than those in the U.S.

Why is the significance rate lower in China? There are two potential explanations. First, it is possible that due to the relatively opaque information environment of the Chinese stock market, the signals we collect might be too noisy or weak to predict returns reliably. Second, the history of the Chinese stock market is relatively short and dominated by two major crises (the 2008 Global Financial Crisis and the 2015 Chinese stock market crash). The short time series and the high level of uncertainty could further reduce the t-statistics and consequently the number of significant strategies.

We are also interested in the relative importance of trading or accounting signals for the

[5] The q^5 model is first introduced in Hou, Mo, Xue, and Zhang (2021) with five factors: market, size, profitability, investment, and expected investment growth. It has been shown to perform better than competing factor models in explaining the cross-section of U.S. stock returns.

[6] After risk adjustments, the significance rate in China drops to 16% under single testing and 5% under multiple testing. Hou, Mo, Xue, and Zhang (2021) do not report the q^5 alphas for equal-weighted strategies in the U.S.

significant strategies in China vs. the U.S. Therefore, in Table 6, we also report the contribution of each group of strategies to the overall significance rate. Panel A of Table 6 shows that the contribution of trading signals in China to the overall significant rate in value-weighted raw returns is 8% out of 23% (a 35% share) under single testing and 2% out of 8% (a 25% share) under multiple testing. By comparison, the contribution of trading signals in the U.S. is 9% out of 35% (a 26% share) under single testing and 7% out of 18% (a 39% share) under multiple testing. After risk adjustments, the contribution of trading signals is 2% out of 5% (a 40% share) in China under single testing and 0% out of 0% under multiple testing, and 1% out of 5% (a 20% share) in the U.S. under single testing and 0% out of 1% (a 0% share) under multiple testing. Overall, trading signals contribute more, and consequently accounting signals contribute less, to the significant strategies in China than in the U.S. The relative importance of trading signals in China is even more evident for the equal-weighted strategies in Panel B of Table 6. For example, the contribution of trading signals in China to the overall significance rate in equal-weighted raw returns is 21% out of 42% (a 50% share) under single testing and 11% out of 24% (a 46% share) under multiple testing, compared to 18% out of 56% (a 32% share) in the U.S. under single testing and 15% out of 47% (a 32% share) under multiple testing.

Why do trading signals play a more important role in China than in the U.S., while the opposite is true for accounting signals? We think the answer again lies in the different features between the two markets. First, more than half of the daily trading volume in China comes from inexperienced and speculative retail investors, who chase past performance, provide liquidity, and create price volatility and noise. While this excessive trading leads to low returns for the retail investors, it also generates predictable return patterns that can be picked up by trading signals. In contrast, trading frictions are much less severe in the U.S., a mature market dominated by institutional investors. Consequently, the relation between trading signals and future stock returns is less pronounced in the U.S. Second, due to China's opaque information environment, accounting signals in China are less informative about firm fundamentals and therefore are less useful in predicting future price movements. On the other hand, accounting signals in the U.S. are more reflective of firm fundamentals and consequently are associated with more significant return predictability.

Overall, the comparison of anomaly patterns between China and the U.S. reveals that the significance rate is lower in China than in the U.S., and that trading signals play a more important role in China than in the U.S.

4.5 Anomalies and Aggregate Market Conditions

The differences between China and the U.S. in anomaly patterns are consistent with evidence from the international anomaly literature that cross-country anomaly patterns heavily depend on country-level macro and market conditions. In this subsection, we directly link anomaly returns in China to these factors.

4.5.1 Aggregate Market-Level Variables

The international anomaly literature has documented three country-level variables that are important in explaining cross-country differences in anomaly returns: trading frictions (Watanabe, Xu, Yao, and Yu, 2013; Jacobs, 2016), financial market development (McLean, Pontiff, and Watanabe, 2009; Titman, Wei, and Xie, 2013), and accounting quality (Watanabe, Xu, Yao, and Yu, 2013). Following these studies, we measure trading frictions in China, FRIC, as the average firm-level idiosyncratic volatility, defined as the standard deviation of residuals obtained from regressing a stock's daily excess returns in a given month on the CH4 model. To measure China's financial market development, we compute DEV as the ratio of total market capitalization to GDP.

We measure China's accounting quality, ACCQ, as the average value across firms of the accounting data quality grades from CNRDS.

The U.S. anomaly literature also relates anomaly returns to investor sentiment and market liquidity. Following Baker, Wugler, and Yuan (2012), we construct China's investor sentiment index, SENT, as the first principal component of market turnover, first-month IPO return, number of IPO firms, and volatility premium. To measure the market liquidity in China, we follow Chordia, Subrahmanyam, and Tong (2014) and compute LIQ as the average monthly share turnover across firms.

Finally, given the strong presence of regulations in the Chinese stock market, we construct a regulation index, REGU, as the average value of the percentages of listed firms that are IPOs, that allow margin trading and short selling, and that have completed the split-share reform, in order to capture important regulatory interventions on investor composition and trading behavior. Details on the construction of the above six aggregate market-level variables are provided in Internet Appendix E.

Panel A of Table 7 reports the summary statistics of the market-level variables. It is worth noting that the correlations among the variables are all positive, and in some cases fairly high. For example, FRIC is highly correlated with LIQ (0.81), and REGU is highly correlated with both DEV (0.84) and ACCQ (0.84). Therefore, we are cautious in interpreting individual coefficients in later discussions.

[Insert Table 7 Here]

4.5.2 Panel Regressions

To understand whether and how the market-level variables influence anomaly returns in China, we follow a similar specification of Jacobs (2016) to estimate the following panel regressions using all 454 univariate strategies as well as trading-based and accounting-based strategies separately:

$$R_{l,t}^{high^*} - R_{l,t}^{low^*} = b_0 + b_1 FRIC_t + b_2 DEV_t + b_3 ACCQ_t + b_4 SENT_t + b_5 LIQ_t + b_6 REGU_t + c_1 MKT_t + c_2 SMB_t + c_3 VMG_t + c_4 PMO_t + e_{l,t}, \quad (4)$$

where $R_{l,t}^{high^*}$ and $R_{l,t}^{low^*}$ represent the returns of decile 10 and decile 1 in month t, respectively, for strategy l, with the decile rankings re-aligned to ensure that the average high-minus-low return is positive over the sample period. We include on the right-hand side the six market-level variables. We also include the CH4 factors to control for systematic risk factors in China. The panel regressions are estimated with strategy fixed effects, and the standard errors are double-clustered by strategy and month.

The regression coefficient estimates are reported in Panel B of Table 7, with columns I-III for value-weighted strategies and columns IV-VI for equal-weighted strategies. Across the six specifications, the coefficients on FRIC are all negative and marginally significant in half of them. The negative coefficients indicate that higher trading frictions are associated with smaller anomaly returns. The coefficients on DEV are also negative in all six specifications and are significant in four, suggesting that more developed financial markets are associated with smaller anomaly returns, especially for trading-based strategies. The coefficients on ACCQ are negative and significant in all specifications, implying that higher accounting quality is also associated with smaller anomaly returns in China. The coefficients on SENT are negative but insignificant in all specifications. The coefficients on LIQ are positive in all specifications and are significant in all but one cases, suggesting that improved market liquidity actually enhances anomaly returns. This might seem surprising because better market liquidity would make it easier for sophisticated investors to arbitrage away anomaly profits. However, in the case of China, the market-level liquidity is likely driven by speculative retail investors who trade too often, and this "excessive" liquidity could exacerbate anomaly returns.[7] Finally, the coefficients on REGU are also positive in all specifications and are significant in all but one specifications, which suggests that regulations in China are associated with larger anomaly returns. This is possibly due to the fact that the regulatory events captured by our index are designed to boost the stock market development, which result in bullish responses from the (mostly retail) market participants and consequently lead to larger anomaly returns.

5. Empirical Results on Composite Strategies

Given the low significance rate of the univariate strategies in China, readers might be wondering whether profitable investment strategies can be constructed in China at all. One possibility is to combine multiple signals and construct composite strategies. That is, with multiple signals available, investors could routinely evaluate the return predictive power of individual signals and dynamically adjust their portfolio exposures to different signals to maximize returns. In this section, we construct composite strategies using four different approaches and evaluate their performance in China.

5.1 Constructing Composite Strategies

To enter the composite strategies, we require a signal to be non-missing for at least 50% of the firm-month observations during the initial training window (July 2000 – June 2010). In

[7] To examine whether retail investors exacerbate anomaly returns in a rigorous way, we decompose market turnover into retail investor turnover and large trader turnover. Following the spirit of Lee and Radhakrishna (2000), Barber, Oden, and Zhu (2009), and Jiang, Liu, Peng, and Wang (2022), which assume that the trades with the largest size are more likely to come from large and sophisticated investors, we estimate large trader turnover as trading volume from the largest trade scaled by free-floating A-shares outstanding. Similarly, we estimate retail investor turnover as trading volume from the other trade groups scaled by free-floating A-shares outstanding. Internet Appendix E reports the details on the construction of these two types of turnover. We replace market turnover in panel regression (4) with retail investor turnover or large trader turnover or both, respectively. The results are shown in Internet Appendix E Table E1. We find that large trader turnover has no significant effect on anomaly returns, whereas retail investor turnover has positive and significant effect on anomaly returns.

addition, it cannot be missing for the entire cross-section of stocks in any month after its coverage starts. Altogether, 173 of the 208 signals we consider satisfy the above requirements. We follow the literature and replace firm-month observations with missing signals with their cross-sectional medians and then rank all stocks each month based on each signal and standardize the rankings to [0, 100].[8]

The first approach we use to construct composite strategies is the composite score method, as in Stambaugh, Yu, and Yuan (2015). For each signal j and each month *t*, we re-align the standardized rankings to ensure that the average return spread between the top and bottom deciles is positive over an expanding window from July 2000 to month *t* (a minimum of ten years). A stock i's re-aligned ranking is denoted $rank(i,j,t)$. The composite score for the stock in month *t*, $score(i,t)$, is the average of the standardized rankings across all available signals, i.e., $score(i,t) = \frac{1}{J}\sum_{j=1}^{J} rank(i,j,t)$. We then sort stocks into deciles based on their composite scores and long stocks in the top composite score decile and short stocks in the bottom decile. We compute both value-weighted and equal-weighted high-minus-low decile returns for month *t*+1 and rebalance the deciles monthly. The advantage of the composite score method is that it is straightforward to compute, and it combines information from individual signals while at the same time diversifies noise in individual signals.

The second approach is the multiple regression method from Lewellen (2015). For each month *t*, we first estimate a cross-sectional regression of stock returns, $R_{i,t}$, on a set of individual signals:[9]

[8] Our results are robust to using percentile rankings instead of standardized rankings. The robustness results are reported in Internet Appendix F Table F2.
[9] We use standardized rankings for returns and individual signals to minimize the impact of outliers on the regression coefficient estimates.

$$R_{i,t} = \theta_{0,t} + \sum_{j=1}^{J} \theta_{j,t} signal^{r}_{i,j,t-1} + e_{i,t}, \tag{5}$$

where $signal^{r}_{i,j,t-1}$ is the standardized ranking of signal j for stock i in month t-1. We obtain the average intercept, $\widehat{avg\theta}_{0,t}$ and the average slope coefficient, $\widehat{avg\theta}_{j,t}$, using the expanding window from July 2000 to month t (a minimum of ten years) and then compute the out-of-sample return forecast for stock i in month t+1 as $\hat{R}_{i,t+1} = \widehat{avg\theta}_{0,t} + \sum_{j=1}^{J} signal^{r}_{i,j,t} \widehat{avg\theta}_{j,t}$, using information available up to month t. Finally, we sort stocks into deciles based on their expected return forecasts and long stocks in the top decile and short stocks in the bottom decile. Again, both value-weighted and equal-weighted high-minus-low decile returns are computed for month t+1, and the deciles are rebalanced monthly. Compared to the composite score method which is nonlinear, the multiple regression method considers multiple signals simultaneously using a linear setup.

Our third method, Lasso, is also based on linear regressions using standardized data. The difference is that it adds a penalty function to overcome model overfitting when there are many signals with potential collinearity. In a nutshell, the Lasso method encourages simple and sparse models (models with less parameters), which can be thought of as a model selection method to manage high dimensionality and avoid overfitting the data. Details on the Lasso method are provided in Internet Appendix F. We follow Gu, Kelly, and Xiu (2020) and estimate a modified version of Equation (5) using panel data over the expanding window from July 2000 to month t (a minimum of ten years) and the Huber loss function as the penalty term. Using the estimated coefficients, we compute the out-of-sample return forecasts for month t+1 and sort stocks into deciles, and long stocks in the top expected return decile and short stocks in the bottom decile.

The final approach we consider, random forest, is a non-parametric method. Random forest is an ensemble method known as bootstrap aggregation or “bagging” that combines forecasts from many different scenarios into a single forecast. We follow the implementation of random forecast in Gu, Kelly, and Xiu (2020) and provide the details in Internet Appendix F. The output of random forest is a return forecast for each stock and each month. We use the expanding window from July 2000 to month *t* (a minimum of ten years) to estimate the return forecasts for month *t*+1. Then, similar to the regression-based methods, we sort stocks into expected return deciles and long stocks in the top decile and short stocks in the bottom decile. Unlike the regression-based methods, however, the non-parametric estimation of random forest is able to capture nonlinearity in the relations between signals and returns, and it allows for interaction effects among signals. In addition, with a robust number of scenarios, it is less likely to overfit the data and thus could result in stable out-of-sample performance.[10]

5.2 Performance of Composite Strategies

Table 8 reports the average high-minus-low raw returns and CH4 alphas of the composite strategies based on the 173 available signals. Note that for all four methods of constructing composite strategies, we use the first 10 years of data to establish the initial estimates. Therefore, the sample period for Table 8 is from July 2010 to December 2020.

[Insert Table 8 Here]

Panel A reports the value-weighted returns of the composite strategies. For the composite score method, the average high-minus-low raw return is 1.66% per month with a highly significant *t*-statistic of 3.40. After risk adjustments, the CH4 alpha decreases to 0.22% per month with a *t*-statistic of 0.61. For the multiple regression method, the average high-minus-low raw return is higher at 2.52% per month with a *t*-statistic of 5.78, and the CH4 alpha is 1.09% per month with a

[10] Lasso and random forest are relatively simple and stable compared to other more advanced machine learning methods, such as neural networks. In this paper, we do not study those methods as our goal is to simply demonstrate that machine learning presents an efficient way of combining signals to improve the performance of anomaly strategies.

t-statistic of 3.47.

For the Lasso method, the average high-minus-low raw return is 2.68% per month with a *t*-statistic of 5.20. The magnitude of the raw return spread is much larger than that of the composite score method and slightly larger than that of the multiple regression method. The improvement in performance is likely due to Lasso overcoming in-sample over-fit with the regularization technique, which achieves dimension reduction by dropping partially redundant and noisy signals. After risk adjustments using the CH4 model, the alpha is 1.12% per month with a *t*-statistic of 2.31, which again is better than the composite score and multiple regression methods.

For the random forest method, the average high-minus-low raw return is 2.86% per month with a *t*-statistic of 6.24, which is larger in magnitude and statistically more significant than the return spreads for the other three methods. After risk adjustments, the CH4 alpha is still a substantial 1.28% per month with a *t*-statistic of 3.81, again the best of the four methods. The superior performance of random forest is consistent with the findings of Gu, Kelly, and Xiu (2020) and suggests that random forest, with its advanced and flexible algorithms that allow for nonlinearity and multiway signal interactions, is more efficient at extracting important and dominant return predicting signals from the noisy information environment in the Chinese stock market.

How does the performance of composite strategies compare to that of univariate strategies? Table 5 and Figure 3 show that, for the 38 significant univariate strategies, the absolute value-weighted high-minus-low raw returns range from 0.46% to 1.56% per month, and their absolute CH4 alphas range from 0.08% to 0.92% per month. Thus, the performance of the composite strategies (1.66–2.86% per month in high-minus-low raw returns and 0.22–1.28% per month in CH4 alphas) largely dominate that of the significant univariate strategies. In fact, the composite strategies based on multiple regression, Lasso, and random forest (2.52%, 2.68%, and 2.86% per month in high-minus-low raw returns and 1.09%, 1.12%, and 1.28% per month in CH4 alphas, respectively) substantially outperform the very best univariate strategies (1.56% per month in high-minus-low raw returns and 0.92% per month in CH4 alphas).

Panel B reports the equal-weighted returns of the composite strategies, which again compare favorably with the univariate strategies. The average high-minus-low raw returns range from 2.03% and 3.06% per month for the composite score and random forest methods, respectively, to 3.16% and 3.24% per month for the multiple regression and Lasso methods, respectively, all with *t*-statistics greater than 6.[11] For comparison, the absolute equal-weighted high-minus-low raw returns for the 38 significant univariate strategies range from 0.33% to 2.10% per month, as shown in Table 5 and Figure 3. The CH4 alphas of the composite strategies range from 0.87% and 1.62% for the composite score and random forest methods, respectively, to 1.82% and 1.86% for the multiple regression and Lasso methods, respectively, with a minimum *t*-statistic of 3.27. For comparison, the absolute CH4 alphas of the significant univariate strategies range from 0.16% to 1.07% per month. Thus, similar to the value-weighted results, the composite strategies largely outperform the significant univariate strategies in both raw and risk-adjusted returns. Again, the composite strategies based on multiple regression, Lasso, and random forest strictly dominate the best-performing univariate strategies in both raw and risk-adjusted returns.

Since composite strategies dynamically shift weights between different signals, we are interested in what types of signals contribute the most to these strategies over time. Here we focus on the better performing composite strategies based on Lasso and random forest. Following Gu, Kelly, and Xiu (2020), we calculate the weight of a signal in Lasso as the squared coefficient on

---

[11] The average equal-weighted high-minus-low raw returns are always larger in magnitude and statistically more significant than their value-weighted counterparts.

that signal divided by the sum of squared coefficients across all signals. For random forest, we calculate the weight of a signal as the average decrease in mean squared forecast errors across regression trees.[12] We normalize the weights to sum to one within a model to allow for easy interpretation across signals.

In Figure 4, we use blue and red bars to plot the aggregate weights of trading and accounting signals, respectively, for the composite strategies based on Lasso (Panel A) and random forest (Panel B). We also plot the top three most important individual signals from each group. Several interesting findings emerge. First, trading signals contribute much more than accounting signals to the composite strategies over the past decade, accounting for more than 80% of the weight on average. Second, for the composite strategy based on Lasso, prior-month variation of daily RMB trading volume (a trading signal) is the most important signal across both groups, with an average weight of 46%. The second and third most important signals are prior-month return and daily Dimson beta (both trading signals), with average weights of 12% and 7%, respectively. By comparison, the weights on individual signals are more evenly distributed for the composite strategy based on random forest, with the three most important signals being prior-month variation of daily RMB trading volume, average daily RMB trading volume, and variation of daily share turnover (all trading signals), with average weights of 20%, 8%, and 7%, respectively. Third, among accounting signals, the three most important ones for the composite strategy based on Lasso are quarterly earnings-to-price ratio, R&D expense-to-market equity, and sales growth, with average weights of 5%, 3%, and 2%, respectively. For the composite strategy based on random forest, the three most important accounting signals are quarterly earnings-to-price ratio, change in

[12] Random forest consists of a set of regression trees. Each regression tree contains a set of internal nodes and leaves. We calculate for each internal node the change in mean squared forecast errors for a signal before and after the split, divide it by the sum of changes across all internal nodes, and then average across regression trees to obtain the weight of the signal.

return on equity, and return on equity, with average weights of 3%, 1%, and 1%, respectively.

[Insert Figure 4 Here]

In sum, the results in this section show that composite strategies compare favorably to, and in the case of those based on the multiple regression, Lasso, and random forest methods significantly outperform, the best-performing univariate strategies. This suggests that investors can improve the profitability of their trading strategies in China by combining individual signals together into composite signals.

6. Including Microcaps

In this section, we conduct robustness checks on the univariate and composite strategies using all Chinese A-share stocks. That is, we now include the smallest 30% of stocks (microcaps), which are previously excluded from our main analysis.

[Insert Table 9 Here]

Panel A of Table 9 reports the results for the univariate strategies under both single testing and multiple testing. For brevity, we focus our discussion on the multiple-testing results. Among the 454 univariate strategies, 28 of them have significant value-weighted raw returns, which translate into a 6% significance rate. These numbers are lower than those for the all-but-micro main sample (38 significant strategies and 8% significance rate, Table 4 Panel A). After including microcaps, the number of significant trading-based strategies increases from 8 to 9, whereas the number of significant accounting-based strategies decreases from 30 to 19. After risk adjustments using the CH4 model, 4 of the value-weighted alphas (all associated with trading-based strategies) are significant, compared with 0 for the all-but-micro main sample. The results for equal-weighted returns are similar. The number of univariate strategies with significant equal-weighted raw returns is lower than that for the main sample (96 vs. 108), but the number of significant equal-weighted

CH4 alphas is higher (47 vs. 21), with the majority of the significant alphas associated with trading-based strategies.

Panel B of Table 9 reports the results for the composite strategies after including microcaps. The average high-minus-low raw returns and CH4 alphas are larger in magnitude than their counterparts for the all-but-micro main sample. For example, the average value-weighted high-minus-low raw returns for the composite strategies based on Lasso and random forest are 2.72% and 3.49% per month with *t*-statistics of 5.34 and 6.54, respectively, compared with 2.68% and 2.86% per month with *t*-statistics of 5.20 and 6.24 (Table 8 Panel A), respectively, for the main sample. The value-weighted CH4 alphas for the composite strategies based on Lasso and random forest are 1.27% and 1.91% per month with *t*-statistics of 2.85 and 5.10, respectively, compared with 1.12% and 1.28% with *t*-statistics of 2.31 and 3.81 (Table 8 Panel A), respectively, for the main sample. Results for equal-weighted returns are similar. Overall, our results are robust after we include microcaps in China.

7. Conclusion

This paper provides a comprehensive study of the cross-section of stock returns in China using 454 strategies over the sample period 2000–2020. Using the single-testing *t*-statistic cutoff of 1.96, 104 strategies have significant value-weighted high-minus-low raw returns, and 189 have significant equal-weighted raw returns. After risk adjustments using the CH3 and CH4 models of Liu, Stambaugh, and Yuan (2019), more than half of the alphas become insignificant. Most of the significant alphas are associated with trading-based strategies, those based on liquidity measures in particular. To address the concern of false discoveries, we recalibrate the *t*-statistic cutoff to 2.85 for the Chinese sample to accommodate multiple-testing adjustments. The higher hurdle rate reduces the number of significant strategies considerably. Only 38 (108) strategies have significant value-weighted (equal-weighted) raw returns, and the number further reduces to 0 (21) after CH4 risk adjustments.

We compare the anomaly patterns in China to those in the U.S. The significance rate is lower in China, suggesting that individual signals might be too weak or noisy to generate significant return predictability. This is potentially due to the fact that the Chinese stock market has an opaque information environment and is dominated by retail investors, symptoms of market inefficiency. In terms of information content, trading signals contribute more to the significant strategies in China than in the U.S., whereas the opposite is true for accounting signals. We also relate anomaly returns in China to aggregate market conditions and find that they comove with financial market development, accounting quality, market liquidity, and regulations.

Finally, we construct composite strategies by combining individual signals together and find that they, especially those based on machine learning methods, generate larger and statistically more significant returns than univariate strategies. Most of the composite strategies can pass the higher hurdle rate under multiple testing, even after risk adjustments.

References

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2009, High idiosyncratic volatility and low returns: International and further U.S. evidence, *Journal of Financial Economics* 91, 1–23.

Baker, Malcolm, Jeffrey Wugler, and Yu Yuan, 2012, Global, local, and contagious investor sentiment, *Journal of Financial Economics* 104, 272–287.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets?, *Review of Financial Studies* 22, 151–186.

Barras, Laurent Richard, Olivier Scaillet, and Russ Wermers, 2010, False discoveries in mutual fund performance: Measuring luck in estimated alphas, *The Journal of Finance* 65, 179–216.

Benjamini, Yoav, and Yosef Hochberg, 1995, Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society Series* 57, 289–300.

Benjamini, Yoav, and Daniel Yekutieli, 2001, The control of the false discovery rate in multiple testing under dependency, *Annals of Statistics* 29, 1165–1188.

Boone, Audra L., and Joshua T. White, 2015, The effect of institutional ownership on firm transparency and information production, *Journal of Financial Economics* 117, 508-533.

Chen, Andrew Y., 2021, The limits of *p*-hacking: Some thought experiments, *The Journal of Finance* 76, 2447–2480.

Chen, Andrew Y., and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 11, 207–264.

Chen, Xuanjuan, Kenneth A. Kim, Tong Yao, and Tong Yu, 2010, On the predictability of Chinese stock returns, *Pacific-Basin Finance Journal* 18, 403–425.

Chordia, Tarun, Amit Goyal, and Alessio Saretto, 2020, Anomalies and false rejections, *Review of Financial Studies* 33, 2134–2179.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong, 2014, Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?, *Journal of Accounting and Economics* 58, 41–58.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2021, Thousands of alpha tests, *Review of Financial Studies* 34, 3456–3496.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2017, The characteristics that provide independent information about average US monthly stock returns, *Review of Financial Studies* 30, 4389–4436.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Harvey, Campbell R., and Yan Liu, 2020, False (and missed) discoveries in financial economics, *The Journal of Finance* 75, 2503–2553.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Heston, Steven L., and Ronnie Sadka, 2008, Seasonality in the cross-section of stock returns, *Journal of Financial Economics* 87, 418–445.

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho, 2011, What factors drive global stock returns?, *Review of Financial Studies* 24, 2527–2574.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An augmented q-factor model with expected growth, *Review of Finance* 25, 1–41.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Hsu, Jason, Vivek Viswanathan, Michael Wang, and Phillip Wool, 2018, Anomalies in Chinese A-shares, *Journal of Portfolio Management* 44, 108–123.

Jacobs, Heiko, 2016, Market maturity and mispricing, *Journal of Financial Economics* 122, 270–287.

Jacobs, Heiko, and Sebastian Muller, 2020, Anomalies across the globe: Once public, no longer existent?, *Journal of Financial Economics* 135, 213–230.

Jansen, Maarten, Laurens Swinkels, and Weili Zhou, 2021, Anomalies in the Chinese A-share market, *Pacific-Basin Finance Journal* 68, 101607.

Jensen, Theis Ingerslev, Bryan T. Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *The Journal of Finance* 78, 2465–2518.

Jiang, Lei, Jinyu Liu, Lin Peng, and Baolian Wang, 2022, Investor attention and asset pricing anomalies, *Review of Finance* 26, 563–593.

Lee, Charles M.C., and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from torq data, *Journal of Financial Markets* 3, 83–111.

Leippold, Markus, Qian Wang, and Wenyu Zhou, 2022, Machine learning in the Chinese stock market, *Journal of Financial Economics* 145, 64–82.

Lewellen, Jonathan, 2015, The cross-section of expected stock returns, *Critical Finance Review* 4, 1–44.

Li, Zhibing, Laura Xiaolei Liu, Xiaoyu Liu, and K.C. John Wei, 2023, Replicating and Digesting Anomalies in the Chinese A-share Market, *Management Science* forthcoming.

Linnainmaa, Juhani T., and Michael R. Roberts, 2018, The history of the cross-section of stock returns, *Review of Financial Studies* 31, 2606–2649.

Liu, Jianan, Robert F. Stambaugh, and Yu Yuan, 2019, Size and value in China, *Journal of Financial Economics* 134, 48–69.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destory stock return predictability?, *The Journal of Finance* 71, 5–32.

McLean, R. David, Jeffrey Pontiff, and Akiko Watanabe, 2009, Share issuance and cross-sectional returns: International evidence, *Journal of Financial Economics* 94, 1–17.

Newey, Whitney K., and Kenneth D. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Sharpe, William F. , 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19, 425–442.

Stambaugh, Robert F., Jianfeng Yu, and Yuan Yu, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *The Journal of Finance* 70, 1903–1948.

Titman, Sheridan, John K. C. Wei, and Feixue Xie, 2013, Market development and the asset growth effect: International evidence, *Journal of Financial and Quantitative Analysis* 48, 1405–1432.

Watanabe, Akiko, Yan Xu, Tong Yao, and Tong Yu, 2013, The asset growth effect: Insights from

international equity markets, *Journal of Financial Economics* 108, 529–563.
Wu, Huihang, Xingkong Wei, and Xiaoyan Zhang, 2021, Are stock returns predictable in China? A machine learning approach, *working paper*.

Table 1. List of Signals

This table lists the 208 individual signals. Based on the information content of the signals, they are separated into two broad groups: trading-based signals and accounting-based signals. The trading signals are further divided into three subgroups: liquidity, risk, and past returns. The accounting signals are also divided into four subgroups: profitability, value, investment, and others. The table reports the number and list of signals for each subgroup. Internet Appendix B provides the detailed definition of each signal.

Group	Subgroup	No. of Signals	List of Signals
Trading-based	Liquidity	35	size, 1-month, 6-month, and 12-month share turnover, 1-month, 6-month, and 12-month variation of share turnover, 1-month, 6-month, and 12-month coefficient of variation of share turnover, abnormal turnover, 1-month, 6-month, and 12-month RMB trading volume, 1-month, 6-month, and 12-month variation of RMB trading volume, 1-month, 6-month, and 12-month coefficient of variation of RMB trading volume, 1-month, 6-month, and 12-month Amihud illiquidity, 1-month, 6-month, and 12-month turnover-adjusted liquidity, return-return, illiquidity-illiquidity, return-illiquidity, liquidity-return, and net liquidity betas, GUBA postings, comments, and readings, web search volume index
	Risk	18	idiosyncratic volatility, idiosyncratic volatility per the CAPM, idiosyncratic volatility per the CH3 factor model, idiosyncratic volatility per the CH4 factor model, total volatility, idiosyncratic skewness per the CAPM, idiosyncratic skewness per the CH3 factor model, idiosyncratic skewness per the CH4 factor model, total skewness, co-skewness, market beta using monthly returns, market beta using daily returns, downside beta, Frazzini-Pedersen beta, Dimson beta, tail risk, firm paper news, firm internet news
	Past Returns	20	cumulative returns from month t-4 to t-2, from t-7 to t-2, from t-10 to t-2, from t-12 to t-2, from month t-36 to t-13, from t-60 to t-13, prior one-month return, industry return, industry lead-lag effect, cumulative return changes, 11-month and 6-month residual returns, 52-week high, maximum daily return, share price, cumulative abnormal returns around earnings announcement dates, seasonality returns in year t-1, non-seasonality returns in year t-1, seasonality returns between year t-2 and t-5, non-seasonality returns between year t-2 and t-5

Group	Subgroup	No. of Signals	List of Signals
Accounting-based	Profitability	35	return on equity, 4-quarter change in return on equity, return in assets, 4-quarter change in return on assets, standard unexpected earnings, revenue surprises, annual and quarterly return on net operating assets, annual and quarterly profit margin, annual and quarterly assets turnover, annual and quarterly capital turnover, gross profits-to-assets, annual and quarterly gross profits-to-lagged assets, operating profits-to-book equity, annual and quarterly operating profits-to-lagged book equity, operating profits-to-assets, annual and quarterly operating profits-to-lagged assets, annual and quarterly taxable income-to-book income, annual and quarterly book leverage, annual and quarterly sales growth, annual and quarterly fundamental score, annual and quarterly Ohlson's O-score, annual and quarterly Altman's Z-score
	Value	24	annual and quarterly book-to-market equity, book-to-June-end-market equity, annual and quarterly debt-to-market equity, annual and quarterly assets-to-market equity, annual and quarterly earnings-to-price, annual and quarterly cash flow-to-price, sales growth rank, annual and quarterly enterprise multiple, annual and quarterly sales-to-price, annual and quarterly operating cash flow-to-price, debt-to-book equity, intangible return, annual and quarterly enterprise book-to-price, annual and quarterly net debt-to-price
	Investment	37	abnormal corporate investment, annual and quarterly investment-to-assets, changes in PPE and inventory-to-assets, net operating assets, changes in net operating assets, 1-year, 2-year, and 3-year investment growth, net share issues, composite equity issuance, composite debt issuance, inventory growth, inventory changes, operating accruals, total accruals, changes in net non-cash working capital, in current operating assets, and in current operating liabilities, changes in net non-current operating assets, in non-current operating assets, and in non-current operating liabilities, changes in net financial assets, in short-term investments, in long-term investments, in financial liabilities, and in book equity, discretionary accruals, percent operating accruals, percent total accruals, percent discretionary accruals, quarterly current asset growth, non-current asset growth, quarterly cash growth, fixed asset growth, non-cash current asset growth, and other asset growth
	Others	39	advertising expense-to-market, growth in advertising expense, annual and quarterly R&D expense-to-market equity, annual and quarterly R&D expense-to-sales, annual and quarterly operating leverage, hiring rate, firm age, % change in sales minus % change in inventory, % change in sales minus % change in accounts receivable, % change in gross margin minus % change in sales, % change in sales minus % change in SG&A, effective tax rate, labor force efficiency, annual and quarterly tangibility, cash flow volatility, cash-to-assets, earnings persistence, earnings predictability, earnings smoothness, value relevance of earnings, earnings timeliness, earnings conservatism, annual and quarterly asset liquidity scaled by total assets, annual and quarterly asset liquidity scaled by market value of assets, tax expense surprises, changes in analyst earnings forecasts, revisions in analyst earnings forecasts, earnings forecast–to-price, analyst coverage, dispersion in analyst forecasts, institutional ownership, SOE indicator, margin trading and short selling indicator

Table 2. Summary Statistics of the Chinese and U.S. Stock Markets

This table reports the summary statistics of the Chinese stock market in comparison to the U.S. over the sample period 2000–2020, including the time-series averages and standard deviations of the number of listed firms and total market capitalization (in trillions of dollars), and pooled firm-level averages and standard deviations of firm size (in billions of dollars), book-to-market ratio (B/M), annual share turnover, and annual stock returns. The number of firms is the number of firms with trading records at the end of each year. The total market capitalization is the sum of all firms' market capitalization at the end of each year. Firm size is the product of closing price and number of shares outstanding at the end of each year. Firm B/M is the annual book-to-market ratio. Firm share turnover is the total number of shares traded in a year divided by the number of shares outstanding at the prior year-end. Annual stock return is the buy-and-hold return for each year.

