ANNALS OF ECONOMICS AND FINANCE **26-2**, 853–890 (2025)

A New Four-factor Model for the Chinese Stock Market*

Heping Xiong, Chao Tang, Jianhui Cao, and Haitao Zhang†

We generate a fundamental signal library containing over 8,000 fundamental signals in the Chinese stock market. Two tests are conducted to identify anomalies within this signal library. Out of these, 142 signals pass both tests. We apply several aggregation techniques to extract information from the signals and find that principal component analysis performs the best. Furthermore, we construct a factor based on the 142 signals and augment the Fama-French three-factor model to create a four-factor model, which performs better than the Fama-French three-factor model, the Carhart four-factor model, the Q4 factor model, the Fama-French five-factor model, and performs at least as well as the Fama-French six-factor model.

Key Words: Data-mining; Anomalies; Cross-Section of Returns; Chinese Stock Market.

JEL Classification Numbers: G11, G12.

1. INTRODUCTION

In the last few decades, scholars have been interested in understanding why different assets yield different returns. In the US market, the empirical asset pricing literature has uncovered a long list of firm characteristics that can predict future stock returns, and this list continues to grow, as noted by Green, Hand, and Zhang (2013); Harvey, Liu, and Zhu (2016); Mclean and Pontiff (2016); Green et al. (2017); and Hou, Xue, and Zhang (2017). As Novy-Marx and Velikov (2016) have pointed out, the incentive to identify these predictors is high for both scholars and practitioners in the financial market.

* The work of Xiong was supported by the National Natural Science Foundation of China,grant 72273099

† Xiong: Department of Finance, Economics and Management School, Wuhan University, Hubei 400072, China. Email: hepingxiong@whu.edu.cn; Tang: Department of Finance, Economics and Management School, Wuhan University, Hubei 400072, China; Cao: Department of Finance, Economics and Management School, Wuhan University, Hubei 400072, China; Zhang: Department of Finance, Economics and Management School, Wuhan University, Hubei 400072, China.

1529-7373/2025

Over the last three decades, China's stock market has grown rapidly, reaching a total market capitalization of more than $7 trillion by the end of 2016 (Hsu et al., 2017), making it the second-largest in the world behind only the US. With increasing expectations to include Chinese stocks in global equity benchmarks, understanding the regulations governing the Chinese stock market is important. However, as a relatively young capital market, it has earned a reputation as a "casino" due to the dominance of retail traders and an uncertain regulatory environment (Carpenter, Lu, & Whitelaw, 2018). Characterized by higher trading costs and short-selling constraints, the Chinese stock market is also less efficient compared to the US market. These microstructural characteristics complicate the understanding of what drives stock returns in China. To date, however, there has been very little published research focused on this area within the Chinese stock market. The purpose of this paper is to provide a systematic analysis of the relationship between fundamental characteristics and cross-sectional stock returns in this emerging, yet already sizable, market.

We focus on fundamental-based variables for clear reasons. First, grounded in the principle that stocks possess intrinsic fair values, fundamental analysis is popular among investors in practice. Additionally, we compare a large number of well-known anomalies, such as the sales-inventories anomaly (Abarbanell & Bushee, 1997) and the return on assets anomaly (Fama & French, 2006), which are based on financial variables. As Hou et al. (2017) report, most published anomalies are relevant to fundamental variables. More importantly, we can generate thousands of fundamental signals by permuting financial statement variables.

Following Yan and Zheng (2017) and Chordia, Goyal, and Saretto (2018), we create a library of fundamental signals by considering various combinations of accounting variables. We start with all 236 accounting variables available in CSMAR and apply filters such as minimum sample size requirements for each variable. Using permutational methods, we then generate a library of 8,288 fundamental signals.

Traditionally, there are two approaches to identifying anomalies. The first is portfolio analysis, which assesses whether the alpha of a hedge portfolio in a time-series regression is significant. One of the most cited studies employing this method is Fama and French (1993). The second approach is the regression-based method, which identifies candidate anomalies by evaluating their predictive power through firm-level regressions; Fama and French (1992) are canonical references. Following Baker, Luo, and Taliaferro (2017), we refer to a candidate signal that passes the first test as a "factor anomaly" and to one that passes the second as a "score anomaly."

In this article, however, we also utilize a bootstrap procedure to set a higher hurdle for factor anomalies. The bootstrap is necessary for three reasons. First, we conduct a Jarque-Bera test (Jarque & Bera, 1980) and find that the proportion of hedge portfolio returns rejecting normality is as high as 79%. Second, the accounting variables are correlated, as are the signals generated from them. Third, there is a multiple comparison problem when evaluating the performance of a large number of signals simultaneously. Following Kosowski et al. (2006) and Fama and French (2010), we perform the bootstrap under the null hypothesis of zero alpha and compare the cross-sectional distribution of t-statistics with the corresponding bootstrapped distribution. The results suggest that all the factor anomalies can pass the bootstrap test; that is, their superior performance is not driven by sampling variation.

After identifying the factor anomalies and score anomalies, we then explore how to synthesize the information contained in these anomalies to build a single, most powerful predictor. First, we use principal component analysis (PCA) to extract information and assess the predictive power of the first principal component for stock returns. Second, we examine whether out-of-sample forecasts from Fama-Macbeth regressions can predict stock returns. Third, we evaluate whether forecast combination methods can efficiently synthesize the information. Finally, we employ partial least squares (PLS) to aggregate information from a large set of predictors.

To assess the performance of these approaches, we sort stocks into five quintile portfolios based on their expected returns and compute relevant statistics for both the quintile portfolios and the spread portfolios. Overall, we find that all methods except the Fama-Macbeth approach effectively aggregate information from multiple firm characteristics, with the forecast combination approach performing the best. These results suggest that forecast combination is a promising method for estimating expected returns from multiple firm characteristics and for constructing profitable trading strategies based on multiple signals.

Finally, we propose a new factor based on all 142 identified anomalies and augment the Fama-French three-factor model (Fama & French, 1992) to develop a four-factor model. First, we create an anomaly index inspired by Engelberg, McLean, and Pontiff (2018) to condense all 142 anomalies. Specifically, we sort stocks into quintiles and construct long and short anomaly portfolios based on each anomaly variable each month. The anomaly index is calculated as the difference between the number of long and short portfolios that a stock belongs to in each month. Next, we create the new factor by sorting stocks into ten decile portfolios and constructing a spread portfolio based on the anomaly index.

We also evaluate the performance of this four-factor model and find that it performs better than the Fama-French three-factor model (Fama & French, 1992), the Carhart four-factor model (Carhart, 1997), the Q4 factor model (Hou et al., 2014), and the Fama-French five-factor model (Fama & French, 2015), and at least as well as the Fama-French six-factor model (Fama & French, 2017) in terms of explaining the hedge portfolio returns of 8,288 fundamental signals.

This paper builds on several previously published studies in the literature. First, it is inspired by emerging research addressing concerns about data mining in finance. The most similar study is by Yan and Zheng (2017), who create a library of 18,000 fundamental signals and examine the severity of data mining on anomalies using this library in the US stock market. They find that the predictive ability of many top-ranked fundamental signals is better explained by mispricing rather than random chance.

This paper differs from Yan and Zheng (2017) in that we construct a library of fundamental signals based on the Chinese stock market and also address the challenge raised by Cochrane (2011) to synthesize the information contained in a large number of identified anomalies. Another related study is by Chordia et al. (2018), who employ a data mining approach to generate over 2 million signals and evaluate the impact of p-hacking in finance. Their results indicate that p-hacking is a serious problem. Conversely, Kogan and Tian (2015) assess the severity of data mining in asset pricing by comparing the ability of simulated factor models and traditional benchmarks to explain cross-sectional returns, finding that model mining is a significant issue.

This paper also contributes to the growing literature on the meta-analysis of firm characteristics that can predict stock returns. For example, Stambaugh, Yu, and Yuan (2012) show that 11 anomalies are stronger following periods of high investor sentiment. Green et al. (2013) provide a comprehensive review of return-predictive signals, analyzing over 330 such signals. Jacobs (2015) replicates 100 anomalies and suggests that sentiment and limits to arbitrage can partly explain market inefficiencies. Harvey et al. (2016) compile a broad set of 313 factors from numerous studies. McLean and Pontiff (2016) evaluate the out-of-sample predictive power of 97 variables. Green et al. (2017) list 94 characteristics, and Hou et al. (2017) expand the factor zoo to 447 anomalies, further examining p-hacking's impact.

Additionally, this paper adds to the asset pricing literature focused on the Chinese stock market. Chen et al. (2010) examine 18 firm characteristics from the US market for their predictive power in China. Jiang et al. (2011) utilize various economic variables to forecast the Chinese market portfolio. Cheung, Hoguet, and Ng (2015) investigate whether value, size, momentum, dividend yield, and volatility premiums are significant in China's A-share market. Cakici, Chan, and Topyan (2015) analyze the effectiveness of 10 return predictors in China. Hsu et al. (2017) systematically test diverse factors for profitability in China, and Carpenter et al. (2018) explore the link between China's stock prices and firm fundamentals.

Jiang, Tang, and Zhou (2018) construct 75 firm characteristic portfolios, demonstrating their significant forecasting power.

Finally, this paper contributes to the emerging literature on handling high-dimensional data in asset pricing through machine learning techniques. Neely et al. (2014) use PCA to extract information from macroeconomic variables and estimate predictive regressions. Kozak, Nagel, and Santosh (2017) investigate PCA-based factor models, finding they can match the explanatory power of popular factor models. Lewellen (2015) creates an expected return estimate using Fama-Macbeth regressions with 15 firm characteristics. Green et al. (2017) run comprehensive Fama-MacBeth regressions with 94 characteristics to identify independent predictors. Forecast combination techniques, as discussed by Rapach, Strauss, and Zhou (2010), are used to improve return predictions. Light, Maslov, and Rytchkov (2017), along with Jiang et al. (2018) and Rytchkov and Zhong (2018), employ partial least squares (PLS) to aggregate information from multiple predictors.

The rest of the paper is organized as follows: Section 2 describes the data and the signal construction process; Section 3 discusses anomaly identification; Section 4 examines the return predictability of different information aggregation techniques; Section 5 evaluates the performance of the new four-factor model (A4); and Section 6 concludes.

2. DATA AND TRADING STRATEGIES

2.1. Data and Sample

We obtained all the data, including financial indicators and stock returns from the CSMAR (Chinese Stock Market and Accounting Research) database. The sample spans from January 2001 to June 2018. The Chinese stock market is characterized by two types of stocks: A-share and B-share. They are denominated in different currencies. The A-shares are denominated in China's renminbi and can only be traded by local Chinese citizens. B-shares, which can only be traded by foreigners, are usually denominated in Hong Kong dollars for those stocks traded on the Shenzhen Security Exchange (SZSE) and denominated in US dollars for those stocks traded on the Shanghai Security Exchange (SHSE). However, after February 19, 2001, the B-share market also opened to Chinese residents. Following Hu, Pan, and Wang (2018), we mainly focus on the A-share market as it contains more stocks, and the liquidity is much higher. The A-share market includes stocks traded on the Shanghai Main Board and Shenzhen Main Board, SMEM (Shenzhen Small and Medium-sized Enterprise Board), and GEM (Growth Enterprise Market) stocks.

2.2. Signals

To construct the long-short portfolios, we first collect all the financial variables with enough data reported in CSMAR. More specifically, it is required that each of the accounting variables should have non-missing values for at least five years from January 2001 to June 2018. In addition, each accounting variable should contain at least 500 firms with non-missing values on average per year. Following Titman, Wei, and Zhao (2017), several standard sample screening procedures are applied to ensure the data calculation quality. First, we filter out stocks with special treatment (ST) and particular transfer (PT) status in the portfolio formation period because these stocks are distressed and lack market liquidity (Jiang, Qi, & Tang, 2016). Second, to mitigate the influence of outliers, we winsorize all variables at 1% level in each period. After having applied these sample screening procedures, 236 fundamental variables can be used to construct long-short portfolios. We split these 236 fundamental variables into two parts. The first part consists of 21 base variables (Y) such as total assets, total liability, and operating profit. The second part compromises the remaining 215 accounting variables (X). The list of these variables is provided in Appendix A1.

Following Yan and Zheng (2017), we scale each of the 215 fundamental variables (X) by 21 different base variables (Y) like ATA (total assets) and BTP (total profit) to generate financial ratios as financial ratios are much more meaningful. On the other hand, financial ratios are more desirable in a cross-sectional analysis since they can effectively put large companies and small companies on an equal playing field.

Besides ratio (X/Y), we compute the change in ratio (Δ in X/Y) and growth of ratio ($\%\Delta$ in X/Y). Finally, we also compute the yearly growth rate in each accounting variable ($\%$ Δ in X), the difference of the yearly growth rate between each variable and the yearly growth of a base variable ($\%\Delta$ in X - $\%\Delta$ in Y), and the change of each accounting variable scaled by a lagged base variable ($\Delta X/lagY$).