	China		U.S.	
	Mean	Stdev	Mean	Stdev
No. of firms	2,103	890	4,292	759
Total market cap (trillions of dollars)	3.81	3.32	18.54	7.29
Firm size (billions of dollars)	1.81	8.09	4.32	24.25
Firm B/M	0.40	0.26	0.84	1.35
Firm annual share turnover	6.53	6.11	2.38	8.63
Annual stock return	0.18	0.70	0.12	0.66

Table 3. Significant Univariate Strategies under Single Testing

The table reports the numbers of univariate strategies with significant value-weighted (Panel A) and equal-weighted (Panel B) high-minus-low raw returns, CAPM alphas, CH3 alphas, and CH4 alphas, over the sample period from July 2000 to December 2020. The statistical significance is based on the Newey-West standard errors with four lags, using the conventional single-testing *t*-statistic cutoff of 1.96.

	Overall	Significance	Trading-based			Accounting-based			
		Rate	Liquidity	Risk	Past Returns	Profitability	Value	Investment	Others
	454		106	52	52	73	44	51	76
Panel A. Value-Weighted Strategies									
Raw return	104	23%	25	6	7	40	2	2	22
CAPM alpha	103	23%	25	6	7	40	2	2	21
CH3 alpha	37	8%	14	3	5	0	0	1	14
CH4 alpha	22	5%	3	2	5	4	0	1	7
Panel B. Equal-Weighted Strategies									
Raw return	189	42%	55	23	17	45	13	9	27
CAPM alpha	188	41%	55	22	17	45	13	9	27
CH3 alpha	91	20%	39	10	11	18	1	4	8
CH4 alpha	72	16%	27	10	12	14	2	3	4

Table 4. Significant Univariate Strategies under Multiple Testing

The table reports the numbers of univariate strategies with significant value-weighted (Panel A) and equal-weighted (Panel B) high-minus-low raw returns, CAPM alphas, CH3 alphas, and CH4 alphas, over the sample period from July 2000 to December 2020. The statistical significance is based on the Newey-West standard errors with four lags, using the multiple-testing t-statistic cutoff of 2.85.

	Overall	Significance	Trading-based			Accounting-based			
		Rate	Liquidity	Risk	Past returns	Profitability	Value	Investment	Others
	454		106	52	52	73	44	51	76
Panel A. Value-Weighted Strategies									
Raw return	38	8%	5	2	1	20	1	0	9
CAPM alpha	38	8%	5	2	1	20	1	0	9
CH3 alpha	4	1%	2	1	1	0	0	0	0
CH4 alpha	0	0%	0	0	0	0	0	0	0
Panel B. Equal-Weighted Strategies									
Raw return	108	24%	34	10	8	37	4	3	12
CAPM alpha	107	24%	34	9	8	37	4	3	12
CH3 alpha	37	8%	20	6	4	6	0	1	0
CH4 alpha	21	5%	8	4	3	6	0	0	0

Table 5. Detailed Results on Significant Univariate Strategies

The table reports the average high-minus-low raw returns and CH4 alphas as well as their associated t-statistics for the 38 univariate strategies with significant value-weighted high-minus-low raw returns under the multiple-testing t-statistic cutoff of 2.85. Details on the construction of the 38 strategies are provided in Internet Appendix B. The complete results for all univariate strategies are reported in Internet Appendix D Table D2.

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw return	t-stat	CH4 alpha	t-stat	Raw return	t-stat	CH4 alpha	t-stat
Liquidity								
turn: share turnover, k=1, m=0, n=1								
turn1-1	-1.26	-2.97	-0.16	-0.46	-1.70	-5.53	-0.68	-2.82
vturn: variation of share turnover, k=1, m=0, n=1								
vturn1-1	-1.56	-3.71	-0.69	-1.88	-1.89	-6.81	-1.01	-5.16
vdtv: variation of RMB trading volume, k=1, m=0, n=1								
vdtv1-1	-1.17	-2.89	-0.39	-2.22	-1.99	-6.38	-0.84	-4.14
Lm: turnover-adjusted number of zero daily trading volume, k=1, m=0, n=1								
Lm1-1	1.22	3.12	0.23	0.60	1.46	5.93	0.62	2.30
GUBA social media coverage, n=1								
post_num1	-1.43	-2.87	-0.64	-1.73	-2.10	-5.77	-1.07	-2.98
Risk								
isc: idiosyncratic skewness per the CAPM, k=1, m=0, n=1								
isc1	-0.67	-3.15	-0.90	-2.78	-0.74	-4.87	-1.01	-5.29
ts: total skewness, k=1, m=0, n=1								
ts1	-0.88	-2.96	-0.52	-1.46	-0.74	-4.48	-0.73	-3.82
Past Returns								
Ra25: seasonality returns between year t-2 and t-5								
Ra25	0.76	2.93	0.89	2.52	0.49	3.13	0.49	2.56
Profitability								
roe: return on equity, n=1, 6								
roe1	1.26	3.73	0.34	0.92	1.18	4.47	0.48	1.78
roe6	0.96	3.05	0.31	1.12	0.78	3.09	0.21	0.86
droe: 4-quarter change in return on equity, n=1, 6, 12								
droe1	1.00	4.23	0.62	1.97	1.10	6.81	0.71	4.01
droe6	0.78	4.25	0.35	1.51	0.75	5.81	0.39	2.72
droe12	0.49	3.12	0.20	1.28	0.44	3.97	0.16	1.25
roa: return on assets, n=1, 6								
roa1	1.23	3.62	0.53	1.38	1.24	4.56	0.64	2.07
roa6	1.01	3.25	0.48	1.47	0.86	3.21	0.35	1.15

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw return	*t*-stat	CH4 alpha	*t*-stat	Raw return	*t*-stat	CH4 alpha	*t*-stat
droa: 4-quarter change in return on assets, n=1, 6								
droa1	0.97	4.56	0.65	2.06	1.09	7.82	0.78	4.54
droa6	0.80	4.72	0.42	1.82	0.76	6.75	0.44	3.26
sue: standard unexpected earnings, n=1								
sue1	0.91	2.96	0.26	0.78	1.13	7.02	0.68	3.79
rna: return on net operating assets, n=1								
rnaq1	0.84	3.05	0.30	0.74	1.12	4.49	0.52	1.87
ato: assets turnover, annual sort								
ato	0.58	3.07	0.32	1.00	0.45	3.30	0.25	1.42
gpla: gross profits-to-lagged assets, n=1, 6								
gplaq1	1.15	3.15	0.53	1.35	1.22	4.45	0.59	1.95
gplaq6	0.97	2.90	0.43	1.23	0.83	3.11	0.29	0.99
ople: operating profits-to-lagged book equity, n=1, 6								
opleq1	1.20	3.70	0.35	1.07	1.21	4.74	0.54	2.29
opleq6	0.92	3.09	0.28	1.05	0.76	3.09	0.18	0.84
opla: operating profits-to-lagged assets, n=1, 6								
oplaq1	1.09	3.16	0.48	1.28	1.24	4.59	0.64	2.19
oplaq6	0.94	2.96	0.42	1.28	0.85	3.24	0.31	1.10
sg: sales growth, n=1, 6								
sgq1	0.91	3.65	0.57	1.79	1.01	5.97	0.73	3.61
sgq6	0.62	3.25	0.56	2.37	0.64	4.46	0.44	2.54
Value								
ep: earnings-to-price, n=1								
epq1	1.23	3.43	0.08	0.38	1.35	5.04	0.39	2.51
Others								
tan: tangibility, n=1, 6, 12								
tan	1.02	3.13	0.63	1.45	0.42	1.63	0.30	0.83
tanq1	0.92	3.26	0.75	1.75	0.33	1.50	0.22	0.76
tanq6	0.97	3.42	0.64	1.56	0.40	1.87	0.26	0.87
tanq12	0.96	3.41	0.63	1.53	0.46	2.09	0.29	0.93
ala: asset liquidity, n=1, 6, 12								
alaq1	0.97	4.05	0.92	2.34	0.52	3.07	0.47	1.88
alaq6	0.81	3.34	0.65	1.81	0.54	3.18	0.39	1.55
alaq12	0.74	3.12	0.59	1.72	0.49	2.82	0.36	1.41
tes: tax expense surprises, n=6, 12								
tes6	0.54	3.40	0.35	1.69	0.46	3.69	0.24	1.79
tes12	0.46	3.16	0.34	2.06	0.33	3.14	0.20	1.81

Table 6. Comparison with U.S. Anomaly Patterns

The table reports the significance rates of the univariate strategies in China and the U.S. under both single testing and multiple testing. The overall significance rate in a country is the number of significant strategies divided by the total number of strategies in that country. The significance rate for a group (trading-based or accounting-based) of strategies is the number of significant strategies for that group divided by the total number of strategies across groups. The multiple-testing t-statistic cutoff is set to 2.85 for Chinese strategies and 2.78 for U.S. strategies. Risk adjustments are based on the CH4 model for Chinese strategies and the q^5 model for U.S. strategies. The U.S. results are tabulated from Hou, Xue, and Zhang (2020) and Hou, Mo, Xue, and Zhang (2021), who do not estimate the q^5 alphas for the equal-weighted strategies in the U.S. Panel A reports the results for the value-weighted strategies. Panel B reports the result for the equal-weighted strategies.

Panel A. Significance Rate of Value-Weighted Strategies

	Overall		Trading-based		Accounting-based	
	Single	Multiple	Single	Multiple	Single	Multiple
China raw return	23%	8%	8%	2%	15%	7%
U.S. raw return	35%	18%	9%	7%	26%	11%
China alpha (CH4)	5%	0%	2%	0%	3%	0%
U.S. alpha (q^5)	5%	1%	1%	0%	4%	1%

Panel B. Significance Rate of Equal-Weighted Strategies

	Overall		Trading-based		Accounting-based	
	Single	Multiple	Single	Multiple	Single	Multiple
China raw return	42%	24%	21%	11%	21%	12%
U.S. raw return	56%	47%	18%	15%	38%	32%
China alpha (CH4)	16%	5%	11%	3%	5%	1%
U.S. alpha (q^5)	-	-	-	-	-	-

Table 7. Anomaly Returns and Aggregate Market Conditions

Panel A reports the summary statistics of six market-level variables: trading friction (FRIC), which is the average firm-level idiosyncratic volatility based on the CH4 model, financial market development (DEV), which is the ratio of total market capitalization to GDP, accounting quality (ACCQ), which is the average firm-level accounting data quality grade from CNRDS, investor sentiment (SENT), which is the first principal component of market turnover, first-month IPO return, the number of IPO firms, and the volatility premium, market liquidity (LIQ), which is the average firm-level monthly share turnover, and regulation (REGU), which is the average value of the percentages of listed firms that are recent IPOs, that allow margin trading and short selling, and that have completed the split-share reform. Panel B reports the regression coefficients and *t*-statistics from regressing the high-minus-low returns of all univariate strategies (and trading-based and accounting-based strategies separately) on the six market-level variables as well as the CH4 factors. The regressions are estimated with strategy fixed effects with standard errors double-clustered by strategy and month.

Panel A. Summary Statistics of Market-Level Variables

	Mean	Stdev	Correlation				
			DEV	ACCQ	SENT	LIQ	REGU
FRIC	0.26	0.07	0.50	0.23	0.21	0.81	0.34
DEV	0.45	0.25		0.64	0.67	0.54	0.84
ACCQ	2.85	0.19			0.52	0.31	0.84
SENT	0.00	1.00				0.28	0.73
LIQ	0.46	0.24					0.43
REGU	0.41	0.24					

Panel B. Panel Regressions

	Value-Weighted Strategies			Equal-Weighted Strategies		
	I	II	III	IV	V	VI
	All	Trading-based	Accounting-based	All	Trading-based	Accounting-based
FRIC	-0.15*	-0.26*	-0.06	-0.12	-0.24*	-0.02
	(-1.66)	(-1.70)	(-1.06)	(-1.41)	(-1.68)	(-0.22)
DEV	-0.16**	-0.33***	-0.02	-0.24**	-0.38***	-0.11
	(-2.11)	(-3.03)	(-0.29)	(-2.43)	(-2.89)	(-0.83)
ACCQ	-0.22***	-0.31***	-0.14*	-0.21***	-0.24**	-0.18**
	(-2.94)	(-2.67)	(-1.93)	(-3.03)	(-2.11)	(-1.96)
SENT	-0.01	-0.02	0.00	-0.05	-0.03	-0.08
	(-0.20)	(-0.23)	(-0.02)	(-1.00)	(-0.28)	(-1.30)
LIQ	0.30***	0.46***	0.16**	0.29***	0.50***	0.11
	(3.02)	(2.69)	(2.54)	(2.96)	(3.01)	(0.90)
REGU	0.35***	0.68***	0.07	0.43***	0.67***	0.23*
	(3.19)	(3.79)	(0.67)	(3.98)	(3.45)	(1.69)
MKT	-0.03***	-0.08***	0.01	-0.03***	-0.08***	0.01
	(-3.46)	(-5.34)	(1.58)	(-3.85)	(-5.87)	(1.19)
SMB	-0.03	0.06	-0.12***	-0.01	0.08**	-0.08***
	(-1.21)	(1.21)	(-5.31)	(-0.27)	(2.09)	(-3.81)
VMG	0.16***	0.10**	0.22***	0.17***	0.13***	0.21***
	(6.21)	(2.46)	(7.97)	(8.87)	(4.17)	(8.07)
PMO	0.06***	0.12***	0.01	0.04**	0.10***	-0.01
	(3.58)	(4.08)	(0.76)	(2.46)	(3.27)	(-0.36)
Observations	105,894	49,212	56,682	105,894	49,212	56,682
Adj.R^2	2.4%	2.3%	6.6%	4.5%	5.4%	9.0%

Table 8. Performance of Composite Strategies

The table reports the average high-minus-low raw returns and CH4 alphas as well as their associated *t*-statistics of the four composite strategies constructed using the composite score, multiple regression, Lasso, and random forest methods, over the period from July 2010 to December 2020. Data from July 2000 to June 2010 are used to establish the initial estimates of the composite signals. Panel A reports the results for the value-weighted composite strategies. Panel B reports the result for the equal-weighted composite strategies.

	Composite Score		Multiple Regression		Lasso		Random Forest	
	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat
Panel A. Value-Weighted Strategies								
Raw return	1.66	3.40	2.52	5.78	2.68	5.20	2.86	6.24
CH4 alpha	0.22	0.61	1.09	3.47	1.12	2.31	1.28	3.81
Panel B. Equal-Weighted Strategies								
Raw return	2.03	6.85	3.16	9.61	3.24	8.10	3.06	8.64
CH4 alpha	0.87	3.27	1.82	7.72	1.86	6.10	1.62	6.19

Table 9. Robustness Results Using All Chinese Firms

The table reports the robustness results using all Chinese firms (including the smallest 30% firms based on market capitalization which are excluded from the main analysis). Panel A reports the numbers of univariate strategies with significant high-minus-low raw returns and CH4 alphas under both single testing and multiple testing. Panel B reports the average high-minus-low raw returns and CH4 alphas as well as their associated t-statistics of the four composite strategies constructed using the composite score, multiple regression, Lasso, and random forest methods.

Panel A. Univariate Strategies

	Overall	Significance	Trading-based			Accounting-based			
		Rate	Liquidity	Risk	Past Returns	Profitability	Value	Investment	Others
	454		106	52	52	73	44	51	76
Value-Weighted Strategies									
Single Testing									
Raw return	98	22%	34	7	9	27	0	0	21
CH4 alpha	31	7%	17	3	5	1	0	0	5
Multiple Testing									
Raw return	28	6%	7	1	1	9	0	0	10
CH4 alpha	4	1%	3	1	0	0	0	0	0
Equal-Weighted Strategies									
Single Testing									
Raw return	172	38%	64	28	22	31	10	1	16
CH4 alpha	92	20%	53	13	12	9	1	0	4
Multiple Testing									
Raw return	96	21%	44	18	7	17	1	0	9
CH4 alpha	47	10%	37	7	1	2	0	0	0

Panel B. Composite Strategies

	Composite Score		Multiple Regression		Lasso		Random Forest	
	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat
Value-Weighted Strategies								
Raw return	1.67	3.64	2.66	5.89	2.72	5.34	3.49	6.54
CH4 alpha	0.24	0.72	1.28	3.87	1.27	2.85	1.91	5.10
Equal-Weighted Strategies								
Raw return	2.08	8.35	3.45	10.04	3.57	8.47	3.52	8.83
CH4 alpha	0.99	3.89	2.19	8.27	2.27	6.98	2.16	6.93

Figure 1. Market Characteristics of China and the U.S.

The figure plots, for both the Chinese and U.S. stock markets, the number of listed firms (Panel A), the total market capitalization in trillions of dollars (Panel B), the average firm-level share turnover (Panel C), and the average firm-level institutional ownership (Panel D), over the 2000–2020 sample period.

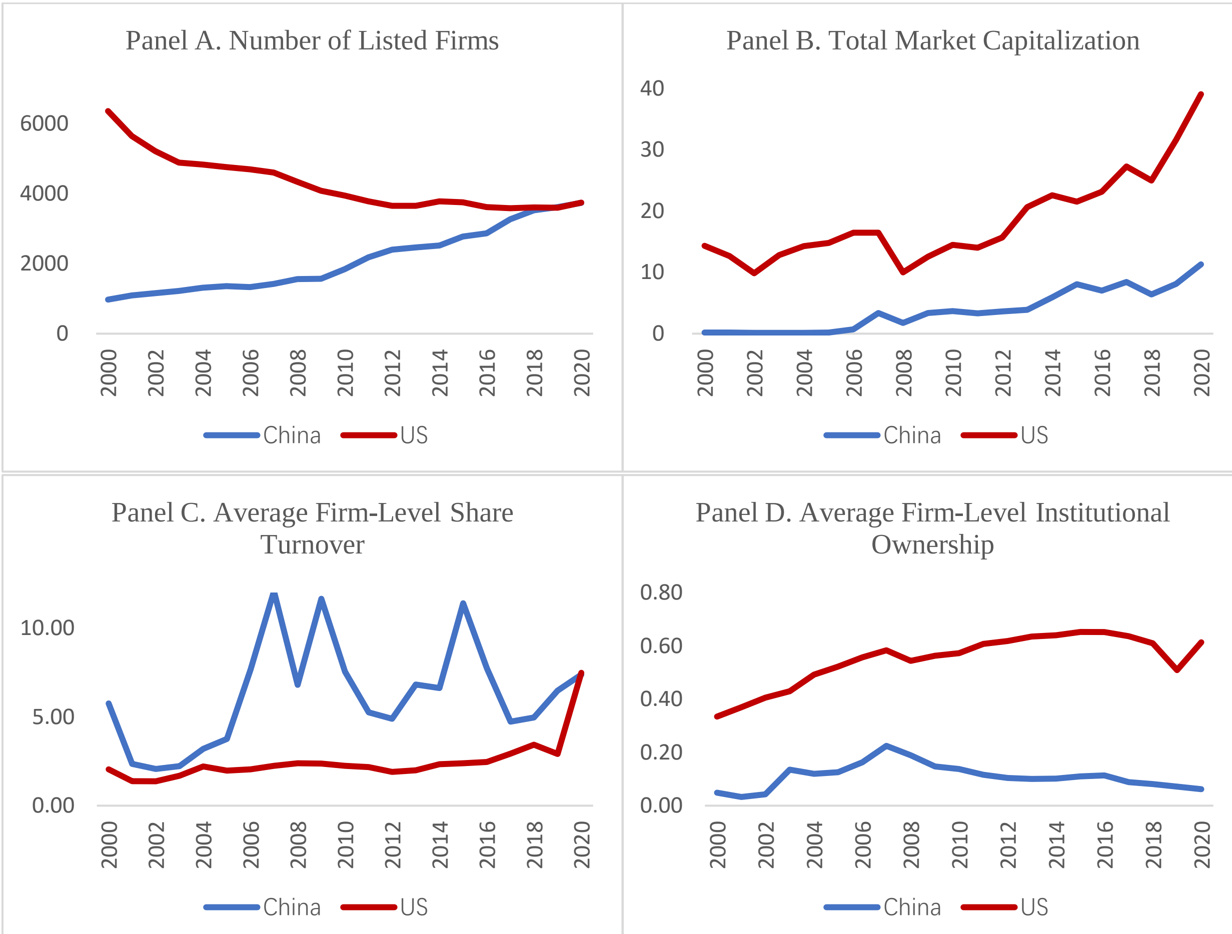
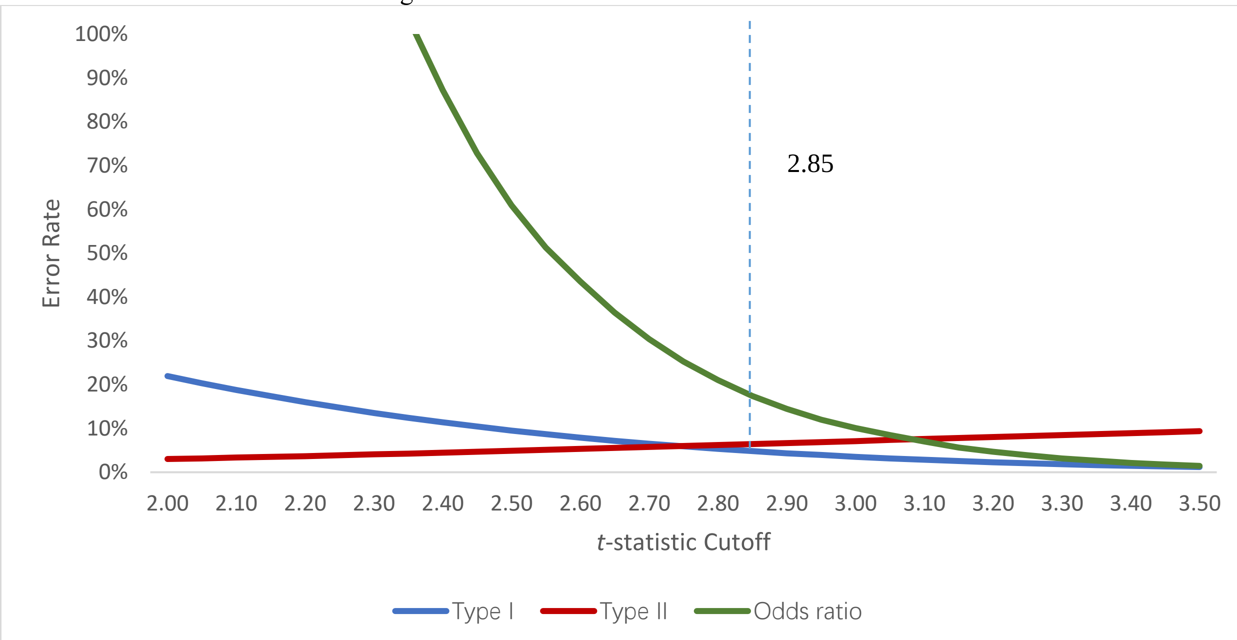


Figure 2. *t*-statistic Cutoff under Multiple Testing

Panel A plots the Type I and Type II error rates and the odds ratio against the *t*-statistic hurdle rate based on applying the multiple-testing procedure of Harvey and Liu (2020) to the 454 univariate strategies in China. p_0, the percentage of strategies that are true, is set to 15%. Panel B plots the *t*-statistic hurdle rate against p_0 varying between 5% and 20%. Details on the multiple-testing procedure are provided in Internet Appendix C2.

Panel A. Error Rate and Odds Ratio against *t*-statistic Hurdle Rate



Panel B. *t*-statistic Cutoff against p_0

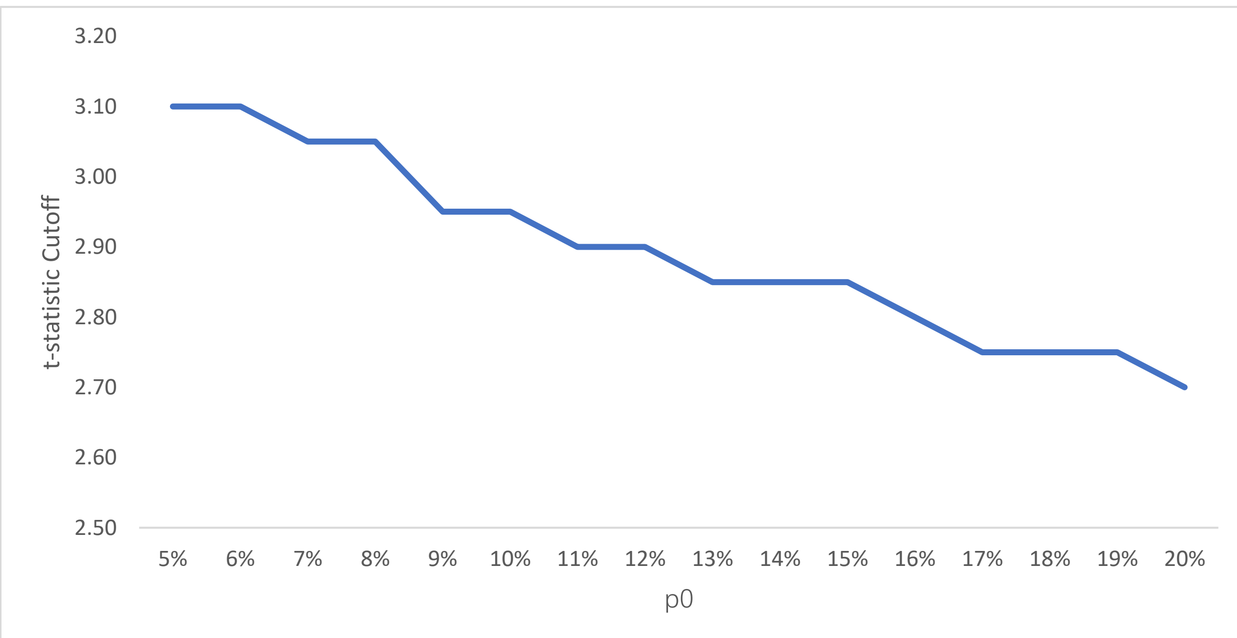
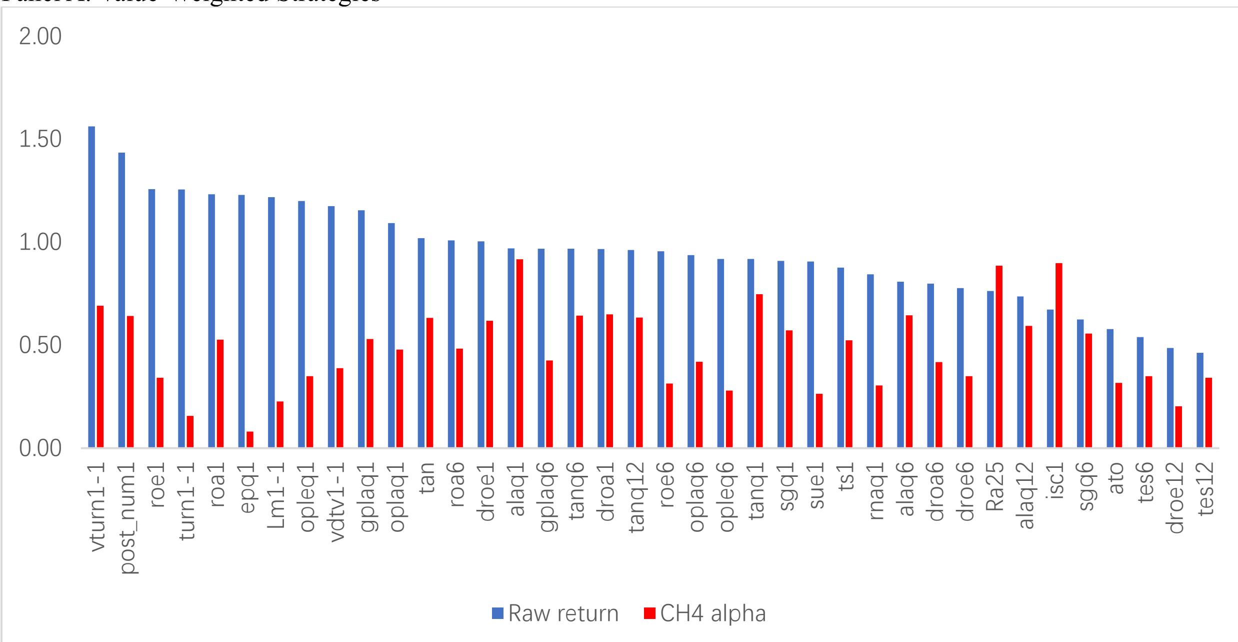


Figure 3. High-Minus-Low Raw Returns and CH4 Alphas of Significant Strategies

The figure plots the magnitude of the value-weighted (Panel A) and equal-weighted (Panel B) high-minus-low raw returns and CH4 alphas for the 38 significant univariate strategies under multiple testing. The blue columns represent the absolute values of the high-minus-low raw returns, and the red columns represent the absolute values of their CH4 alphas. The 38 strategies are ordered based on their absolute high-minus-low raw returns.

Panel A. Value-Weighted Strategies



Panel B. Equal-Weighted Strategies

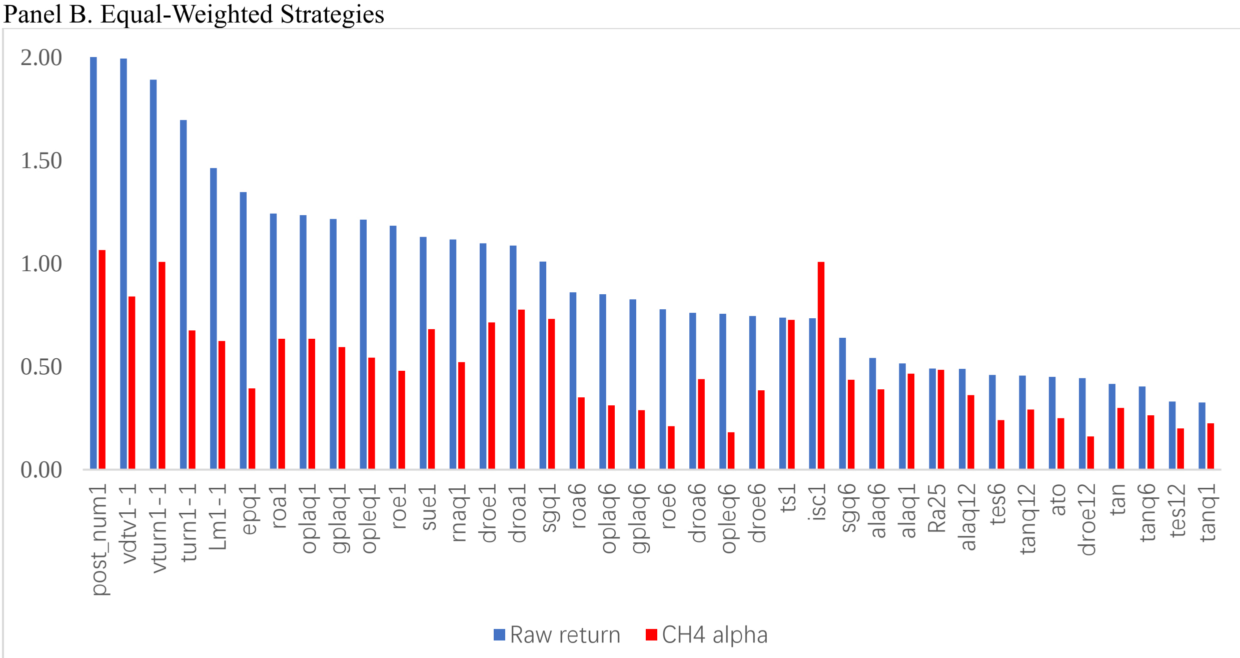
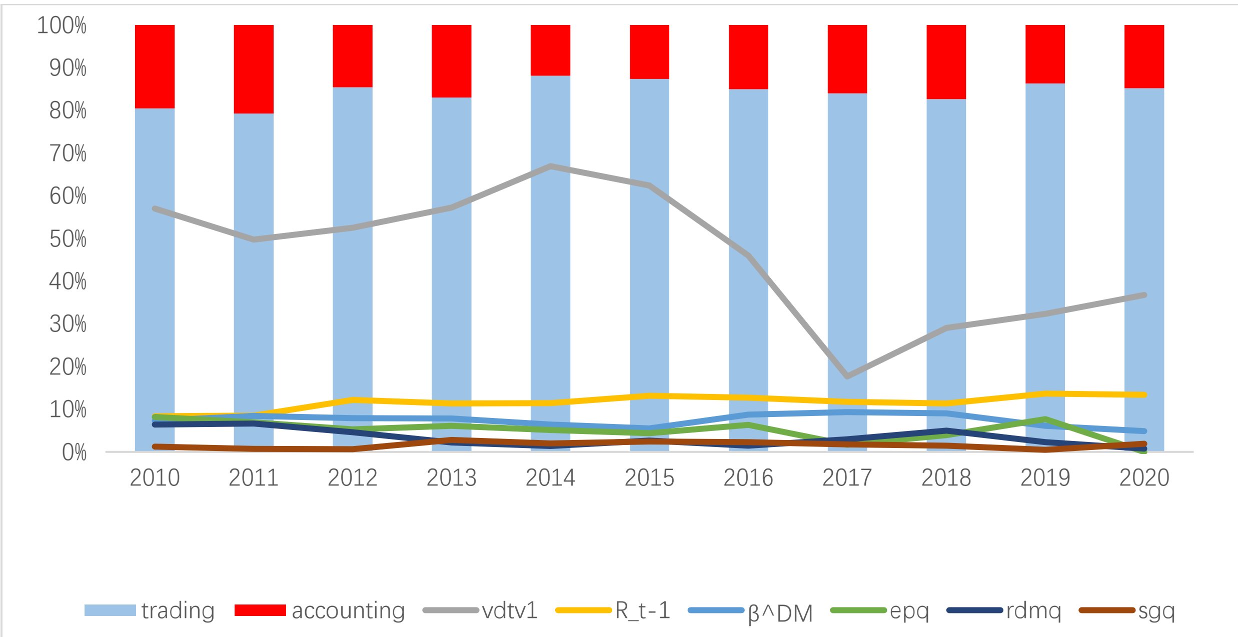


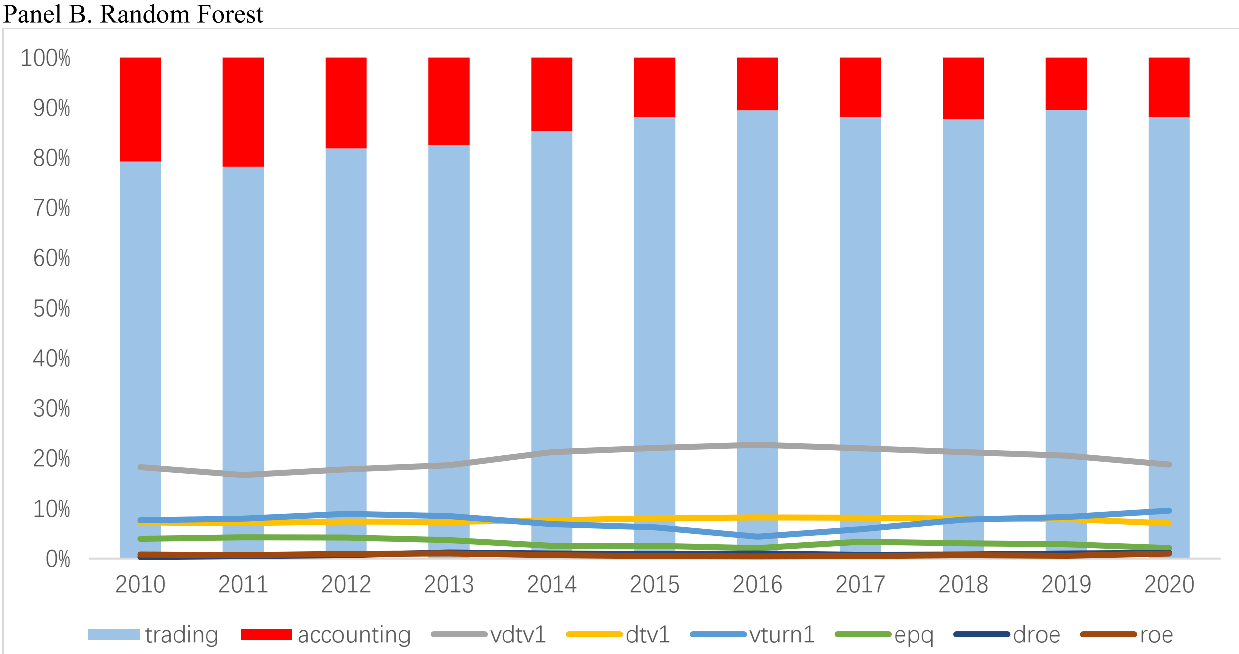
Figure 4. Signal Weights in Composite Strategies

The figure plots, for each year from 2010 to 2020, the weights of individual signals in the composite strategies based on Lasso (Panel A) and random forest (Panel B). The weight of a signal in Lasso is the squared coefficient on that signal divided by the sum of squared coefficients across all signals. The weight of a signal in random forest is the average decrease in mean squared forecast errors across regression trees. Signal weights are normalized to sum to one within a model. The blue and red bars represent the aggregate weights of trading and accounting signals, respectively. The colored lines represent the top three most important trading and accounting signals.

Panel A. Lasso



Panel B. Random Forest



Finding Anomalies in China

—Internet Appendix

A. Additional Details on Data

We collect daily stock trading data from WIND, which includes stock code, trading date, closing price, trading volume, RMB trading volume, turnover, return with dividends, the number of free-float A-shares outstanding, the number of total A-shares outstanding, and the number of total shares outstanding.

Three different types of shares coexist in the Chinese stock market: A shares, B shares, and H shares. A shares are the RMB-denominated shares of Chinese firms listed on the Shanghai and Shenzhen Stock Exchanges, and can be traded by domestic investors and foreign investors under QFII/RQFII/Stock Connect programs. B shares are the USD-denominated and HKD-denominated shares listed on the Shanghai and Shenzhen Stock Exchanges, respectively, and can be traded by domestic investors with appropriate currency accounts and foreign investors. H shares are the HKD-denominated shares listed on the Hong Kong Stock Exchange, and can be traded by domestic investors under QDII/Stock Connect programs and foreign investors. We follow the literature and focus on A shares in our study.

We obtain firm financial statement data from WIND. The balance sheet and income statement data start from 1990, while the cash flow statement data start from 1997. Before January 2002, firms report their financial statements semi-annually (for accounting periods ending in June and December), and after January 2002, they report quarterly (for accounting periods ending in March, June, September, and December). To keep the data structure consistent throughout our sample period, we convert the semi-annual data before January 2002 to quarterly data by filling in the quarterly balance sheet data for March and September with the latest available semi-annual data, and quarterly income statement and cash flow statement data for March and September with one half of the latest available semi-annual data. Note that in China, the income statement and cash flow statement data are reported in cumulative values semi-annually before January 2002 and quarterly after January 2002.

We obtain institutional ownership data from CSMAR, which includes the number and proportion of shares held by mutual funds, brokers, insurance companies, security funds, entrusts, etc. In China, institutions are required to report all of their holdings semi-annually and their top 10 holdings quarterly. Hedge funds are not required to report their holdings.

We obtain analyst earnings forecast data from SUNTIME, which contains 1.6 million reports produced by more than 22,000 analysts from 383 institutions and includes information on the analyst's name, institution, forecast period, forecast release date, forecasted earnings per share, etc.

We also obtain social media coverage from GUBA, web search index from WSVI, and news data from CFND via the Chinese Research Data Services (CNRDS) platform. GUBA data contains the numbers of daily postings, comments, and readings from the biggest retail investor forum in China, the East Money Stock Forum. The WSVI index is a daily web search volume index of Chinese firms. CFND contains the daily numbers of printed news and internet news from more than 400 newspapers and journals and more than 500 internet media sites.

B. Definitions of Signals

B.1 Trading-Based Signals

B.1.1 Liquidity

B.1.1.1 Firm Size (size, size1, size6, and size12) Firm size is calculated as the closing price (unadjusted) times the number of total A-shares outstanding, following Liu, Stambaugh, and Yuan (2019). For annual sorting, at the end of June of each year *t*, we sort stocks into deciles based on the June-end firm size and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1. For monthly sorting, at the beginning of each month *t*, we sort stocks into deciles based on firm size at the end of month *t*-1 and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to size1, size6, and size12, respectively).

B.1.1.2 Share Turnover (turn1-1, turn1-6, turn1-12, turn6-1, turn6-6, turn6-12, turn12-1, turn12-6, and turn12-12) Following Datar, Naik, and Radcliffe (1998), we calculate a stock's share turnover (turn) as the average value of daily share turnover. Daily share turnover is calculated as the trading volume on a given day divided by the number of free-float A-shares outstanding. At the beginning of each month *t*, we sort stocks into deciles based on turn estimated with daily data over months [*t*-*k*, *t*-1] (*k*=1, 6, and 12, corresponding to turn1, turn6, and turn12, respectively). We require a minimum of 75% of daily trading records for the estimation period. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to turn1-1, turn1-6, turn1-12, turn6-1, turn6-6, turn6-12, turn12-1, turn12-6, and turn12-12, respectively).

B.1.1.3 Variation of Share Turnover (vturn1-1, vturn1-6, vturn1-12, vturn6-1, vturn6-6, vturn6-12, vturn12-1, vturn12-6, and vturn12-12) Following Chordia, Subrahmanyam, and Anshuman (2001), we measure a stock's variation of share turnover (vturn) as the standard deviation of daily share turnover. At the beginning of each month *t*, we sort stocks into deciles based on vturn estimated with daily data over months [*t*-*k*, *t*-1] (*k*=1, 6, and 12, corresponding to vturn1, vturn6, and vturn12, respectively). We require a minimum of 75% of daily trading records for the estimation period. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to vturn1-1, vturn1-6, vturn1-12, vturn6-1, vturn6-6, vturn6-12, vturn12-1, vturn12-6, and vturn12-12, respectively).

B.1.1.4 Coefficient of Variation of Share Turnover (cvturn1-1, cvturn1-6, cvturn1-12, cvturn6-1, cvturn6-6, cvturn6-12, cvturn12-1, cvturn12-6, and cvturn12-12) Following Chordia, Subrahmanyam, and Anshuman (2001), we measure a stock's coefficient of variation of share turnover (cvturn) as the ratio of the standard deviation to the mean for daily share turnover. At the beginning of each month *t*, we sort stocks into deciles based on cvturn estimated with daily data over months [*t*-*k*, *t*-1] (*k*=1, 6, and 12, corresponding to cvturn1, cvturn6, and cvturn12, respectively). We require a minimum of 75% of daily trading records for the estimation period. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to cvturn1-1, cvturn1-6, cvturn1-12, cvturn6-1, cvturn6-6, cvturn6-12, cvturn12-1, cvturn12-6, and cvturn12-12, respectively).

B.1.1.5 Abnormal Turnover (abturn1, abturn6, and abturn12) Following Liu, Stambaugh, and Yuan (2019), at the beginning of each month *t*, we estimate abnormal turnover (abturn) as the ratio of the average daily turnover in month *t*-1 to the average daily turnover in the prior 12 months from month *t*-12 to *t*-1. We require a minimum of 75% of daily trading records in month *t*-1 and in the prior 12 months. At the beginning of each month *t*, we sort stocks into deciles based on abturn and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to abturn1, abturn6, and abturn12, respectively).

B.1.1.6 RMB Trading Volume (dtv1-1, dtv1-6, dtv1-12, dtv6-1, dtv6-6, dtv6-12, dtv12-1,

dtv12-6, and dtv12-12) Following Brennan, Chordia, and Subrahmanyam (1998), we calculate a stock's RMB trading volume (dtv) as the average value of daily RMB trading volume. At the beginning of each month *t*, we sort stocks into deciles based on dtv estimated with daily data over months [*t-k, t*-1] (*k*=1, 6, and 12, corresponding to dtv1, dtv6, and dtv12, respectively). We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t, t*+*n*-1] (*n*=1, 6, and 12, corresponding to dtv1-1, dtv1-6, dtv1-12, dtv6-1, dtv6-6, dtv6-12, dtv12-1, dtv12-6, and dtv12-12, respectively).

B.1.1.7 Variation of RMB Trading Volume (vdtv1-1, vdtv1-6, vdtv1-12, vdtv6-1, vdtv6-6, vdtv6-12, vdtv12-1, vdtv12-6, and vdtv12-12) Following Chordia, Subrahmanyam, and Anshuman (2001), we measure a stock's variation of RMB trading volume (vdtv) as the standard deviation of daily RMB trading volume. At the beginning of each month *t*, we sort stocks into deciles based on vdtv estimated with daily data over months [*t-k, t*-1] (*k*=1, 6, and 12, corresponding to vdtv1, vdtv6, and vdtv12, respectively). We require a minimum of 75% of daily trading records for the estimation period. We then calculate monthly decile portfolio returns over months [*t, t*+*n*-1] (*n*=1, 6, and 12, corresponding to vdtv1-1, vdtv1-6, vdtv1-12, vdtv6-1, vdtv6-6, vdtv6-12, vdtv12-1, vdtv12-6, and vdtv12-12, respectively).

B.1.1.8 Coefficient of Variation of RMB Trading Volume (cvdtv1-1, cvdtv1-6, cvdtv1-12, cvdtv6-1, cvdtv6-6, cvdtv6-12, cvdtv12-1, cvdtv12-6, and cvdtv12-12) Following Chordia, Subrahmanyam, and Anshuman (2001), we measure a stock's coefficient of variation of RMB trading volume (cvdtv) as the ratio of the standard deviation to the mean for daily RMB trading volume. At the beginning of each month *t*, we sort stocks into deciles based on cvdtv estimated with daily data over months [*t-k, t*-1] (*k*=1, 6, and 12, corresponding to cvdtv1, cvdtv6, and cvdtv12, respectively). We require a minimum of 75% of daily trading records for the estimation period. We then calculate monthly decile portfolio returns over months [*t, t*+*n*-1] (*n*=1, 6, and 12, corresponding to cvdtv1-1, cvdtv1-6, cvdtv1-12, cvdtv6-1, cvdtv6-6, cvdtv6-12, cvdtv12-1, cvdtv12-6, and cvdtv12-12, respectively).

B.1.1.9 Amihud Illiquidity (Ami1-1, Ami1-6, Ami1-12, Ami6-1, Ami6-6, Ami6-12, Ami12-1, Ami12-6, and Ami12-12) We calculate the Amihud's (2002) illiquidity measure (Ami) as the ratio of absolute daily stock return to daily RMB trading volume, averaged over the prior *k* months. At the beginning of each month *t*, we sort stocks into deciles based on Ami estimated with daily data over months [*t-k, t*-1] (*k*=1, 6, and 12, corresponding to Ami1, Ami6, and Ami12, respectively). We require a minimum of 75% of daily trading records for the estimation period. We then calculate monthly decile portfolio returns over months [*t, t*+*n*-1] (*n*=1, 6, and 12, corresponding to Ami1-1, Ami1-6, Ami1-12, Ami6-1, Ami6-6, Ami6-12, Ami12-1, Ami12-6, and Ami12-12, respectively).

B.1.1.10 Turnover-adjusted Number of Zero Daily Trading Volume (Lm1-1, Lm1-6, Lm1-12, Lm6-1, Lm6-6, Lm6-12, Lm12-1, Lm12-6, and Lm12-12) Following Liu (2006), we calculate the standardized turnover-adjusted number of zero daily trading volume over the prior *k* months (Lm) as follows:

$$Lm^k=\left[\text{Number of days with volumes}<150{,}000\text{ in prior }k\text{ months}+\frac{\frac{1}{k\text{-month turnover}}}{Deflator}\right]\frac{21k}{NoTD}, \qquad \text{(B1)}$$

where *k*-month turnover is the sum of daily turnover over the prior *k* months, *NoTD* is the total number of trading days over the prior *k* months. We set the deflator to $\max\left\{\frac{1}{k-\text{month turnover}}\right\}+1$, in which the maximization is taken across all stocks each month. The choice of the deflator ensures that (1/(*k*-month turnover))/Deflator is between zero and one for all stocks. Since there are many trading suspensions in China, we replace the number of days with zero volume in the prior

k months with the number of days with volume less than 150,000 shares. We require a minimum of 75% of daily trading records for the estimation period. At the beginning of each month t, we sort stocks into deciles based on Lm estimated with daily data over months [t-k, t-1] (k=1, 6, and 12, corresponding to LM1, LM6, and LM12, respectively). We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to Lm1-1, Lm1-6, Lm1-12, Lm6-1, Lm6-6, Lm6-12, Lm12-1, Lm12-6, and Lm12-12, respectively).

B.1.1.11 Liquidity Betas (return-return, β^{ret}1, β^{ret}6, β^{ret}12, illiquidity-illiquidity, β^{lcc}1, β^{lcc}6, β^{lcc}12, return-illiquidity, β^{lrc}1, β^{lrc}6, β^{lrc}12, liquidity-return, β^{lcr}1, β^{lcr}6, β^{lcr}12, and net, β^{net}1, β^{net}6, β^{net}12) Following Acharya and Pedersen (2005), we measure illiquidity using the Amihud's (2002) measure, Ami. For stock i in month t, Ami_t^i is the average ratio of absolute daily return to daily RMB trading volumes (in millions). We require a minimum of 75% of daily trading records. The market illiquidity, Ami_t^M, is the value-weighted average of $\min(\mathrm{Ami}_t^i, (30-0.25)/(0.30P_{t-1}^M))$, in which P_{t-1}^M is the ratio of the total market capitalization of CSI 300 index at the end of month t-1 to its value at the end of January 1998. We measure market illiquidity innovations, ϵ_{Mt}^c, as the residual from the regression below:

$$\begin{aligned}(0.25+0.3Ami_t^M P_{t-1}^M) = a_0 + a_1(0.25+0.30Ami_{t-1}^M P_{t-1}^M)\\ +a_2(0.25+0.3Ami_{t-2}^M P_{t-1}^M)+\epsilon_{M,t}^c.\end{aligned} \tag{B2}$$

Innovations to individual stock's illiquidity, $\epsilon_{i,t}^c$, are measured analogously by replacing Ami^M with $\min\ (Ami_t^i,\ (30-0.25)/(0.30P_{t-1}^M))$ in Equation (B2). Finally, we measure innovations to the market return, ϵ_{Mt}^r, as the residual from the second-order autoregression of market returns. Following Acharya and Pedersen (2005), we define five measures of liquidity betas:

$$\text{Return-return: } \beta_i^{ret} = \frac{Cov(r_{i,t}, \epsilon_{M,t}^r)}{var(\epsilon_{M,t}^r - \epsilon_{M,t}^c)}, \tag{B3}$$

$$\text{Illiquidity-illiquidity: } \beta_i^{lcc} = \frac{Cov(\epsilon_{i,t}^c,\ \epsilon_{M,t}^c)}{var(\epsilon_{M,t}^r - \epsilon_{M,t}^c)}, \tag{B4}$$

$$\text{Return-illiquidity: } \beta_i^{lrc} = \frac{Cov(r_{i,t}, \epsilon_{M,t}^c)}{var(\epsilon_{M,t}^r - \epsilon_{M,t}^c)}, \tag{B5}$$

$$\text{Illiquidity-return: } \beta_i^{lcr} = \frac{Cov(\epsilon_{i,t}^c, \epsilon_{M,t}^r)}{var(\epsilon_{M,t}^r - \epsilon_{M,t}^c)}, \tag{B6}$$

$$\text{Net: } \beta_i^{net} = \beta_i^{ret} + \beta_i^{lcc} - \beta_i^{lrc} - \beta_i^{lcr}. \tag{B7}$$

At the beginning of each month t, we sort stocks into deciles based on $\beta_i^{ret}, \beta_i^{lcc}, \beta_i^{lrc}, \beta_i^{lcr}$, and β_i^{net} estimated with the past 60 months from month t-60 to t-1. We require a minimum of 75% of monthly returns for the estimation period. We calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to β^{ret}1, β^{ret}6, β^{ret}12, β^{lcc}1, β^{lcc}6, β^{lcc}12, β^{lrc}1, β^{lrc}6, β^{lrc}12, β^{lcr}1, β^{lcr}6, β^{lcr}12, β^{net}1, β^{net}6, and β^{net}12, respectively).

B.1.1.12 GUBA Social Media Coverage (post_num1, post_num6, post_num12, com_num1, com_num6, com_num12, read_num1, read_num6, and read_num12) We obtain the numbers of GUBA postings, comments, and readings for Chinese listed firms from CNRDS. At the beginning of each month t, we sort stocks into deciles based on their average number of daily GUBA postings (post_num), comments (com_num), and readings (read_num) in month t-1. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to post_num1, post_num6, post_num12, com_num1, com_num6, com_num12, read_num1, read_num6, and read_num12, respectively).

B.1.1.13 Web Search Volume Index (wsvi1, wsvi6, and wsvi12) We obtain the web search volume index of Chinese listed firms from CNRDS. At the beginning of each month t, we sort stocks into deciles based on their average daily web search volume index (wsvi) in month t-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to wsvi1, wsvi6, and wsvi12, respectively).

B.1.2 Risk

B.1.2.1 Idiosyncratic Volatility (iv) Following Ali, Hwang, and Trombley (2003), we calculate idiosyncratic volatility (iv) as the residual volatility from regressing a stock's daily excess returns on the market excess returns over the prior year. At the end of June of each year *t*, we sort stocks into deciles based on iv estimated with daily returns from July of year *t*-1 to June of year *t*. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.1.2.2 Idiosyncratic Volatility per the CAPM (ivc1, ivc6, and ivc12) We calculate idiosyncratic volatility relative to the CAPM (ivc) as the residual volatility from regressing a stock's daily excess returns on the market excess returns over the prior month. At the beginning of each month *t*, we sort stocks into deciles based on ivc estimated with daily returns from month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to ivc1, ivc6, and ivc12, respectively).

B.1.2.3 Idiosyncratic Volatility per the CH3 Factor Model (ivch3-1, ivch3-6, and ivch3-12) We calculate idiosyncratic volatility relative to the Liu, Stambaugh, and Yuan (2019) Chinese three-factor model (CH3) (ivch3) as the residual volatility from regressing a stock's daily excess returns on the CH3 factors over the prior month. At the beginning of each month *t*, we sort stocks into deciles based on ivch3 estimated with daily returns from month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to ivch3-1, ivch3-6, and ivch3-12, respectively).

B.1.2.4 Idiosyncratic Volatility per the CH4 Factor Model (ivch4-1, ivch4-6, and ivch4-12) We calculate idiosyncratic volatility relative to the Liu, Stambaugh, and Yuan (2019) Chinese four-factor model (CH4) (ivch4) as the residual volatility from regressing a stock's daily excess returns on the CH4 factors over the prior month. At the beginning of each month *t*, we sort stocks into deciles based on ivch4 estimated with daily returns from month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to ivch4-1, ivch4-6, and ivch4-12, respectively).

B.1.2.5 Total Volatility (tv1, tv6, and tv12) At the beginning of each month *t*, we sort stocks into deciles based on total volatility (tv) estimated as the volatility of a stock's daily returns from month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to tv1, tv6, and tv12, respectively).

B.1.2.6 Idiosyncratic Skewness per the CAPM (isc1, isc6, and isc12) At the beginning of each month *t*, we sort stocks into deciles based on idiosyncratic skewness (isc) estimated as the skewness of the residuals from regressing a stock's excess returns on the market excess returns using daily observations from month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to isc1, isc6, and isc12, respectively).

B.1.2.7 Idiosyncratic Skewness per the CH3 Factor Model (isch3-1, isch3-6, and isch3-12) At the beginning of each month *t*, we sort stocks into deciles based on idiosyncratic skewness relative to the CH3 factor model (isch3) estimated as the skewness of the residuals from regressing a stock's excess returns on the CH3 factors using daily observations from month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to isch3-1, isch3-6, and isch3-12, respectively).

B.1.2.8 Idiosyncratic Skewness per the CH4 Factor Model (isch4-1, isch4-6, and isch4-12) At the beginning of each month t, we sort stocks into deciles based on idiosyncratic skewness relative to the CH4 factor model (isch4) estimated as the skewness of the residuals from regressing a stock's excess returns on the CH4 factors using daily observations from month t-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to isch4-1, isch4-6, and isch4-12, respectively).

B.1.2.9 Total Skewness (ts1, ts6, and ts12) At the beginning of each month t, we sort stocks into deciles based on total skewness (ts) calculated with daily returns from month t-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to ts1, ts6, and ts12, respectively).

B.1.2.10 Co-skewness (cs1, cs6, and cs12) Following Harvey and Siddique (2000), we measure co-skewness (cs) as

$$cs = \frac{E[\epsilon_i \epsilon_m^2]}{\sqrt{E[\epsilon_i^2]}E[\epsilon_m^2]}, \tag{B8}$$

in which ϵ_i is the residual from regressing stock i's excess return on the market excess return, and ϵ_m is the demeaned market excess return. At the beginning of each month t, we sort stocks into deciles based on cs estimated with daily returns from month t-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to cs1, cs6, and cs12, respectively).

B.1.2.11 Market Beta Using Monthly Returns ($\boldsymbol{\beta^m}$1, $\boldsymbol{\beta^m}$6, and $\boldsymbol{\beta^m}$12) At the beginning of each month t, we sort stocks into deciles based on their market beta (β^m), which is estimated with monthly returns from month t-60 to t-1. We require a minimum of 75% of monthly returns in the prior 60 months. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to β^m1, β^m6, and β^m12, respectively).

B.1.2.12 Market Beta Using Daily Returns ($\boldsymbol{\beta}$1, $\boldsymbol{\beta}$6, and $\boldsymbol{\beta}$12) At the beginning of each month t, we sort stocks into deciles based on their market beta (β), which is estimated with daily returns from month t-12 to t-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to β1, β6, and β12, respectively).

B.1.2.13 Downside Beta ($\boldsymbol{\beta^-}$1, $\boldsymbol{\beta^-}$6, and $\boldsymbol{\beta^-}$12) Following Ang, Chen, and Xing (2006), we calculate downside beta (β^-) as

$$\beta^- = \frac{Cov(r_i, r_m | r_m < \mu_m)}{Var(r_m | r_m < \mu_m)}, \tag{B9}$$

where r_i and r_m are stock i's and market excess returns, respectively, and μ_m is the average market excess return. At the beginning of each month t, we sort stocks into deciles based on β^-, which is estimated with daily returns from month t-12 to t-1. We only use daily observations with $r_m < \mu_m$. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to β^-1, β^-6, and β^-12, respectively).

B.1.2.14 The Frazzini-Pedersen Beta ($\boldsymbol{\beta^{FP}}$1, $\boldsymbol{\beta^{FP}}$6, and $\boldsymbol{\beta^{FP}}$12) Following Frazzini and Pedersen (2014), we estimate the market beta for stock i (β^{FP}) as $\hat{\rho}\frac{\hat{\sigma}_i}{\hat{\sigma}_m}$, where $\hat{\sigma}_i$ and $\hat{\sigma}_m$ are the estimated return volatilities for the stock and the market, respectively, and $\hat{\rho}$ is their return correlation. To estimate return volatilities, we compute the standard deviation of daily log returns over a one-year rolling window (with a minimum of 75% of daily trading records). To estimate return correlations, we use overlapping three-day log returns, $r_{i,t}^{3d} = \sum_{k=0}^{2} log\left(1 + R_{i,t+k}\right)$, over a five-year rolling window (with a minimum of 75% of daily trading records). At the beginning of each month *t*, we sort stocks into deciles based on β^{FP} estimated at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to $\beta^{FP}1$, $\beta^{FP}6$, and $\beta^{FP}12$, respectively).

B.1.2.15 The Dimson Beta ($\beta^{DM}1$, $\beta^{DM}6$, and $\beta^{DM}12$) Following Dimson (1979), we use the current as well as the lead and lag of the market return when estimating the market beta:

$$r_{i,d} - r_{f,d} = \alpha_i + \beta_{i1}\left(r_{m,d-1} - r_{f,d-1}\right) + \beta_{i2}\left(r_{m,d} - r_{f,d}\right) + \beta_{i3}\left(r_{m,d+1} - r_{f,d+1}\right) + \epsilon_{i,d}, \quad \text{(B10)}$$

where $r_{i,d}$ is stock *i*'s return on day *d*, $r_{m,d}$ is the market return on day *d*, and $r_{f,d}$ is the risk-free rate (the one-year deposit rate). The Dimson beta for stock *i* (β^{DM}) is calculated as $\hat{\beta}_{i1} + \hat{\beta}_{i2} + \hat{\beta}_{i3}$. At the beginning of each month *t*, we sort stocks into deciles based on β^{DM} estimated with daily returns from month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to $\beta^{DM}1$, $\beta^{DM}6$, and $\beta^{DM}12$, respectively).

B.1.2.16 Tail Risk (tail1, tail6, and tail12) Following Kelly and Jiang (2014), we estimate common tail risk, λ_t, by pooling daily returns for all stocks in month *t,* as follows:

$$\lambda_t = \frac{1}{K_t}\sum_{k=1}^{K_t} \log\left(\frac{r_{k,t}}{\mu_t}\right), \quad \text{(B11)}$$

where μ_t is the fifth percentile of all daily returns in month *t*, $r_{k,t}$ is the *k*-th daily return that is below μ_t, and K_t is the total number of daily returns that are below μ_t. At the beginning of each month *t*, we sort stocks into deciles based on tail risk (tail) estimated as the slope from regressing a stock's monthly excess returns on one-month-lagged common tail risk over the most recent 120 months from month *t*-120 to *t*-1. We require a minimum of 75% of monthly returns. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to tail1, tail6, and tail12, respectively).

B.1.2.17 Firm News (paper_news1, paper_news6, paper_news12, inter_news1, inter_news6, and inter_news12) At the beginning of each month *t*, we sort stocks into deciles based on their average numbers of daily reported paper news (paper_news) and internet news (inter_news) in month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to paper_news1, paper_news6, paper_news12, inter_news1, inter_news6, and inter_news12, respectively).

B.1.3 Past Returns

B.1.3.1 Short-Term Prior *k*-month Cumulative Returns ($R_{t-4,t-2}1$, $R_{t-4,t-2}6$, $R_{t-4,t-2}12$, $R_{t-7,t-2}1$, $R_{t-7,t-2}6$, $R_{t-7,t-2}12$, $R_{t-10,t-2}1$, $R_{t-10,t-2}6$, $R_{t-10,t-2}12$, $R_{t-12,t-2}1$, $R_{t-12,t-2}6$, and $R_{t-12,t-2}12$) At the beginning of each month *t*, we sort stocks into deciles based on their *k*-month cumulative returns over months [*t*-*k*-1, *t*-2] (*k*=3, 6, 9, 11, corresponding to $R_{t-4,t-2}$, $R_{t-7,t-2}$, $R_{t-10,t-2}$, $R_{t-12,t-2}$, respectively). We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to $R_{t-4,t-2}1$, $R_{t-4,t-2}6$, $R_{t-4,t-2}12$, $R_{t-7,t-2}1$, $R_{t-7,t-2}6$, $R_{t-7,t-2}12$, $R_{t-10,t-2}1$, $R_{t-10,t-2}6$, $R_{t-10,t-2}12$, $R_{t-12,t-2}1$, $R_{t-12,t-2}6$, and $R_{t-12,t-2}12$, respectively).