Using permutational arguments, we should obtain a total of 22,790 ($5 \times 215 \times 21 + 215$) signals with the above process. The final number of fundamental signals included in the analysis is 8,288, which is smaller than 22,790 since many of the ratios do not have sufficient samples. The complete list of the configurations is listed in Appendix A2.

2.3. Raw Returns of Hedge Portfolios

In line with Fama and French (1996), we allocate the stocks into equal-weighted deciles at the end of June of each year t based on each of the fundamental signals. The equal-weighted returns of these portfolios are calculated from July of year t to June of year $t+1$. We create a long-short portfolio by going long on stocks falling into the top decile and shorting stocks falling into the bottom decile. Different from Chordia et al. (2018), we force each of the hedge portfolios to be positive to facilitate the model comparison since we can always short the top decile and long the bottom decile to achieve a positive return for those long-short portfolios that obtain a negative average return.

Table 1 reports the descriptive statistics of the return of hedge portfolios from January 2001 to June 2018. Specifically, panel A reports the mean, median, standard deviation, minimum, maximum, number, and ratio of portfolios crossing given thresholds. We also report the corresponding statistics for t-statistics in panel B.

TABLE 1.

Descriptive statistics of hedge portfolios

	Panel A: Average return									
	N	Mean	Median	Std	Min	Max	$\|return\| >$ 0.5%		$\|return\|$ 1.0%	
Category							#	%	#	%
X/Y	1780	0.07	0.05	0.24	−0.48	1.65	71	3.99	8	0.45
Δ in X/Y	1558	0.01	0.01	0.16	−0.57	0.47	12	0.77	0	0.00
$\%\Delta$ in X/Y	1468	0.00	0.00	0.13	−0.51	0.42	1	0.07	0	0.00
$\%\Delta$ in X - $\%\Delta$ in Y	1804	0.05	0.05	0.12	−0.30	0.62	1	0.06	0	0.00
$\Delta X/lagY$	1591	−0.08	−0.06	0.16	−0.69	0.42	32	2.01	0	0.00
$\%\Delta$ in X	87	−0.11	−0.10	0.13	−0.44	0.19	0	0.00	0	0.00
all	8288	0.01	0.01	0.18	−0.69	1.65	117	1.41	8	0.10
	Panel B: Average return t-statistic									
	N	Mean	Median	Std	Min	Max	$\|t\| > 1.96$		$\|t\| > 2.57$	
Category							#	%	#	%
X/Y	1780	0.33	0.31	1.11	−2.83	3.57	143	8.03	33	1.85
Δ in X/Y	1558	0.16	0.09	1.16	−3.06	3.30	167	10.72	38	2.44
$\%\Delta$ in X/Y	1468	0.01	0.01	1.00	−3.99	2.79	71	4.84	17	1.16
$\%\Delta$ in X - $\%\Delta$ in Y	1804	0.38	0.39	0.89	−2.36	3.33	78	4.32	14	0.78
$\Delta X/lagY$	1591	−0.53	−0.46	1.00	−3.64	2.67	159	9.99	59	3.71
$\%\Delta$ in X	87	−0.64	−0.59	0.79	−3.03	1.80	3	3.45	1	1.15
all	8288	0.08	0.06	1.09	−3.99	3.57	621	7.49	162	1.95

Panel A of Table 1 indicates that the mean and median of the cross-sectional average return of the hedge portfolios are approximately zero. The overall standard deviation of hedge portfolio returns at 0.18% shows a large number of hedge portfolios with high absolute returns. For instance, the number of portfolios with returns larger than 0.5% is 117. And when the threshold is 1.0% it is 8. The result is similar to that of the U.S. market (see Chordia et al. (2018)).

Panel B of Table 1 indicates that a great many hedge portfolios are crossing the traditional statistical significance hurdle. Specifically, there are

621 hedge portfolios with absolute t-statistics larger than 1.96 accounting for 7.49% of the total 8288 hedge portfolios. And 162 of them clear a higher hurdle, with an absolute t-statistic greater than 2.57. Although it only accounts for 1.95 % of the whole number of the hedge portfolios, this ratio is two times as large as 1.06% for the U.S. market (see Chordia et al. (2018)).

3. IDENTIFY ANOMALIES

3.1. Portfolio Analysis

To identify the anomalies, we begin by estimating the alpha of each hedge portfolio with the Fama-French five-factor model (Fama & French, 2015). It is modeled as:

$$r_{i,t} = \alpha_i + \beta_i MKT_t + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{i,t} \quad (1)$$

TABLE 2.

Descriptive statistics of portfolio abnormal returns

	Panel A: Average abnormal return									
	N	Mean	Median	Std	Min	Max	$\|\alpha\| > 0.5\%$		$\|\alpha\| > 1.0\%$	
Category							#	%	#	%
X/Y	1780	0.05	0.06	0.17	−0.62	0.64	9	0.51	0	0
Δ in X/Y	1558	0.03	0.03	0.15	−0.47	0.56	5	0.32	0	0
$\%\Delta$ in X/Y	1468	0.03	0.02	0.12	−0.46	0.47	0	0.00	0	0
$\%\Delta$ in X - $\%\Delta$ in Y	1804	0.00	0.00	0.12	−0.45	0.43	0	0.00	0	0
$\Delta X/lagY$	1591	−0.01	0.00	0.15	−0.60	0.58	11	0.69	0	0
$\%\Delta$ in X	87	0.06	0.07	0.11	−0.19	0.37	0	0.00	0	0
all	8288	0.02	0.02	0.15	−0.62	0.64	25	0.30	0	0
	Panel B: Average t-statistic of abnormal return									
	N	Mean	Median	Std	Min	Max	$\|t\| > 1.96$		$\|t\| > 2.57$	
Category							#	%	#	%
X/Y	1780	0.30	0.32	1.03	−3.23	4.51	113	6.35	30	1.69
Δ in X/Y	1558	0.21	0.22	1.04	−3.45	3.33	104	6.68	18	1.16
$\%\Delta$ in X/Y	1468	0.18	0.19	0.92	−3.45	3.10	54	3.68	9	0.61
$\%\Delta$ in X - $\%\Delta$ in Y	1804	0.00	0.01	0.96	−3.38	3.13	76	4.21	14	0.78
$\Delta X/lagY$	1591	−0.09	−0.04	1.01	−3.56	2.64	102	6.41	29	1.82
$\%\Delta$ in X	87	0.50	0.50	0.88	−1.38	3.28	6	6.90	1	1.15
all	8288	0.12	0.15	1.00	−3.56	4.51	455	5.49	101	1.22

Table 2 reports the summary statistics of the abnormal returns relative to the Fama and French five-factor model. Panel A of Table 2 shows that there are 25 monthly alphas larger than 0.5%. Panel B indicates that the cross-sectional distribution of t-statistics for alpha has a mean of nearly 0.12 and a standard deviation as high as 1.0 meaning that there are a lot of t-statistics distributed in the tails. As panel B shows, a total of 455 t-statistics cross the threshold of 1.96. And among them, 101 t-statistics surpass the hurdle of 2.57.

3.2. Fama and Macbeth Regression

The main benefit of evaluating the alpha of the hedge portfolio is that portfolio analysis is a nonparametric method that does not require the relationship between the variables being investigated to meet certain assumptions. However, it is difficult for portfolio analysis to control multiple controls. Different from portfolio analysis, the Fama-Macbeth regression (Fama & MacBeth, 1973) can control several variables at the same time. Another drawback with portfolio analysis, as mentioned by Chordia et al. (2018), is that the alpha of the hedge portfolio only reflects the efficacy of 40% of the sample. With the Fama-Macbeth regression, we can evaluate a signal's predictability in the cross-section of stocks with the entire sample. Particularly, we run the following regression to evaluate the cross-predictability of the signal.

$$R_{i,t} = \gamma_{0,t} + \gamma_{1,t} X_{i,t-1} + \lambda_{2,t} Z_{i,t-1} + \varepsilon_{i,t} \tag{2}$$

where $R_{i,t}$ is the return of stock i at time t and $Z_{i,t-1}$ represents the control variables of stock i at time $t-1$. Z includes controls for the commonly used variables known to predict cross-sectional stock returns, size, relative valuations, profitability, and short-horizon past performance, measured here using the natural logarithm of the firm's market capitalizations, the natural logarithm of the book-to-market ratio, gross profitability calculated as that of Fama and French (2015), and stocks' return of months spanning from month $t-11$ to $t-1$.

Table 3 presents the descriptive statistics of t-statistics from Fama-Macbeth regressions. The table reveals that more than 10% of the absolute t-statistics are larger than 1.96 and around 3.6% are larger than 2.57.

3.3. Bootstrap

We also employ a bootstrap procedure to evaluate the performance of the fundamental signals. As a nonparametric approach, bootstrap is a proper inference for many reasons. First, long-short returns are not often normally distributed. Following Jarque and Bera (1980), we run a Jarque-Bera test on all the 8,288 fundamental signals in the untabulated analysis and find that for more than 79% of them, normality is rejected. Second, accounting variables are strongly correlated. The bootstrap provides a general approach for dealing with unknown time-series dependencies that are due to serial correlation in the residuals from performance regressions.

TABLE 3.

Descriptive statistics of Fama-Macbeth regression t-statistics

	N	Mean	Median	Std	Min	Max	$\lvert t\rvert > 1.96$		$\lvert t\rvert > 2.57$	
Category							#	%	#	%
X/Y	1780	0.25	0.19	1.18	−3.34	4.27	203	11.40	55	3.09
Δ in X/Y	1558	−0.05	−0.04	1.09	−3.44	2.54	110	7.06	23	1.48
$\%\Delta$ in X/Y	1468	−0.42	−0.42	1.13	−3.95	3.18	147	10.01	46	3.13
$\%\Delta$ in X - $\%\Delta$ in Y	1804	−0.02	−0.03	1.29	−3.41	4.22	223	12.36	79	4.38
$\Delta X/lagY$	1591	0.34	0.30	1.26	−3.87	3.77	202	12.70	90	5.66
$\%\Delta$ in X	87	0.81	0.66	1.23	−1.70	4.71	13	14.94	6	6.9
all	8288	0.04	0.02	1.23	−3.95	4.71	898	10.83	299	3.61

Third, it involves a multiple comparison problem as mentioned by Harvey et al. (2016) when we simultaneously evaluate the performance of a large number of signals.

Based on the work of Kosowski et al. (2006), Fama and French (2010), Harvey and Liu (2016), and Yan and Zheng (2017), we conduct the bootstrap with the following steps. First, we regress the long-short return of signal i on the Fama-French five-factor model and save the estimated regression coefficients, alpha, as well as the time series of residuals. Second, we draw the saved residuals with replacement to generate a "fake" time series of residuals. In this step, we follow Fama and French (2010) to draw the residual and the corresponding benchmark factors at that point in time when I sample a particular period. Third, we generate a simulated time series of long-short returns for each of the signals with the "fake" residual and benchmark factors while imposing an assumption of zero alpha. Fourth, we re-estimate the benchmark model with the simulated time series of long-short returns and factors and save the estimated alpha and the corresponding t-statistics. Finally, we calculate alpha and t-statistics for various percentiles. Repeating steps 2-4 10,000 times, we can get the distribution for various percentiles of alphas and t-statistics.

Since we are only interested in whether the 455 long-short portfolios that pass the first test are driven by data mining, we obtain the t-statistic of the alpha for each of the 455 long-short portfolios that pass the first test under the null hypothesis of zero alpha[1]. Following Fama and French (2010), Yan and Zheng (2017), and Chordia et al. (2018), we compare the t-statistics from the real sample with the corresponding t-statistics from the simulations for each of the 455 factor anomalies.

[1]We also compare the t-statistics and p-values of the real sample with that of simulations for the selected percentiles as documented in Appendix A3.

Table 4 documents the number and ratio of 455 factor anomalies that pass the bootstrap hurdle and the Fama-Macbeth hurdle. A factor anomaly can be said to pass the bootstrap hurdle if the bootstrapped p-value of the alpha of the corresponding long/short portfolio is less than 5%. The table reveals that all the 455 factor anomalies pass the bootstrap hurdle and 142 of them pass the Fama-Macbeth hurdle, accounting for more than 31.2% of the entire group of factor anomalies.

TABLE 4.

Description of the anomalies

	alpha			Bootstrap						Fama Macbeth					
	+	−	all	+		−		all		+		−		all	
Category				#	%	#	%	#	%	#	%	#	%	#	%
X/Y	74	39	113	74	100%	39	100%	113	100%	39	52.70%	5	12.82%	44	38.94%
Δ in X/Y	73	31	104	73	100%	31	100%	104	100%	4	5.48%	12	38.71%	16	15.38%
$\%\Delta$ in X/Y	25	29	54	25	100%	29	100%	54	100%	5	20.00%	17	58.62%	22	40.74%
$\%\Delta$ in X - $\%\Delta$ in Y	39	37	76	39	100%	37	100%	76	100%	18	46.15%	11	29.73%	29	38.16%
$\Delta X/lagY$	24	78	102	24	100%	78	100%	102	100%	16	66.67%	10	12.82%	26	25.49%
$\%\Delta$ in X	6	0	6	6	100%				100%	5	83.33%				83.33%
All	241	214	455	241	100%	214	100%	455	100%	87	36.10%	55	25.70%	142	31.21%

3.4. Anomalies That Survive Hurdles

Table 5 presents summary statistics for those 142 strategies that survive the bootstrap and Fama-Macbeth hurdle. It includes the mean, FF5 alpha, bootstrapped p-value, and t-statistics for the alpha and the FM coefficient.