B.1.3.2 Prior One-month Return ($R_{t-1}1$, $R_{t-1}6$, and $R_{t-1}12$) At the beginning of each month *t*, we sort stocks into deciles based on their prior-month return (R_{t-1}). We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to $R_{t-1}1$, $R_{t-1}6$, and $R_{t-1}12$, respectively).

B.1.3.3 Long-Term Prior *k*-month Cumulative Returns ($R_{t-60,t-13}1$, $R_{t-60,t-13}6$, $R_{t-60,t-13}12$, $R_{t-36,t-13}1$, $R_{t-36,t-13}6$, and $R_{t-36,t-13}12$) At the beginning of each month *t*, we sort stocks into deciles based on their cumulative returns from month *t*-60 to *t*-13 ($R_{t-60,t-13}$), and separately, on their cumulative returns from month *t*-36 to *t*-13 ($R_{t-36,t-13}$). We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to $R_{t-60,t-13}1$, $R_{t-60,t-13}6$, $R_{t-60,t-13}12$, $R_{t-36,t-13}1$, $R_{t-36,t-13}6$, and $R_{t-36,t-13}12$, respectively).

B.1.3.4 Industry Returns (inr1, inr6, and inr12) We have 27 Shenwanhongyuan industries after excluding financial firms. At the beginning of each month *t*, we sort industries based on their prior six-month value-weighted returns (inr) from month *t*-6 to *t*-1. We form nine portfolios (9×3=27), each of which contains three different industries. We define the return of a given portfolio as the simple average of the three industry returns within the portfolio. We then calculate monthly portfolio returns for the nine portfolios over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to inr1, inr6, and inr12, respectively).

B.1.3.5 Industry Lead-Lag Effect (ilr1, ilr6, and ilr12) We have 27 Shenwanhongyuan industries after excluding financial firms. At the beginning of each month *t*, we sort industries based on the value-weighted return of the portfolio consisting of the 30% largest firms within a given industry (ilr) in month *t*-1. We form nine portfolios (9×3=27), each of which contains three different industries. We define the return of a given portfolio as the simple average of the three value-weighted industry returns within the portfolio. We then calculate monthly portfolio returns for the nine portfolios over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to ilr1, ilr6, and ilr12, respectively).

B.1.3.6 Cumulative Return Changes (crchg1, crchg6, and crchg12) At the beginning of each month *t*, we sort stocks into deciles based on their cumulative return changes (crchg) calculated as the cumulative returns from month *t*-7 to *t*-2 minus the cumulative returns from month *t*-13 to *t*-8. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to crchg1, crchg6, and crchg12, respectively).

B.1.3.7 Prior *k*-month Residual Returns ($RR_{t-12,t-2}1$, $RR_{t-12,t-2}6$, $RR_{t-12,t-2}12$, $RR_{t-7,t-2}1$, $RR_{t-7,t-2}6$, and $RR_{t-7,t-2}12$) At the beginning of each month *t*, we sort stocks into deciles based on the cumulative residual returns from month *t*-12 to *t*-2 ($RR_{t-12,t-2}$) or from month *t*-7 to *t*-2 ($RR_{t-7,t-2}$), scaled by their standard deviation over the same period. Residual returns are estimated from regressing monthly stock excess returns on the CH4 factors over the prior 36 months from month *t*−36 to *t*−1. We require monthly returns to be available for all prior 36 months. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to $RR_{t-12,t-2}1$, $RR_{t-12,t-2}6$, $RR_{t-12,t-2}12$, $RR_{t-7,t-2}1$, $RR_{t-7,t-2}6$, and $RR_{t-7,t-2}12$, respectively).

B.1.3.8 52-Week High (52wh1, 52wh6, and 52wh12) At the beginning of each month *t*, we sort stocks into deciles based on their 52-week high (52wh), which is the ratio of its split-adjusted price per share at the end of month *t*-1 to its highest (daily) split-adjusted price per share during the prior one-year period from month *t*-12 to *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to 52wh1, 52wh6, and 52wh12, respectively).

B.1.3.9 Maximum Daily Return (mdr1, mdr6, and mdr12) Due to the daily 10% price limit rule implemented after 1996 in China, we estimate maximum daily return (mdr) as the average of the 5 highest daily returns of a given stock from the prior month, following Bali, Brown, Murray, and Tang (2017). At the beginning of month *t*, we sort stocks into deciles based on mdr estimated with daily returns in month *t*-1. We require a minimum of 75% of daily trading records. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to mdr1, mdr6, and mdr12, respectively).

B.1.3.10 Share Price (pps1, pps6, and pps12) At the beginning of each month *t*, we sort stocks into deciles based on share price (pps) at the end of month *t*-1. Share price is adjusted for splitting and delisting. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to pps1, pps6, and pps12, respectively).

B.1.3.11 Cumulative Abnormal Returns around Earnings Announcement Dates (abr1, abr6, and abr12) We calculate cumulative abnormal stock returns (abr) around the latest quarterly earnings announcement date, following Chan, Jegadeesh, and Lakonishok (1996).

$$abr_i = \sum_{d=-2}^{+1}(r_{i,d} - r_{m,d}), \tag{B12}$$

where $r_{i,d}$ is stock *i*'s return on day *d* (with the earnings announced on day 0), and $r_{m,d}$ is the value-weighted market index return. We cumulate returns until one (trading) day after the announcement date to account for the one-day-delayed reaction to earnings news.

At the beginning of each month *t*, we sort stocks into deciles based on their most recent abr. For a firm to enter portfolio sorts, we require the end of the fiscal quarter that corresponds to its most recent abr to be within 12 months prior to the portfolio formation. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to abr1, abr6, and abr12, respectively).

B.1.3.12 Seasonality (Ra1, Rn1, Ra25, and Rn25) Following Heston and Sadka (2008), at the beginning of each month *t*, we sort stocks into deciles based on various measures of past performance, including seasonality returns in year *t*-1 (returns in month *t*-12) (Ra1), non-seasonality returns in year *t*-1 (average returns from month *t*-11 to *t*-1) (Rn1), seasonality returns between year *t*-2 and *t*-5 (average returns across month *t*-24, *t*-36, *t*-48, and *t*-60) (Ra25), and non-seasonality returns between year *t*-2 and *t*-5 (average returns from month *t*-60 to *t*-13, excluding returns in month *t*-24, *t*-36, *t*-48, and *t*-60) (Rn25). We then calculate monthly decile portfolio returns for month *t*.

B.2 Ac**counting-Based Signals**

B.2.1 Profitability

B.2.1.1 Return on Equity (roe1, roe6, and roe12) Following Hou, Xue, and Zhang (2015), we estimate return on equity (roe) as quarterly net income minus nonrecurrent gains/losses divided by one-quarter-lagged book equity. Book equity is total shareholders' equity minus the book value of preferred stocks (zero if missing). At the beginning of each month *t*, we sort stocks into deciles based on their most recently released roe and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to roe1, roe6, and roe12, respectively).

B.2.1.2 4-quarter Change in Return on Equity (droe1, droe6, and droe12) We estimate change in return on equity (droe) as return on equity minus its value from four quarters ago. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released droe and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to droe1, droe6, and droe12, respectively).

B.2.1.3 Return on Assets (roa1, roa6, and roa12) Following Balakrishnan, Bartov, and

Faurel (2010), we measure return on assets (roa) as quarterly net income minus nonrecurrent gains/losses divided by one-quarter-lagged total assets. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released roa and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to roa1, roa6, and roa12, respectively).

B.2.1.4 4-quarter Change in Return on Assets (droa1, droa6, and droa12) We estimate change in return on assets (droa) as return on assets minus its value from four quarters ago. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released droa and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to droa1, droa6, and droa12, respectively).

B.2.1.5 Standard Unexpected Earnings (sue1, sue6, and sue12) Following Foster, Olsen, and Shevlin (1984), we calculate standard unexpected earnings (sue) as the change in split-adjusted quarterly earnings per share from its value four quarters ago divided by the standard deviation of this change over the prior eight quarters (with a minimum of six quarters). At the beginning of each month *t*, we sort stocks into deciles based on their most recently released sue and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to sue1, sue6, and sue12, respectively).

B.2.1.6 Revenue Surprises (rs1, rs6, and rs12) Following Jegadeesh and Livnat (2006), we measure revenue surprises (rs) as the change in revenue per share from its value four quarters ago divided by the standard deviation of this change over the prior eight quarters (with a minimum of six quarters). At the beginning of each month *t*, we sort stocks into deciles based on their most recently released rs and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to rs1, rs6, and rs12, respectively).

B.2.1.7 Return on Net Operating Assets (rna), Profit Margin (pm), and Assets Turnover (ato) Soliman (2008) decomposes roe as $\text{roe} = \text{rna} + \text{flev} \times \text{spread}$, where roe is return on equity, rna is return on net operating assets, flev is financial leverage, and spread is the difference between return on net operating assets and borrowing costs. We further decompose rna as $\text{pm} \times \text{ato}$, where pm is profit margin, and ato is assets turnover.

At the end of June of year *t*, we measure rna as operating income for the fiscal year ending in calendar year *t*-1 divided by net operating assets (noa) for the fiscal year ending in calendar year *t*-2. Net operating assets (noa) are calculated as operating assets minus operating liabilities. Operating assets are total assets minus cash and short-term investments (zero if missing). Operating liabilities are calculated as total assets minus debt in current liabilities (zero if missing), minus long-term debt (zero if missing), minus minority interests (zero if missing), minus preferred stocks (zero if missing), and minus common equity. pm is calculated as operating income divided by sales for the fiscal year ending in calendar year *t*-1, and ato is calculated as sales for the fiscal year ending in calendar year *t*-1 divided by noa for the fiscal year ending in year *t*-2. At the end of June of each year *t*, we sort stocks into deciles based on rna, pm, and ato, and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.1.8 Quarterly Return on Net Operating Assets (rnaq1, rnaq6, and rnaq12), Profit Margin (pmq1, pmq6, and pmq12), and Assets Turnover (atoq1, atoq6, and atoq12) We measure quarterly return on net operating assets (rnaq) as quarterly operating income divided by one-quarter-lagged net operating assets. Net operating assets (noa) are operating assets minus operating liabilities. Operating assets are total assets minus cash and short-term investments (zero if missing). Operating liabilities are total assets minus debt in current liabilities (zero if missing), minus long-term debt (zero if missing), minus minority interests (zero if missing), minus preferred stocks (zero if missing), and minus common equity. Quarterly profit margin (pmq) is quarterly operating income divided by quarterly sales. Quarterly assets turnover (atoq) is quarterly sales divided by one-quarter-lagged noa. At the beginning of each month t, we sort stocks into deciles based on their most recently released rnaq, pmq, and atoq, and calculate monthly decile portfolio returns over months $[t, t+n-1]$ (n=1, 6, and 12, corresponding to rnaq1, rnaq6, rnaq12, pmq1, pmq6, pmq12, atoq1, atoq6, and atoq12, respectively).

B.2.1.9 Capital Turnover (ct) At the end of June of each year t, we sort stocks into deciles based on capital turnover (ct) measured as sales for the fiscal year ending in calendar year t-1 divided by total assets for the fiscal year ending in year t-2. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.1.10 Quarterly Capital Turnover (ctq1, ctq6, and ctq12) We measure quarterly capital turnover (ctq) as quarterly sales scaled by one-quarter-lagged total assets. At the beginning of each month t, we sort stocks into deciles based on their most recently released ctq and calculate monthly decile portfolio returns over months $[t, t+n-1]$ (n=1, 6, and 12, corresponding to ctq1, ctq6, and ctq12, respectively).

B.2.1.11 Gross Profits-to-Assets (gpa) Following Novy-Marx (2013), we measure gross profits-to-assets (gpa) as total revenue minus cost of goods sold divided by total assets. At the end of June of each year t, we sort stocks into deciles based on gpa for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.1.12 Gross Profits-to-Lagged Assets (gpla) We measure gross profits-to-lagged assets (gpla) as total revenue minus cost of goods sold divided by one-year-lagged total assets. At the end of June of each year t, we sort stocks into deciles based on gpla for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.1.13 Quarterly Gross Profits-to-Lagged Assets (gplaq1, gplaq6, and gplaq12) We measure quarterly gross profits-to-lagged assets (gplaq) as quarterly total revenue minus cost of goods sold divided by one-quarter-lagged total assets. At the beginning of each month t, we sort stocks into deciles based on their most recently released gplaq and calculate monthly decile portfolio returns over months $[t, t+n-1]$ (n=1, 6, and 12, corresponding to gplaq1, gplaq6, and gplaq12, respectively).

B.2.1.14 Operating Profits-to-Book Equity (ope) Following Fama and French (2015), we measure operating profits-to-book equity (ope) as total revenue minus cost of goods sold, minus selling, general, and administrative expenses, and minus interest expense, scaled by book equity. Book equity is total shareholders' equity minus the book value of preferred stocks (zero if missing). At the end of June of each year t, we sort stocks into deciles based on ope for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.1.15 Operating Profits-to-Lagged Book Equity (ople) We estimate operating profits-to-lagged book equity (ople) as total revenue minus cost of goods sold, minus selling, general, and administrative expenses, and minus interest expense, scaled by one-year-lagged book equity. Book equity is total shareholders' equity minus the book value of preferred stocks (zero if missing). At the end of June of each year t, we sort stocks into deciles based on ople for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.1.16 Quarterly Operating Profits-to-Lagged Book Equity (opleq1, opleq6, and opleq12) We measure quarterly operating profits-to-lagged book equity (opleq) as quarterly total revenue minus cost of goods sold, minus selling, general, and administrative expenses, and minus interest expense, scaled by one-quarter-lagged book equity. Book equity is total shareholders' equity minus the book value of preferred stocks (zero if missing). At the beginning of each month *t*, we sort stocks into deciles based on their most recently released opleq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to opleq1, opleq6, and opleq12, respectively).

B.2.1.17 Operating Profits-to-Assets (opa) Following Ball, Gerakos, Linnaimma, and Nikolaev (2015), we measure operating profits-to-assets (opa) as total revenue minus cost of goods sold, minus selling, general, and administrative expenses, and minus interest expense, scaled by total assets. At the end of June of each year *t*, we sort stocks into deciles based on opa for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.1.18 Operating Profits-to-Lagged Assets (opla) We measure operating profits-to-lagged assets (opla) as total revenue minus cost of goods sold, minus selling, general, and administrative expenses, and minus interest expense, scaled by one-year-lagged total assets. At the end of June of each year *t*, we sort stocks into deciles based on opla for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.1.19 Quarterly Operating Profits-to-Lagged Assets (oplaq1, oplaq6, and oplaq12) We measure quarterly operating profits-to-lagged assets (oplaq) as quarterly total revenue minus cost of goods sold, minus selling, general, and administrative expenses, and minus interest expense, scaled by one-quarter-lagged total assets. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released oplaq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to oplaq1, oplaq6, and oplaq12, respectively).

B.2.1.20 Taxable Income-to-Book Income (tbi) Following Green, Hand, and Zhang (2013), we measure taxable income-to-book income (tbi) as pretax income divided by net income. At the end of June of each year *t*, we sort stocks into deciles based on tbi for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.1.21 Quarterly Taxable Income-to-Book Income (tbiq1, tbiq6, and tbiq12) We measure quarterly taxable income-to-book income (tbiq) as quarterly pretax income divided by net income. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released tbiq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to tbiq1, tbiq6, and tbiq12, respectively).

B.2.1.22 Book Leverage (bl) Following Fama and French (1992), we measure book leverage (bl) as total assets divided by book equity. Book equity is total shareholders' equity minus the book value of preferred stocks (zero if missing). At the end of June of each year *t*, we sort stocks into deciles based on bl for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.1.23 Quarterly Book Leverage (blq1, blq6, and blq12) We measure quarterly book leverage (blq) as total assets divided by book equity. Book equity is total shareholders' equity minus the book value of preferred stocks (zero if missing). At the beginning of each month *t*, we sort stocks into deciles based on their most recently released blq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to blq1, blq6, and blq12, respectively).

B.2.1.24 Sales Growth (sg) Following Lakonishok, Shleifer, and Vishny (1994), we estimate sales growth (sg) as the growth in annual sales from the fiscal year ending in calendar year *t*-2 to the fiscal year ending in calendar year *t*-1. We exclude firms with non-positive sales. At the end of June of each year *t*, we sort stocks into deciles based on sg for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.1.25 Quarterly Sales Growth (sgq1, sgq6, and sgq12) We compute quarterly sales growth (sgq) as quarterly sales divided by its value four quarters ago. We exclude firms with non-positive sales. At the beginning of each month t, we sort stocks into deciles based on their most recently released sgq and calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to sgq1, sgq6, and sgq12, respectively).

B.2.1.26 Fundamental Score (F) Piotroski (2000) classifies each fundamental signal as either good or bad, depending on the signal's implication for future stock prices and profitability. An indicator variable for a particular signal is 1 if its realization is good and 0 if it is bad. The aggregate signal, denoted F, is the sum of the eight binary signals. F is designed to measure the overall quality, or strength, of the firm's financial position. The eight fundamental signals are chosen to measure three areas of a firm's financial condition: profitability, liquidity, and operating efficiency.

We use four variables to measure profitability: (1) roa is net income minus nonrecurrent gains/losses scaled by one-year-lagged total assets. If the firm's roa is positive, the indicator variable F_{roa} equals one and zero otherwise. (2) Cf/A is cash flows from operations scaled by one-year-lagged total assets. If the firm's Cf/A is positive, the indicator variable $F_{Cf/A}$ equals one and zero otherwise. (3) droa is the current year's roa minus the prior year's roa. If droa is positive, the indicator variable F_{droa} equals one and zero otherwise. (4) the indicator F_{Acc} equals one if Cf/A > roa and zero otherwise.

We use two variables to measure changes in capital structure and a firm's ability to meet debt obligations. Piotroski (2000) assumes that an increase in leverage, a deterioration in liquidity, or the use of external financing is a bad signal about financial risk. (1) dlever is the change in the ratio of total long-term debt to the average of current and one-year-lagged total assets. F_{dlever} equals one if the firm's leverage ratio falls (dlever<0) and zero otherwise. (2) dliquid measures the change in a firm's current ratio from the prior year, in which the current ratio is the ratio of current assets to current liabilities. An improvement in liquidity (dliquid>0) is a good signal regarding the firm's ability to service debt obligations. The indicator $F_{dliquid}$ equals one if the firm's liquidity improves and zero otherwise.

The remaining two signals are designed to measure changes in the efficiency of the firm's operations that reflect two key constructs underlying the decomposition of return on assets. (1) dmargin is the firm's current gross margin ratio, measured as gross margin scaled by sales, minus the prior year's gross margin ratio. An improvement in margins signifies a potential improvement in factor costs, a reduction in inventory costs, or a rise in the price of the firm's product. The indictor $F_{dmargin}$ equals one if dmargin>0 and zero otherwise. (2) dturn is the firm's current year asset turnover ratio, measured as total sales scaled by one-year-lagged total assets, minus the prior year's asset turnover ratio. An improvement in asset turnover ratio signifies greater productivity from the asset base. The indicator, F_{dturn}, equals one if dturn>0 and zero otherwise.

Piotroski (2000) forms a composite score (F) as the sum of the individual binary signals:

$$F = F_{roa} + F_{droa} + F_{Cf/A} + F_{Acc} + F_{dmargin} + F_{dturn} + F_{dlever} + F_{diquid}. \tag{B13}$$

At the end of June of each year t, we sort stocks based on F for the fiscal year ending in calendar year $t-1$ to form six portfolios: low (F = 0, 1, 2), 3, 4, 5, 6, and high (F = 7, 8). Because extreme F scores are rare, we combine scores 0, 1, and 2 into the low portfolio and scores 7 and 8 into the high portfolio. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.1.27 Quarterly Fundamental Score (Fq1, Fq6, and Fq12) We use quarterly accounting data and the same binary variables from Piotroski (2000) to measure quarterly F-score. Four variables used to measure profitability are: (1) roa is quarterly net income minus nonrecurrent gains/losses scaled by one-quarter-lagged total assets. If the firm's roa is positive, the indicator variable F_{roa} equals one and zero otherwise. (2) Cf/A is cash flows from operation scaled by one-quarter-lagged total assets. If the firm's Cf/A is positive, the indicator variable $F_{Cf/A}$ equals one and zero otherwise. (3) droa is the current quarter's roa minus the roa from four quarters ago. If droa is positive, the indicator variable F_{droa} equals one and zero otherwise. (4) the indicator F_{Acc} equals one if Cf/A > roa and zero otherwise.

We use two variables to measure changes in capital structure and a firm's ability to meet debt obligations. (1) dlever is the change in the ratio of total long-term debt to the average of current and one-quarter-lagged total assets. F_{dLever} equals one if the firm's leverage ratio falls (dlever<0) and zero otherwise. (2) dliquid measures the change in the current ratio from four quarters ago, in which the current ratio is the ratio of current assets to current liabilities. An improvement in liquidity (dliquid>0) is a good signal about the firm's ability to service debt obligations. The indicator $F_{dliquid}$ equals one if the firm's liquidity improves and zero otherwise.

The remaining two signals are designed to measure changes in the efficiency of the firm's operations that reflect two key constructs underlying the decomposition of return on assets. (1) dmargin is the current gross margin ratio, measured as gross margin scaled by sales, minus gross margin ratio four quarters ago. An improvement in margins signifies a potential improvement in factor costs, a reduction in inventory costs, or a rise in the price of the firm's product. The indictor $F_{dmargin}$ equals one if dmargin>0 and zero otherwise. (2) dturn is the current year's asset turnover ratio, measured as total sales scaled by one-quarter-lagged total assets, minus the asset turnover ratio four quarters ago. An improvement in asset turnover ratio signifies greater productivity from the asset base. The indicator, F_{dturn}, equals one if dturn>0 and zero otherwise. The composite score (Fq) is the sum of the individual binary variables:

$$Fq = F_{roa} + F_{droa} + F_{Cf/A} + F_{Acc} + F_{dmargin} + F_{dturn} + F_{dlever} + F_{dliquid}. \tag{B14}$$

At the beginning of each month *t*, we sort stocks based on their most recently available Fq. We form six portfolios: low (F = 0, 1, 2), 3, 4, 5, 6, and high (F = 7, 8). Because extreme F scores are rare, we combine scores 0, 1, and 2 into the low portfolio and scores 7 and 8 into the high portfolio. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to Fq1, Fq6, and Fq12, respectively).

B.2.1.28 Ohlson's O-score (O) We follow Ohlson (1980) to construct the O-score (O):

$$O = -1.32 - 0.407\,log(TA) + 6.03TLTA - 1.43WCTA + 0.076CLCA - 1.72OENEG - 2.37NITA - 1.83FUTL + 0.285IN2 - 0.521CHIN, \tag{B15}$$

in which TA is total assets, TLTA is the leverage ratio defined as total debt divided by total assets, WCTA is working capital, measured as current assets minus current liabilities divided by total assets, CLCA is current liabilities divided by current assets, OENEG equals one if total liabilities exceed total assets and zero otherwise, NITA is net income divided by total assets, FUTL is the fund provided by operations divided by total liabilities, IN_2 equals one if net income is negative for the last two years and zero otherwise, CHIN is $(NI_s - NI_{s-1})/(|NI_s| + |NI_{s-1}|)$, in which NI_s and NI_{s-1} are net income for the current and prior years. We winsorize all nondummy variables on the right-hand side of Equation (B15) at the 1^{st} and 99^{th} percentiles of their distributions each year. At the end of June of each year *t*, we sort stocks into deciles based on the O-score for the fiscal year ending in year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.1.29 Quarterly O-score (Oq1, Oq6, and Oq12) We use quarterly accounting data to construct the quarterly O-score (Oq) as:

$$Oq = -1.32 - 0.407\,log(TA) + 6.03TLTA - 1.43WCTA + 0.076CLCA - 1.72OENEG - 2.37NITA - 1.83FUTL + 0.285IN2 - 0.521CHIN, \quad \text{(B16)}$$

in which TA is total assets, TLTA denotes the leverage ratio defined as total debt divided by total assets, WCTA is working capital, measured as current assets minus current liabilities divided by total assets, CLCA is current liabilities divided by current assets, OENEG equals one if total liabilities exceed total assets and zero otherwise, NITA is net income divided by total assets, FUTL is the fund provided by operations divided by total liabilities, IN_2 equals one if net income is negative for the current quarter and four quarters ago, and zero otherwise, CHIN is $(NI_s - NI_{s-4})/(|NI_s| + |NI_{s-4}|)$, in which NI_s and NI_{s-4} are the net income for the current quarter and four quarters ago. We winsorize all nondummy variables on the right-hand side of Equation (B16) at the 1st and 99th percentiles of their distributions each month. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released Oq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to Oq1, Oq6, and Oq12, respectively).

B.2.1.30 Altman's Z-score (Z) We follow Altman (1968) to construct the Z-score (Z):

$$Z = 1.2WCTA + 1.4RETA + 3.3EBITTA + 0.6METL + SALETA, \quad \text{(B17)}$$

in which WCTA is working capital divided by total assets, RETA is retained earnings divided by total assets, EBITTA is earnings before interest and taxes divided by total assets, METL is market equity at fiscal year-end divided by total liabilities, and SALETA is sales divided by total assets. We winsorize all nondummy variables on the right-hand side of Equation (B17) at the 1st and 99th percentiles of their distributions each year. At the end of June of each year *t*, we sort stocks into deciles based on Z for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.1.31 Quarterly Z-score (Zq1, Zq6, and Zq12) We use quarterly accounting data to construct the quarterly Z-score (Zq) as:

$$Zq = 1.2WCTA + 1.4RETA + 3.3EBITTA + 0.6METL + SALETA, \quad \text{(B18)}$$

in which WCTA is working capital divided by total assets, RETA is retained earnings divided by total assets, EBITTA is earnings before interest and taxes divided by total assets, METL is market equity at fiscal quarter end divided by total liabilities, and SALETA is sales divided by total assets. We winsorize all nondummy variables on the right-hand side of Equation (B18) at the 1st and 99th percentiles of their distributions each month. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released Zq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to Zq1, Zq6, and Zq12, respectively).

B.2.2 Value

B.2.2.1 Book-to-Market Equity (bm) Following Fama and French (1992), we estimate book-to-market equity (bm) as total shareholders' equity minus the book value of preferred stocks (zero if missing) for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of December of year *t*-1. Market equity is calculated as the closing price (unadjusted) multiplied by the number of total A-shares outstanding. At the end of June of each year *t*, we sort stocks into deciles based on bm for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.2 Book-to-June-end Market Equity (bmj) Following Asness and Frazzini (2013), we estimate book-to-June-end market equity (bmj) as total shareholders' equity minus the book value of preferred stocks (zero if missing) for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of June of year *t*. Market equity is calculated as the closing price (unadjusted) multiplied by the number of total A-shares outstanding. At the end of June of each year *t*, we sort stocks into deciles based on bmj for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.3 Quarterly Book-to-Market Equity (bmq1, bmq6, and bmq12) At the beginning of each month *t*, we sort stocks into deciles based on quarterly book-to-market equity (bmq) estimated as total shareholders' equity minus the book value of preferred stocks (zero if missing) for the most recently released quarter divided by the market equity at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to bmq1, bmq6, and bmq12, respectively).

B.2.2.4 Debt-to-Market Equity (dm) At the end of June of each year *t*, we sort stocks into deciles based on debt-to-market equity (dm) estimated as total liabilities for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of December of year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.5 Quarterly Debt-to-Market Equity (dmq1, dmq6, and dmq12) At the beginning of each month *t*, we sort stocks into deciles based on quarterly debt-to-market equity (dmq) estimated as total liabilities for the most recently released quarter divided by the market equity at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to dmq1, dmq6, and dmq12, respectively).

B.2.2.6 Assets-to-Market Equity (am) At the end of June of each year *t*, we sort stocks into deciles based on assets-to-market equity (am) estimated as total assets for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of December of year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.7 Quarterly Assets-to-Market Equity (amq1, amq6, and amq12) At the beginning of each month *t*, we sort stocks into deciles based on quarterly assets-to-market equity (amq) estimated as total assets for the most recently released quarter divided by the market equity at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to amq1, amq6, and amq12, respectively).

B.2.2.8 Earnings-to-Price (ep) At the end of June of each year *t*, we sort stocks into deciles based on earnings-to-price (ep) measured as income minus nonrecurrent gains/losses for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of December of year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.9 Quarterly Earnings-to-Price (epq1, epq6, and epq12) At the beginning of each month *t*, we sort stocks into deciles based on quarterly earnings-to-price (epq) estimated as income minus nonrecurrent gains/losses for the most recently released quarter divided by the market equity at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to epq1, epq6, and epq12, respectively).

B.2.2.10 Cash Flow-to-Price (cfp) At the end of June of each year *t*, we sort stocks into deciles based on cash flow-to-price (cfp) estimated as the net change in cash or cash equivalents between two most recent cash flow statements for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of December of year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.11 Quarterly Cash Flow-to-Price (cfpq1, cfpq6, and cfpq12) At the beginning of each month *t*, we sort stocks into deciles based on quarterly cash flow-to-price (cfpq) estimated as the net change in cash or cash equivalents between two most recent cash flow statements for the most recently released quarter divided by the market equity at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to cfpq1, cfpq6, and cfpq12, respectively).

B.2.2.12 5-year Sales Growth Rank (sr) Following Lakonishok, Shleifer, and Vishny (1994), we measure the five-year sales growth rank (sr) at the end of June of year *t* as the weighted average of the annual sales growth ranks for the prior five years: $\sum_{j=1}^{5}(6-j)\times Rank(t-j)$. The sales growth for year $t-j$ is the growth rate in sales from the fiscal year ending in $t-j-1$ to the fiscal year ending in $t-j$. Only firms with data for all the prior five years are used to determine the annual sales growth rank, and we exclude firms with nonpositive sales. At the end of June of each year *t*, we sort stocks into deciles based on sr for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.13 Enterprise Multiple (em) Following Loughran and Wellman (2011), we estimate enterprise multiple (em) as enterprise value divided by operating income before depreciation. Enterprise value is calculated as market equity plus total debt plus the book value of preferred stocks (zero if missing) minus cash and short-term investments. At the end of June of each year *t*, we sort stocks into deciles based on em for the fiscal year ending in year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.14 Quarterly Enterprise Multiple (emq1, emq6, and emq12) We estimate quarterly enterprise multiple (emq) as enterprise value divided by operating income before depreciation. Enterprise value is calculated as market equity plus total debt plus the book value of preferred stocks (zero if missing) minus cash and short-term investments. At the beginning of each month *t*, we sort stocks into deciles based on the most recently released emq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to emq1, emq6, and emq12, respectively).

B.2.2.15 Sales-to-Price (sp) At the end of June of each year *t*, we sort stocks into deciles based on sales-to-price (sp), which is calculated as operating revenue for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of December of year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.16 Quarterly Sales-to-Price (spq1, spq6, and spq12) At the beginning of each month *t*, we sort stocks into deciles based on quarterly sales-to-price (spq), which is calculated as quarterly operating revenue for the most recently released quarter divided by the market equity at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to spq1, spq6, and spq12, respectively).

B.2.2.17 Operating Cash Flow-to-Price (ocfp) At the end of June of each year *t*, we sort stocks into deciles based on operating cash flow-to-price (ocfp), which is calculated as operating cash flows for the fiscal year ending in calendar year *t*-1 divided by the market equity at the end of December of year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.2.18 Quarterly Operating Cash Flow-to-Price (ocfpq1, ocfpq6, and ocfpq12) At the beginning of each month *t*, we sort stocks into deciles based on quarterly operating cash flow-to-price (ocfpq), which is calculated as operating cash flows for the most recently released quarter divided by the market equity at the end of month *t*-1. We then calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to ocfpq1, ocfpq6, and ocfpq12, respectively).

B.2.2.19 Debt-to-Book Equity (de) At the end of June of each year *t*, we sort stocks into deciles based on debt-to-book equity (de) estimated as total liabilities for the fiscal year ending in calendar year *t*-1 divided by total shareholders' equity minus the book value of preferred stocks (zero if missing) for the fiscal year ending in calendar year *t*-1. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.2.20 Intangible Return (ir) Following Daniel and Titman (2006), at the end of June of each year t, we compute intangible return (ir) as the residuals from estimating the cross-sectional regression of each firm's past five-year stock returns on its five-year-lagged log book-to-market and five-year log book returns:

$$r(t-5,t) = \gamma_0 + \gamma_1 bm_{t-5} + \gamma_2 r^B(t-5,t) + \mu_t, \quad \text{(B19)}$$

in which $r(t-5,t)$ is the past five-year log stock returns from the end of year t-6 to the end of t-1, bm_{t-5} is the five-year-lagged log book-to-market, and $r^B(t-5,t)$ is the five-year log book returns. The five-year-lagged log book-to-market is calculated as $bm_{t-5} = log\left(\frac{B_{t-5}}{M_{t-5}}\right)$, in which B_{t-5} is the book equity for the fiscal year ending in calendar year t-6, and M_{t-5} is the market equity at the end of December of year t-6. We compute the five-year log book returns as $r^B(t-5,t) = log\left(\frac{B_t}{B_{t-5}}\right) + \sum_{s=t-5}^{t-1}(r_s - log\,(\frac{P_s}{P_{s-1}}))$, in which B_t is the book equity for the fiscal year ending in calendar year t-1, r_s is the stock return from the end of year s-1 to the end of year s, and P_s is the stock price per share at the end of year s. Book equity is total shareholders' equity minus the book value of preferred stocks (zero if missing). At the end of June of each year t, we sort stocks into deciles based on ir for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.2.21 Enterprise Book-to-Price (ebp) and Net Debt-to-Price (ndp) Following Penman, Richardson, and Tuna (2007), we measure enterprise book-to-price (ebp) as the ratio of the book value of net operating assets (net debt plus book equity) to the market value of net operating assets (net debt plus market equity). Net debt-to-price (ndp) is the ratio of net debt to market equity. Net debt is calculated as financial liabilities minus financial assets. Financial liabilities are measured as the sum of long-term debt, debt in current liabilities, and the carrying value of preferred stocks (zero if missing). Financial assets are measured as cash and short-term investments. At the end of June of each year t, we sort stocks into deciles based on ebp and ndp for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.2.22 Quarterly Enterprise Book-to-Price (ebpq1, ebpq6, and ebpq12) and Quarterly Net Debt-to-Price (ndpq1, ndpq6, and ndpq12) We measure quarterly enterprise book-to-price (ebpq) as the ratio of the book value of net operating assets (net debt plus book equity) to the market value of net operating assets (net debt plus market equity). Quarterly net debt-to-price (ndpq) is the ratio of net debt to market equity. Net debt is calculated as financial liabilities minus financial assets. Financial liabilities are measured as the sum of long-term debt, debt in current liabilities, and the carrying value of preferred stocks (zero if missing). Financial assets are measured as cash and short-term investments. At the beginning of each month t, we sort stocks into deciles based on their most recently released ebpq and ndpq. The market equity is measured at the end of month t-1. We exclude firms with nonpositive book or market value of net operating assets. For the ndpq portfolios, we exclude firms with nonpositive net debt. We then calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to ebpq1, ebpq6, ebpq12, ndpq1, ndpq6, and ndpq12, respectively).

B.2.3 Investment

B.2.3.1 Abnormal Corporate Investment (aci) At the end of June of year t, we measure abnormal corporate investment (aci) as $\frac{Ce_{t-1}}{\left[\frac{Ce_{t-2}+Ce_{t-3}+Ce_{t-4}}{3}\right]} - 1$, in which Ce_{t-j} is capital expenditure scaled by sales for the fiscal year ending in calendar year t-j. We exclude firms with negative sales. We measure capital expenditure as cash flows paid out for purchasing fixed assets, intangible assets, and other long-term investment. At the end of June of each year *t*, we sort stocks into deciles based on aci for the fiscal year ending in year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.2 Investment-to-Assets (ag) At the end of June of each year *t*, we sort stocks into deciles based on investment-to-assets (ag), which is estimated as total assets from the fiscal year ending in calendar year *t*-1 divided by total assets for the fiscal year ending in calendar year *t*-2 minus one. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.3 Quarterly Investment-to-Assets (agq1, agq6, and agq12) We measure quarterly investment-to-assets (agq) as quarterly total assets divided by four-quarter-lagged total assets minus one. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released agq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to agq1, agq6, and agq12, respectively).

B.2.3.4 Changes in PPE and Inventory-to-Assets (dpia) We define changes in PPE and inventory-to-assets (dpia) as the annual change in gross property, plant, and equipment (fixed assets) plus the annual change in inventory, scaled by one-year-lagged total assets. At the end of June of each year *t,* we sort stocks into deciles based on dpia for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.5 Net Operating Assets (noa) and Changes in Net Operating Assets (dnoa) We measure net operating assets as operating assets minus operating liabilities. Operating assets are total assets minus cash and short-term investments. Operating liabilities are total assets minus debt included in current liabilities (zero if missing), minus long-term debt (zero if missing), minus minority interests, minus the book value of preferred stocks (zero if missing), and minus common equity. noa is net operating assets scaled by one-year-lagged total assets. Changes in net operating assets (dnoa) is the annual change in net operating assets scaled by one-year-lagged total assets. At the end of June of each year *t,* we sort stocks into deciles based on noa and dnoa for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.6 x-year Investment Growth (ig, ig2, and ig3) At the end of June of each year *t*, we sort stocks into deciles based on x-year investment growth, which is the growth rate in capital expenditure from the fiscal year ending in calendar year *t*-*x*-1 to the fiscal year ending in year *t*-1 (*x*=1, 2, 3, corresponding to ig, ig2, and ig3, respectively). We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.7 Net Stock Issues (nsi) At the end of June of each year *t*, we sort stocks into deciles based on net stock issues (nsi) measured as the natural logarithm of the ratio of the split-adjusted total A-shares outstanding for the fiscal year ending in calendar year *t*-1 to the split-adjusted total A-shares outstanding for the fiscal year ending in calendar year *t*-2. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.8 Composite Equity Issuance (cei) At the end of June of each year *t*, we sort stocks into deciles based on composite equity issuance (cei), which is the log growth rate in the market equity not attributed to stock returns from year *t*-5 to *t*. That is calculated as $cei = log\left(\frac{ME_t}{ME_{t-5}}\right) - r(t-5,t)$, where $r(t-5,t)$ is the cumulative log stock returns from the last trading day of June in year *t*-5 to the last trading day of June in year *t*, and ME_t and ME_{t-5} are the market equity on the last trading day of June in year *t* and year *t*-5, respectively. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.9 Composite Debt Issuance (cdi) At the end of June of each year *t*, we sort stocks into deciles based on composite debt issuance (cdi), which is the log growth rate of total liabilities from the fiscal year ending in year *t*-6 to the fiscal year ending in year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.10 Inventory Growth (ivg) At the end of June of each year *t*, we sort stocks into deciles based on inventory growth (ivg), which is the annual growth rate in inventory from the fiscal year ending in year *t*-2 to the fiscal year ending in year *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.11 Inventory Changes (ivchg) At the end of June of each year *t*, we sort stocks into deciles based on inventory changes (ivchg), which is the annual change in inventory from the fiscal year ending in year *t*-2 to *t*-1 scaled by the average of total assets for the fiscal years ending in calendar years *t*-2 and *t*-1. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.12 Operating Accruals (oa) Following Hribar and Collins (2002), we measure operating accruals (oa) as net profits minus operating cash flows scaled by one-year-lagged total assets. At the end of June of each year *t,* we sort stocks into deciles based on oa for the fiscal year ending in year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.13 Total Accruals (ta) We measure total accruals (ta) as net income minus cash flows scaled by one-year-lagged total assets. At the end of June of each year *t*, we sort stocks into deciles based on ta for the fiscal year ending in year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.14 Changes in Net Noncash Working Capital (dwc), Current Operating Assets (dcoa), and Current Operating Liabilities (dcol) Net noncash working capital (wc) is calculated as current operating assets (coa) minus current operating liabilities (col). Current operating assets (coa) are current assets minus cash and short-term investments. Current operating liabilities (col) are current liabilities minus debt in current liabilities (zero if missing). At the end of June of each year *t*, we sort stocks into deciles based on changes in wc, coa, and col (dwc, dcoa, and dcol) from the fiscal year ending in calendar year *t*-2 to the fiscal year ending in calendar year *t*-1 scaled by total assets for the fiscal year ending in calendar year *t*-2. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.15 Changes in Net Noncurrent Operating Assets (dnco), Noncurrent Operating Assets (dnca), and Noncurrent Operating Liabilities (dncl) Net noncurrent operating assets (nco) are calculated as noncurrent operating assets minus noncurrent operating liabilities. Noncurrent operating assets (nca) are total assets minus current assets, minus long-term investments (zero if missing). Long-term investments are measured as the sum of held-to-maturity investments, long-term equity investments, investment in real estate, and fixed deposit. Noncurrent operating liabilities (ncl) are total liabilities minus current liabilities, and minus long-term debt (zero if missing). At the end of June of each year *t*, we sort stocks into deciles based on changes in nco, nca, and ncl (dnco, dnca, and dcl) from the fiscal year ending in calendar year *t*-2 to the fiscal year ending in calendar year *t*-1 scaled by total assets for the fiscal year ending in calendar year *t*-2. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.3.16 Changes in Net Financial Assets (dfin), Short-Term Investments (dsti), Long-Term Investments (dlti), Financial Liabilities (dfnl), and Book Equity (dbe) Net financial assets are calculated as financial assets minus financial liabilities. We measure financial assets as short-term investments plus long-term investments (zero if missing). Long-term investments are measured as the sum of held-to-maturity investments, long-term equity investments, investment in real estate, and fixed deposit. Financial liabilities are the sum of long-term debt (zero if missing), debt in current liabilities (zero if missing), and the book value of preferred stocks (zero if missing). At the end of June of year t, we measure dfin, dsti, dlti, and dfnl as the annual change in net financial assets, short-term investments, long-term investments, and financial liabilities from the fiscal year ending in calendar year t-2 to the fiscal year ending in calendar year t-1 scaled by total assets for the fiscal year ending in calendar year t-2. We measure dbe as the change in book equity for the fiscal year ending in calendar year t-1 scaled by total assets for the fiscal year ending in calendar year t-2. At the end of June of each year t, we sort stocks into deciles based on dfin, dsti, dlti, dfnl, and dbe for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.3.17 Discretionary Accruals (dac) We measure discretionary accruals (dac) using the following model:

$$\frac{oa_{i,t}}{A_{i,t-1}} = \alpha_1 \frac{1}{A_{i,t-1}} + \alpha_2 \frac{dSALE_{i,t} - dREC_{i,t}}{A_{i,t-1}} + \alpha_3 \frac{PPE_{i,t}}{A_{i,t-1}} + e_{i,t}, \tag{B20}$$

in which $oa_{i,t}$ represents operating accruals for firm i, $A_{i,t-1}$ is total assets at the end of year t-1, $dSALE_{i,t}$ is the annual change in sales from year t-1 to t, $dREC_{i,t}$ is the annual change in net accounts receivable from year t-1 to t, and $PPE_{i,t}$ is gross property, plant, and equipment (fixed assets) at the end of year t. We winsorize the variables on the right-hand side of Equation (B20) at the 1st and 99th percentiles of their distributions each year. We estimate the cross-sectional regression (B20) for each Shenwanhongyuan industry each year and require at least six firms for each regression. The discretionary accruals for stock i are defined as the residual from the regression, $e_{i,t}$. At the end of June of each year t, we sort stocks into deciles based on dac for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.3.18 Percent Operating Accruals (poa) At the end of June of each year t, we sort stocks into deciles based on percent operating accruals (poa) measured as operating accruals for the fiscal year ending in calendar year t-1 divided by the absolute value of net income for the fiscal year ending in calendar year t-1. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.3.19 Percent Total Accruals (pta) At the end of June of each year t, we sort stocks into deciles based on percent total accruals (pta) measured as total accruals divided by the absolute value of net income for the fiscal year ending in calendar year t-1. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.3.20 Percent Discretionary Accruals (pda) At the end of June of each year t, we sort stocks into deciles based on percent discretionary accruals (pda) measured as the discretionary accruals for the fiscal year ending in calendar year t-1 multiplied by total assets for the fiscal year ending in calendar year t-2, and then scaled by the absolute value of net income for the fiscal year ending in calendar year t-1. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.3.21 Quarterly Current Asset Growth (cagq1, cagq6, and cagq12) and Noncurrent Asset Growth (ncagq1, ncagq6, and ncagq12) Total assets are the sum of current assets and noncurrent assets, so total asset growth can be decomposed into current asset growth and noncurrent asset growth. We measure quarterly current asset growth (cagq) as quarterly current assets minus four-quarter-lagged current assets, divided by four-quarter-lagged total assets. We measure quarterly noncurrent asset growth (ncagq) as quarterly noncurrent assets minus four-quarter-lagged noncurrent assets, divided by four-quarter-lagged total assets. At the beginning of each month t, we sort stocks into deciles based on their most recently released cagq and ncagq, and calculate monthly decile portfolio returns over months [t, $t+n-1$] (n=1, 6, and 12, corresponding to cagq1, cagq6, cagq12, ncagq1, ncagq6, and ncagq12, respectively).

B.2.3.22 Quarterly Cash Growth (cashgq1, cashgq6, and cashgq12), Fixed Asset Growth (fagq1, fagq6, and fagq12), Noncash Current Asset Growth (nccgq1, nccgq6, and nccgq12), and Other Asset Growth (oagq1, oagq6, and oagq12) According to Cooper, Gulen, and Schill (2008), asset growth can be decomposed as follows: Total asset growth = cash growth + noncash current asset growth + property, plant, and equipment growth + other asset growth. We measure quarterly cash growth (cashgq) as quarterly cash minus four-quarter-lagged cash, divided by total assets four quarters ago. We measure quarterly noncash current asset growth (nccagq) as the annual change in quarterly current assets minus cash, divided by four-quarter-lagged total assets. We measure quarterly property, plant, and equipment growth (fagq) as quarterly fixed assets minus four-quarter-lagged fixed assets, divided by total assets four quarters ago. We measure quarterly other asset growth (oagq) as the annual change in quarterly other assets, divided by four-quarter-lagged total assets. Other assets are equal to noncurrent assets minus fixed assets. At the beginning of each month t, we sort stocks into deciles based on their most recently released cashgq, fagq, nccgq, and oagq, and calculate monthly decile portfolio returns over months [t, $t+n-1$] (n=1, 6, and 12, corresponding to cashgq1, cashgq6, cashgq12, fagq1, fagq6, fagq12, nccgq1, nccgq6, nccgq12, oagq1, oagq6, and oagq12, respectively).

B.2.4 Others

B.2.4.1Advertising Expense-to-Market (adm) At the end of June of each year t, we sort stocks into deciles based on advertising expense-to-market (adm), measured as advertising (selling) expense for the fiscal year ending in calendar year t-1 divided by the market equity at the end of December of t-1. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.2 Growth in Advertising Expense (gad) At the end of June of each year t, we sort stocks into deciles based on growth in advertising expense (gad), which is the growth rate of advertising (selling) expense from the fiscal year ending in calendar year t-2 to the fiscal year ending in calendar year t-1. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.3 R&D Expense-to-Market (rdm) At the end of June of each year t, we sort stocks into deciles based on R&D expense-to-market (rdm), measured as R&D (administrative) expense for the fiscal year ending in calendar year t-1 divided by the market equity at the end of December of t-1. We only keep firms with positive R&D expense. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.4 Quarterly R&D Expense-to-Market (rdmq1, rdmq6, and rdmq12) At the beginning of each month t, we sort stocks into deciles based on quarterly R&D expense-to-market (rdmq), measured as the most recently released quarterly R&D (administrative) expense divided by the market equity at the end of month t-1. We only keep firms with positive R&D expense. We then calculate monthly decile portfolio returns over months [t, $t+n-1$] (n=1, 6, and 12, corresponding to rdmq1, rdmq6, and rdmq12, respectively).

B.2.4.5 R&D Expense-to-Sales (rds) At the end of June of each year t, we sort stocks into deciles based on R&D expense-to-sales (rds), which is R&D (administrative) expense for the fiscal year ending in calendar year t-1 divided by sales for the fiscal year ending in calendar year t-1. We only keep firms with positive R&D expense. We then calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.6 Quarterly R&D Expense-to-Sales (rdsq1, rdsq6, and rdsq12) We measure quarterly R&D expense-to-sales (rdsq) as quarterly R&D (administrative) expense divided by quarterly sales. We only keep firms with positive R&D expense. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released rdsq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to rdsq1, rdsq6, and rdsq12, respectively).

B.2.4.7 Operating Leverage (ol) Following Novy-Marx (2011), we estimate operating leverage (ol) as operating costs divided by total assets. At the end of June of each year *t*, we sort stocks into deciles based on ol for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.8 Quarterly Operating Leverage (olq1, olq6, and olq12) We measure quarterly operating leverage (olq) as quarterly operating costs divided by total assets. At the beginning of each month *t*, we sort stocks into deciles based on their most recently released olq and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to olq1, olq6, and olq12, respectively).

B.2.4.9 Hiring Rate (hn) At the end of June of each year *t*, we sort stocks into deciles based on the hiring rate (hn), measured as $(N_{t-1} - N_{t-2})/(0.5N_{t-1} + 0.5N_{t-2})$, in which N_{t-j} is the number of employees for the fiscal year ending in calendar year *t-j*. We exclude firms with zero hn. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.10 Firm Age (age1, age6, and age12) Following Jiang, Lee, and Zhang (2005), we compute firm age (age) as the number of months between the portfolio formation date and the firm's IPO date. At the beginning of each month *t*, we sort stocks into deciles based on age at the end of month *t*-1 and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to age1, age6, and age12, respectively).

B.2.4.11 % Change in Sales minus % Change in Inventory (dsi) Following Abarbanell and Bushee (1998), we define the %d(.) operator as the percentage change in the variable in parentheses from its average over the prior two years. For example, %d(Sales)=[Sales(*t*)-E[Sales(*t*)]]/E[Sales(*t*)], in which E[Sales(*t*)]=[Sales(*t*-1)+Sales(*t*-2)]/2. dsi is calculated as %d(Sales)-%d(Inventory), in which sales are operating revenue, and inventory is net inventory. We exclude firms with negative sales. At the end of June of each year *t*, we sort stocks into deciles based on dsi for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.12 % Change in Sales minus % Change in Accounts Receivable (dsa) Following Abarbanell and Bushee (1998), we define the %d(.) operator as the percentage change in the variable in parentheses from its average over the prior two years. For example, %d(Sales)=[Sales(*t*)-E[Sales(*t*)]]/E[Sales(*t*)], in which E[Sales(*t*)]=[Sales(*t*-1)+Sales(*t*-2)]/2. dsa is calculated as %d(Sales)-%d(Accounts receivable), in which sales are operating revenue, and accounts receivable is total receivables. We exclude firms with negative sales. At the end of June of each year *t*, we sort stocks into deciles based on dsa for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.13 % Change in Gross Margin minus % Change in Sales (dgs) Following Abarbanell and Bushee (1998), we define the %d(.) operator as the percentage change in the variable in parentheses from its average over the prior two years. For example, %d(Sales)=[Sales(*t*)-E[Sales(*t*)]]/E[Sales(*t*)], in which E[Sales(*t*)]=[Sales(*t*-1)+Sales(*t*-2)]/2. dgs is calculated as %d(Gross margin)-%d(Sales), in which sales are operating revenue, and gross margin is sales minus cost of goods sold. We exclude firms with negative sales. At the end of June of each year t, we sort stocks into deciles based on dgs for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.14 % Change in Sales minus % Change in SG&A (dss) Following Abarbanell and Bushee (1998), we define the %d(.) operator as the percentage change in the variable in parentheses from its average over the prior two years. For example, %d(Sales)=[Sales(t)-E[Sales(t)]]/E[Sales(t)], in which E[Sales(t)]=[Sales(t-1)+Sales(t-2)]/2. dss is calculated as %d(Sales)-%d(SG&A), in which sales are operating revenue, and SG&A are selling, general, and administrative expenses. We exclude firms with negative sales. At the end of June of each year t, we sort stocks into deciles based on dss for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.15 Effective Tax Rate (etr) Following Abarbanell and Bushee (1998), we measure the effective tax rate (etr) as

$$etr(t) = \left[\frac{TaxExpense(t)}{EBT(t)} - \frac{1}{3}\sum_{\tau=1}^{3}\frac{TaxExpense(t-\tau)}{EBT(t-\tau)}\right] \times dEPS(t), \tag{B21}$$

in which TaxExpense(t) is the total income taxes paid in year t, EBT(t) is EBIT minus interest expense, and dEPS is the change in split-adjusted earnings per share between year $t-1$ and t, divided by the closing price at the end of year $t-1$. At the end of June of each year t, we sort stocks into deciles based on etr for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.16 Labor Force Efficiency (lfe) Following Abarbanell and Bushee (1998), we measure labor force efficiency (lfe) as

$$lfe(t) = \left[\frac{Sales(t)}{Employees(t)} - \frac{Sales(t-1)}{Employees(t-1)}\right] / \frac{Sales(t-1)}{Employees(t-1)}, \tag{B22}$$

In which Sales(t) is net sales in year t, and Employees(t) is the number of employees. At the end of June of each year t, we sort stocks into deciles based on lfe for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.17 Tangibility (tan) We measure tangibility (tan) as cash holdings + 0.715×accounts receivable + 0.547×inventory + 0.535×gross property, plant, and equipment (fixed assets), all scaled by total assets. At the end of June of each year t, we sort stocks into deciles based on tan for the fiscal year ending in calendar year t-1 and calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.18 Quarterly Tangibility (tanq1, tanq6, and tanq12) We measure quarterly tangibility (tanq) as cash holdings (zero if missing) + 0.715×accounts receivable (zero if missing) + 0.547×inventory + 0.535×gross property, plant, and equipment (fixed assets), all scaled by total assets. At the beginning of each month t, we sort stocks into deciles based on the most recently released tanq and calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to tanq1, tanq6, and tanq12, respectively).

B.2.4.19 Cash Flow Volatility (vcf1, vcf6, and vcf12) We measure cash flow volatility (vcf) as the standard deviation of the ratio of operating cash flows to sales during the past 16 quarters (with a minimum of eight nonmissing quarters). At the beginning of each month t, we sort stocks into deciles based on the most recently released vcf and calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to vcf1, vcf6, and vcf12, respectively).

B.2.4.20 Cash-to-Assets (cta1, cta6, and cta12) We measure cash-to-assets (cta) as cash and cash equivalents divided by total assets. At the beginning of each month t, we sort stocks into deciles based on the most recently released cta and calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to cta1, cta6, and cta12, respectively).

B.2.4.21 Earnings Persistence (eper) and Earnings Predictability (eprd) Following

Francis, LaFond, Olsson, and Schipper (2004), we estimate earnings persistence (eper) and earnings predictability (eprd) from a first-order autoregressive model for annual earnings per share. Earnings per share are net profits minus nonrecurrent gains/losses scaled by the number of total A-shares outstanding. At the end of June of each year *t*, we estimate the autoregressive model in the 10-year rolling window up to the fiscal year ending in calendar year *t*−1. Only firms with a complete 10-year history are included. eper is measured as the slope coefficient, and eprd is measured as the residual volatility. At the end of June of each year *t*, we sort stocks into deciles based on eper and eprd, and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.22 Earnings Smoothness (esm) Following Francis, LaFond, Olsson, and Schipper (2004), we measure earnings smoothness (esm) as the ratio of the standard deviation of earnings scaled by one-year-lagged total assets to the standard deviation of cash flow from operations scaled by one-year-lagged total assets. Earnings are net profits minus nonrecurrent gains/losses. At the end of June of each year *t*, we sort stocks into deciles based on esm estimated over the 10-year rolling window up to the fiscal year ending in calendar year *t*−1. Only firms with a complete 10-year history are included. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.23 Value Relevance of Earnings (evr) Following Francis, LaFond, Olsson, and Schipper (2004), we measure the value relevance of earnings (evr) as the R^2 from the following rolling-window regression:

$$R_{i,t} = \delta_{i,0} + \delta_{i,1} EARN_{i,t} + \delta_{i,2} dERARN_{i,t} + \epsilon_{i,t}, \qquad \text{(B23)}$$

in which $R_{i,t}$ is firms *i*'s 15-month stock return ending three months after the end of fiscal year ending in calendar year *t*. $EARN_{i,t}$ is net profits minus nonrecurrent gains/losses for the fiscal year ending in calendar year *t*, scaled by the fiscal year-end market equity. $dERARN_{i,t}$ is the one-year change in earnings scaled by market equity. At the end of June of each year *t*, we sort stocks into deciles based on evr estimated over the 10-year rolling window up to the fiscal year ending in calendar year *t*−1. Only firms with a complete 10-year history are included. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.24 Earnings Timeliness (etl) and Earnings Conservatism (ecs) Following Francis, LaFond, Olsson, and Schipper (2004), we measure earnings timeliness (etl) and earnings conservatism (ecs) from the following rolling-window regression:

$$EARN_{i,t} = \alpha_{i,0} + \alpha_{i,1} NEG_{i,t} + \beta_{i,1} R_{i,t} + \beta_{i,2} NEG_{i,t} \times R_{i,t} + \epsilon_{i,t}, \qquad \text{(B24)}$$

in which $EARN_{i,t}$ is net profits minus nonrecurrent gains/losses for the fiscal year ending in calendar year *t* scaled by the fiscal year-end market equity, $R_{i,t}$ is firm *i*'s 15-month stock return ending three months after the end of fiscal year ending in calendar year *t*, $NEG_{i,t}$ equals one if $R_{i,t}<0$, and zero otherwise. We measure etl as the R^2 and ecs as $(\beta_{i,1} + \beta_{i,2})/\beta_{i,1}$ from the regression in (B24). At the end of June of each year *t*, we sort stocks into deciles based on etl and ecs, both of which are estimated over the 10-year rolling window up to the fiscal year ending in calendar year *t*−1. Only firms with a complete 10-year history are included. We then calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.25 Asset Liquidity (ala and alm) We measure asset liquidity as cash + 0.75 × noncash current assets + 0.50 × tangible fixed assets. Cash is defined as cash and short-term investments, noncash current assets are current assets minus cash, and tangible fixed assets are total assets minus current assets, minus goodwill (zero if missing), and minus intangibles (zero if missing). ala is asset liquidity scaled by one-year-lagged total assets. alm is asset liquidity scaled by one-year-lagged market value of assets. The market value of assets is total assets plus market equity minus book equity. At the end of June of each year *t*, we sort stocks into deciles based on ala and alm for the fiscal year ending in calendar year *t*-1 and calculate monthly decile portfolio returns from July of year *t* to June of *t*+1.

B.2.4.26 Quarterly Asset Liquidity (alaq1, alaq6, alaq12, almq1, almq6, and almq12) We measure quarterly asset liquidity as cash + 0.75 × noncash current assets + 0.50 × tangible fixed assets. Cash is defined as cash and short-term investments, noncash current assets are current assets minus cash, and tangible fixed assets are total assets minus current assets, minus goodwill (zero if missing), and minus intangibles (zero if missing). alaq is quarterly asset liquidity scaled by one-quarter-lagged total assets. almq is quarterly asset liquidity scaled by one-quarter-lagged market value of assets. The market value of assets is total assets plus market equity minus book equity. At the beginning of each month *t*, we sort stocks into deciles based on the most recently released alaq and almq, and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to alaq1, alaq6, alaq12, almq1, almq6, and almq12, respectively).

B.2.4.27 Tax Expense Surprises (tes1, tes6, and tes12) Following Thomas and Zhang (2011), we measure tax expense surprises (tes) as the change in total tax expense, which is tax expense in quarter *q* minus tax expense in quarter *q*-4, scaled by total assets in quarter *q*-4. At the beginning of each month *t*, we sort stocks into deciles based on the most recently released tes and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to tes1, tes6, and tes12, respectively).

B.2.4.28 Changes in Analyst Earnings Forecasts (def1, def6, and def12) Following Hawkins, Chamberlin, and Daniel (1984), we measure changes in analyst earnings forecasts (def) as:

$$def = (f_{i,t-1} - f_{i,t-2})/(0.5|f_{i,t-1}| + 0.5|f_{i,t-2}|), \tag{B25}$$

where $f_{i,t-1}$ is the consensus mean forecast issued in month *t*−1 for firm *i*'s current fiscal year earnings, respectively. At the beginning of each month *t*, we sort stocks into deciles based on def and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to def1, def6, and def12, respectively).

B.2.4.29 Revisions in Analyst Earnings Forecasts (re1, re6, and re12) Following Chan, Jegadeesh, and Lakonishok (1996), we measure revisions in analyst earnings forecasts (re) as the six-month moving average of past changes in analysts' forecasts:

$$re_{i,t} = \sum_{\tau=1}^{6} \frac{f_{i,t-\tau} - f_{i,t-\tau-1}}{P_{i,t-\tau-1}}, \tag{B26}$$

where $f_{i,t-\tau}$ is the consensus mean forecast issued in month $t-\tau$ for firm *i*'s current fiscal year earnings, and $P_{i,t-\tau-1}$ is the closing price in the prior month. We adjust for any stock splits and require a minimum of four monthly forecast changes when constructing re. At the beginning of each month *t*, we sort stocks into deciles based on re and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to re1, re6, and re12, respectively).

B.2.4.30 Earnings Forecast–to-Price (efp1, efp6, and efp12) Following Elgers, Lo, and Pfeiffer (2001), we measure analyst earnings forecast-to-price (efp) as the consensus median forecasts for the current fiscal year divided by the closing price. At the beginning of each month *t*, we sort stocks into deciles based on efp estimated in month *t*-1 and calculate monthly decile portfolio returns over months [*t*, *t*+*n*-1] (*n*=1, 6, and 12, corresponding to efp1, efp6, and efp12, respectively).

B.2.4.31 Analyst Coverage (ana1, ana6, and ana12) We measure analyst coverage (ana) as the number of analyst earnings forecasts for the current fiscal year. At the beginning of each month *t*, we sort stocks into deciles based on ana estimated in month *t*-1 and calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to ana1, ana6, and ana12, respectively).

B.2.4.32 Dispersion in Analyst Forecasts (dis1, dis6, and dis12) We measure dispersion in analyst earnings forecasts (dis) as the ratio of the standard deviation of earnings forecasts for the current fiscal year to the absolute value of the consensus mean forecasts. At the beginning of each month t, we sort stocks into deciles based on dis estimated in month t−1 and calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to dis1, dis6, and dis12, respectively).

B.2.4.33 Institutional Ownership (io) At the end of June of each year t, we sort stocks into deciles based on institutional ownership (io) for the fiscal year ending in calendar year t-1. Firm-level institutional ownership is calculated as the sum of the fractions of free-float A-shares held by mutual funds, brokers, insurance companies, security funds, entrusts, etc. We calculate monthly decile portfolio returns from July of year t to June of t+1.

B.2.4.34 Quarterly Institutional Ownership (ioq1, ioq6, and ioq12) At the beginning of each month t, we sort stocks into deciles based on the most recently released quarterly institutional ownership (ioq). Firm-level institutional ownership is calculated as the sum of the fractions of free-float A-shares held by mutual funds, brokers, insurance companies, security funds, entrusts, etc. We calculate monthly decile portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to ioq1, ioq6, and ioq12, respectively).

B.2.4.35 SOE Indicator (soe1, soe6, and soe12) We create a dummy variable, soe, for state-owned enterprises. If a firm is a state-owned enterprise, soe equals 1, and 0 otherwise. At the beginning of each month t, we sort stocks into two groups based on soe in month t-1 and calculate monthly portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to soe1, soe6, and soe12, respectively).

B.2.4.36 Margin Trading and Short Selling Indicator (margin1, margin6, and margin12) We create a dummy variable, margin, for firms that are approved for margin trading and short selling. If a firm is on the margin trading and short selling approved list, margin equals 1, and 0 otherwise. At the beginning of each month t, we sort stocks into two groups based on margin in month t-1 and calculate monthly portfolio returns over months [t, t+n-1] (n=1, 6, and 12, corresponding to margin1, margin6, and margin12, respectively).