This table presents the 142 anomalies that survive the bootstrap and Fama-Macbeth hurdle. We say an anomaly passes the bootstrap hurdle if the bootstrapped p-value of the alpha of the corresponding long-short portfolio is less than 5%. An anomaly is said to pass the Fama-Macbeth hurdle if the absolute t-value of $\gamma_{1,t}$ in equation 2 passes 1.96. Specifically, we present the mean, alpha, t-value of alpha, and bootstrapped p-value of the t-statistic and t-value of Fama-Macbeth regression in the table. The mean and α are reported in percentage. We construct the variables as in appendix A2, and the sample period spans from January of 2001 to June of 2018.

4. INFORMATION AGGREGATION

In this section, we consider several techniques to aggregate the information from the 142 anomalies identified earlier. First, we use principal component analysis to extract information and test the predictive ability of the first principal component. Second, we investigate whether the out-of-

TABLE 5.

Anomalies that survive the hurdles

	name	mean	α	$t\alpha$	p	tm
1	%Δ in CCPTAO	0.17	0.37	3.28	99.80%	3.46
2	AOP/BMI	0.22	0.31	3.19	100.00%	2.09
3	%Δ in AEBP - %Δ in ATNA	0.28	0.34	3.13	99.90%	3.93
4	CVTP/ATSE	0.21	0.4	3.08	100.00%	2.33
5	CCROAB/BMI	0.21	0.3	3	99.80%	2.62
6	CCPFGA/BMI	0.17	0.28	2.95	100.00%	2.49
7	ANIA/BMI	0.16	0.31	2.92	99.90%	2.57
8	%Δ in CCPTAO - %Δ in ATNA	0.29	0.31	2.91	100.00%	3.31
9	CPFB/BMI	0.16	0.28	2.83	99.80%	2.04
10	CVTP/ATA	0.34	0.39	2.77	99.60%	2.5
11	CVTP/ATLASS	0.34	0.39	2.76	99.60%	2.5
12	BNE/ASC	0.09	0.34	2.74	99.80%	2.48
13	AEBP/BMI	0.08	0.3	2.71	99.80%	2.34
14	%Δ in AEBP - %Δ in ASC	0.28	0.3	2.68	99.70%	2.95

sample forecasts from Fama-Macbeth regressions including 142 predictors can simultaneously forecast stock returns. Third, we try to use the forecast combination to synthesize the information from the 142 anomalies. Finally, we extract a common factor from the 142 firm characteristics with a partial least squares (PLS) framework and use this common factor to forecast stock returns.

4.1. Principal Component Analysis

A common approach to condense information from multiple characteristics is to use principal component analysis (PCA), which proscribes to use the first principal component of the characteristics to predict future stock returns. To obtain the estimates of expected returns at time t, this approach consists of three steps. First, we compute the coefficients λ_s^a by applying PCA to the demeaned and standardized characteristics $X_{it}^a, i = 1, \ldots, N$, in each month s, $s \leq t$. Second, we average the λ_s^a over a given tperiodτ (the last 12 months, past 24 months, or past 36 months) to get $\overline{\lambda}_t^a = \frac{1}{\tau}\sum_{s=1}^{\tau}\lambda_s^a$. Finally, we compute the predictor as $\hat{\mu}_{it} = \sum_{a=1}^{A}\overline{\lambda}_t^a X_{it}^a$.

To evaluate the performance of the PCA-based approach, we sort stocks into 5 quintile portfolios based on $\hat{\mu}_{it}$ and compute statistics of the quintile portfolios as well as the spread portfolios. Table 6 reports the alphas of each quintile portfolio with respect to CAPM, FF3, FFC, Q4, FF5, and FF6 regressions as well as the alphas of the spread portfolios in the left side. Also reported are the corresponding t-statistics on the right side. Panel A

TABLE 5—*Continued*

	name	mean	α	$t\alpha$	p	tm
15	%Δ in CVTP - %Δ in ATNL	0.14	0.28	2.68	99.80%	2.58
16	$\Delta AAR/LAGACM$	0.41	0.55	2.64	99.60%	2.21
17	Δ in ASB/BMI	0.28	0.27	2.58	99.80%	2.18
18	BAE/ATEATO	0.36	0.32	2.53	99.70%	3.68
19	CVTP/ATCL	0.48	0.5	2.51	99.90%	2.76
20	%Δ in CEBOCA - %Δ in ATLL	0.04	0.33	2.5	99.90%	3.16
21	%Δ in ADTA	-0.1	0.37	2.5	99.40%	2.52
22	CNCFFO/ALD	0.32	0.4	2.48	99.50%	2.02
23	CVTP/ATNA	0.35	0.35	2.48	99.10%	3.48
24	%Δ in CCPFDO - %Δ in ATNA	0.22	0.23	2.47	99.60%	2.17
25	CNIOCA/ATEATO	0.19	0.32	2.45	99.10%	2.23
26	%Δ in AEBP - %Δ in ATNL	0.16	0.27	2.45	98.90%	2.38
27	ANI/BMI	0.11	0.22	2.39	99.10%	2.09
28	CCRFSO/BTOC	0.23	0.35	2.39	99.50%	2.66
29	%Δ in ATP	0.19	0.26	2.36	99.40%	2.7
30	%Δ in AEBP - %Δ in ATCL	0.24	0.29	2.36	98.40%	3.01
31	Δ in BNI/BMI	0.21	0.22	2.36	99.90%	2.08
32	CVTP/ATEATO	0.15	0.31	2.33	99.50%	2.18
33	BNE/ATCL	0.17	0.29	2.32	98.40%	3.33
34	$\Delta ANIP/LAGBNP$	0.18	0.49	2.31	99.60%	1.99
35	$\Delta COCPRT/LAGATCL$	0.17	0.3	2.29	99.90%	2.47
36	%Δ in ATP/ATA	0.23	0.23	2.29	99.50%	2.03
37	COCPRT/BMI	0.16	0.24	2.29	99.10%	2.81
38	BNE/ATNA	0.12	0.25	2.26	99.60%	4.27
39	%Δ in ATP/ATLASS	0.22	0.22	2.26	99.30%	2.01
40	%Δ in ATP/ATEATO	0.2	0.22	2.24	99.10%	2
41	ANP/BMI	0.07	0.22	2.23	99.30%	1.98
42	%Δ in AEBP - %Δ in ATL	0.26	0.26	2.21	99.10%	2.63
43	BAE/ATA	0.54	0.42	2.21	99.10%	2.78
44	%Δ in AEBP - %Δ in BTOR	0.19	0.25	2.21	99.10%	2.72
45	BAE/ATLASS	0.53	0.42	2.2	99.10%	2.78
46	$\Delta ATP/LAGBTOC$	0.12	0.23	2.2	98.80%	2.95
47	$\Delta AAR/LAGATL$	0.19	0.26	2.2	97.90%	2.24
48	$\Delta AAR/LAGATA$	0.22	0.27	2.19	98.50%	2.15
49	$\Delta AAR/LAGATLASS$	0.22	0.27	2.19	98.50%	2.15
50	$\Delta AAR/LAGBOR$	0.21	0.24	2.19	98.50%	2.46
51	$\Delta AAR/LAGBTOR$	0.2	0.24	2.19	98.50%	2.5
52	CNCFFO/ATA	0.22	0.36	2.18	98.90%	2.03
53	CNCFFO/ATLASS	0.22	0.36	2.18	98.90%	2.03
54	%Δ in CCPTAO - %Δ in ATNL	0.15	0.23	2.17	98.90%	2.1
55	BAE/BMI	0.1	0.22	2.17	97.90%	2.28
56	%Δ in AEBP	0.07	0.26	2.17	97.80%	4.71

TABLE 5—*Continued*

	name	mean	α	$t\alpha$	p	tm
57	ATP/BMI	0.06	0.23	2.16	98.70%	2.14
58	$\Delta AAR/LAGBTOC$	0.19	0.24	2.14	97.70%	2.41
59	$\Delta COCPRT/LAGATL$	0.2	0.28	2.13	99.40%	2.59
60	BNE/ATA	0.14	0.27	2.12	98.90%	2.91
61	BAE/ATSE	0.35	0.28	2.12	97.80%	3.97
62	%Δ in AEBP - %Δ in BNPATO	0.29	0.22	2.12	97.20%	2.57
63	BNE/ATLASS	0.14	0.27	2.11	98.90%	2.91
64	Δ in BITE/ATEATO	0.25	0.25	2.11	98.00%	2.3
65	%Δ in CVTP	0.02	0.22	2.11	99.50%	2.12
66	$\Delta AEBP/LAGATNA$	0.2	0.27	2.1	98.20%	2.55
67	CCPTAO/ATNA	0.65	0.4	2.1	99.10%	2.34
68	%Δ in ADTA - %Δ in ATLL	0.03	0.29	2.1	98.80%	3.8
69	%Δ in AEBP - %Δ in BNP	0.3	0.22	2.1	98.40%	2.6
70	CCRFSO/BMI	0.1	0.22	2.09	99.30%	2.17
71	$\Delta AAR/LAGATCL$	0.16	0.25	2.09	97.10%	2.39
72	%Δ in BAE - %Δ in ATNA	0.2	0.26	2.08	99.20%	2.49
73	$\Delta AAR/LAGBOC$	0.19	0.23	2.08	97.70%	2.43
74	%Δ in AEBP - %Δ in ATA	0.29	0.25	2.07	98.00%	2.95
75	%Δ in ATP/ATSE	0.18	0.2	2.07	98.40%	1.99
76	%Δ in AEBP - %Δ in ATLASS	0.29	0.25	2.07	98.10%	2.93
77	ADTA/ATCL	0.34	0.32	2.05	98.70%	3
78	Δ in ANAR/BMI	0.17	0.22	2.04	98.60%	2.04
79	BNE/ATL	0.24	0.27	2.04	97.30%	3.17
80	$\Delta ANIP/LAGBNPATO$	0.11	0.42	2.03	98.70%	2.02
81	%Δ in BNI/BMI	0.16	0.24	2.01	99.30%	2.24
82	CNCFFO/ATCL	0.29	0.35	2.01	98.30%	2.27
83	%Δ in AEBP - %Δ in BOR	0.22	0.23	2.01	98.30%	2.17
84	CNIOCA/ATSE	0.18	0.28	2.01	98.30%	2.27
85	CVTP/ATL	0.57	0.42	2	98.20%	2.32
86	BNI/BMI	0.07	0.21	1.99	98.60%	2.95
87	$\Delta AEBP/LAGASC$	0.08	0.25	1.99	98.50%	2.5
88	%Δ in ASB - %Δ in BTOC	0.01	-0.2	-2	2.20%	-2.2
89	CNCFFF/BTP	-0.2	-0.3	-2	2.00%	-2.6
90	Δ in COCRRT3/ATLASS	-0.3	-0.4	-2	2.70%	-3.4
91	$\Delta BFE/LAGBNPATO$	-0.2	-0.2	-2	1.70%	-2.8
92	CNCFFF/BNPATO	-0.3	-0.3	-2.1	2.90%	-2.5
93	CNCFFF/BNP	-0.2	-0.3	-2.1	1.90%	-2.4
94	$\Delta CCROAB/LAGBMI$	-0.3	-0.2	-2.1	2.50%	-3.1
95	%Δ in ANI/BTOR	-0.2	-0.2	-2.2	1.70%	-2.7
96	Δ in ALPE/BOC	-0.2	-0.3	-2.2	2.40%	-2
97	Δ in ANI/BTOC	-0.2	-0.2	-2.2	1.20%	-2.8
98	%Δ in ANAR - %Δ in ATL	-0.2	-0.3	-2.2	2.60%	-2.8
99	$\Delta CCPFDO/LAGBNP$	-0.2	-0.2	-2.3	1.30%	-2.5