C. Additional Details on Methodologies

C.1 Liu, Stambaugh, and Yuan's (2019) Chinese Three-/Four-Factor Model

The CH3 model has three factors. The first factor is the market factor, which is the same as in the Chinese CAPM. We then independently sort stocks into two size groups and three value groups to construct the size and value factors. The breakpoint for the size groups is the median market capitalization of the largest 70% of A-share stocks, and the breakpoints for the value groups are the 30th and 70th percentiles of the earnings-to-price ratio (E/P) for the largest 70% of A-share stocks. We obtain six portfolios at the intersections of the size and value groups. The size factor, SMB, is computed as the difference between the simple average of the returns on the three portfolios of small stocks (below the median market capitalization) and the simple average of the returns on the three portfolios of large stocks (above the median market capitalization). The value factor, VMG, is computed as the difference between the simple average of the returns on the two portfolios of value stocks (above the 70th percentile of E/P) and the simple average of the returns on the two portfolios of growth stocks (below the 30th percentile of E/P).

The CH4 model adds a liquidity factor, PMO, to the CH3 model. The liquidity factor is based on abnormal turnover, which is the ratio of average daily turnover in the past month to average daily turnover in the past year. Similar to the construction of VMG, we construct PMO by independently sorting stocks into two size groups and three abnormal turnover groups. PMO is then computed as the difference between the simple average of the returns on the two portfolios of low abnormal turnover stocks (below the 30th percentile of abnormal turnover) and the simple average of the returns on the two portfolios of high abnormal turnover stocks (above the 70th percentile of abnormal turnover).

C.2 Multiple Tests

C.2.1 Harvey and Liu's (2020) Multiple-Testing Method

We describe the Harvey and Liu's (2020) multiple-testing method in the context of testing the performance of many trading strategies. There are *N* univariate strategies and *D* time periods. We arrange the high-minus-low return series into a *D*×*N* matrix, X_0. Suppose one believes that a fraction p_0 of the *N* strategies are true. We develop a simulation-based framework to evaluate the error rates related to testing multiple hypotheses for a given p_0. The choice of p_0 is inherently subjective and depends on both previous experience and data summary statistics.

For a given p_0, we start by choosing $p_0 \times N$ strategies that are deemed to be true. A simple way to choose these strategies is to first rank the strategies by their absolute *t*-statistics and then select the top $p_0 \times N$ strategies with the highest absolute *t*-statistics. While this approach aligns with the idea that strategies with higher absolute *t*-statistics are more likely to be true, it ignores the sampling uncertainty inherent in ranking the strategies. To account for this uncertainty, we follow the steps below to perturb the data, rank the strategies based on the perturbed data, and evaluate the error rates:

Step 1: We bootstrap the time periods of X_0 and create an alternative panel of high-minus-low returns, X_i. For X_i, we calculate the corresponding 1×*N* vector of absolute *t*-statistics, t_i.

Step 2: We rank the strategies based on their absolute *t*-statistics, t_i. For the top $p_0 \times N$ strategies with the highest absolute *t*-statistics, we find the corresponding strategies in X_0. We adjust these strategies in X_0 so that their means equal the means of the top $p_0 \times N$ strategies in X_i. We denote the return matrix of these adjusted strategies by $X_{0,1}^{(i)}$. For the remaining strategies in X_0, we adjust their returns to have zero means and denote the return matrix for these adjusted strategies by $X_{0,0}^{(i)}$. We combine $X_{0,1}^{(i)}$ and $X_{0,0}^{(i)}$ into a new return matrix Y_i by concatenating them.

Step 3: For Y_i, we bootstrap the time periods *J* times to evaluate the error rates for a fixed *t*-statistic threshold. We know which strategies in Y_i are believed to be true (and false). This allows us to summarize the testing outcomes ($TN^{i,j}$, $FP^{i,j}$, $TP^{i,j}, FN^{i,j}$) for the j-th bootstrap iteration, where $TN^{i,j}$ is the number of false strategies correctly identified as false, $FP^{i,j}$ is the number of false strategies incorrectly identified as true, $FN^{i,j}$ is the number of true strategies incorrectly identified as false, and $TP^{i,j}$ is the number of true strategies correctly identified as true. Table C1 illustrates these four testing outcomes.

Table C1. Summary of Testing Outcomes

Decision	Null Hypothesis H_0: high-minus-low return is equal to zero	Alternative Hypothesis H_1: high-minus-low return is not equal to zero
Reject	False discovery (Type I error) $FP^{i,j}$	True discovery $TP^{i,j}$
Accept	True non discovery $TN^{i,j}$	Miss discovery (Type II error) $FN^{i,j}$

With these testing outcomes, we construct three types of error rates. The first one is the realized false discovery rate (RFDR): $RFDR^{i,j} = \frac{FP^{i,j}}{FP^{i,j}+TP^{i,j}}$, which is the proportion of false discoveries ($FP^{i,j}$) divided by all discoveries ($FP^{i,j} + TP^{i,j}$). The second type of error rate is the realized rate of misses (RMISS): $RMISS^{i,j} = \frac{FN^{i,j}}{FN^{i,j}+TN^{i,j}}$, which is the proportion of misses ($FN^{i,j}$) divided by all strategies that are declared insignificant ($FN^{i,j} + TN^{i,j}$). The third type of error rate is the odds ratio (RRATIO): $RRATIO^{i,j} = \frac{FP^{i,j}}{FN^{i,j}}$, which is the realized ratio of false discoveries ($FP^{i,j}$) to misses ($FN^{i,j}$).

Step 4: We repeat Steps 1-3 above *I* times. That is, we bootstrap the data X_0 I times, and each time we adjust it to a new return matrix Y_i, and generate *J* bootstrapped random samples. We account for the sampling uncertainty in ranking the strategies and the uncertainty in generating the realized error rates for each ranking by averaging across both *i* and *j*. We calculate the final bootstrapped error rates as

$$\begin{aligned} TYPE1 &= \frac{1}{I*J}\sum_{i=1}^{I}\sum_{j=1}^{J} RFDR^{i,j}, \\ TYPE2 &= \frac{1}{I*J}\sum_{i=1}^{I}\sum_{j=1}^{J} RMISS^{i,j}, \\ ORATIO &= \frac{1}{I*J}\sum_{i=1}^{I}\sum_{j=1}^{J} RRATIO^{i,j}, \end{aligned} \tag{C1}$$

where $TYPE1$ denotes the Type I error rate, $TYPE2$ denotes the Type II error rate, and $ORATIO$ denotes the odds ratio between false discoveries and misses. Note that our estimated $TYPE1$, $TYPE2$, and $ORATIO$ implicitly depend on the significance threshold. We set *I*=100 and *J*=1000.

C.2.2 Other Multiple-Testing Methods

We also consider other multiple-testing methods, including Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001), and Barras, Scaillet, and Wermers (2010), to control for false discoveries. Suppose *N* denotes the number of univariate strategies. The null hypothesis is H_0: $R_l^{high} - R_l^{low} = 0$, and the alternative hypothesis is H_1: $R_l^{high} - R_l^{low} \neq 0$, for *l*=1, 2,…, *N*. The significance level α is the probability of Type I error (e.g., 5%). The procedures for the aforementioned multiple-testing methods are as follows:

First, we calculate the two-sided *p*-values based on the *t*-statistics of the high-minus-low return series. Second, we rank the original *p*-values from low to high, such that $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(l)} \leq \cdots \leq p_{(N)}$, for index *l*=1, 2,…, *N*.

Third, we calculate the adjusted *p*-values (denoted by *adj-p*). BH, BHY, and BSW provide different methods to calculate the adjusted *p*-values. For BH, the adjusted *p*-value is $adj\_p = \frac{l*\alpha}{N}$. For BHY, the adjusted *p*-value is $adj\_p = \frac{l*\alpha}{N*c(N)}$, where $c(N) = \sum_{l=1}^{N}\frac{1}{l}$. For BSW, the adjusted *p*-value is $adj\_p = \frac{l*\alpha}{N_0}$, where N_0 $(\pi_0 * N)$ is the number of *p*-values expected to satisfy the null hypothesis. We estimate $\hat{\pi}_0$ as follows. We know that the vast majority of *p*-values (assuming uniform distribution between (0, 1)) larger than a sufficiently high $\lambda \in (0,1)$ are likely to come from the null hypothesis. After selecting λ, we expect a proportion $\frac{\sum_{l=1}^{N}\{p_{(l)}>\lambda\}}{N}$ of *p*-values exceeding λ, where $\sum_{l=1}^{N}\{p_{(l)} > \lambda\}$ is the number of strategies with *p*-values exceeding λ. Extrapolating the (λ, 1) interval over the entire (0, 1) interval, we have

$$\hat{\pi}_0(\lambda) = \frac{\hat{N}_0}{N} = \frac{\sum_{l=1}^{N}\{p_{(l)} > \lambda\}}{N} * \frac{1}{1-\lambda}. \quad \text{(C2)}$$

To choose the optimal λ, we apply the bootstrap procedure proposed by Storey (2002). This resampling procedure selects λ in three steps from the data to minimize an estimate of the mean-squared error (MSE) of $\hat{\pi}_0(\lambda)$:

(1) We compute $\hat{\pi}_0(\lambda)$ across a range of λ values ($\lambda = 0.3, 0.35, \ldots, 0.7$).

(2) For each possible λ, we generate 1000 bootstrap replications of $\hat{\pi}_0(\lambda)$ by drawing with replacement from an $N\times 1$ vector of p-values. These replications are denoted by $\hat{\pi}_0^b(\lambda)$ for b=1, 2, …,1000.

(3) We compute the estimated MSE for each possible λ:

$$\widehat{MSE}(\lambda) = \frac{1}{1000}\sum_{b=1}^{1000}\left[\hat{\pi}_0^b(\lambda) - \min_{\lambda}\hat{\pi}_0(\lambda)\right]^2. \quad \text{(C3)}$$

We choose the optimal λ such that $\lambda = \arg\min_{\lambda} \widehat{MSE}(\lambda)$.

Finally, we find the maximum index k such that $p_{(l)} \leq adj\_p$ for $l \leq k$. This means that for $l > k$, $p_{(l)} > adj\_p$. Then $p_{(k)}$ is the new p-value under multiple testing. We convert the new p-value to the two-sided t-statistic, which serves as the t-cutoff under multiple testing. Table C2 reports the t-statistic cutoffs under multiple testing. Panel A presents the cutoffs for different values of λ for the BSW method. Panel B reports the cutoffs for the BH, BHY, and BSW methods.

Table C2. Multiple-Testing *t*-statistic Cutoffs

Panel A. Multiple-testing *t*-statistic cutoffs for different λ in the BSW method

	λ=0.3	λ= 0.6	λ= 0.65	λ= 0.8
t-cutoff	2.69	2.69	2.69	2.69

Panel B. Multiple-testing *t*-statistic cutoffs for different methods (significance level $\alpha = 5\%$)

	BHY	BH	BSW(λ= 0.70)
t-cutoff	4.05	2.87	2.69

D. Additional Results on Univariate Strategies

Table D1. Significant Long-leg Returns under Single Testing

The table reports the numbers of univariate strategies with significant value-weighted (Panel A) and equal-weighted (Panel B) long-leg raw returns, CAPM alphas, CH3 alphas, and CH4 alphas, over the sample period from July 2000 to December 2020. Long-leg portfolios are decile 1 or decile 10, which has the highest raw returns. The statistical significance is based on the Newey-West standard errors with four lags, using the conventional single-testing *t*-statistic cutoff of 1.96.

	Overall	Significance	Trading-based			Accounting-based			
		Rate	Liquidity	Risk	Past Returns	Profitability	Value	Investment	Others
	454		106	52	52	73	44	51	76
Panel A. Value-Weighted Strategies									
Raw return	2	0%	2	0	0	0	0	0	0
CAPM alpha	2	0%	2	0	0	0	0	0	0
CH3 alpha	2	0%	2	0	0	0	0	0	0
CH4 alpha	1	0%	1	0	0	0	0	0	0
Panel B. Equal-Weighted Strategies									
Raw return	20	4%	3	0	0	5	1	0	11
CAPM alpha	20	4%	3	0	0	5	1	0	11
CH3 alpha	15	3%	3	0	0	5	0	0	7
CH4 alpha	13	3%	3	0	0	5	0	0	5

Table D2. Detailed Results on Univariate Strategies

This table reports the average value-weighted and equal-weighted high-minus-low raw returns, CH4 alphas, and their associated *t*-statistics for the 454 univariate strategies. Details on the construction of the 454 strategies are provided in Internet Appendix B.

Panel A. Liquidity

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw	*t*-stat	CH4	*t*-stat	Raw	*t*-stat	CH4	*t*-stat
size: firm size, n=1, 6, 12								
size	-0.38	-0.82	-0.38	-2.90	-0.38	-0.96	-0.30	-1.70
size1	-0.70	-1.41	-0.50	-3.63	-0.62	-1.51	-0.35	-2.37
size6	-0.34	-0.74	-0.32	-2.54	-0.27	-0.70	-0.18	-1.17
size12	-0.34	-0.76	-0.32	-2.91	-0.30	-0.79	-0.21	-1.45
turn: share turnover, k=1, 6, 12, m=0, n=1, 6, 12								
turn1-1	-1.26	-2.97	-0.16	-0.46	-1.70	-5.53	-0.68	-2.82
turn1-6	-0.62	-1.56	0.12	0.39	-0.70	-2.59	-0.07	-0.29
turn1-12	-0.21	-0.59	0.30	1.12	-0.42	-1.74	-0.01	-0.03
turn6-1	-0.81	-1.82	0.03	0.08	-0.91	-2.94	-0.15	-0.55
turn6-6	-0.25	-0.61	0.31	1.03	-0.43	-1.59	0.04	0.17
turn6-12	-0.09	-0.26	0.35	1.36	-0.35	-1.41	0.00	-0.01
turn12-1	-0.37	-0.88	0.30	0.90	-0.67	-2.38	-0.10	-0.42
turn12-6	-0.11	-0.29	0.44	1.54	-0.42	-1.65	-0.04	-0.18
turn12-12	-0.05	-0.15	0.40	1.40	-0.35	-1.36	-0.06	-0.23
vturn: variation of share turnover, k=1, 6, 12, m=0, n=1, 6, 12								
vturn1-1	-1.56	-3.71	-0.69	-1.88	-1.89	-6.81	-1.01	-5.16
vturn1-6	-0.64	-1.64	-0.08	-0.26	-0.70	-2.85	-0.18	-0.86
vturn1-12	-0.26	-0.73	0.15	0.62	-0.43	-1.97	-0.07	-0.37
vturn6-1	-1.09	-2.51	-0.32	-0.91	-1.14	-3.74	-0.44	-1.76
vturn6-6	-0.53	-1.32	0.03	0.10	-0.62	-2.30	-0.12	-0.55
vturn6-12	-0.23	-0.63	0.17	0.73	-0.42	-1.76	-0.10	-0.56
vturn12-1	-0.76	-1.78	-0.03	-0.10	-0.90	-3.18	-0.24	-1.09
vturn12-6	-0.35	-0.90	0.17	0.66	-0.58	-2.26	-0.19	-0.90
vturn12-12	-0.15	-0.43	0.24	0.95	-0.40	-1.59	-0.16	-0.72
cvturn: coefficient of variation of share turnover, k=1, 6, 12, m=0, n=1, 6, 12								
cvturn1-1	-0.72	-1.99	-0.79	-1.87	-1.32	-6.67	-1.24	-5.55
cvturn1-6	-0.28	-1.42	-0.35	-1.59	-0.27	-2.76	-0.32	-3.24
cvturn1-12	-0.16	-0.92	-0.22	-1.30	-0.12	-1.43	-0.18	-1.93
cvturn6-1	-0.76	-2.30	-0.79	-2.06	-0.79	-3.56	-0.87	-3.42
cvturn6-6	-0.58	-2.28	-0.43	-1.42	-0.39	-2.16	-0.44	-2.07
cvturn6-12	-0.31	-1.35	-0.26	-1.12	-0.20	-1.23	-0.29	-1.42
cvturn12-1	-0.88	-2.38	-0.60	-1.51	-0.77	-3.33	-0.64	-2.26
cvturn12-6	-0.45	-1.59	-0.36	-1.14	-0.28	-1.37	-0.31	-1.21
cvturn12-12	-0.25	-0.97	-0.23	-0.85	-0.16	-0.83	-0.26	-1.09
abturn: abnormal turnover, k=12, m=0, n=1, 6, 12								
abturn1	-0.86	-2.32	-0.03	-0.12	-1.49	-5.24	-0.59	-3.73
abturn6	-0.32	-1.37	0.16	0.77	-0.33	-2.07	0.15	1.03
abturn12	-0.09	-0.48	0.12	0.80	-0.20	-1.60	0.07	0.57
dtv: RMB trading volume, k=1, 6, 12, m=0, n=1, 6, 12								
dtv1-1	-0.86	-2.10	-0.18	-1.07	-1.55	-4.80	-0.48	-2.46
dtv1-6	-0.39	-0.99	-0.01	-0.05	-0.78	-2.68	-0.13	-0.67
dtv1-12	-0.34	-0.88	-0.10	-0.73	-0.65	-2.34	-0.20	-1.16
dtv6-1	-0.52	-1.26	0.01	0.05	-0.92	-2.90	-0.17	-0.76
dtv6-6	-0.36	-0.91	-0.06	-0.39	-0.64	-2.10	-0.19	-0.97
dtv6-12	-0.39	-0.99	-0.18	-1.20	-0.65	-2.21	-0.31	-1.67
dtv12-1	-0.54	-1.36	-0.14	-0.85	-0.89	-2.98	-0.38	-1.87
dtv12-6	-0.47	-1.22	-0.22	-1.45	-0.71	-2.40	-0.35	-1.84

dtv12-12	-0.42	-1.10	-0.23	-1.48	-0.61	-2.10	-0.30	-1.60
vdtv: variation of RMB trading volume, k=1, 6, 12, m=0, n=1, 6, 12								
vdtv1-1	-1.17	-2.89	-0.39	-2.22	-1.99	-6.38	-0.84	-4.14
vdtv1-6	-0.47	-1.25	-0.09	-0.60	-0.88	-3.43	-0.24	-1.36
vdtv1-12	-0.37	-1.01	-0.14	-1.04	-0.70	-2.86	-0.25	-1.70
vdtv6-1	-0.73	-1.79	-0.15	-0.88	-1.16	-3.85	-0.38	-1.66
vdtv6-6	-0.43	-1.11	-0.11	-0.72	-0.77	-2.78	-0.30	-1.59
vdtv6-12	-0.42	-1.12	-0.22	-1.52	-0.72	-2.68	-0.38	-2.30
vdtv12-1	-0.80	-2.03	-0.33	-2.11	-1.19	-4.12	-0.54	-2.72
vdtv12-6	-0.62	-1.63	-0.36	-2.42	-0.89	-3.20	-0.47	-2.64
vdtv12-12	-0.51	-1.35	-0.32	-2.02	-0.73	-2.69	-0.40	-2.34
cvdtv: coefficient of variation of RMB trading volume, k=1, 6, 12, m=0, n=1, 6, 12								
cvdtv1-1	-0.60	-1.49	-0.63	-1.46	-1.42	-7.00	-1.28	-5.91
cvdtv1-6	-0.32	-1.50	-0.33	-1.56	-0.36	-3.38	-0.35	-3.38
cvdtv1-12	-0.18	-0.92	-0.21	-1.25	-0.18	-1.94	-0.19	-2.05
cvdtv6-1	-0.89	-2.42	-0.62	-1.57	-0.97	-3.65	-0.73	-2.66
cvdtv6-6	-0.61	-2.24	-0.36	-1.23	-0.47	-2.35	-0.34	-1.56
cvdtv6-12	-0.29	-1.16	-0.23	-1.00	-0.23	-1.26	-0.23	-1.07
cvdtv12-1	-1.08	-2.80	-0.58	-1.47	-0.89	-3.26	-0.54	-1.75
cvdtv12-6	-0.68	-2.21	-0.44	-1.55	-0.41	-1.84	-0.27	-0.97
cvdtv12-12	-0.36	-1.27	-0.28	-1.09	-0.22	-1.06	-0.21	-0.85
Ami: Amihud illiquidity, k=1, 6, 12, m=0, n=1, 6, 12								
Ami1-1	0.48	1.09	0.19	1.31	1.01	2.78	0.47	2.61
Ami1-6	0.27	0.65	0.09	0.71	0.50	1.55	0.18	1.06
Ami1-12	0.23	0.58	0.14	1.17	0.43	1.37	0.19	1.28
Ami6-1	0.17	0.40	0.03	0.21	0.43	1.28	0.20	1.01
Ami6-6	0.21	0.52	0.12	0.92	0.31	0.96	0.16	0.93
Ami6-12	0.17	0.44	0.12	0.92	0.33	1.04	0.17	1.06
Ami12-1	0.12	0.30	0.18	1.08	0.37	1.11	0.29	1.48
Ami12-6	0.17	0.44	0.21	1.39	0.31	0.96	0.20	1.13
Ami12-12	0.10	0.26	0.16	1.07	0.24	0.74	0.08	0.49
Lm: turnover-adjusted number of zero daily trading volume, k=1, 6, 12, m=0, n=1, 6, 12								
Lm1-1	1.22	3.12	0.23	0.60	1.46	5.93	0.62	2.30
Lm1-6	0.54	1.54	-0.16	-0.50	0.70	3.29	0.21	0.85
Lm1-12	0.21	0.64	-0.31	-1.14	0.42	2.23	0.10	0.46
Lm6-1	0.00	-0.01	-0.47	-1.38	0.37	1.91	-0.03	-0.13
Lm6-6	-0.15	-0.68	-0.54	-2.00	0.20	1.31	-0.03	-0.17
Lm6-12	-0.12	-0.60	-0.53	-2.32	0.12	0.95	-0.03	-0.19
Lm12-1	0.06	0.30	-0.13	-0.50	0.23	1.32	0.03	0.16
Lm12-6	0.12	0.63	-0.26	-1.15	0.19	1.32	0.01	0.06
Lm12-12	0.00	0.01	-0.36	-1.49	0.01	0.06	-0.13	-0.80
Liquidity betas, k=60, m=0, n=1, 6, 12								
$\beta^{ret}1$	-0.39	-1.12	0.28	0.63	-0.46	-1.80	-0.09	-0.32
$\beta^{ret}6$	-0.30	-0.89	0.22	0.51	-0.38	-1.62	-0.16	-0.56
$\beta^{ret}12$	-0.28	-0.86	0.24	0.58	-0.36	-1.57	-0.20	-0.74
$\beta^{lcc}1$	-0.12	-0.29	-0.03	-0.20	0.15	0.49	0.10	0.58
$\beta^{lcc}6$	-0.08	-0.20	-0.02	-0.13	0.10	0.32	-0.01	-0.03
$\beta^{lcc}12$	-0.11	-0.28	-0.07	-0.47	0.04	0.12	-0.10	-0.67
$\beta^{lrc}1$	-0.01	-0.03	-0.43	-1.19	0.08	0.36	0.01	0.05
$\beta^{lrc}6$	0.00	-0.01	-0.38	-1.07	0.22	0.98	0.15	0.68
$\beta^{lrc}12$	0.04	0.13	-0.40	-1.18	0.22	1.02	0.14	0.65
$\beta^{lcr}1$	-0.03	-0.09	-0.19	-1.23	-0.27	-0.96	-0.22	-1.31
$\beta^{lcr}6$	-0.06	-0.16	-0.18	-1.22	-0.19	-0.67	-0.12	-0.74
$\beta^{lcr}12$	-0.06	-0.15	-0.16	-1.06	-0.15	-0.52	-0.07	-0.44
$\beta^{net}1$	-0.37	-0.98	-0.24	-1.42	-0.01	-0.03	0.00	-0.02
$\beta^{net}6$	-0.27	-0.72	-0.18	-1.25	-0.05	-0.16	-0.11	-0.78

β^{net}12	-0.26	-0.68	-0.17	-1.17	-0.09	-0.30	-0.17	-1.20
GUBA social media coverage, k=1, m=0, n=1, 6, 12								
post_num1	-1.43	-2.87	-0.64	-1.73	-2.10	-5.77	-1.07	-2.98
post_num6	-1.01	-2.29	-0.45	-1.48	-1.45	-4.99	-0.75	-2.47
post_num12	-0.86	-2.08	-0.48	-1.70	-1.25	-4.61	-0.75	-2.81
read_num1	-1.07	-2.33	-0.24	-0.68	-1.73	-4.70	-0.62	-1.68
read_num6	-0.81	-2.01	-0.26	-0.92	-1.20	-4.31	-0.46	-1.54
read_num12	-0.75	-1.97	-0.34	-1.33	-1.06	-4.16	-0.54	-2.12
com_num1	-1.16	-2.55	-0.33	-0.94	-1.84	-5.50	-0.81	-2.57
com_num6	-0.82	-2.13	-0.30	-1.08	-1.24	-4.58	-0.58	-2.09
com_num12	-0.72	-1.98	-0.35	-1.36	-1.09	-4.41	-0.62	-2.55
wsvi: web search volume index, k=1, m=0, n=1, 6, 12								
wsvi1	-0.56	-0.93	-0.77	-2.48	-1.02	-2.20	-0.69	-2.12
wsvi6	-0.32	-0.57	-0.60	-2.04	-0.60	-1.46	-0.46	-1.67
wsvi12	-0.19	-0.34	-0.52	-1.82	-0.39	-0.97	-0.33	-1.32

Panel B. Risk

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw	*t*-stat	CH4	*t*-stat	Raw	*t*-stat	CH4	*t*-stat
iv: idiosyncratic volatility, annual sort								
iv	-0.57	-1.29	-0.23	-0.72	-0.77	-2.61	-0.49	-1.61
ivc: idiosyncratic volatility per the CAPM, k=1, m=0, n=1, 6, 12								
ivc1	-0.93	-2.41	0.50	1.21	-1.43	-5.66	-0.30	-0.94
ivc6	-0.49	-1.38	0.23	0.61	-0.59	-2.72	0.00	0.01
ivc12	-0.17	-0.50	0.26	0.85	-0.42	-2.12	-0.06	-0.25
ivch3: idiosyncratic volatility per the CH3 factor model, k=1, m=0, n=1, 6, 12								
ivch3-1	-0.98	-2.63	0.40	0.98	-1.74	-7.27	-0.72	-2.03
ivch3-6	-0.55	-1.60	0.17	0.45	-0.65	-3.32	-0.16	-0.60
ivch3-12	-0.22	-0.67	0.21	0.70	-0.47	-2.59	-0.16	-0.71
ivch4: idiosyncratic volatility per the CH4 factor model, k=1, m=0, n=1, 6, 12								
ivch4-1	-0.90	-2.54	0.43	1.38	-1.76	-7.54	-0.73	-2.25
ivch4-6	-0.48	-1.41	0.21	0.61	-0.65	-3.32	-0.17	-0.66
ivch4-12	-0.22	-0.66	0.23	0.81	-0.46	-2.57	-0.16	-0.69
tv: total volatility, k=1, m=0, n=1, 6, 12								
tv1	-0.79	-1.90	0.69	1.54	-0.90	-3.18	0.22	0.68
tv6	-0.56	-1.52	0.30	0.79	-0.60	-2.46	0.09	0.30
tv12	-0.26	-0.77	0.26	0.91	-0.47	-2.16	-0.01	-0.06
isc: idiosyncratic skewness per the CAPM, k=1, m=0, n=1, 6, 12								
isc1	-0.67	-3.15	-0.90	-2.78	-0.74	-4.87	-1.01	-5.29
isc6	-0.22	-1.79	-0.40	-3.06	-0.25	-3.28	-0.36	-4.13
isc12	-0.08	-0.78	-0.10	-0.88	-0.10	-1.60	-0.15	-1.93
isch3: idiosyncratic skewness per the CH3 factor model, k=1, m=0, n=1, 6, 12								
isch3-1	-0.32	-1.32	-0.48	-1.79	-0.35	-2.62	-0.39	-2.87
isch3-6	-0.18	-1.25	-0.27	-2.15	-0.15	-2.09	-0.18	-2.69
isch3-12	-0.04	-0.34	-0.08	-0.77	-0.04	-0.68	-0.07	-1.12
isch4: idiosyncratic skewness per the CH4 factor model, k=1, m=0, n=1, 6, 12								
isch4-1	-0.39	-1.91	-0.59	-2.69	-0.32	-2.32	-0.41	-3.69
isch4-6	-0.11	-0.77	-0.25	-2.29	-0.11	-1.65	-0.15	-2.31
isch4-12	0.00	0.00	-0.04	-0.38	0.00	0.05	-0.02	-0.36
ts: total skewness, k=1, m=0, n=1, 6, 12								
ts1	-0.88	-2.96	-0.52	-1.46	-0.74	-4.48	-0.73	-3.82
ts6	-0.23	-1.51	-0.14	-0.98	-0.25	-2.79	-0.23	-2.39
ts12	-0.15	-1.10	-0.05	-0.45	-0.16	-2.03	-0.12	-1.64
cs: co-skewness, k=1, m=0, n=1, 6, 12								
cs1	-0.38	-1.06	-0.33	-0.72	-0.23	-1.02	-0.12	-0.43
cs6	-0.24	-1.07	-0.23	-0.91	-0.11	-0.89	-0.02	-0.11

cs12	-0.16	-0.83	-0.17	-1.04	-0.11	-1.10	-0.04	-0.38
β^m: market beta using monthly returns, k=60, m=0, n=1, 6, 12								
$\beta^m 1$	-0.46	-1.28	0.40	0.90	-0.56	-1.97	-0.06	-0.21
$\beta^m 6$	-0.44	-1.24	0.16	0.36	-0.49	-1.88	-0.18	-0.62
$\beta^m 12$	-0.36	-1.03	0.20	0.48	-0.45	-1.76	-0.22	-0.78
β: market beta using daily returns, k=12, m=0, n=1, 6, 12								
$\beta 1$	-0.49	-1.13	0.13	0.29	-0.37	-1.24	0.02	0.05
$\beta 6$	-0.34	-0.87	0.16	0.40	-0.41	-1.49	-0.07	-0.25
$\beta 12$	-0.21	-0.55	0.28	0.74	-0.32	-1.25	0.02	0.07
β^-: downside beta, k=12, m=0, n=1, 6, 12								
$\beta^- 1$	-0.29	-0.73	0.00	0.00	-0.13	-0.52	-0.13	-0.47
$\beta^- 6$	-0.02	-0.04	0.15	0.48	-0.16	-0.68	-0.20	-0.92
$\beta^- 12$	0.07	0.19	0.27	0.99	-0.05	-0.24	-0.09	-0.50
β^{FP}: the Frazzini-Pedersen beta, k=12, m=0, n=1, 6, 12								
$\beta^{FP} 1$	-0.63	-1.51	0.05	0.13	-0.54	-2.08	-0.02	-0.06
$\beta^{FP} 6$	-0.34	-0.82	0.04	0.10	-0.45	-1.85	-0.17	-0.73
$\beta^{FP} 12$	-0.21	-0.55	0.16	0.47	-0.35	-1.51	-0.11	-0.48
β^{DM}: the Dimson beta, k=1, m=0, n=1, 6, 12								
$\beta^{DM} 1$	1.04	2.74	1.56	3.41	0.91	3.31	1.19	4.39
$\beta^{DM} 6$	-0.01	-0.06	0.27	1.01	0.05	0.39	0.21	1.18
$\beta^{DM} 12$	-0.06	-0.29	0.13	0.64	-0.02	-0.20	0.10	0.71
tail: tail risk, k=120, m=0, n=1, 6, 12								
tail1	0.29	0.98	-0.01	-0.03	0.32	1.81	0.11	0.61
tail6	0.18	0.67	-0.06	-0.20	0.17	1.08	0.00	0.03
tail12	0.13	0.55	-0.13	-0.46	0.13	0.88	-0.04	-0.22
Firm news, k=1, m=0, n=1, 6, 12								
paper_news1	-0.41	-1.03	-0.23	-1.13	-0.23	-1.11	0.01	0.08
paper_news6	-0.27	-0.81	-0.24	-1.52	-0.13	-0.71	-0.10	-0.80
paper_news12	-0.19	-0.55	-0.16	-1.02	-0.01	-0.04	0.08	0.84
inter_news1	-0.17	-0.53	-0.28	-1.81	0.00	0.00	-0.11	-1.22
inter_news6	-0.15	-0.47	-0.17	-1.03	-0.01	-0.07	0.06	0.63
inter_news12	-0.16	-0.54	-0.27	-1.73	-0.02	-0.12	-0.08	-0.95