TABLE 5—*Continued*

	name	mean	α	$t\alpha$	p	tm
100	Δ in ANI/BOC	−0.2	−0.3	−2.3	90.00%	−2.9
101	%Δ in ANI/BOR	−0.2	−0.2	−2.3	1.70%	−2.7
102	%Δ in ANAR/ATCL	−0.3	−0.3	−2.3	1.80%	−2.7
103	%Δ in ANAR - %Δ in ATEATO	−0.1	−0.3	−2.3	1.00%	−2.2
104	$\Delta CCPFDO/LAGBTP$	−0.3	−0.2	−2.3	1.40%	−2.8
105	%Δ in ANAR/ATEATO	−0.3	−0.3	−2.3	1.40%	−2.2
106	$\Delta BFE/LAGBOP$	−0.2	−0.3	−2.3	1.00%	−3.9
107	Δ in ANAR/BTOR	−0.3	−0.4	−2.4	1.00%	−2.1
108	%Δ in ANAR - %Δ in ATSE	−0.1	−0.3	−2.4	0.60%	−2.3
109	%Δ in ATCA/BOR	−0.3	−0.3	−2.4	0.70%	−2.1
110	%Δ in ATCA/BTOR	−0.3	−0.3	−2.4	0.60%	−2.1
111	Δ in ATCA/BOC	−0.3	−0.3	−2.5	1.00%	−2
112	%Δ in ANAR/ATNA	−0.4	−0.3	−2.5	0.40%	−2.5
113	$\Delta CCPFDO/LAGBNPATO$	−0.3	−0.3	−2.5	1.10%	−3.1
114	Δ in ATCA/BOR	−0.3	−0.3	−2.5	0.70%	−2.1
115	Δ in ATCA/BTOR	−0.3	−0.3	−2.5	0.60%	−2.1
116	%Δ in ANAR/ATLASS	−0.3	−0.3	−2.5	1.50%	−2.7
117	%Δ in ANAR/ATA	−0.3	−0.3	−2.5	1.60%	−2.7
118	CNCFFF/BOC	−0.2	−0.3	−2.6	0.60%	−2.3
119	%Δ in ANAR/ATL	−0.3	−0.3	−2.6	0.50%	−3
120	%Δ in ANAR/ATSE	−0.3	−0.3	−2.6	1.00%	−2.2
121	%Δ in ANAR - %Δ in BOC	−0.3	−0.3	−2.6	0.50%	−2.2
122	Δ in BAIL/BOR	−0.4	−0.4	−2.7	0.80%	−2.1
123	%Δ in ATCA/BOC	−0.3	−0.3	−2.7	0.40%	−2.1
124	$\Delta BFE/LAGBTP$	−0.3	−0.3	−2.7	0.20%	−3.6
125	%Δ in ANAR - %Δ in ATCL	−0.3	−0.3	−2.8	0.30%	−2.7
126	%Δ in ANAR - %Δ in BOR	−0.3	−0.3	−2.8	0.20%	−2.5
127	$\Delta BFE/LAGBNP$	−0.3	−0.4	−2.8	0.20%	−3.5
128	$\Delta AAP/LAGBTP$	−0.3	−0.3	−2.8	0.00%	−2.2
129	CNCFFF/ATCL	−0.3	−0.4	−2.8	0.30%	−2.2
130	Δ in ATCA/BTOC	−0.4	−0.4	−2.8	0.30%	−2.1
131	%Δ in ANAR - %Δ in ATA	−0.2	−0.4	−2.9	0.10%	−2.5
132	%Δ in ANAR - %Δ in ATLASS	−0.2	−0.4	−2.9	0.10%	−2.5
133	$\Delta AAP/LAGBNP$	−0.3	−0.3	−2.9	0.00%	−2
134	%Δ in ANAR/ASC	−0.4	−0.4	−3	0.10%	−2.6
135	%Δ in ANAR - %Δ in BTOR	−0.3	−0.3	−3	0.00%	−2.4
136	Δ in ANAR/BTOC	−0.4	−0.5	−3.3	0.20%	−2.5
137	%Δ in ANAR/BOC	−0.4	−0.4	−3.4	0.10%	−3.8
138	%Δ in ANAR/BTOR	−0.5	−0.5	−338	0.00%	−4
139	%Δ in ANAR - %Δ in BTOC	−0.3	−0.4	−3.4	0.10%	−2.4
140	%Δ in ANAR/BOR	−0.5	−0.5	−3.4	0.00%	−3.8
141	%Δ in ANAR/BTOC	−0.4	−0.4	−3.5	0.00%	−3.9
142	Δ in ANAR/BOC	−0.4	−0.5	−3.5	0.20%	−2.3

of Table 6 shows that, to some extent, the PCA-based approach provides estimates for expected returns. The alphas of the spread portfolios range from 0.30% to 0.45% and the corresponding t-statistics range from 2.60 to 3.88. For other averaging schemes with $\tau = 24$ months and $\tau = 36$ months, the results remain consistent.

TABLE 6.

Alphas and t-statistics on quintile PCA portfolios

	Alpha						t					
	Low	2	3	4	High	H-L	Low	2	3	4	High	H-L
Panel A: $\tau = 12$ months												
CAPM	1.11	1.28	1.32	1.42	1.48	0.37	1.75	2.05	2.09	2.26	2.37	3.27
FF3	0.50	0.68	0.70	0.80	0.86	0.36	1.64	2.21	2.26	2.66	2.99	3.08
FF5	−0.11	0.07	0.10	0.25	0.34	0.45	−0.84	0.54	0.79	1.82	2.34	3.88
FF6	−0.03	0.12	0.11	0.25	0.33	0.36	−0.24	0.85	0.75	1.66	2.03	3.07
FFC	−0.01	0.10	0.06	0.23	0.29	0.30	−0.07	0.68	0.43	1.51	1.76	2.60
Q4	−0.05	0.06	0.05	0.23	0.29	0.34	−0.40	0.45	0.42	1.68	1.99	3.10
Panel B: $\tau = 24$ months												
CAPM	1.22	1.39	1.41	1.51	1.56	0.34	1.86	2.13	2.14	2.32	2.41	2.95
FF3	0.51	0.68	0.69	0.81	0.85	0.34	1.57	2.07	2.15	2.57	2.85	2.90
FF5	−0.09	0.08	0.12	0.27	0.36	0.44	−0.61	0.59	0.90	1.98	2.45	3.81
FF6	0.01	0.10	0.12	0.25	0.34	0.32	0.08	0.64	0.77	1.59	2.04	2.79
FFC	0.04	0.06	0.09	0.22	0.30	0.26	0.24	0.43	0.56	1.36	1.82	2.36
Q4	−0.02	0.05	0.09	0.24	0.31	0.33	−0.15	0.39	0.65	1.70	2.13	3.03
Panel C: $\tau = 36$ months												
CAPM	1.02	1.15	1.20	1.34	1.39	0.37	1.52	1.73	1.81	2.02	2.12	3.20
FF3	0.51	0.64	0.67	0.82	0.87	0.37	1.51	1.83	1.99	2.44	2.80	3.01
FF5	−0.07	0.05	0.11	0.30	0.40	0.47	−0.48	0.34	0.82	2.10	2.64	3.92
FF6	0.02	0.03	0.08	0.25	0.38	0.35	0.15	0.18	0.50	1.55	2.18	2.91
FFC	0.06	0.00	0.06	0.22	0.35	0.29	0.36	0.02	0.35	1.31	2.01	2.53
Q4	0.01	0.02	0.09	0.26	0.37	0.36	0.04	0.16	0.63	1.77	2.40	3.22

4.2. Fama-Macbeth Regression

An alternative way to aggregate information is to use the out-of-sample forecasts from Fama-Macbeth regressions based on multiple lagged characteristics and slope estimates available in real time (e.g., Haugen and Baker (1996), Lewellen (2015), Light et al. (2017), Green et al. (2017), and Rytchkov and Zhong (2018)). This approach prescribes the following steps to find the expected return $\hat{\mu}_{it}$ on stock i at time t. We first run Fama-Macbeth regression of realized return r_{it} on an intercept and lagged characteristics $X^a_{it-1}, a = 1, \ldots, A$, in each period $i = 2, \ldots, t$. Then we take the average of the estimated intercepts α_i and slopes β^a_i over a given tperiodτ (past 12 months, past 24 months, or past 36 months) to get $\overline{\alpha}_t$ and $\overline{\beta}_t^a = \frac{1}{\tau}\sum_{P=t-\tau}^{t} \beta_p^a,\ a = 0, \ldots, A$. Finally, we compute the predicted return as $\hat{\mu}_{it} = \hat{\alpha}_t + \sum_{a=1}^{A} \overline{\beta}_t^a X_{it}^a$. For each month, we sort the stocks into 5 quintile portfolios based on the predicted returns and rebalance the quintile portfolios monthly.

Table 7 shows that the forecasts from the Fama-Macbeth regression are a poor proxy of future returns. When the averaging window τ takes 12 months, most of the alphas are insignificant with the alphas of the spread portfolios floating around -0.2 and the corresponding t-statistics vary from -0.91 to -2.1. The results remain unchanged if τ takes a value of 24 months or 36 months as shown in panel B and panel C of Table 7. Hence, the Fama-Macbeth regression approach underperforms the PCA approach. There are several possible explanations for the poor performance of the Fama-Macbeth approach as mentioned by Light et al. (2017). First, the predicted returns are imprecise and even nonexistent when there is a long list of characteristics on the right-hand side of the Fama-Macbeth regression. Second, the high cross-sectional correlation between fundamental indicators can generate a multicollinearity problem in the regression.

4.3. Forecast Combination

Another approach to aggregate information and estimate expected returns is forecast combination. In the same manner, as asset diversification improves portfolio performance, forecast combination can produce a better forecast than the best individual model by combining forecasts across models (Timmermann, 2006). One of the most cited studies is that of Rapach and Zhou (2013) who employ this approach to improve equity premium forecasts and find that forecast combination is well-suited for finance. More recently, Han et al. (2018) have applied forecast combinations to estimate high-dimensional linear regressions and find that this method provides informative forecasts of cross-sectional returns in the US stock market.

The forecast combination approach is implemented in three steps: First of all, we run the bivariate predictive regression $r_{t+1} = \alpha_i + \beta_i X_{i,t} + \varepsilon_{i,t+1},\ i = 1, \ldots, K$ in each period $\tau = 2, \ldots, t$. Then we average the estimated alphas and slopes over a given period τ (the last 12 months, past 24 months, or past 36 months) to get $\overline{\alpha}_{i,t} = \frac{1}{\tau}\sum_{P=t-\tau}^{t} \alpha_P$ and $\overline{\beta}_{i,t} = \frac{1}{\tau}\sum_{P=t-\tau}^{t} \beta_P$ for each characteristic i from 1 to K. Next, we calculate the predicted return based on $\mu_{i,t+1} = \overline{\alpha}_{i,t} + \overline{\beta}_{i,t} X_{i,t}$ for each fundamental signals $i = 1, \ldots, K$. Finally, we compute the combination forecast as $\hat{\mu}_{t+1}^{POOL} = \sum_{i=1}^{K} \overline{r}_{i,t+1}$. Similar to the Fama-Macbeth approach, we sort the stocks into 5 quintile portfolios depending on the combination forecast in

TABLE 7.

Alphas and t-statistics on quintile FM portfolios

	Alpha						t					
	Low	2	3	4	High	H-L	Low	2	3	4	High	H-L
Panel A: $\tau = 12$ months												
CAPM	1.26	1.20	1.21	1.24	1.10	−0.15	1.89	1.76	1.78	1.82	1.60	−1.11
FF3	0.76	0.70	0.71	0.74	0.60	−0.15	2.37	2.06	2.02	2.14	1.74	−1.10
FF5	0.22	0.11	0.12	0.19	0.10	−0.13	1.42	0.75	0.88	1.36	0.637	−0.91
FF6	0.26	0.16	0.11	0.16	0.10	−0.16	1.50	0.99	0.68	0.98	0.57	−1.13
FFC	0.30	0.14	0.08	0.12	0.04	−0.26	1.73	0.88	0.48	0.77	0.24	−1.95
Q4	0.30	0.10	0.09	0.15	0.03	−0.27	1.91	0.71	0.65	1.07	0.20	−2.10
Panel B: $\tau = 24$ months												
CAPM	1.12	1.12	1.23	1.26	1.20	0.08	1.61	1.57	1.70	1.75	1.65	0.53
FF3	0.66	0.65	0.75	0.77	0.71	0.05	1.89	1.76	2.02	2.11	1.98	0.34
FF5	0.16	0.13	0.26	0.29	0.30	0.14	0.97	0.82	1.72	2.13	1.95	0.91
FF6	0.19	0.16	0.29	0.24	0.24	0.04	1.07	0.92	1.74	1.56	1.38	0.28
FFC	0.22	0.15	0.27	0.20	0.19	−0.03	1.20	0.83	1.60	1.29	1.08	−0.22
Q4	0.22	0.13	0.25	0.24	0.23	0.01	1.31	0.80	1.62	1.74	1.49	0.06
Panel C: $\tau = 36$ months												
CAPM	1.33	1.46	1.42	1.37	1.29	−0.04	1.79	1.95	1.92	1.82	1.71	−0.36
FF3	0.69	0.82	0.79	0.72	0.64	−0.05	1.83	2.12	2.04	1.85	1.71	−0.45
FF5	0.15	0.25	0.25	0.18	0.19	0.04	0.94	1.51	1.73	1.22	1.20	0.34
FF6	0.21	0.27	0.23	0.17	0.17	−0.04	1.16	1.45	1.37	1.01	0.93	−0.36
FFC	0.21	0.24	0.20	0.15	0.12	−0.09	1.14	1.29	1.21	86.00	0.68	−0.76
Q4	0.18	0.22	0.23	0.16	0.14	−0.04	1.07	1.35	1.56	1.09	0.84	−0.37

each month. We present the results for the forecast combination approach in Table 8.

Table 8 indicates that forecast combination can aggregate information in the anomalies to some extent, especially when we take $\tau = 36$ months. Most of the t statistics for spread portfolios (H-L) are larger than 2 and the table also implies that the longer the τ takes, the more significant the alphas are for the spread portfolios. In other words, the results with $\tau = 12$ months and $\tau = 24$ months are nosier than that of $\tau = 36$ months. Thus, the results corroborate with Rapach and Zhou (2013) who emphasize that a combination of individual forecasts can stabilize individual forecasts, reduce forecasting risk, and improve forecasting performance.

4.4. PLS-based Approach

Finally, we employ another technique denoted as partial least squares (PLS) to extract information from a large set of predictors. Following Light et al. (2017), we implement the PLS method with the following

TABLE 8.

Alphas and t-statistics on quintile FC portfolios

	Alpha						t					
	Low	2	3	4	High	H-L	Low	2	3	4	High	H-L
Panel A: $\tau = 12$ months												
CAPM	1.19	1.31	1.45	1.47	1.63	0.44	1.91	2.10	2.31	2.39	2.65	2.58
FF3	0.57	0.70	0.84	0.85	1.01	0.44	1.89	2.30	2.78	2.92	3.50	2.49
FF5	−0.12	0.01	0.13	0.17	0.39	0.51	−0.78	0.09	1.02	1.40	2.49	2.81
FF6	0.00	0.10	0.18	0.16	0.35	0.36	−0.02	0.67	1.25	1.14	2.02	1.92
FFC	0.09	0.12	0.16	0.10	0.22	0.13	0.58	0.85	1.06	0.74	1.35	0.87
Q4	0.05	0.08	0.11	0.11	0.21	0.17	0.32	0.57	0.89	0.86	1.44	1.17
Panel B: $\tau = 24$ months												
CAPM	1.10	1.28	1.39	1.50	1.59	0.49	1.71	1.99	2.13	2.32	2.49	2.52
FF3	0.44	0.62	0.73	0.84	0.94	0.50	1.43	1.99	2.29	2.65	3.03	2.49
FF5	−0.16	0.04	0.18	0.30	0.50	0.66	−1.02	0.31	1.33	2.30	3.19	3.36
FF6	−0.04	0.08	0.21	0.27	0.39	0.42	0.23	0.52	1.35	1.76	2.21	2.19
FFC	0.03	0.09	0.18	0.22	0.28	0.25	0.18	0.61	1.16	1.47	1.66	1.45
Q4	−0.04	0.08	0.16	0.25	0.34	0.38	−0.25	0.60	1.16	1.89	2.31	2.26
Panel C: $\tau = 36$ months												
CAPM	1.00	1.27	1.34	1.45	1.55	0.56	1.51	1.92	2.02	2.21	2.40	2.82
FF3	0.41	0.69	0.74	0.86	0.96	0.55	1.25	2.09	2.23	2.62	3.04	2.70
FF5	−0.27	0.05	0.15	0.29	0.56	0.83	−1.72	0.40	1.09	2.19	3.57	4.44
FF6	−0.15	0.08	0.14	0.27	0.45	0.60	−0.89	0.53	0.88	1.72	2.55	3.18
FFC	−0.10	0.09	0.12	0.22	0.36	0.46	−0.58	0.57	0.74	1.40	2.13	2.76
Q4	−0.16	0.09	0.13	0.23	0.43	0.59	−1.04	0.63	0.94	1.70	2.83	3.61

three steps to obtain the estimates of expected returns at time t. First, we run a cross-sectional regression of realized returns r_{is} on individual lagged characteristic $X^a_{i,s-1},\ a = 1, \ldots, A$ for time i from 2 to t. Second, we take the average of the obtained slopes λ^a_s over the past τ months (τ = 12, 24 or 36) to obtain $\overline{\lambda}^a_t = \frac{1}{\tau}\sum_{s=t-\tau}^{t}\lambda^a_s$. Third, we regress the characteristics X^a_{it} on $\lambda^a_s,\ a = 1, \ldots, A$ for each stock i. And we use the obtained slopes $\hat{\mu}_{i,t}$ as proxies for expected returns.

To evaluate the performance of PLS empirically, we sort stocks into 5 quintile portfolios based on the expected returns and compute the statistics for quintile portfolios as well as the spread portfolios. The results are presented in Table 9, which shows that most of the alphas are significant and the alpha remains constant for different τ values. Overall, the PLS approach can, to some extent, aggregate information from multiple characteristics. However, it does not produce a better proxy than PCA and the forecast combination-based approach.

TABLE 9.

Alphas and t-statistics on quintile PLS portfolios

	Alpha						t					
	Low	2	3	4	High	H-L	low	2	3	4	high	H-L
Panel A: $\tau = 12$ months												
CAPM	1.11	1.28	1.42	1.50	1.70	0.59	1.76	2.02	2.26	2.40	2.74	2.95
FF3	0.51	0.68	0.83	0.91	1.12	0.61	1.69	2.25	2.77	3.03	3.84	2.98
FF5	−0.20	−0.05	0.11	0.17	0.49	0.69	−1.21	−0.36	0.89	1.33	2.97	3.28
FF6	−0.09	0.03	0.15	0.17	0.42	0.51	−0.50	0.24	1.09	1.16	2.27	2.36
FFC	0.04	0.06	0.14	0.10	0.26	0.22	0.24	0.41	0.94	0.66	1.55	1.37
Q4	0.00	0.01	0.11	0.09	0.29	0.29	0.00	0.11	0.82	0.67	1.88	1.82
Panel B: $\tau = 24$ months												
CAPM	1.01	1.17	1.34	1.49	1.62	0.61	1.60	1.81	2.11	2.31	2.56	2.73
FF3	0.44	0.59	0.77	0.92	1.07	0.63	1.50	1.93	2.55	2.91	3.40	2.78
FF5	−0.24	−0.0970.14	0.28	0.53	0.76	−1.50	−0.69	1.07	2.1373.13	3.45		
FF6	-0.10	−0.03	0.15	0.26	0.42	0.52	−0.60	−0.18	1.02	1.68	2.27	2.37
FFC	0.00	0.00	0.14	0.19	0.27	0.27	−0.01	−0.03	0.92	1.27	1.58	1.49
Q4	−0.08	−0.04	0.14	0.21	0.33	0.40	−0.50	−0.30	1.02	1.57	2.11	2.20
Panel C: $\tau = 36$ months												
CAPM	1.14	1.23	1.44	1.51	1.68	0.55	1.76	1.90	2.20	2.30	2.61	2.42
FF3	0.49	0.58	0.78	0.84	1.04	0.55	1.62	1.86	2.49	2.60	3.22	2.35
FF5	−0.20	−0.09	0.16	0.25	0.54	0.74	−1.24	−0.63	1.19	1.84	3.20	3.34
FF6	−0.07	−0.04	0.15	0.22	0.43	0.50	−0.41	−0.25	0.99	1.45	2.34	2.29
FFC	0.01	−0.01	0.12	0.17	0.29	0.27	0.08	−0.06	0.80	1.09	1.67	1.46
Q4	−0.06	−0.03	0.13	0.19	0.34	0.40	−0.36	−0.19	0.97	1.37	2.21	2.15

5. A NEW FOUR-FACTOR MODEL

Inspired by Stambaugh, Yu, and Yuan (2015); Stambaugh and Yuan (2016); and Engelberg et al. (2018), we create an anomaly index to reflect all of the 142 anomalies that survive adjustment for the five factors of Fama and French (2015) and surpass the hurdle of Fama and MacBeth (1973) and bootstrap. The objective of combining all the anomalies together is to construct a single measure that can mitigate the noise in each anomaly and therefore improve the precision.

Our method to construct an anomaly index is simple. For each anomaly, we sort stocks into quintiles based on the given anomaly variable and construct long and short anomaly portfolios each month. We measure the anomaly index as the difference between the number of long and short portfolios a stock belongs to in each month. For example, an anomaly index equal to 5 means that the stock belongs to 5 more long portfolios than short portfolios in a given month.

Then we create a composite factor with the anomaly index by sorting stocks into 5 decile portfolios and constructing a spread portfolio. More specifically, for each month, we sort stocks by anomaly index, split them into 5 portfolios, and calculate the return of each portfolio. The composite factor is the return spread between the long and short portfolios. We augment the Fama-French three-factor model (Fama & French, 1992) with this factor to form a new four-factor model.

Following Fama and French (2014); Stambaugh and Yuan (2016); and Liu, Stambaugh, and Yuan (2018), we take a left-hand-side (LHS) approach (Fama & French, 2017) to evaluate the four-factor model relative to FF3, FFC, FF5, Q4, and FF6. Numerous papers in empirical finance such as Fama and French (1993), Fama and French (2015), Fama and French (2016), Hou and Loh (2016), and Harvey and Liu (2016) have adopted this method.

Following Hou et al. (2018), we confront the new four-factor model with a large group of testing portfolios to evaluate its pricing ability. More specifically, we run a large-scale empirical horse race by pricing all 8,288 factors mentioned earlier to compare the relative pricing power of different factor models with the new four-factor model.

TABLE 10.

Alphas and t-statistics on quintile RANK portfolios

	Alpha						t					
	Low	2	3	4	High	H-L	Low	2	3	4	High	H-L
CAPM	1.33	1.59	1.63	1.74	1.84	0.51	2.10	2.56	2.66	2.89	3.19	3.55
FF3	0.48	0.74	0.80	0.91	1.02	0.55	1.58	2.58	2.80	3.37	4.10	3.83
FF5	−0.27	0.06	0.15	0.33	0.59	0.86	−2.16	0.50	1.16	2.59	4.18	7.31
FF6	−0.16	0.10	0.20	0.36	0.58	0.74	−1.18	0.74	1.40	2.58	3.72	6.31
FFC	−0.15	0.09	0.19	0.35	0.55	0.70	−1.09	0.68	1.33	2.46	3.52	6.09
Q4	−0.21	0.07	0.16	0.33	0.55	0.76	-1.69	0.59	1.21	2.49	3.80	6.63

5.1. Overall Performance

Table 11 presents the overall performance of the competing factor models in explaining the 8,288 spread portfolios. In general, the new four-factor model is the best performer. The number of significant high-minus-low alphas with respect to $|t| \geq 1.96$ is 424, dropping dramatically from 869 in the FF3. The number of significant high-minus-low alphas for $|t| \geq 2.57$ is 83, which is the lowest among the competing models. The Fama-French five-factor model and six-factor model perform well. However, both underperform the new four-factor model from the perspective of the number of high-minus-low alphas with $|t| \geq 2.56$. The CAPM, FF3, FFC, and Q4 perform poorly. The numbers of significant high-minus-low alphas with

$|t| \geq 1.96$ and $|t| \geq 2.57$ are significantly higher than that of FF5, FF6, and A4

TABLE 11.

Overall performance of factor models in explaining 8,288 hedge portfolios

	$\|t\| \geq 1.96$		$\|t\| \geq 2.57$	
Models	#	%	#	%
CAPM	584	7.05%	118	1.42%
FF3	869	10.49%	282	3.40%
FF5	455	5.49%	101	1.22%
FF6	423	5.10%	102	1.23%
FFC	516	6.23%	134	1.62%
Q4	1238	14.94%	458	5.53%
A4	424	5.00%	83	1.00%

6. CONCLUSION

We construct a library of 8,288 fundamental signals in the Chinese stock market based on all accounting variables with sufficient data from CSMAR. We identify the anomalies by examining the alphas with portfolio analysis as well as regression coefficients with Fama-Macbeth regression from this library. The results indicate that hundreds of signals can pass both the test of portfolio analysis and Fama-Macbeth regression even after accounting for data mining. We also apply several techniques to aggregate the information from these significant signals (anomalies) and find that Forecast Combination is the most efficient one for aggregating the information from all those identified anomalies.

Based on these anomalies, we also construct a new factor and obtain a new four-factor model. The four-factor model performs better than the Fama and French three-factor model (Fama & French, 1992), Carhart four-factor model (Carhart, 1997), Q4 factor model of Hou et al. (2014), and Fama and French five-factor model (Fama & French, 2015) and at least as well as the Fama and French six-factor model (Fama & French, 2017) in terms of the ability to accommodate all the hedge portfolio returns of 8288 fundamental signals.

APPENDIX

Table A1 presents the 215 accounting variables used in the analysis. The sample period spans from January of 2001 to June of 2018. To begin with, we collect all the financial variables with enough amount of data reported in CSMAR. Next, we filter out the variables with no-missing values fewer than five years or fewer than 500 firms on average per year. We also exclude ATLASS (total liabilities and shareholder's equity), ATA (total assets), ATNA (total non-current assets), ACM (construction material), ASC(share capital), ATSE (total shareholders' equity), ATEATO (total equity attributable to owners of the parent company), ATL (total liabilities), ATCL (total current liabilities), ALD (long-term debts), ATLL (total long-term liabilities), ATNL (total non-current liabilities), BTP (total profit), BNP (net profit), BOP (operating profit), BTOR (total operating revenue), BOR (operating revenue), BTOC (total operating costs), BOC (operating costs), BNPATO (net profit attributable to owners of the parent company), and BMI (Minority Interests), respectively.