Panel C. Past Returns

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw	*t*-stat	CH4	*t*-stat	Raw	*t*-stat	CH4	*t*-stat
Short-term prior k-month cumulative returns, k=11, 9, 6, 3, m=1, n=1, 6, 12								
$R_{t-12,t-2}1$	0.61	1.43	0.84	1.41	0.46	1.26	0.72	1.50
$R_{t-12,t-2}6$	0.42	1.12	0.26	0.46	0.29	0.91	0.20	0.46
$R_{t-12,t-2}12$	0.32	1.00	0.08	0.17	0.11	0.39	-0.06	-0.15
$R_{t-10,t-2}1$	0.35	0.82	1.02	1.87	0.33	1.00	0.61	1.34
$R_{t-10,t-2}6$	0.57	1.60	0.53	1.05	0.38	1.24	0.33	0.78
$R_{t-10,t-2}12$	0.27	0.89	0.12	0.27	0.10	0.40	-0.05	-0.13
$R_{t-7,t-2}1$	0.22	0.52	1.22	2.55	0.31	1.00	0.84	2.15
$R_{t-7,t-2}6$	0.62	1.86	0.87	1.94	0.53	1.92	0.65	1.75
$R_{t-7,t-2}12$	0.32	1.20	0.33	0.89	0.20	0.86	0.13	0.43
$R_{t-4,t-2}1$	0.26	0.72	1.17	2.60	0.08	0.27	0.62	1.83
$R_{t-4,t-2}6$	0.47	1.73	0.95	2.81	0.42	2.07	0.68	2.40
$R_{t-4,t-2}12$	0.39	1.89	0.56	2.16	0.24	1.45	0.28	1.21
Prior one-month return, k=1, m=0, n=1, 6, 12								
$R_{t-1}1$	-0.73	-1.88	-0.14	-0.34	-1.33	-4.46	-0.56	-1.96
$R_{t-1}6$	-0.03	-0.22	0.43	2.40	-0.15	-1.14	0.26	1.64
$R_{t-1}12$	0.23	1.82	0.39	2.42	0.11	1.08	0.25	1.85
Long-term prior k-month cumulative returns, k=24, 48, m=12, n=1, 6, 12								
$R_{t-36,t-13}1$	-0.51	-1.53	-0.62	-1.47	-0.52	-1.93	-0.70	-2.09
$R_{t-36,t-13}6$	-0.34	-1.07	-0.33	-0.93	-0.27	-1.10	-0.33	-1.16

$R_{t-36.t-13}12$	-0.30	-0.96	-0.25	-0.73	-0.27	-1.16	-0.24	-0.90
$R_{t-60.t-13}1$	-0.55	-1.43	-0.54	-1.35	-0.45	-1.48	-0.46	-1.35
$R_{t-60.t-13}6$	-0.53	-1.37	-0.45	-1.10	-0.44	-1.54	-0.32	-1.03
$R_{t-60.t-13}12$	-0.51	-1.37	-0.30	-0.78	-0.48	-1.72	-0.25	-0.89
inr: industry return, k=6, m=0, n=1, 6, 12								
inr1	0.76	2.15	1.55	3.76	0.76	2.15	1.55	3.76
inr6	0.56	1.80	1.11	2.68	0.56	1.80	1.11	2.68
inr12	0.41	1.59	0.62	1.99	0.41	1.59	0.62	1.99
ilr: industry lead-lag effect, k=1, m=0, n=1, 6, 12								
ilr1	0.58	1.94	0.79	2.44	0.58	1.94	0.79	2.44
ilr6	0.16	1.11	0.62	3.50	0.16	1.11	0.62	3.50
ilr12	0.18	1.49	0.41	2.73	0.18	1.49	0.41	2.73
crchg: cumulative return changes, k=12, m=1, n=1, 6, 12								
crchg1	-0.18	-0.52	0.88	2.50	-0.02	-0.08	0.57	2.18
crchg6	0.52	2.69	0.94	4.09	0.46	2.93	0.74	3.72
crchg12	0.18	1.22	0.38	2.29	0.14	1.30	0.25	1.84
Prior k-month residual returns, k=11, 6, m=0, n=1, 6, 12								
$RR_{t-12.t-2}1$	0.55	2.18	0.89	2.41	0.63	3.21	0.65	2.42
$RR_{t-12.t-2}6$	0.23	0.99	0.29	1.02	0.29	1.66	0.29	1.26
$RR_{t-12.t-2}12$	0.15	0.80	0.13	0.59	0.12	0.81	0.06	0.31
$RR_{t-7.t-2}1$	0.38	1.54	0.89	2.93	0.41	2.17	0.65	2.96
$RR_{t-7.t-2}6$	0.37	1.85	0.77	2.93	0.42	2.77	0.59	2.90
$RR_{t-7.t-2}12$	0.16	0.89	0.33	1.67	0.17	1.21	0.22	1.29
52wh: 52-week high, k=12, m=0, n=1, 6, 12								
52wh1	0.58	1.24	0.74	1.49	0.11	0.25	0.41	0.90
52wh6	0.78	2.15	0.78	1.64	0.72	2.32	0.77	1.94
52wh12	0.71	2.31	0.51	1.26	0.61	2.26	0.50	1.44
mdr: maximum daily return, k=1, m=0, n=1, 6, 12								
mdr1	-0.66	-1.71	0.71	1.78	-1.08	-4.12	0.07	0.21
mdr6	-0.28	-0.84	0.62	1.85	-0.49	-2.47	0.21	0.86
mdr12	0.02	0.05	0.57	2.07	-0.32	-1.79	0.10	0.48
pps: share price, k=1, m=0, n=1, 6, 12								
pps1	0.06	0.19	0.65	1.93	-0.40	-1.65	0.06	0.18
pps6	0.17	0.54	0.60	1.74	-0.19	-0.80	0.13	0.41
pps12	0.14	0.47	0.50	1.45	-0.20	-0.86	0.11	0.35
abr: cumulative abnormal returns around earnings announcement dates, n=1, 6, 12								
abr1	0.69	2.58	0.74	2.45	0.75	4.38	0.86	4.45
abr6	0.24	1.16	0.43	1.86	0.43	3.30	0.51	3.26
abr12	0.21	1.24	0.35	2.20	0.30	2.73	0.34	2.94
Ra1: seasonality returns in year *t*-1								
Ra1	0.52	1.67	0.27	0.72	0.51	2.25	0.31	1.26
Rn1: non-seasonality returns in year *t*-1								
Rn1	0.05	0.10	0.80	1.38	-0.20	-0.56	0.40	0.87
Ra25: seasonality returns between year *t*-2 and *t*-5								
Ra25	0.76	2.93	0.89	2.52	0.49	3.13	0.49	2.56
Rn25: seasonality returns between year *t*-2 and *t*-5								
Rn25	-0.72	-1.79	-0.57	-1.19	-0.91	-2.97	-0.76	-2.02

Panel D. Profitability

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw	*t*-stat	CH4	*t*-stat	Raw	*t*-stat	CH4	*t*-stat
roe: return on equity, n=1, 6, 12								
roe1	1.26	3.73	0.34	0.92	1.18	4.47	0.48	1.78
roe6	0.96	3.05	0.31	1.12	0.78	3.09	0.21	0.86
roe12	0.74	2.41	0.26	1.14	0.53	2.25	0.08	0.37
droe: 4-quarter change in return on equity, n=1, 6, 12								

droe1	1.00	4.23	0.62	1.97	1.10	6.81	0.71	4.01
droe6	0.78	4.25	0.35	1.51	0.75	5.81	0.39	2.72
droe12	0.49	3.12	0.20	1.28	0.44	3.97	0.16	1.25
roa: return on assets, n=1, 6, 12								
roa1	1.23	3.62	0.53	1.38	1.24	4.56	0.64	2.07
roa6	1.01	3.25	0.48	1.47	0.86	3.21	0.35	1.15
roa12	0.76	2.51	0.38	1.32	0.61	2.40	0.20	0.69
droa: 4-quarter change in return on assets, n=1, 6, 12								
droa1	0.97	4.56	0.65	2.06	1.09	7.82	0.78	4.54
droa6	0.80	4.72	0.42	1.82	0.76	6.75	0.44	3.26
droa12	0.36	2.46	0.12	0.67	0.46	4.60	0.21	1.76
sue: standard unexpected earnings, n=1, 6, 12								
sue1	0.91	2.96	0.26	0.78	1.13	7.02	0.68	3.79
sue6	0.67	2.79	0.16	0.52	0.72	4.85	0.29	1.81
sue12	0.50	2.35	0.07	0.30	0.49	3.57	0.13	0.93
rs: revenue surprises, n=1, 6, 12								
rs1	0.47	1.78	0.27	1.15	0.89	6.08	0.57	3.55
rs6	0.46	2.13	0.27	1.16	0.62	5.25	0.39	2.73
rs12	0.45	2.28	0.32	1.76	0.45	3.61	0.23	1.74
rna: return on net operating assets, n=1, 6, 12								
rna	0.55	1.89	-0.04	-0.11	0.45	1.79	0.08	0.28
rnaq1	0.84	3.05	0.30	0.74	1.12	4.49	0.52	1.87
rnaq6	0.66	2.38	0.15	0.38	0.79	3.34	0.27	1.07
rnaq12	0.53	2.17	0.09	0.24	0.60	2.67	0.18	0.76
pm: profit margin, n=1, 6, 12								
pm	0.42	1.07	0.04	0.17	0.28	1.10	0.13	0.59
pmq1	0.75	2.08	-0.03	-0.13	0.71	2.92	0.29	1.56
pmq6	0.44	1.31	-0.16	-0.87	0.35	1.53	0.06	0.34
pmq12	0.42	1.30	-0.08	-0.45	0.25	1.12	0.01	0.04
ato: assets turnover, n=1, 6, 12								
ato	0.58	3.07	0.32	1.00	0.45	3.30	0.25	1.42
atoq1	0.52	2.46	0.25	0.75	0.62	4.00	0.25	1.21
atoq6	0.49	2.40	0.23	0.65	0.55	3.44	0.23	1.11
atoq12	0.44	2.39	0.22	0.66	0.54	3.64	0.29	1.50
ct: capital turnover, n=1, 6, 12								
ct	0.17	0.69	0.35	1.23	0.29	1.63	0.20	0.94
ctq1	0.32	1.20	0.46	1.51	0.66	3.58	0.33	1.42
ctq6	0.24	0.96	0.35	1.18	0.55	3.00	0.29	1.29
ctq12	0.21	0.83	0.32	1.11	0.49	2.67	0.28	1.24
gpa: gross profits-to-assets, annual sort								
gpa	0.51	1.39	0.38	1.04	0.44	1.59	0.17	0.60
gpla: gross profits-to-lagged assets, n=1, 6, 12								
gpla	0.55	1.52	0.33	0.92	0.42	1.46	0.12	0.43
gplaq1	1.15	3.15	0.53	1.35	1.22	4.45	0.59	1.95
gplaq6	0.97	2.90	0.43	1.23	0.83	3.11	0.29	0.99
gplaq12	0.74	2.34	0.37	1.21	0.63	2.48	0.21	0.76
ope: operating profits-to-book equity, annual sort								
ope	0.63	1.73	0.12	0.43	0.37	1.45	-0.02	-0.09
ople: operating profits-to-lagged book equity, n=1, 6, 12								
ople	0.57	1.57	0.22	0.79	0.30	1.12	-0.07	-0.35
opleq1	1.20	3.70	0.35	1.07	1.21	4.74	0.54	2.29
opleq6	0.92	3.09	0.28	1.05	0.76	3.09	0.18	0.84
opleq12	0.73	2.59	0.28	1.26	0.52	2.26	0.05	0.27
opa: operating profits-to-assets, annual sort								
opa	0.60	1.59	0.46	1.42	0.41	1.50	0.16	0.59
opla: operating profits-to-lagged assets, n=1, 6, 12								

opla	0.54	1.52	0.38	1.24	0.36	1.26	0.10	0.36
oplaq1	1.09	3.16	0.48	1.28	1.24	4.59	0.64	2.19
oplaq6	0.94	2.96	0.42	1.28	0.85	3.24	0.31	1.10
oplaq12	0.73	2.41	0.35	1.22	0.64	2.56	0.23	0.86
tbi: taxable income-to-book income, n=1, 6, 12								
tbi	0.06	0.19	0.21	0.74	0.00	0.03	-0.01	-0.07
tbiq1	-0.04	-0.15	0.24	0.78	0.14	1.59	0.11	0.93
tbiq6	-0.11	-0.47	0.19	0.79	0.06	0.76	0.03	0.27
tbiq12	-0.11	-0.57	0.09	0.44	-0.01	-0.11	-0.07	-0.77
bl: book leverage, n=1, 6, 12								
bl	-0.19	-0.53	-0.70	-2.00	-0.27	-1.01	-0.48	-1.54
blq1	-0.14	-0.37	-0.61	-1.77	-0.09	-0.35	-0.27	-0.89
blq6	-0.12	-0.32	-0.60	-1.72	-0.10	-0.40	-0.36	-1.23
blq12	-0.14	-0.39	-0.58	-1.64	-0.19	-0.73	-0.43	-1.45
sg: sales growth, n=1, 6, 12								
sg	0.19	0.75	0.35	1.15	0.07	0.46	0.02	0.11
sgq1	0.91	3.65	0.57	1.79	1.01	5.97	0.73	3.61
sgq6	0.62	3.25	0.56	2.37	0.64	4.46	0.44	2.54
sgq12	0.46	2.71	0.48	2.45	0.45	3.58	0.29	1.78
F: Fundamental score, n=1, 6, 12								
F	0.21	0.97	-0.11	-0.54	0.24	1.40	-0.10	-0.60
Fq1	0.68	2.42	-0.33	-1.06	0.88	5.90	0.32	2.13
Fq6	0.44	2.46	-0.24	-1.30	0.64	5.11	0.25	2.00
Fq12	0.48	2.83	-0.07	-0.45	0.50	4.06	0.17	1.38
O: Ohlson’s (1980) O-score, n=1, 6, 12								
O	-0.43	-1.30	-0.35	-1.02	-0.36	-1.57	-0.33	-1.11
Oq1	-0.40	-1.41	-0.36	-1.37	-0.47	-2.30	-0.42	-1.62
Oq6	-0.21	-0.76	-0.19	-0.72	-0.32	-1.60	-0.31	-1.17
Oq12	-0.29	-1.08	-0.29	-1.03	-0.34	-1.69	-0.37	-1.35
Z: Altman’s (1968) Z-score, n=1, 6, 12								
Z	0.06	0.15	0.32	0.58	-0.07	-0.21	0.16	0.36
Zq1	-0.03	-0.07	0.21	0.35	-0.12	-0.36	0.07	0.15
Zq6	-0.02	-0.06	0.19	0.32	-0.10	-0.30	0.12	0.27
Zq12	0.01	0.04	0.25	0.47	-0.05	-0.16	0.17	0.39

Panel E. Value

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw	*t*-stat	CH4	*t*-stat	Raw	*t*-stat	CH4	*t*-stat
bmj: book-to-June-end market equity, annual sort								
bmj	0.39	1.00	0.03	0.06	0.54	1.69	0.21	0.44
bm: book-to-market equity, n=1, 6, 12								
bm	0.43	0.98	0.03	0.05	0.54	1.66	0.12	0.26
bmq1	0.59	1.34	-0.16	-0.25	0.86	2.40	0.19	0.38
bmq6	0.32	0.74	-0.14	-0.21	0.46	1.33	0.04	0.08
bmq12	0.33	0.80	-0.15	-0.24	0.47	1.41	0.07	0.15
dm: debt-to-market equity, n=1, 6, 12								
dm	0.12	0.30	-0.43	-0.93	0.17	0.51	-0.15	-0.34
dmq1	0.06	0.15	-0.60	-1.31	0.51	1.41	0.08	0.17
dmq6	0.01	0.02	-0.57	-1.24	0.27	0.79	-0.10	-0.23
dmq12	0.03	0.08	-0.51	-1.09	0.22	0.66	-0.12	-0.26
am: assets-to-market equity, n=1, 6, 12								
am	0.25	0.59	-0.33	-0.66	0.30	0.85	-0.12	-0.26
amq1	0.19	0.46	-0.74	-1.54	0.71	1.84	0.09	0.17
amq6	0.12	0.28	-0.51	-1.02	0.39	1.06	-0.04	-0.09
amq12	0.19	0.48	-0.37	-0.73	0.38	1.06	-0.01	-0.03
ep: earnings-to-price, n=1, 6, 12								

ep	0.49	1.28	-0.05	-0.19	0.50	1.76	0.00	0.00
epq1	1.23	3.43	0.08	0.38	1.35	5.04	0.39	2.51
epq6	0.79	2.41	-0.10	-0.54	0.81	3.35	0.07	0.48
epq12	0.64	1.93	-0.07	-0.35	0.63	2.63	0.01	0.09
cfp: cash flow-to-price, n=1, 6, 12								
cfp	0.17	0.80	0.03	0.12	0.25	2.18	0.09	0.75
cfpq1	-0.06	-0.28	-0.03	-0.14	0.13	1.35	0.19	1.97
cfpq6	0.09	0.60	0.16	0.98	0.16	2.50	0.18	2.44
cfpq12	0.02	0.17	0.11	0.88	0.07	1.39	0.06	1.09
sr: 5-year sales growth rank, annual sort								
sr	-0.41	-1.24	-0.52	-1.48	-0.21	-1.06	-0.38	-1.61
em: enterprise multiple, n=1, 6, 12								
em	0.14	0.51	-0.04	-0.10	0.01	0.08	0.16	1.32
emq1	-0.06	-0.34	-0.07	-0.29	0.06	0.61	0.10	0.91
emq6	0.03	0.25	0.11	0.75	0.02	0.20	0.06	0.61
emq12	0.01	0.06	-0.02	-0.12	0.00	0.06	0.01	0.06
sp: sales-to-price, n=1, 6, 12								
sp	0.52	1.44	-0.16	-0.33	0.45	1.80	-0.08	-0.26
spq1	0.58	1.65	-0.47	-1.15	0.99	3.84	0.22	0.73
spq6	0.51	1.48	-0.35	-0.83	0.67	2.90	0.07	0.24
spq12	0.47	1.43	-0.25	-0.58	0.63	2.75	0.06	0.21
ocfp: operating cash flow-to-price, n=1, 6, 12								
ocfp	0.30	1.12	-0.10	-0.45	0.49	2.76	0.12	0.73
ocfpq1	0.14	0.61	-0.24	-1.09	0.36	2.60	0.11	0.78
ocfpq6	0.21	1.11	0.02	0.12	0.32	2.62	0.18	1.46
ocfpq12	0.13	0.78	0.04	0.27	0.28	2.46	0.10	0.85
de: debt-to-book equity, annual sort								
de	-0.16	-0.44	-0.71	-2.03	-0.26	-1.00	-0.51	-1.59
ir: intangible return, annual sort								
ir	-0.35	-0.93	-0.44	-0.90	-0.42	-1.45	-0.48	-1.20
ebp: enterprise book-to-price, n=1, 6, 12								
ebp	0.16	0.40	-0.08	-0.13	0.26	0.81	-0.06	-0.13
ebpq1	0.42	1.00	0.12	0.19	0.65	1.71	0.22	0.41
ebpq6	0.17	0.41	0.00	0.01	0.26	0.73	-0.02	-0.04
ebpq12	0.06	0.17	-0.13	-0.22	0.23	0.66	-0.04	-0.07
ndp: net debt-to-price, n=1, 6, 12								
ndp	0.01	0.05	-0.12	-0.35	0.06	0.28	-0.02	-0.09
ndpq1	0.15	0.44	-0.30	-0.63	0.26	0.93	-0.10	-0.29
ndpq6	0.05	0.15	-0.23	-0.53	0.11	0.43	-0.09	-0.25
ndpq12	0.10	0.34	-0.15	-0.36	0.16	0.66	0.00	0.00

Panel F. Investment

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw	*t*-stat	CH4	*t*-stat	Raw	*t*-stat	CH4	*t*-stat
aci: abnormal corporate investment, annual sort								
aci	-0.03	-0.17	-0.02	-0.05	0.19	1.63	0.16	1.00
ag: investment-to-assets, n=1, 6, 12								
ag	-0.03	-0.10	0.23	0.75	0.05	0.27	0.15	0.72
agq1	0.65	2.44	0.64	2.18	0.67	3.33	0.55	2.50
agq6	0.44	1.70	0.54	1.89	0.41	2.08	0.34	1.51
agq12	0.26	1.02	0.42	1.55	0.21	1.19	0.19	0.88
dpia: changes in PPE and inventory-to-assets, annual sort								
dpia	0.06	0.21	0.36	1.34	0.04	0.26	0.14	0.77
noa: net operating assets, annual sort								
noa	-0.36	-1.23	-0.01	-0.05	-0.23	-1.37	-0.05	-0.24
dnoa: changes in net operating assets, annual sort								

dnoa	-0.18	-0.80	-0.15	-0.52	0.01	0.05	0.13	0.70
ig: 1/2/3-year investment growth, annual sort								
ig	0.06	0.34	-0.03	-0.11	0.25	2.22	0.16	1.26
ig2	0.04	0.20	0.26	1.01	0.13	1.12	0.12	0.72
ig3	-0.24	-1.19	0.13	0.48	0.08	0.65	0.17	1.08
nsi: net stock issues, annual sort								
nsi	-0.12	-0.50	0.07	0.24	-0.24	-1.74	-0.08	-0.46
cei: composite equity issuance, annual sort								
cei	0.47	1.46	0.17	0.46	0.48	1.86	0.17	0.56
cdi: composite debt issuance, annual sort								
cdi	-0.11	-0.42	0.31	1.12	-0.08	-0.47	0.24	1.29
ivg: inventory growth, annual sort								
ivg	0.07	0.39	0.36	1.74	0.02	0.14	0.13	0.84
ivchg: inventory changes, annual sort								
ivchg	0.02	0.11	0.09	0.44	-0.07	-0.50	-0.05	-0.33
oa: operating accruals, annual sort								
oa	-0.10	-0.39	0.12	0.45	-0.19	-1.44	-0.06	-0.39
ta: total accruals, annual sort								
ta	0.09	0.45	-0.23	-0.96	0.22	1.59	0.06	0.34
dwc: changes in net non-cash working capital, annual sort								
dwc	-0.02	-0.15	0.04	0.23	-0.07	-0.54	0.09	0.66
dcoa: changes in current operating assets, annual sort								
dcoa	-0.04	-0.18	0.01	0.03	-0.09	-0.56	-0.07	-0.35
dcol: changes in current operating liabilities, annual sort								
dcol	0.23	1.19	0.22	0.86	0.10	0.73	0.11	0.68
dnco: changes in net non-current operating assets, annual sort								
dnco	0.05	0.30	0.21	1.02	0.12	0.86	0.17	1.11
dnca: changes in non-current operating assets, annual sort								
dnca	-0.07	-0.39	0.25	1.12	0.11	0.74	0.22	1.22
dncl: changes in non-current operating liabilities, annual sort								
dncl	-0.07	-0.36	0.06	0.31	-0.06	-0.50	0.04	0.31
dfin: changes in net financial assets, annual sort								
dfin	0.19	0.95	0.06	0.27	0.12	0.86	-0.01	-0.06
dsti: changes in short-term investment, annual sort								
dsti	0.06	0.30	0.02	0.09	-0.02	-0.14	-0.06	-0.52
dlti: changes in long-term investment, annual sort								
dlti	-0.04	-0.14	0.15	0.54	0.08	0.72	0.14	1.03
dfnl: changes in financial liabilities, annual sort								
dfnl	-0.28	-1.47	-0.06	-0.30	-0.12	-0.90	-0.03	-0.18
dbe: changes in common equity, annual sort								
dbe	0.18	0.56	0.17	0.50	0.16	0.76	-0.01	-0.05
dac: discretionary accruals, annual sort, n=12								
dac	-0.26	-1.08	0.08	0.35	-0.14	-1.28	0.08	0.65
poa: percent operating accruals, annual sort								
poa	-0.29	-1.28	-0.13	-0.59	-0.33	-2.42	-0.19	-1.16
pta: percent total accruals, annual sort								
pta	-0.13	-0.58	0.00	0.01	-0.09	-0.84	0.14	1.20
pda: percent discretionary accruals, annual sort								
pda	-0.21	-0.88	0.38	1.35	-0.27	-1.98	0.06	0.41
cag: current asset growth, n=1, 6, 12								
cagq1	0.56	2.36	0.51	1.58	0.58	3.40	0.44	2.56
cagq6	0.35	1.60	0.30	1.14	0.31	1.88	0.18	1.01
cagq12	0.22	1.03	0.19	0.81	0.16	1.03	0.05	0.30
ncag: non-current asset growth, n=1, 6, 12								
ncagq1	0.11	0.53	0.36	1.50	0.37	2.11	0.33	1.63
ncagq6	0.03	0.13	0.28	1.23	0.16	1.01	0.16	0.78

ncagq12	0.00	0.00	0.35	1.74	0.09	0.62	0.18	0.93
cashg: cash growth, n=1, 6, 12								
cashgq1	0.41	1.84	0.25	1.38	0.25	1.93	0.17	1.52
cashgq6	0.27	1.38	0.18	0.97	0.15	1.19	0.04	0.41
cashgq12	0.11	0.61	0.06	0.37	0.02	0.20	-0.09	-0.89
fag: fixed asset growth, n=1, 6, 12								
fagq1	0.25	1.11	0.50	1.78	0.40	2.48	0.37	1.88
fagq6	0.14	0.65	0.47	1.90	0.27	1.83	0.36	2.02
fagq12	0.04	0.22	0.47	2.04	0.22	1.62	0.32	1.81
nccag: non-cash current asset growth, n=1, 6, 12								
nccagq1	0.51	1.94	0.34	1.02	0.57	3.21	0.44	2.08
nccagq6	0.40	1.73	0.29	1.08	0.30	1.83	0.19	1.00
nccagq12	0.26	1.15	0.24	1.03	0.16	1.08	0.13	0.71
oag: other asset growth, n=1, 6, 12								
oagq1	-0.06	-0.41	0.05	0.26	0.18	1.37	0.17	1.06
oagq6	-0.25	-1.80	-0.12	-0.73	-0.05	-0.42	-0.03	-0.20
oagq12	-0.21	-1.54	-0.03	-0.18	-0.09	-0.85	-0.01	-0.10

Panel G. Others

	Value-Weighted Strategies				Equal-Weighted Strategies			
	Raw	*t*-stat	CH4	*t*-stat	Raw	*t*-stat	CH4	*t*-stat
adm: advertising expense-to-market, annual sort								
adm	0.48	2.53	0.04	0.23	0.47	3.55	0.13	0.87
gad: growth in advertising expense, annual sort								
gad	0.07	0.41	0.17	0.76	0.15	1.24	0.23	1.41
rdm: R&D expense-to-market equity, n=1, 6, 12								
rdm	0.43	1.51	-0.04	-0.12	0.51	2.66	0.15	0.65
rdmq1	0.88	2.56	-0.02	-0.04	1.10	4.62	0.54	1.72
rdmq6	0.64	2.05	0.10	0.24	0.73	3.36	0.36	1.22
rdmq12	0.54	1.86	0.03	0.07	0.69	3.38	0.32	1.19
rds: R&D expense-to-sales, n=1, 6, 12								
rds	-0.06	-0.23	-0.08	-0.28	0.07	0.41	0.42	1.90
rdsq1	-0.17	-0.62	-0.23	-0.77	-0.04	-0.22	0.33	1.61
rdsq6	-0.07	-0.26	-0.13	-0.46	-0.04	-0.31	0.29	1.71
rdsq12	-0.14	-0.58	-0.19	-0.74	-0.03	-0.23	0.27	1.53
ol: operating leverage, n=1, 6, 12								
ol	-0.03	-0.13	-0.28	-1.02	0.17	1.19	-0.10	-0.58
olq1	0.18	0.83	-0.10	-0.46	0.39	2.75	0.01	0.07
olq6	0.06	0.30	-0.16	-0.66	0.30	2.29	0.01	0.07
olq12	0.09	0.44	-0.13	-0.55	0.31	2.31	0.01	0.08
hn: hiring rate, annual sort								
hn	-0.17	-0.70	0.18	0.60	0.03	0.18	0.25	1.39
age: firm age, n=1, 6, 12								
age1	0.14	0.50	0.14	0.37	-0.11	-0.48	0.09	0.29
age6	0.05	0.19	0.06	0.20	-0.19	-0.93	0.01	0.02
age12	-0.07	-0.29	-0.09	-0.31	-0.30	-1.45	-0.15	-0.50
dsi: % change in sales minus % change in inventory, annual sort								
dsi	0.10	0.57	-0.06	-0.28	0.11	1.10	0.00	-0.01
dsa: % change in sales minus % change in accounts receivable, annual sort								
dsa	0.14	0.81	0.16	0.83	0.01	0.08	-0.10	-0.77
dgs: % change in gross margin minus % change in sales, annual sort								
dgs	0.05	0.22	0.05	0.26	0.11	0.82	0.07	0.52
dss: % change in sales minus % change in SG&A, annual sort								
dss	-0.03	-0.16	-0.14	-0.51	0.01	0.04	0.03	0.18
etr: effective tax rate, annual sort								
etr	-0.26	-1.23	0.07	0.22	-0.08	-0.77	-0.02	-0.13

lfe: labor force efficiency, annual sort								
lfe	0.04	0.17	0.08	0.30	-0.03	-0.26	-0.23	-1.27
tan: tangibility, n=1, 6, 12								
tan	1.02	3.13	0.63	1.45	0.42	1.63	0.30	0.83
tanq1	0.92	3.26	0.75	1.75	0.33	1.50	0.22	0.76
tanq6	0.97	3.42	0.64	1.56	0.40	1.87	0.26	0.87
tanq12	0.96	3.41	0.63	1.53	0.46	2.09	0.29	0.93
vcf: cash flow volatility, n=1, 6, 12								
vcf1	-0.11	-0.47	-0.24	-0.93	-0.42	-2.15	-0.25	-1.07
vcf6	-0.09	-0.35	-0.22	-0.76	-0.42	-2.10	-0.38	-1.63
vcf12	-0.06	-0.24	-0.20	-0.67	-0.45	-2.31	-0.42	-1.86
cta: cash-to-assets, n=1, 6, 12								
cta1	0.75	2.78	0.86	2.44	0.50	2.32	0.41	1.36
cta6	0.68	2.46	0.89	2.74	0.49	2.37	0.45	1.53
cta12	0.62	2.18	0.81	2.41	0.46	2.28	0.43	1.46
eper: earnings persistence, annual sort								
eper	0.44	1.30	0.42	1.19	0.39	1.79	0.36	1.46
eprd: earnings predictability, annual sort								
eprd	0.49	1.34	0.85	2.56	0.19	0.73	0.11	0.46
esm: earnings smoothness, annual sort								
esm	-0.22	-0.53	-0.20	-0.56	-0.10	-0.43	-0.01	-0.07
evr: value relevance of earnings, annual sort								
evr	0.51	1.97	0.67	2.16	0.30	1.52	0.48	2.34
etl: earnings timeliness, annual sort								
etl	0.09	0.33	0.25	0.68	-0.03	-0.20	-0.09	-0.44
ecs: earnings conservatism, annual sort								
ecs	-0.10	-0.43	-0.06	-0.22	-0.10	-0.65	-0.10	-0.57
ala: asset liquidity, n=1, 6, 12								
ala	0.37	1.35	0.44	1.21	0.17	0.92	0.24	1.04
alaq1	0.97	4.05	0.92	2.34	0.52	3.07	0.47	1.88
alaq6	0.81	3.34	0.65	1.81	0.54	3.18	0.39	1.55
alaq12	0.74	3.12	0.59	1.72	0.49	2.82	0.36	1.41
alm	0.44	1.23	-0.09	-0.20	0.28	1.17	-0.19	-0.67
almq1	0.50	1.28	0.00	-0.01	0.70	2.23	0.37	0.86
almq6	0.54	1.41	0.05	0.10	0.61	2.05	0.23	0.58
almq12	0.52	1.34	0.01	0.01	0.50	1.73	0.08	0.21
tes: tax expense surprises, n=1, 6, 12								
tes1	0.64	2.83	0.55	2.00	0.62	4.35	0.46	2.67
tes6	0.54	3.40	0.35	1.69	0.46	3.69	0.24	1.79
tes12	0.46	3.16	0.34	2.06	0.33	3.14	0.20	1.81
def: changes in analyst earnings forecast, n=1, 6, 12								
def1	0.81	2.11	0.50	1.42	0.71	3.46	0.43	2.18
def6	0.31	1.84	0.21	0.99	0.42	3.69	0.32	2.26
def12	0.33	2.43	0.23	1.93	0.33	3.47	0.21	2.21
re: revisions in analyst earnings forecast, n=1, 6, 12								
re1	-0.21	-0.52	-0.39	-0.84	0.50	1.64	0.30	0.85
re6	-0.05	-0.18	-0.49	-1.36	0.34	1.29	-0.02	-0.07
re12	0.07	0.33	-0.31	-1.39	0.12	0.58	-0.22	-1.05
efp: analyst earnings forecast-to-price, n=1, 6, 12								
efp1	0.84	1.60	-0.87	-1.91	1.06	2.35	-0.13	-0.29
efp6	0.30	0.57	-1.01	-2.26	0.49	1.14	-0.46	-1.11
efp12	0.06	0.12	-0.86	-1.89	0.26	0.63	-0.47	-1.09
ana: analyst coverage, n=1, 6, 12								
ana1	0.27	0.56	0.49	1.31	0.26	0.80	0.56	2.12
ana6	0.21	0.51	0.36	1.20	0.21	0.70	0.39	1.57
ana12	0.22	0.57	0.32	1.17	0.15	0.55	0.32	1.37

dis: dispersion in analyst forecasts, n=1, 6, 12								
dis1	-0.05	-0.14	0.47	1.00	-0.15	-0.84	0.12	0.52
dis6	-0.02	-0.09	0.30	1.10	-0.06	-0.45	0.06	0.46
dis12	-0.10	-0.44	0.17	0.84	-0.10	-0.67	-0.04	-0.30
io: institutional ownership, n=1, 6, 12								
io	0.48	1.65	-0.02	-0.05	0.19	0.72	0.01	0.03
io1	0.76	2.55	0.41	1.34	0.43	1.75	0.13	0.46
io6	0.65	2.58	0.17	0.68	0.24	1.05	0.01	0.06
io12	0.53	2.16	0.08	0.32	0.14	0.60	-0.04	-0.15
soe: SOE indicator, n=1, 6, 12								
soe1	-0.22	-0.94	-0.34	-1.58	-0.05	-0.34	-0.13	-0.73
soe6	-0.18	-0.78	-0.31	-1.43	-0.07	-0.49	-0.13	-0.77
soe12	-0.16	-0.69	-0.25	-1.10	-0.08	-0.50	-0.13	-0.77
margin: margin trading and short selling indicator, n=1, 6, 12								
margin1	0.29	1.85	0.48	2.93	0.01	0.04	-0.04	-0.27
margin6	0.23	1.41	0.40	2.35	-0.07	-0.30	-0.10	-0.65
margin12	0.14	0.95	0.26	1.60	-0.10	-0.46	-0.16	-1.08

E. Anomalies and Aggregate Market-Level Variables

E.1 Details on Aggregate Market-Level Variables

Trading Frictions

Following Watanabe, Xu, Yao, and Yu (2013) and Jacobs (2016), we use idiosyncratic volatility as a proxy for trading frictions. For each stock, we regress its daily stock excess returns in a given month on the CH4 model and calculate idiosyncratic volatility as the standard deviation of the regression residuals. We then measure market-level trading frictions as the cross-sectional average of firm-level idiosyncratic volatility.