Table A1: Accounting variables used in the analysis

#	Variable	Description	#	Variable	Description
1	ACACE	Cash and Cash Equivalents	109	AMI	Minority Interests
2	AIDFC	Including Deposits rom Customers	110	BNII	Net Interest Income
3	ASP	Settlement Provisions	111	BII	Interest Income
4	AISPFC	Including Settlement Provisions from Customers	112	BIE	Interest Expenses
5	ACADFC	Cash and Due from Central Bank	113	BPE	Premiums Earned
6	ADFBAO	Due from Banks and Other Financial Institutions	114	BIUI	Insurance Underwriting Income
7	APM	Precious Metal	115	BIRPI	Including Reinsurance Premium Income
8	ANLTBA	Net Lendings to Banks and Other Financial Institutions	116	BLPCTR	Less Premium Ceded to Reinsurers
9	ATFA	Trading Financial Assets	117	BLPFUP	Less Provision for Unearned Premium Reserves
10	ADFA	Derivative Financial Assets	118	BNFACI	Net Fees and Commissions Income
11	ANSI	Net Short-Term Investments	119	BINIFS	Including Net Income from Securities Broker
12	ANNR	Net Notes Receivable	120	BINIFA	Including Net Income from Asset Management for Customer
13	ANAR	Net Accounts Receivable	121	BFACI	Fees and Commissions Income
14	ANP	Net Prepayments	122	BFACE	Fees and Commissions Expenses

Continued

#	Variable	Description	#	Variable	Description
15	ANPR	Net Premiums Receivable	123	BOOI	Other Operating Income
16	ANRR	Net Reinsurance Receivable	124	BPOS	Payments on Surrenders
17	ANSR	Net Subrogation Receivable	125	BNCE	Net Claim Expenses
18	ANRRR	Net Reinsurance Reserves Receivable	126	BCE	Claim Expenses
19	AIRFCU	Including Receivable from Ceded Unearned Premium Reserves	127	BLCRFR	Less Claims Recoverable from Reinsurers
20	AIRFCC	Including Receivable from Ceded Claim Reserves	128	BNPFIR	Net Provision for Insurance Reserves
21	AIRFCL	Including Receivable from Ceded Life Insurance Reserves	129	BPFIR	Provision for Insurance Reserves
22	ANIR	Net Interest Receivable	130	BLIRRF	Less Insurance Reserves Recoverable from Reinsurance
23	ANDR	Net Dividends Receivable	131	BPD	Policyholder Dividends
24	ANOR	Net Other Receivables	132	BEFRA	Expenses for Reinsurance Accepted
25	ANAPUA	Net Assets Purchased Under Agreements to Resell	133	BSTAEC	Sales Tax and Extra Charges
26	ANI	Net Inventories	134	BBAME	Business and Management Expenses
27	ANADWO	Non-Current Assets Due Within One Year	135	BLERFR	Less Expenses Recoverable from Reinsurers
28	ARD	Refundable Deposits	136	BSE	Selling Expenses
29	AOCA	Other Current Assets	137	BAE	Administrative Expenses
30	ATCA	Total Current Assets	138	BFE	Finance Expenses
31	ANPL	Net Policyholder Loans	139	BAIL	Asset Impairment Losses
32	ATD	Term Deposits	140	BOOC	Other Operating Costs
33	ANLAR	Net Loans and Receivables	141	BIFCIF	Income from Changes in Fair Value
34	AAFA	Available-For-Sale Financial Assets	142	BIG	Investment Gains
35	AHI	Held-To-Maturity Investments	143	BIIGFA	Including Investment Gains from Associates and Joint Ventures
36	ANLEI	Net Long-Term Equity Investments	144	BFEG	Foreign Exchange Gains

Continued

#	Variable	Description	#	Variable	Description
37	ANLDI	Net Long-Term Debt Investments	145	BPFOO	Profit from Other Operations
38	ANLI	Net Long-Term Investment	146	BNI	Non-Operating Income
39	ADFCR	Deposits for Capital Recognizance	147	BIEFDO	Including Earnings from Disposal of Non-Current Assets
40	ASAA	Separate Account Assets	148	BNE	Non-Operating Expenses
41	ANIP	Net Investment Properties	149	BINLFD	Including Net Loss from Disposal of Non-Current Assets
42	ANFA	Net Fixed Assets	150	BILFDO	Including Loss from Disposal of Non-Current Assets
43	ANCIP	Net Construction in Progress	151	BITE	Income Tax Expenses
44	ADOFA	Disposal of Fixed Assets	152	BUIL	Unconfirmed Investment Loss
45	ANBBA	Net Bearer Biological Assets	153	BOIANP	Other Items Affecting Net Profit
46	ANOAGA	Net Oil and Gas Assets	154	BBEPS	Basic Earnings Per Share
47	ANIA	Net Intangible Assets	155	BDEPS	Diluted Earnings Per Share
48	AITSF	Including Trading Seat Fee	156	BOCI	Other Consolidated Income (Loss)
49	AR&DE	Research & Development Expenses	157	BTCI	Total Consolidated Income
50	ANG	Net Goodwill	158	BCIATO	Consolidated Income Attributable to Owners of the Parent Company
51	ALPE	Long-Term Prepaid Expenses	159	BCIATM	Consolidated Income Attributable to Minority Shareholders
52	ADTA	Deferred Tax Assets	160	CCRFSO	Cash Received from Sales of Goods or Rendering Services
53	AONA	Other Non-Current Assets	161	CNIICD	Net Increase in Customer Deposits and Due to Banks and Other Financial Institutions
54	AOA	Other Assets	162	CNCBFC	Net Cash Borrowings from Central Bank
55	ASB	Short-Term Borrowings	163	CIIBFO	Increase In Borrowings from Other Financial Institutions
56	AIML	Including Mortgage Loan	164	CPR	Premiums Received
57	ABFCB	Borrowings from Central Bank	165	CNCRFR	Net Cash Received from Reinsurance

Continued

#	Variable	Description	#	Variable	Description
58	ADADTB	Deposits and Due to Banks and Other Financial Institutions	166	CIIPDA	Increase in Policyholders Deposits and Investment
59	AIDTBA	Including Due to Banks and Other Financial Institutions	167	CCRFDO	Cash Received from Disposal of Trading Financial Assets
60	AIDTC	Including Due to Customers	168	CIFACR	Interests, Fees, and Commissions Received
61	ABFBAO	Borrowings from Banks and Other Financial Institutions	169	CIIBFB	Increase in Borrowings from Banks and Other Financial Institutions
62	ATFL	Trading Financial Liabilities	170	CIIR	Increase in Repo
63	ADFL	Derivative Financial Liabilities	171	CTR	Tax Refund
64	AAP	Accounts Payable	172	COCRRT	Other Cash Received Relating to Operating Activities
65	AAR	Advance Receipts	173	CCPFGA	Cash Paid for Goods and Services
66	AASUAT	Assets Sold Under Agreements to Repurchase	174	CIILTC	Increase in Loan to Customers
67	AFACP	Fees and Commissions Payable	175	CNIIDF	Net Increase in Due from Central Bank and Financial Institutions
68	AEBP	Employee Benefits Payable	176	CCP	Claims Paid
69	ATP	Taxes Payable	177	CIFACP	Interests, Fees, and Commissions Paid
70	AIP	Interests Payable	178	CPDP	Policy Dividends Paid
71	ADP	Dividends Payable	179	CCPTAO	Cash Paid to and on Behalf of Employees
72	ACP	Claims Payable	180	CVTP	Various Taxes Paid
73	APDP	Policy Dividends Payable	181	COCPRT	Other Cash Paid Relating to Operating Activities
74	APDAI	Policyholder Deposits and Investments	182	CNCFFO	Net Cash Flow from Operating Activities
75	ARFIC	Reserves For Insurance Contract	183	CCRFRO	Cash Received from Returns on Investments
76	AIUPR	Including Unearned Premium Reserves	184	CNCRFD	Net Cash Received from Disposals Of Fixed Assets, Intangible Assets and Other Long-Term Assets

Continued

#	Variable	Description	#	Variable	Description
77	AIRFOL	Including Reserves for Outstanding Losses	185	CCPTAA	Cash Paid to Acquire and Construct Fixed Asset, Intangible Assets and Other Long-Term Assets
78	AIRFLI	Including Reserves for Life Insurance Liabilities	186	CCPFI	Cash Paid for Investments
79	AILHIP	Including Long-Term Health Insurance Policyholders Reserves	187	CNIIPL	Net Increase in Policyholder Loans
80	AOP	Other Payables	188	CNCPFA	Net Cash Paid for Acquisition of Subsidiaries And Other Operating Units
81	ARP	Reinsurance Payable	189	CNCFFI	Net Cash Flow from Investing Activities
82	ASTAB	Securities Trading as Brokerage	190	CCRFAO	Cash Received from Absorption Of Investments
83	APFSU	Proceeds from Securities Underwriting	191	CPFIS	Proceeds from Issuing Shares
84	APRIA	Premium Received in Advance	192	CIPFIS	Including Proceeds from Issuing Shares to Minority Investments by Subsidiaries
85	ANLDWO	Non-Current Liabilities Due Within One Year	193	CPFIB	Proceeds from Issuing Bonds
86	AOCL	Other Current Liabilities	194	CPFB	Proceeds from Borrowings
87	ADIL	Deferred Income-Current Liabilities	195	CCROAB	Cash Repayments of Amounts Borrowed
88	ASAL	Separate Account Liabilities	196	CCPFDO	Cash Paid for Distribution of Dividends, Profits, or Interest Payments
89	ABP	Bonds Payable	197	CIDAPD	Including Dividends and Profits Distributed Minority Shareholders by Subsidiaries
90	ALP	Long-Term Payables	198	CNCFFF	Net Cash Flow from Financing Activities
91	APFSP	Payables for Special Projects	199	CEOERC	Effect of Exchange Rate Changes on Cash and Cash Equivalents
92	ACL	Contingency Liabilities	200	CEOOIO	Effect of Other Items on Cash
93	ADTL	Deferred Tax Liabilities	201	CNIOCA	Net Increase of Cash and Cash Equivalents

Continued

#	Variable	Description	#	Variable	Description
94	AONL	Other Non-Current Liabilities	202	CBBOCA	Beginning Balance of Cash and Cash Equivalents
95	AOL	Other Liabilities	203	CEBOCA	Ending Balance of Cash and Cash Equivalents
96	AOEI	Other Equity Instruments	204	AIRFCL1	Including Receivables from Ceded Long-Duration Health Insurance Liability Reserves
97	AIPS	Including Preference Share	205	ANLAR1	Net Long-term Accounts Receivable
98	AIPB	Including Perpetual Bond	206	ANP1	Notes Payable
99	AIO	Including Others	207	ADIL1	Deferred Income-Non-Current Liabilities
100	ACR	Capital Reserves	208	ASR1	Special Reserves
101	AITS	Including Treasury Stock	209	BINIFS1	Including Net Income from Securities Underwriting
102	ASR	Surplus Reserves	210	CCRFDO1	Cash Received from Disposal of Investments
103	AGRR	General Risk Reserves	211	CNCRFD1	Net Cash Received from Disposals of Subsidiaries and Other Operating Units
104	ARE	Retained Earnings	212	COCRRT1	Other Cash Received Relating to Investing Activities
105	AFTD	Foreign Translation Difference	213	COCPRT1	Other Cash Paid Relating to Investing Activities
106	AUIL	Unconfirmed Investment Loss	214	COCRRT2	Other Cash Received Relating to Financing Activities
107	ATRR	Trading Risk Reserves	215	COCRRT3	Other Cash Received Relating to Financing Activities
108	AOCI	Other Consolidated Income			

Table A2 presents the 106 configurations used in the study. The sample period spans from January of 2001 to June of 2018. We first collect all the financial variables with enough amount of data reported in CSMAR. Then, we filter out the variables with no-missing values fewer than five years or fewer than 500 firms on average per year. X denotes the 215 accounting variables listed in Appendix A1. Y denotes the 21 base variables. They are ATLASS (total liabilities and shareholder's equity), ATA (total assets), ATNA (total non-current assets), ACM (construction material), ASC(share capital), ATSE (total shareholders' equity), ATEATO (total equity attributable to owners of the parent company), ATL (total liabilities), ATCL (total current liabilities), ALD (long-term debts), ATLL (total long-term liabilities), ATNL (total non-current liabilities), BTP (total profit), BNP (net profit), BOP (operating profit), BTOR (total operating revenue), BOR (operating revenue), BTOC (total operating costs), BOC (operating costs), BNPATO (net profit attributable to owners of the parent company), and BMI (Minority Interests).

Table A2: Configurations used in the study.

#	Description	#	Description	#	Description	#	Description
1	X/ACM	28	Δ in X/ATNL	55	%Δ in X/BTOR	82	$\Delta X/LAGBNP$
2	X/ATNA	29	Δ in X/ATL	56	%Δ in X/BOR	83	$\Delta X/LAGBNPATO$
3	X/ATA	30	Δ in X/ASC	57	%Δ in X/BTOC	84	$\Delta X/LAGBMI$
4	X/ATCL	31	Δ in X/ATEATO	58	%Δ in X/BOC	85	%Δ in X - %Δ in ACM
5	X/ALD	32	Δ in X/ATSE	59	%Δ in X/BOP	86	%Δ in X - %Δ in ATNA
6	X/ATLL	33	Δ in X/ATLASS	60	%Δ in X/BTP	87	%Δ in X - %Δ in ATA
7	X/ATNL	34	Δ in X/BTOR	61	%Δ in X/BNP	88	%Δ in X - %Δ in ATCL
8	X/ATL	35	Δ in X/BOR	62	%Δ in X/BNPATO	89	%Δ in X - %Δ in ALD
9	X/ASC	36	Δ in X/BTOC	63	%Δ in X/BMI	90	%Δ in X -%Δ in ATLL
10	X/ATEATO	37	Δ in X/BOC	64	$\Delta X/LAGACM$	91	%Δ in X - %Δ in ATNL
11	X/ATSE	38	Δ in X/BOP	65	$\Delta X/LAGATNA$	92	%Δ in X - %Δ in ATL
12	X/ATLASS	39	Δ in X/BTP	66	$\Delta X/LAGATA$	93	%Δ in X - %Δ in ASC
13	X/BTOR	40	Δ in X/BNP	67	$\Delta X/LAGATCL$	94	%Δ in X - %Δ in ATEATO
14	X/BOR	41	Δ in X/BNPATO	68	$\Delta X/LAGALD$	95	%Δ in X - %Δ in ATSE
15	X/BTOC	42	Δ in X/BMI	69	$\Delta X/LAGATLL$	96	%Δ in X - %Δ in ATLASS
16	X/BOC	43	%Δ in X/ACM	70	$\Delta X/LAGATNL$	97	%Δ in X -%Δ in BTOR
17	X/BOP	44	%Δ in X/ATNA	71	$\Delta X/LAGATL$	98	%Δ in X - %Δ in BOR
18	X/BTP	45	%Δ in X/ATA	72	$\Delta X/LAGASC$	99	%Δ in X - %Δ in BTOC
19	X/BNP	46	%Δ in X/ATCL	73	$\Delta X/LAGATEATO$	100	%Δ in X - %Δ in BOC
20	X/BNPATO	47	%Δ in X/ALD	74	$\Delta X/LAGATSE$	101	%Δ in X - %Δ in BOP
21	X/BMI	48	%Δ in X/ATLL	75	$\Delta X/LAGATLASS$	102	%Δ in X - %Δ in BTP
22	Δ in X/ACM	49	%Δ in X/ATNL	76	$\Delta X/LAGBTOR$	103	%Δ in X - %Δ in BNP
23	Δ in X/ATNA	50	%Δ in X/ATL	77	$\Delta X/LAGBOR$	104	%Δ in X - %Δ in BNPATO
24	Δ in X/ATA	51	%Δ in X/ASC	78	$\Delta X/LAGBTOC$	105	%Δ in X - %Δ in BMI
25	Δ in X/ATCL	52	%Δ in X/ATEATO	79	$\Delta X/LAGBOC$	106	%Δ in X
26	Δ in X/ALD	53	%Δ in X/ATSE	80	$\Delta X/LAGBOP$		
27	Δ in X/ATLL	54	%Δ in X/ATLASS	81	$\Delta X/LAGBTP$		

Table A3 reports selected percentiles of alpha, t-statistics, and p-value for hedge portfolio alphas of 8,288 fundamental signals as described in the text. For each percentile, the table also presents the fraction of simulations where the bootstrapped percentile was smaller than the actual percentile for alpha, t-statistics, and p-value. The sample period is from January of 2001 to June of 2018. The results are based on 10,000 simulations.

Table A3: Actual and bootstrapped distribution

	alpha	pvalue_alpha	alpha_t	pvalue_alpha_t	alpha_p	pvalue_alpha_p
0	−0.0062	81.20%	−3.5629	54.90%	0.0000	8.90%
1	−0.0036	62.00%	−2.4447	27.70%	0.0083	29.10%
2	−0.0031	53.60%	−2.0608	43.60%	0.0171	30.80%
3	−0.0028	50.60%	−1.8673	47.30%	0.0281	36.00%
4	−0.0025	58.00%	−1.7169	52.00%	0.0375	35.90%
5	−0.0023	59.60%	−1.5861	57.20%	0.0466	35.70%
10	−0.0016	74.40%	−1.1596	71.70%	0.0962	38.40%
20	−0.0010	80.20%	−0.6738	82.50%	0.1962	39.30%
30	−0.0005	83.40%	−0.3591	85.30%	0.3033	47.40%
40	−0.0001	85.10%	−0.0967	85.60%	0.4122	56.20%
50	0.0002	85.60%	0.1451	85.10%	0.5139	59.50%
60	0.0006	85.00%	0.3796	81.90%	0.6116	58.60%
70	0.0009	83.30%	0.6357	80.00%	0.7105	60.50%
80	0.0013	79.50%	0.9521	79.70%	0.8063	59.30%
90	0.0020	79.00%	1.3961	79.60%	0.9024	55.00%
95	0.0026	72.10%	1.7321	76.10%	0.9545	77.90%
96	0.0027	67.10%	1.8306	75.40%	0.9656	89.80%
97	0.0030	66.30%	1.9474	71.90%	0.9758	95.50%
98	0.0033	58.60%	2.1163	71.70%	0.9850	98.50%
99	0.0037	48.20%	2.3564	65.00%	0.9928	98.50%
100	0.0064	27.40%	4.5106	94.90%	1.0000	95.90%

Table A4 reports descriptive statistics on monthly average abnormal returns of hedge portfolios. The hedge portfolios are constructed as mentioned in the text and the sample period is from January of 2001 to June of 2018. We present cross-sectional mean, median, standard deviation, minimum, maximum, number, and ratio of abnormal returns larger than 0.5% and 1.0% in panel A. We calculate abnormal returns relative CAPM (Sharpe, 1963; Linter, 1965) one-factor model. We report mean, median, standard deviation, minimum, and ratio of abnormal returns larger than 0.5% or 1.0% in percentage term. We also present the corresponding t-statistics in panel B.

Table A4: Descriptive statistics of portfolio abnormal returns relative to CAPM

	Panel A: Average return									
	N	Mean	Median	Std	Min	Max	$\lvert return \rvert > 0.5\%$		$\lvert return \rvert > 1.0\%$	
Category							#	%	#	%
ratio	1780	0.06	0.05	0.24	−0.49	1.66	65	3.65	6	0.34
ratio_chg	1558	0.01	0.01	0.16	−0.63	0.48	14	0.9	0	0
ratio_growth	1468	0	0	0.13	−0.48	0.47	0	0	0	0
ratio_growth_dif	1804	0.05	0.05	0.13	−0.36	0.59	1	0.06	0	0
ratio_x_chg_over_lag_y	1591	−0.08	−0.06	0.16	−0.79	0.41	38	2.39	0	0
x_growth	87	−0.12	−0.1	0.13	−0.43	0.2	0	0	0	0
all	8288	0.01	0.01	0.18	−0.79	1.66	118	1.42	6	0.07
	Panel B: Average return t-statistic									
	N	Mean	Median	Std	Min	Max	$\lvert t \rvert > 1.96$		$\lvert t \rvert > 2.75$	
Category							#	%	#	%
ratio	1780	0.3	0.33	1.08	−2.66	3.17	122	6.85	24	1.35
ratio_chg	1558	0.14	0.07	1.13	−2.83	3.11	154	9.88	22	1.41
ratio_growth	1468	0.02	0.01	1	−3.73	2.8	73	4.97	15	1.02
ratio_growth_dif	1804	0.34	0.36	0.93	−2.49	3.4	92	5.1	14	0.78
ratio_x_chg_over_lag_y	1591	−0.49	−0.43	1	−3.9	2.5	140	8.8	42	2.64
x_growth	87	−0.65	−0.65	0.77	−2.9	1.87	3	3.45	1	1.15
all	8288	0.07	0.07	1.07	−3.9	3.4	584	7.05	118	1.42

Table A5 reports descriptive statistics on monthly average abnormal returns of hedge portfolios. The hedge portfolios are constructed as mentioned in the text and the sample period is from January of 2001 to June of 2018. We present cross-sectional mean, median, standard deviation, minimum, maximum, number, and ratio of abnormal returns larger than 0.5% and 1.0% in panel A. We calculate abnormal returns relative to the Fama and French (Fama & French, 1993) three-factor model (FF3). We report mean, median, standard deviation, minimum, and ratio of abnormal returns larger than 0.5% or 1.0% in percentage term. We also present the corresponding t-statistics in panel B.

Table A5: Descriptive statistics of portfolio abnormal returns relative to FF3

	Panel A: Average return									
	N	Mean	Median	Std	Min	Max	$\lvert return\rvert > 0.5\%$		$\lvert return\rvert > 1.0\%$	
Category							#	%	#	%
ratio	1780	0.07	0.08	0.23	−0.69	0.79	69	3.88	0	0
ratio_chg	1558	0.01	0.02	0.15	−0.46	0.65	5	0.32	0	0
ratio_growth	1468	0.04	0.04	0.14	−0.46	0.56	2	0.14	0	0
ratio_growth_dif	1804	−0.02	−0.02	0.16	−0.56	0.48	1	0.06	0	0
ratio_x_chg_over_lag_y	1591	0.02	0.02	0.17	−0.64	0.55	6	0.38	0	0
x_growth	87	−0.02	−0.04	0.14	−0.4	0.32	0	0	0	0
all	8288	0.02	0.02	0.18	−0.69	0.79	83	1	0	0
	Panel B: Average return t-statistic									
	N	Mean	Median	Std	Min	Max	$\lvert t\rvert > 1.96$		$\lvert t\rvert > 2.75$	
Category							#	%	#	%
ratio	1780	0.38	0.47	1.31	−3.87	5.38	272	15.28	104	5.84
ratio_chg	1558	0.12	0.12	1.07	−3.18	3.36	95	6.1	29	1.86
ratio_growth	1468	0.3	0.33	1.06	−3.88	3.22	113	7.7	25	1.7
ratio_growth_dif	1804	−0.2	−0.2	1.27	−4.27	3.77	232	12.86	84	4.66
ratio_x_chg_over_lag_y	1591	0.12	0.1	1.13	−2.97	3.03	151	9.49	38	2.39
x_growth	87	−0.12	−0.33	1.07	−2.87	2.82	6	6.9	2	2.3
all	8288	0.14	0.15	1.2	−4.27	5.38	869	10.49	282	3.4

Table A6 reports descriptive statistics on monthly average abnormal returns of hedge portfolios. The hedge portfolios are constructed as mentioned in the text and the sample period is from January of 2001 to June of 2018. We present cross-sectional mean, median, standard deviation, minimum, maximum, number, and ratio of abnormal returns larger than 0.5% and 1.0% in panel A. We calculate abnormal returns relative to the Carhart (1997) four-factor model (FFC). We report mean, median, standard deviation, minimum, and ratio of abnormal returns larger than 0.5% or 1.0% in percentage term. We also present the corresponding t-statistics in panel B.

Table A6: Descriptive statistics of portfolio abnormal returns relative to FFC

	Panel A: Average return									
	N	Mean	Median	Std	Min	Max	$\lvert return \rvert > 0.5\%$		$\lvert return \rvert > 1.0\%$	
Category							#	%	#	%
ratio	1780	0.05	0.06	0.19	−0.68	0.74	21	1.18	0	0
ratio_chg	1558	0.02	0.02	0.14	−0.45	0.69	7	0.45	0	0
ratio_growth	1468	0.02	0.03	0.12	−0.43	0.49	0	0	0	0
ratio_growth_dif	1804	−0.02	−0.03	0.14	−0.58	0.42	3	0.17	0	0
ratio_x_chg_over_lag_y	1591	0	−0.01	0.15	−0.61	0.65	12	0.75	0	0
x_growth	87	0.01	−0.01	0.12	−0.3	0.3	0	0	0	0
all	8288	0.01	0.01	0.15	−0.68	0.74	43	0.52	0	0
	Panel B: Average return t-statistic									
	N	Mean	Median	Std	Min	Max	$\lvert t \rvert > 1.96$		$\lvert t \rvert > 2.75$	
Category							#	%	#	%
ratio	1780	0.3	0.33	1.13	−3.55	4.96	155	8.71	51	2.87
ratio_chg	1558	0.17	0.15	0.99	−2.89	3.55	82	5.26	17	1.09
ratio_growth	1468	0.19	0.21	0.95	−3.42	3.09	53	3.61	14	0.95
ratio_growth_dif	1804	−0.22	−0.26	1.11	−3.93	3.46	139	7.71	35	1.94
ratio_x_chg_over_lag_y	1591	−0.04	−0.05	0.98	−2.97	2.68	83	5.22	16	1.01
x_growth	87	0.06	−0.07	0.92	−2.22	2.58	4	4.6	1	1.15
all	8288	0.08	0.07	1.06	−3.93	4.96	516	6.23	134	1.62

Table A7 reports descriptive statistics on monthly average abnormal returns of hedge portfolios. The hedge portfolios are constructed as mentioned in the text and the sample period is from January of 2001 to June of 2018. We present cross-sectional mean, median, standard deviation, minimum, maximum, number, and ratio of abnormal returns larger than 0.5% and 1.0% in panel A. We calculate abnormal returns relative to Q4 four-factor model (Hou et al., 2014). We report mean, median, standard deviation, minimum, and ratio of abnormal returns larger than 0.5% or 1.0% in percentage term. We also present the corresponding t-statistics in panel B.