Financial Market Development

Following McLean, Pontiff, and Watanabe (2009) and Titman, Wei, and Xie (2013), we measure financial market development as the ratio of total market capitalization at the end of a given month to annual GDP from the prior year.

Accounting Quality

We collect firm-level annual accounting data quality grades (4, 3, 2, and 1 for Excellent, Good, Qualified, and Bad, respectively) from CNRDS. We measure market-level accounting quality as the average value across firms.

Sentiment Index

Following Baker, Wugler, and Yuan (2012), we construct the sentiment index as the first principal component of market turnover, first-month IPO return, the number of IPO firms, and volatility premium. Market turnover (TURN) is the natural log of total market turnover (total RMB volume over a year divided by total market capitalization at the end of the previous year), detrended with a five-year moving average. First-month IPO return (RIPO) is the equal-weighted average of first-month returns of IPOs in a year. Because IPOs in China are subject to daily price limits, we use first-month IPO returns instead of first-day IPO returns as in Baker, Wugler, and Yuan (2012). Number of IPO firms (NIPO) is the natural log of the number of IPOs in a year. Volatility premium (PVOL) is the month-end natural log of the ratio of the value-weighted average market-to-book ratio of high volatility stocks to that of low volatility stocks. High and low volatility stocks are those in the top and bottom three deciles, respectively, based on the variance of monthly returns in the previous year. We standardize the four sentiment proxies to have zero mean and unit standard deviation and then take the first principal component to construct the sentiment index:

$$Sentiment_t = 0.19 * TURN_t + 0.59 * PVOL_t + 0.47 * NIPO_t + 0.63 * RIPO_t, \qquad \text{(E1)}$$

where the first principal component explains about 43% of total variance.

Market Liquidity

Following Chordia, Subrahmanyam, and Tong (2014), we first calculate monthly firm-level share turnover as the total trading volume for a given month divided by the number of free-float A-shares outstanding at the end of the previous month. We then measure market liquidity as the average share turnover across firms.

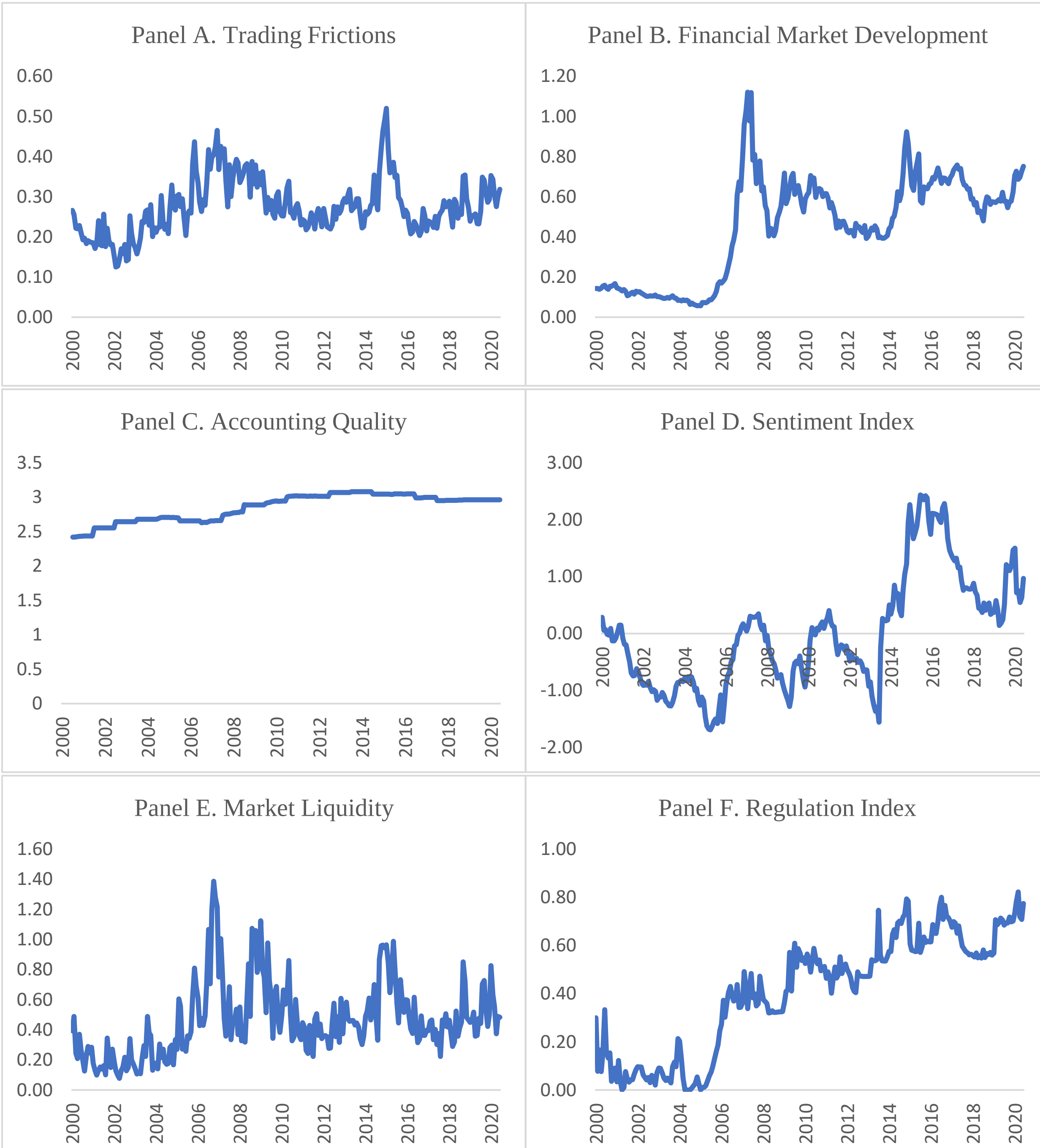
To decompose market turnover into retail investor turnover and large trader turnover, we obtain daily stock order imbalance data from the China Stock Market and Accounting Research (CSMAR) database. Based on Lee and Ready (1991), the dataset classifies total trading volume into buying-initiated and selling-initiated volume for very large trades (greater than RMB1,000,000), large trades (greater than RMB200,000 million but less than RMB1,000,000), medium trades (greater than RMB40,000 RMB but less than RMB200,000), and small trades (less than RMB40,000). The sample period is from January 2003 to December 2020. Following the spirit of Lee and Radhakrishna (2000), Barber, Odean, and Zhu (2009), and Jiang, Liu, Peng, and Wang (2022), which assume that trades with the largest size are more likely to come from large and sophisticated investors, we use the largest size trades as a proxy for institutional investors and trades from the other groups as a proxy for retail investors. For retail investor turnover, we first calculate monthly firm-level retail investor turnover as the total trading volume (sum of buying-initiated and selling-initiated volume) from small, medium, and large trades for a given month divided by the number of free-float A-shares outstanding at the end of the previous month. We then measure retail investor turnover as the average value across firms. For large trader turnover, we first calculate monthly firm-level large trader turnover as the total trading volume (sum of buying-initiated and selling-initiated trading volume) from very large trades for a given month divided by the number of free-float A-shares outstanding at the end of the previous month. We then measure large trader turnover as the average value across firms.

Regulation Index

Given the strong presence of regulations in the Chinese stock market, we construct a regulation index to capture important regulatory interventions on investor composition and trading behavior. The first regulation we consider is the split-share structure reform, which converts non-floating shares to floating shares issued to the general public. For a given month, we measure the percentage of firms that have completed the reform. The second important regulation is the pilot program on margin trading and short selling, designed to deepen and diversify investor participation. Starting from March 2010, regulators expanded the number of stocks included in the pilot program five times. We measure the percentage of firms included in the program in a month. Finally, we use the percentage of firms that are IPOs in a month to measure the government's desires to further stock market development versus concerns about overheating and speculation in the stock market. We scale the three regulation proxies to [0, 1] and then take the simple average to construct the regulation index.

Figure E1. Time-Series Plots of Market-Level Variables

This figure plots the time series of the six aggregate market-level variables, including trading frictions, financial market development, accounting quality, sentiment index, market liquidity, and regulation index, from 2000 to 2020.



E.2 Panel Regressions

Table E1. Anomaly Returns and Different Types of Turnover

This Table reports the regression coefficients and *t*-statistics from regressing the high-minus-low returns of all univariate strategies (and trading-based and accounting-based strategies separately) on the six market-level variables as well as the CH4 factors. The six market-level variables include: trading frictions (FRIC), which are the average firm-level idiosyncratic volatility based on the CH4 model, financial market development (DEV), which is the ratio of total market capitalization to GDP, accounting quality (ACCQ), which is the average firm-level accounting data quality grade from CNRDS, investor sentiment (SENT), which is the first principal component of market turnover, first-month IPO return, the number of IPO firms, and the volatility premium, market liquidity (LIQ), which is the average firm-level monthly retail or large trader turnover, and regulation (REGU), which is the average value of the percentages of listed firms that are recent IPOs, that allow margin trading and short selling, and that have completed the split-share reform. The regressions are estimated with strategy fixed effects and standard errors double-clustered by strategy and month. In Panel A, retail investor turnover (LIQ_Retail) is constructed with trading volume from small, medium, and large trades. In Panel B, large trader turnover (LIQ_Large) is measured with trading volume from very large trades. Panel C includes both retail investor turnover and large trader turnover.

Panel A. Retail Investor Turnover

	Value-Weighted Strategies			Equal-Weighted Strategies		
	I	II	III	IV	V	VI
	All	Trading-based	Accounting-based	All	Trading-based	Accounting-based
FRIC	-0.11	-0.16	-0.07	-0.07	-0.13	-0.03
	(-1.14)	(-0.99)	(-1.28)	(-0.72)	(-0.80)	(-0.25)
DEV	-0.14*	-0.30***	-0.01	-0.21**	-0.34**	-0.10
	(-1.86)	(-2.65)	(-0.13)	(-2.19)	(-2.51)	(-0.77)
ACCQ	-0.22**	-0.28*	-0.16**	-0.22**	-0.21	-0.23**
	(-2.33)	(-1.91)	(-1.98)	(-2.43)	(-1.41)	(-1.96)
SENT	0.00	-0.03	0.02	-0.04	-0.04	-0.04
	(-0.06)	(-0.29)	(0.49)	(-0.74)	(-0.41)	(-0.67)
LIQ_Retail	0.26***	0.37**	0.16**	0.23**	0.41**	0.08
	(2.58)	(2.17)	(2.55)	(2.21)	(2.38)	(0.68)
REGU	0.34***	0.67***	0.04	0.41***	0.65***	0.21
	(2.92)	(3.52)	(0.41)	(3.54)	(3.14)	(1.46)
MKT	-0.03***	-0.07***	0.01	-0.03***	-0.08***	0.01
	(-2.88)	(-4.62)	(1.57)	(-3.16)	(-5.12)	(1.32)
SMB	-0.03	0.06	-0.12***	-0.01	0.08**	-0.08***
	(-1.18)	(1.21)	(-5.28)	(-0.24)	(2.05)	(-3.71)
VMG	0.16***	0.11***	0.21***	0.17***	0.14***	0.20***
	(6.06)	(2.60)	(7.61)	(8.66)	(4.41)	(7.65)
PMO	0.06***	0.13***	0.01	0.04**	0.10***	-0.01
	(3.55)	(4.03)	(0.77)	(2.40)	(3.23)	(-0.37)
Observations	96,186	44,520	51,666	96,186	44,520	51,666
Adj.R^2	2.3%	2.3%	6.4%	4.4%	5.5%	8.7%

Panel B. Large Trader Turnover

	Value-Weighted Strategies			Equal-Weighted Strategies		
	I	II	III	IV	V	VI
	All	Trading-based	Accounting-based	All	Trading-based	Accounting-based
FRIC	-0.01	-0.02	-0.01	-0.03	0.00	-0.06
	(-0.15)	(-0.13)	(-0.18)	(-0.46)	(-0.02)	(-0.69)
DEV	-0.16**	-0.33***	-0.02	-0.24**	-0.38***	-0.12
	(-2.11)	(-2.99)	(-0.31)	(-2.41)	(-2.85)	(-0.88)
ACCQ	-0.24**	-0.31**	-0.18**	-0.27***	-0.27*	-0.28**
	(-2.41)	(-2.01)	(-2.03)	(-2.93)	(-1.78)	(-2.43)
SENT	-0.08	-0.14	-0.03	-0.14***	-0.18*	-0.11
	(-1.38)	(-1.38)	(-0.60)	(-2.68)	(-1.93)	(-1.63)
LIQ_Large	0.14	0.21	0.09	0.21**	0.28*	0.15
	(1.59)	(1.49)	(1.38)	(2.28)	(1.93)	(1.53)
REGU	0.43***	0.81***	0.10	0.51***	0.81***	0.25*
	(3.75)	(4.37)	(0.91)	(4.43)	(4.15)	(1.74)
MKT	-0.02**	-0.06***	0.02**	-0.03***	-0.07***	0.01
	(-2.38)	(-4.75)	(2.04)	(-3.19)	(-5.52)	(1.31)
SMB	-0.03	0.08	-0.11***	0.00	0.10**	-0.08***
	(-0.89)	(1.51)	(-5.03)	(0.08)	(2.53)	(-3.82)
VMG	0.16***	0.11***	0.21***	0.17***	0.15***	0.20***
	(6.06)	(2.65)	(7.57)	(8.59)	(4.54)	(7.63)
PMO	0.06***	0.13***	0.01	0.04***	0.10***	-0.01
	(3.65)	(4.14)	(0.78)	(2.59)	(3.43)	(-0.40)
Observations	96,186	44,520	51,666	96,186	44,520	51,666
Adj.R^2	2.3%	2.2%	6.4%	4.4%	5.4%	8.7%

Panel C. Retail and Large Trader Turnover

	Value-Weighted Strategies			Equal-Weighted Strategies		
	I	II	III	IV	V	VI
	All	Trading-based	Accounting-based	All	Trading-based	Accounting-based
FRIC	-0.13	-0.19	-0.08	-0.12	-0.17	-0.07
	(-1.28)	(-1.11)	(-1.28)	(-1.33)	(-1.13)	(-0.70)
DEV	-0.14*	-0.30***	-0.01	-0.23**	-0.36***	-0.11
	(-1.88)	(-2.72)	(-0.16)	(-2.27)	(-2.59)	(-0.87)
ACCQ	-0.23**	-0.30**	-0.17**	-0.26***	-0.25*	-0.27**
	(-2.39)	(-1.98)	(-1.98)	(-2.86)	(-1.73)	(-2.44)
SENT	-0.02	-0.06	0.01	-0.10*	-0.10	-0.10
	(-0.40)	(-0.55)	(0.13)	(-1.70)	(-0.95)	(-1.43)
LIQ_Retail	0.24**	0.34*	0.15**	0.17	0.35*	0.02
	(2.13)	(1.80)	(2.03)	(1.38)	(1.76)	(0.16)
LIQ_Large	0.05	0.08	0.03	0.14	0.15	0.15
	(0.47)	(0.47)	(0.41)	(1.22)	(0.82)	(1.22)
REGU	0.35***	0.69***	0.05	0.45***	0.69***	0.25*
	(2.94)	(3.57)	(0.48)	(3.64)	(3.34)	(1.71)
MKT	-0.03***	-0.07***	0.01	-0.03***	-0.08***	0.01
	(-2.96)	(-4.73)	(1.55)	(-3.45)	(-5.34)	(1.22)
SMB	-0.03	0.06	-0.12***	-0.01	0.08**	-0.08***
	(-1.18)	(1.21)	(-5.27)	(-0.23)	(2.07)	(-3.77)
VMG	0.16***	0.10***	0.21***	0.17***	0.14***	0.20***
	(6.04)	(2.59)	(7.59)	(8.59)	(4.37)	(7.65)
PMO	0.06***	0.13***	0.01	0.04**	0.10***	-0.01
	(3.59)	(4.07)	(0.77)	(2.52)	(3.30)	(-0.40)
Observations	96,186	44,520	51,666	96,186	44,520	51,666
Adj.R^2	2.3%	2.3%	6.4%	4.4%	5.5%	8.7%

F. Composite Strategies

F.1 Machine Learning Algorithms

Lasso

Following Gu, Kelly, and Xiu (2020), we first estimate a linear regression of stock returns on a set of lagged signals using panel data,

$$R_{i,t} = \theta_0 + \sum_{j=1}^{J} \theta_j signal_{i,j,t-1}^{r} + \epsilon_{i,t}, \quad t \in \tau, \tag{F1}$$

where $R_{i,t}$ is the standardized ranking of stock i's return in month t, $signal_{i,j,t-1}$ is the standardized ranking of signal j for stock i in month t-1, and τ indicates the regression model estimation period. To deal with heavy-tailed observations, we follow Gu, Kelly, and Xiu (2020) and use the Huber robust objective function, which is defined as

$$L_H(\theta) = \frac{1}{N} \sum_{t\in\tau} \sum_{i=1}^{N_t} H(R_{i,t} - \theta_0 - \sum_{j=1}^{J} \theta_j signal_{i,j,t-1}^{r}, \xi), \tag{F2}$$

where

$$H(z,\xi) = \begin{cases} z^2, & \text{if } |z| \leq \xi; \\ 2\xi|z| - \xi^2, & \text{if } |z| > \xi. \end{cases}.$$

The Huber loss, $H(.)$, incorporates hybrid squared losses for relatively small errors and absolute losses for relatively large errors, where the threshold is determined by a tuning parameter, ξ. N is the total number of observations across all firms in the model estimation period, and θ is the parameter vector containing the parameters $\{\theta_0, \theta_1, \ldots, \theta_J\}$. Different from multiple linear regression models which can overfit data in the presence of many signals, Lasso appends a parameter penalty to the original Huber loss function in equation (F2) to avoid overfitting. Thus, the Lasso Huber loss function becomes

$$L_H^{Lasso}(\theta) = L_H(\theta) + \lambda \sum_{j=1}^{J} |\theta_j|, \tag{F3}$$

where $\lambda \sum_{j=1}^{J} |\theta_j|$ is the penalty function based on absolute parameter values, and λ is the nonnegative hyper-parameter (tuning parameter). A large λ indicates a big penalty for the parameters, which would reduce more estimated parameters towards zero.

We follow the literature for model estimation, hyper-parameter selection, and performance evaluation. We divide our sample period into three subperiods: five years for the training period (July 2000 – June 2005), five years for the validation period (July 2005 – June 2010), and 11 years for the out-of-sample testing period (July 2010 – December 2020). We use the training period to estimate the model parameters subject to some pre-specified hyper-parameters. The validation period is used to optimize the hyper-parameters. We iteratively search for the hyper-parameters that minimize mean squared forecast errors for the validation period. The out-of-sample testing period is used to estimate expected returns and evaluate model performance. Since machine learning methods are computationally intensive, we follow Gu, Kelly, and Xiu (2020) and Leippold, Wang, and Zhou (2022) to refit the model every year instead of every month. Each time we refit the model, we extend the training period by one year, maintain the length of the validation period, but roll it forward to include the most recent 12 months of data.

Random Forest

Following Gu, Kelly, and Xiu (2020), we repeatedly draw random samples from the original data with replacements, resulting in samples having the same number of observations as the original data. For each bootstrapped sample, we then construct a regression tree using a random subset of signals. The regression tree consists of a set of decision rules, which categorize stocks into multiple disjoint subgroups ("leaves") that behave similarly to each other (more details below). Using a random subset of signals ensures low correlation among regression trees, further improving variance reduction relative to standard bagging. For each regression tree, we calculate the average return of the leaves that a stock is clustered into and obtain its predicted return. We repeat the above procedure for each of *B* bootstrapped samples of the data. The final expected return estimate, $\hat{R}_{i,t+1}$, is the average of the predicted returns from *B* regression trees.

Regression Tree

Each regression tree contains a set of internal nodes and leaves. At each node, the tree chooses a splitting variable to generate two disjoint branches based on a split point. The regression tree "grows" by sequentially developing branches until it reaches the leaves (terminal nodes). The decision rule of a regression tree is based on clustering observations into one of the "leaves". Mathematically, a regression tree with *K* leaves and depth *L* can be represented as:

$$g\left(x_{i,t};\theta,L,K\right)=\sum_{k=1}^{K}\theta_k 1_{\{x_{i,t}\in C_k(L)\}}, \tag{F4}$$

where $g\left(x_{i,t};\theta,L,K\right)$ is the conditional expectation of stock *i*'s excess return, $x_{i,t}$ is the vector of signals $\{signal_{i,j,t}^{r}\}$, $C_k(L)$ is the *k*-th partition of the data, and the depth *L* is the largest number of nodes in a complete branch (from the top node to any terminal node). Suppose we cluster stock *i* with signals $x_{i,t}$ into the *k*-th leaf of the tree, then the predicted stock return from the regression tree is θ_k, which is the average value of outcomes within the *k*-th leaf. We refer to James, Witten, Hastie, and Tibshirani (2013) for an excellent description on selecting splitting variables and split points for the basic tree model.

Table F1. Hyper-parameters for Lasso and Random Forest

	Lasso	Random Forest
Parameters	$\lambda\in(10^{-4},2*10^{-2})$, Huber loss $\xi\in(0.1,0.2)$.	# depth L= 2-6, # trees B = 100-150, # features M=3-50.

F.2 Additional Results on Composite Strategies

Table F2. Composite Strategies Using Percentile Rankings

This table reports the average high-minus-low raw returns, CH4 alphas, and their associated *t*-statistics of the four composite strategies constructed using percentile rankings instead of standardized rankings, over the period from July 2010 to December 2020. Data from July 2000 to June 2010 are used to establish the initial estimates of the composite signals. Panel A presents the results for the all-but-micro main sample, and Panel B reports the results for all Chinese firms.

Panel A. All-But-Micro Main Sample

	Composite Score		Multiple Regression		Lasso		Random Forest	
	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat
Value-Weighted Strategies								
Raw return	1.68	3.47	2.54	6.00	2.68	5.39	2.87	5.98
CH4 alpha	0.19	0.55	1.12	3.70	1.15	2.47	1.35	4.05
Equal-Weighted Strategies								
Raw return	2.06	6.98	3.11	9.39	3.23	8.03	3.05	8.13
CH4 alpha	0.88	3.35	1.77	7.32	1.79	5.66	1.66	6.05

Panel B. All Chinese Firms

	Composite Score		Multiple Regression		Lasso		Random Forest	
	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat
Value-Weighted Strategies								
Raw return	1.75	3.83	2.73	6.39	2.89	5.32	3.35	6.47
CH4 alpha	0.30	0.91	1.37	4.34	1.47	3.00	1.79	4.91
Equal-Weighted Strategies								
Raw return	2.13	8.50	3.38	10.15	3.55	8.19	3.41	8.64
CH4 alpha	1.02	3.99	2.13	8.31	2.20	6.43	2.06	6.63

References

Abarbanell, Jeffery S., and Brian J. Bushee, 1998, Abnormal returns to a fundamental analysis strategy, *The Accounting Review* 73, 19–45.

Acharya, Viral V., and Lasse Heje Pedersen, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics* 77, 375–410.

Ali, Ashiq, Lee-Seok Hwang, and Mark A. Trombley, 2003, Arbitrage risk and the book-to-market anomaly, *Journal of Financial Economics* 69, 355–373.

Altman, Edward I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance* 23, 589–609.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Ang, Andrew, Joseph Chen, and Yuhang Xing, 2006, Downside risk, *Review of Financial Studies* 19, 1191–1239.

Asness, Clifford, and Andrea Frazzini, 2013, The devil in HML's details, *Journal of Portfolio Management* 39, 49–68.

Baker, Malcolm, Jeffrey Wugler, and Yu Yuan, 2012, Global, local, and contagious investor sentiment, *Journal of Financial Economics* 104, 272–287.

Balakrishnan, Karthik, Eli Bartov, and Lucile Faurel, 2010, Post loss/profit announcement drift, *Journal of Accounting and Economics* 50, 20–41.

Bali, Turan G., Stephen J. Brown, Scott Murray, and Yi Tang, 2017, A lottery-demand-based explanation of the beta anomaly, *Journal of Financial and Quantitative Analysis* 52, 2369–2397.

Ball, Ray, Joseph Gerakos, Juhani Linnaimma, and Valeri Nikolaev, 2015, Deflating profitability, *Journal of Financial Economics* 117, 225–248.

Barber, Brad M., Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets?, *Review of Financial Studies* 22, 151–186.

Barras, Laurent Richard, Olivier Scaillet, and Russ Wermers, 2010, False discoveries in mutual fund performance: Measuring luck in estimated alphas, *The Journal of Finance* 65, 179–216.

Benjamini, Yoav, and Yosef Hochberg, 1995, Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society Series* 57, 289–300.

Benjamini, Yoav, and Daniel Yekutieli, 2001, The control of the false discovery rate in multiple testing under dependency, *Annals of Statistics* 29, 1165–1188.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics* 49, 345–373.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *The Journal of Finance* 51, 1681–1713.

Chordia, Tarun, Avanidhar Subrahmanyam, and V. Ravi Anshuman, 2001, Trading activity and expected stock returns, *Journal of Financial Economics* 59, 3–32.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong, 2014, Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?, *Journal of Accounting and Economics* 58, 41–58.

Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *The Journal of Finance* 63, 1609–1651.

Daniel, Kent D., and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *The Journal of Finance* 61, 1605–1643.

Datar, Vinay T., Narayan Y. Naik, and Robert Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1, 203–219.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 197–226.

Elgers, Pieter T., May H. Lo, and Ray J. Pfeiffer, 2001, Delayed security price adjustments to financial analysts' forecasts of annual earnings, *The Accounting Review* 76, 613–632.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Foster, George, Chris Olsen, and Terry Shevlin, 1984, Earnings releases, anomalies, and the behavior of security returns, *The Accounting Review* 59, 574–603.

Francis, Jennifer, Ryan LaFond, Per M. Olsson, and Katherine Schipper, 2004, Cost of equity and earnings attributes, *The Accounting Review* 79, 967–1010.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2013, The supraview of return predictive signals, *Review of Accounting Studies* 18, 692–730.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Harvey, Campbell R., and Yan Liu, 2020, False (and missed) discoveries in financial economics, *The Journal of Finance* 75, 2503–2553.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *The Journal of Finance* 55, 1263–1295.

Hawkins, Eugene H., Stanley C. Chamberlin, and Wayne E. Daniel, 1984, Earnings expectations and security prices, *Financial Analysts Journal* 40, 24–38.

Heston, Steven L., and Ronnie Sadka, 2008, Seasonality in the cross-section of stock returns, *Journal of Financial Economics* 87, 418–445.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Hribar, Paul, and Daniel W. Collins, 2002, Errors in estimating accruals: Implications for empirical research, *Journal of Accounting Research* 40, 105–134.

Jacobs, Heiko, 2016, Market maturity and mispricing, *Journal of Financial Economics* 122, 270–287.

James, Gareth, Daniela Witten, Trevor Hastie, and Rober Tibshirani, 2013, An introduction to statistical learning, Springer.

Jegadeesh, Narasimhan, and Joshua Livnat, 2006, Revenue surprises and stock returns, *Journal of Accounting and Economics* 41, 147–171.

Jiang, Guohua, Charles M. C. Lee, and Yi Zhang, 2005, Information uncertainty and expected returns, *Review of Accounting Studies* 10, 185–221.

Jiang, Lei, Jinyu Liu, Lin Peng, and Baolian Wang, 2022, Investor attention and asset pricing anomalies, *Review of Finance* 26, 563–593.

Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *Review of Financial Studies* 27, 2841–2871.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *The Journal of Finance* 49, 1541–1578.

Lee, Charles M. C., and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from torq data, *Journal of Financial Markets* 3, 83–111.

Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733–746.

Leippold, Markus, Qian Wang, and Wenyu Zhou, 2022, Machine learning in the Chinese stock market, *Journal of Financial Economics* 145, 64–82.

Liu, Weimin, 2006, A liquidity-augmented capital asset pricing model, *Journal of Financial Economics* 82, 631–671.

Liu, Jianan, Robert F. Stambaugh, and Yu Yuan, 2019, Size and value in China, *Journal of Financial Economics* 134, 48–69.

Loughran, Tim, and Jay W. Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Analysis* 46, 16–29.

McLean, R. David, Jeffrey Pontiff, and Akiko Watanabe, 2009, Share issuance and cross-sectional returns: International evidence, *Journal of Financial Economics* 94, 1–17.

Novy-Marx, Robert, 2011, Operating leverage, *Review of Finance* 15, 103–134.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Ohlson, James A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18, 109–131.

Penman, Stephen H., Scott A. Richardson, and Irem Tuna, 2007, The book-to-price effect in stock returns: Accounting for leverage, *Journal of Accounting Research* 45, 427–467.

Piotroski, Joseph D., 2000, Value investing: The use of historical financial statement information to separate winners from losers, *Journal of Accounting Research* 38, 1–41.

Soliman, Mark T., 2008, The use of dupont analysis by market participants, *The Accounting Review* 83, 823–853.

Storey, John D., 2002, A direct approach to false discovery rates, *Journal of the Royal Statistical Society* 64, 479–498.

Thomas, Jacob, and Frank X. Zhang, 2011, Tax expense momentum, *Journal of Accounting Research* 49, 791–821.

Titman, Sheridan, John K. C. Wei, and Feixue Xie, 2013, Market development and the asset growth effect: International evidence, *Journal of Financial and Quantitative Analysis* 48, 1405–1432.

Watanabe, Akiko, Yan Xu, Tong Yao, and Tong Yu, 2013, The asset growth effect: Insights from international equity markets, *Journal of Financial Economics* 108, 529–563.