Table A7: Descriptive statistics of portfolio abnormal returns relative to Q4

	Panel A: Average return									
	N	Mean	Median	Std	Min	Max	$\|return\| > 0.5\%$		$\|return\| > 1.0\%$	
Category							#	%	#	%
ratio	1780	0.07	0.07	0.3	−1.19	1.15	185	10.39	9	0.51
ratio_chg	1558	0.06	0.05	0.22	−1.04	0.65	47	3.02	1	0.06
ratio_growth	1468	0.06	0.06	0.18	−0.52	0.64	14	0.95	0	0
ratio_growth_dif	1804	−0.05	−0.06	0.19	−0.64	0.68	16	0.89	0	0
ratio_x_chg_over_lag_y	1591	0.07	0.08	0.23	−1.05	0.76	75	4.71	1	0.06
x_growth	87	0.25	0.25	0.21	−0.18	0.87	10	11.49	0	0
all	8288	0.04	0.04	0.23	−1.19	1.15	347	4.19	11	0.13
	Panel B: Average return t-statistic									
	N	Mean	Median	Std	Min	Max	$\|t\| > 1.96$		$\|t\| > 2.75$	
Category							#	%	#	%
ratio	1780	0.43	0.32	1.45	−3.55	6.37	375	21.07	172	9.66
ratio_chg	1558	0.36	0.37	1.21	−3.84	3.41	207	13.29	61	3.92
ratio_growth	1468	0.42	0.42	1.21	−3.47	4.4	193	13.15	60	4.09
ratio_growth_dif	1804	−0.35	−0.44	1.28	−3.83	4.1	243	13.47	84	4.66
ratio_x_chg_over_lag_y	1591	0.44	0.52	1.17	−3.02	3.73	194	12.19	65	4.09
x_growth	87	1.54	1.48	1.25	−1.29	5.09	26	29.89	16	18.39
all	8288	0.26	0.24	1.32	−3.84	6.37	1238	14.94	458	5.53

Table A8 reports descriptive statistics on monthly average abnormal returns of hedge portfolios. The hedge portfolios are constructed as mentioned in the text, and the sample period is from January of 2001 to June of 2018. We present cross-sectional mean, median, standard deviation, minimum, maximum, number and ratio of abnormal returns larger than 0.5% and 1.0% in panel A. We calculate abnormal returns relative to the Fama and French (Fama & French, 2017) six-factor model (FF6). We report mean, median, standard deviation, minimum and ratio of abnormal returns larger than 0.5% or 1.0% in percentage term. We also present the corresponding t-statistics in panel B.

Table A8: Descriptive statistics of portfolio abnormal returns relative to FF6

	Panel A: Average return									
	N	Mean	Median	Std	Min	Max	$\lvert return\rvert > 0.5\%$		$\lvert return\rvert > 1.0\%$	
Category							#	%	#	%
ratio	1780	0.04	0.04	0.17	−0.66	0.6	8	0.45	0	0
ratio_chg	1558	0.03	0.03	0.15	−0.48	0.61	7	0.45	0	0
ratio_growth	1468	0.02	0.02	0.12	−0.43	0.45	0	0	0	0
ratio_growth_dif	1804	−0.01	−0.01	0.12	−0.5	0.36	0	0	0	0
ratio_x_chg_over_lag_y	1591	−0.02	−0.01	0.15	−0.57	0.69	16	1.01	0	0
x_growth	87	0.08	0.08	0.11	−0.18	0.4	0	0	0	0
all	8288	0.01	0.01	0.14	−0.66	0.69	31	0.37	0	0
	Panel B: Average return t-statistic									
	N	Mean	Median	Std	Min	Max	$\lvert t\rvert > 1.96$		$\lvert t\rvert > 2.75$	
Category							#	%	#	%
ratio	1780	0.25	0.23	1.02	−3.21	4.45	107	6.01	30	1.69
ratio_chg	1558	0.24	0.23	1.01	−3.23	3.11	104	6.68	22	1.41
ratio_growth	1468	0.14	0.14	0.9	−3.18	2.74	47	3.2	5	0.34
ratio_growth_dif	1804	−0.05	−0.05	0.95	−3.34	2.78	70	3.88	12	0.67
ratio_x_chg_over_lag_y	1591	−0.14	−0.1	0.97	−3.85	2.74	87	5.47	31	1.95
x_growth	87	0.6	0.62	0.87	−1.38	3.14	8	9.2	2	2.3
all	8288	0.09	0.09	0.98	−3.85	4.45	423	5.1	102	1.23

REFERENCES

Abarbanell, J. S., and B. J. Bushee, 1997. Fundamental Analysis, Future Earnings, and Stock Prices. *Journal of Accounting Research* **35**, 1-24.

Baker, M., P. Luo, and R. Taliaferro, 2017. Detecting Anomalies: The Relevance and Power of Standard Asset Pricing Tests. Working paper.

Cakici, N., K. Chan, and K. Topyan, 2015. Cross-sectional stock return predictability in China. *The European Journal of Finance* **0**, 1-25.

Carhart, M. M., 1997. On persistence in mutual fund performance. *The Journal of Finance* **52**, 57-82.

Carpenter, J. N., F. Lu, and R. Whitelaw, 2021. The Real Value of China's Stock Market. *Journal of Financial Economics* **139**, 679-696.

Chen, X., K. A. Kim, T. Yao, and T. Yu, 2010. On the predictability of Chinese stock returns. *Pacific-Basin Finance Journal* **18**, 403-425.

Cheung, C., G. Hoguet, and S. Ng, 2015. Value, Size, Momentum, Dividend Yield, and Volatility in China's A-Share Market. *The Journal of Portfolio Management* **41**, 57-70.

Chordia, T., A. Goyal, and A. Saretto, 2018. Anomalies and Multiple Hypothesis Testing: Evidence from Two Million Trading Strategies (Tech. Rep.).

Cochrane, J. H., 2011. Presidential Address: Discount Rates. *The Journal of Finance* **66**, 1047-1108.

Engelberg, J., R. D. McLean, and J. Pontiff, 2020. Analysts and Anomalies. *Journal of Accounting and Economics* **69**, 1-13.

Fama, E. F., and K. R. French, 1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance* **47**, 427-465.

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

Fama, E. F., and K. R. French, 1996. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance* **51**, 55-84.

Fama, E. F., and K. R. French, 2006. Profitability, investment, and average returns. *Journal of Financial Economics* **82**, 491-518.

Fama, E. F., and K. R. French, 2010. Luck versus Skill in the Cross-Section of Mutual Fund Returns. *The Journal of Finance* **65**, 1915-1947.

Fama, E. F., and K. R. French, 2016. Dissecting Anomalies with a Five-Factor Model. *The Review of Financial Studies* **29**, 69-103.

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

Fama, E. F., and K. R. French, 2016. Dissecting Anomalies with a Five-Factor Model. *Review of Financial Studies* **29**, 69-103.

Fama, E. F., and K. R. French, 2018. Choosing Factors. *Journal of Financial Economics* **128**, 234-252.

Fama, E. F., and J. D. MacBeth, 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* **81**, 607-636.

Green, J., J. R. M. Hand, and X. F. Zhang, 2013. The supraview of return predictive signals. *Review of Accounting Studies* **18**, 692-730.

Green, J., J. R. M. Hand, and X. F. Zhang, 2017. The Characteristics That Provide Independent Information About Average U.S. Monthly Stock Returns. *The Review of Financial Studies.*

Han, Y., A. He, D. Rapach, and G. Zhou, 2018. How Many Firm Characteristics Drive US Stock Returns? (SSRN Scholarly Paper). Rochester, NY.

Harvey, C. R., and Y. Liu, 2021. Lucky factors. *Journal of Financial Economics* **141**, 413-435.

Harvey, C. R., Y. Liu, and H. Zhu, 2016. . . . and the Cross-Section of Expected Returns. *The Review of Financial Studies* **29**.

Haugen, R. A., and N. L. Baker, 1996. Commonality in the determinants of expected stock returns. *Journal of Financial Economics* **41**, 401-439.

Hou, K., and R. K. Loh, 2016. Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics* **121**, 167-194.

Hou, K., H. Mo, C. Xue, and L. Zhang, 2021. An Augmented q-Factor Model with Expected Growth. *Review of Finance* **25**, 1-41.

Hou, K., H. Mo, C. Xue, and L. Zhang, 2017. The Economics of Value Investing (Working Paper The Ohio State University).

Hou, K., C. Xue, and L. Zhang, 2014. Digesting Anomalies: An Investment Approach. *Review of Financial Studies* **28**, 650-705.

Hou, K., C. Xue, and L. Zhang, 2020. Replicating Anomalies. *The Review of Financial Studies* **33**, 2019-2133.

Hsu, J. C., V. Viswanathan, C. Wang, and P. Wool, 2018. Anomalies in Chinese A-Shares. *Journal of Portfolio Management* **44**, 108-123.

Hu, G. X., J. Pan, and J. Wang, 2021. Chinese Capital Market: An Empirical Overview. *Critical Finance Review* **10**, 125-206.

Jacobs, H., 2015. What explains the dynamics of 100 anomalies? *Journal of Banking and Finance* **57**, 65-85.

Jarque, C. M., and A. K. Bera, 1980. Efficient tests for normality, homoscedasticity, and serial independence of regression residuals. *Economics Letters* **6**, 255-259.

Jiang, F., X. Qi, and G. Tang, 2018. Q-Theory, Mispricing, and Profitability Premium: Evidence from China. *Journal of Banking and Finance* **87**, 135-149.

Jiang, F., G. Tang, and G. Zhou, 2018. Firm Characteristics and Chinese Stocks. *Journal of Management Science and Engineering* **3**, 259-283.

Jiang, F., J. Tu,, D. Rapach, J. K. Strauss, and G. Zhou, 2011. How Predictable is the Chinese Stock Market? *Journal of Financial Research*, 107-121.

Kogan, L., and M. H. Tian, 2021. Firm Characteristics and Empirical Factor Models: A Model-Mining Experiment. *The Review of Financial Studies* **34**, 6087-6125.

Kosowski, R., A. Timmermann, R. Wermers, and H. White, 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *The Journal of Finance* **61**, 2551-2595.

Kozak, S., S. Nagel, and S. Santosh, 2018. Interpreting Factor Models. *The Journal of Finance* **73**, 1183-1223.

Lewellen, J., 2015. The Cross-section of Expected Stock Returns. *Critical Finance Review* **4**, 1-44.

Li, Y., 2023. Asset Pricing and Microcaps. *Annals of Economics and Finance* **24-1**, 119-140.

Light, N., D. Maslov, and O. Rytchkov, 2017. Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach. *The Review of Financial Studies* **30**, 1339-1381.

Lin, Q., 2018. Technical analysis and stock return predictability: An aligned approach. *Journal of Financial Markets* **38**, 103-123.

Liu, J., R. F. Stambaugh, and Y. Yuan, 2019. Size and Value in China. *Journal of Financial Economics* **134**, 148-69.

Mclean, R. D., and J. Pontiff, 2016. Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance* **71**, 5-32.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou, 2014. Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Management Science* **60**, 1772-1791.

Novy-Marx, R., and M. Velikov, 2016. A Taxonomy of Anomalies and Their Trading Costs. *Review of Financial Studies* **29**, 104-147.

Rapach, D., and G. Zhou, 2013. Chapter 6 — Forecasting Stock Returns. In G. Elliott & A. Timmermann (Eds.), Handbook of Economic Forecasting (Vol. 2, p. 328 - 383).

Rapach, D. E., J. K. Strauss, and G. Zhou, 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies* **23**, 821-862.

Rytchkov, O., and X. Zhong, 2020. Information Aggregation and P-hacking. *Management Science* **66**, 1605-1626.

Stambaugh, R. F., J. Yu, and Y. Yuan, 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* **104**, 288-302.

Stambaugh, R. F., J. Yu, and Y. Yuan, 2015. Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle. *The Journal of Finance* **70**, 1903-1948.

Stambaugh, R. F., and Y. Yuan, 2016. Mispricing factors. *The Review of Financial Studies* **30**, 1270-1315.

Timmermann, A., 2006. Chapter 4 Forecast Combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), (Vol. 1).

Titman, S., C. Wei, and B. Zhao, 2017. Is the Chinese Stock Market Really Different? Evidence from Stock Splits in the U.S. and China. Working paper, University of Texas at Austin.

Yan, X. S., and L. Zheng, 2017. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. *The Review of Financial Studies* **30**, 1382-1423.