Pacific-Basin Finance Journal 82 (2023) 102192

Contents lists available at ScienceDirect

Pacific-Basin Finance Journal

journal homepage: www.elsevier.com/locate/pacfin

Expectation disarray: Analysts' growth forecast anomaly in China☆

Laura Xiaolei Liu [a], Yandi Zhu [b], Xinyu Zhang [a], Yingguang Zhang [a,*]

[a] Peking University, Guanghua School of Management, China
[b] Peking University, School of Economics, China

ARTICLE INFO

JEL codes:
G11
G12
G14
D83
D84

Keywords:
Stock return
Analyst forecast
China
Expectation

ABSTRACT

Analysts' growth forecasts positively predict stock returns in China, opposite to the results found in the US. Strategies that buy stocks with high growth forecasts and sell those with low growth forecasts earn annual abnormal returns of up to 20% (with *t*-values exceeding three). These results are stronger for longer-horizon forecasts and ex-ante more informative forecasts. In addition, the deviation of analysts' forecasts from unbiased forecasts based on statistical models positively predicts abnormal returns. Although the relationship between analysts' forecasts and returns in China is opposite to that in the US, these forecasts positively predict actual growth and are often too extreme, as in the US. Our results suggest that investors in China overlook valuable information contained in analysts' forecasts.

1. Introduction

Numerous studies indicate that analysts' growth forecasts tend to negatively predict stock returns in the US stock market. Researchers attribute this phenomenon mainly to biases in analysts' and investors' expectations (e.g., Bordalo et al., 2019; Da and Warachka, 2011; De Bondt and Thaler, 1990; La Porta, 1996). However, we find the exact opposite pattern by using data from China's stock market: analysts' growth forecasts *positively* predict abnormal stock returns. Strategies that buy stocks with high growth forecasts and sell those with low growth forecasts generate abnormal returns, reaching as high as 20% (with *t*-values exceeding three) on an annual basis.

Two reasons potentially explain this discrepancy between the two markets: (1) differences in analysts and (2) differences in investors. We explore these two interpretations separately. Analysts' forecasts in the US have been shown to be both highly informative regarding future earnings and predictably biased such as being overly optimistic (Richardson et al., 2004) and extrapolative (Bordalo

☆ We appreciate the helpful comments and suggestions from Ling Cen (discussant), Allaudeen Hameed, Bing Han, Narasimhan Jegadeesh, Sheridan Titman, K.C. John Wei, and other participants at 2023 Pacific-Basin Finance Journal Special Issue Conference. Laura Xiaolei Liu acknowledges the financial support from the National Natural Science Foundation of China (NSFC Grant No. 72273006 and 71872006). We acknowledge the financial support from Guanghua Thought Leadership Platform in Peking University.

* Corresponding author.
E-mail addresses: laura.xi-aolei.liu@gsm.pku.edu.cn (L.X. Liu), yandi.zhu@stu.pku.edu.cn (Y. Zhu), xinyu_michelle@stu.pku.edu.cn (X. Zhang), yingguang.zhang@gsm.pku.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.pacfin.2023.102192
Received 11 October 2023; Accepted 15 October 2023
Available online 18 October 2023
0927-538X/

et al., 2019; La Porta, 1996).[1] We obtain similar results in China: analysts' earnings growth forecasts predict actual growth in a pooled regression with a coefficient of 0.9 for the current fiscal year and a coefficient of 0.3 for the subsequent two years. This finding suggests that analysts' forecasts are informative but also predictably too extreme. Thus, although firms with optimistic forecasts indeed experience faster growth, they systematically miss these forecasts and experience downward forecast revisions. These patterns concerning analysts' forecasts are consistent with the findings in the US market.

After ruling out the difference-in-analysts interpretation, we focus on understanding investors' behavior. We find that stocks with optimistic growth forecasts yield significantly positive abnormal returns even though these firms are more likely to miss analysts' forecasts during earnings announcements and experience downward forecast revisions. Importantly, the positive association between growth forecasts and subsequent stock returns is more pronounced for longer-horizon forecasts and when forecasts are ex-ante more accurate (proxied by past accuracy, institutional ownership, and analyst coverage). This evidence aligns with the interpretation that investors in China tend to overlook the fundamental information contained in analysts' growth forecasts.

How to reconcile the similarities and differences between the US and China evidence? On the one hand, analysts' primary duties in both markets are to produce research reports that deliver key information for investment decisions such as earnings forecasts, recommendations, and target prices. They also receive similar professional training and work for financial institutions (e.g., investment banks and brokerage firms) with similar organizational setups. Thus, the forecasts in the two markets should have similar information content. On the other hand, the stock market environment in China substantially differs from that in the US in terms of the composition of market participants. Retail investors dominate China's stock market, whereas institutional investors have more influence in the US market. This difference can make analysts' forecasts have drastically different asset pricing implications in the two markets. In particular, institutional investors are more likely to consume analysts' reports on a regular basis or to adopt a mindset similar to analysts because of their similar professional training. In contrast, retail investors typically do not have good access to analyst reports, and even when they do, they may not incorporate the information as fully as institutional investors would.

To further explore this explanation, we examine whether analysts' private information is fully incorporated into prices. We measure the deviations of analysts' forecasts from statistical model-based unbiased forecasts. These deviations capture both analysts' private information and their biases (de Silva and Thesmar, 2021). If analysts' private information is fully incorporated into prices, these deviations should not predict subsequent stock returns. If analysts' biases are also embedded in prices, these deviations should negatively predict stock returns. Evidence from the US indicates that this deviation strongly and negatively predicts subsequent stock returns, suggesting that the private information in analysts' forecasts is well incorporated into prices, along with the biases in these forecasts. However, in China, we again find the opposite pattern: deviations in analysts' forecasts from statistical benchmarks *positively* predict abnormal returns. This finding further supports the argument that analysts' growth forecasts in China contain private information not yet reflected in the stock market.

Our primary contribution is documenting that in China's stock market, investors appear to largely overlook analysts' information. Firms with optimistic analyst forecasts significantly outperform in the stock market, despite these firms consistently *missing* growth forecasts and experiencing downward forecast revisions. In contrast to the prominent findings in the US that analysts' long-horizon growth forecasts *negatively* predict subsequent stock returns (Bordalo et al., 2019; Da and Warachka, 2011; La Porta, 1996), we find that in China, longer-horizon growth forecasts more strongly and *positively* predict returns.

These results have important implications for researchers seeking to understand market expectations and price efficiency. Existing literature suggests that analysts' forecasts are predictably biased and that these biases distort stock prices in the US market, leading to prominent market anomalies, such as short-term underreactions and long-term overreactions in prices (e.g., Bordalo et al., 2019; Bouchaud et al., 2019; Da and Warachka, 2011; De Bondt and Thaler, 1990; Engelberg et al., 2020; Guo et al., 2020; Jegadeesh et al., 2004; La Porta, 1996). The close correspondence between forecast biases and anomalous pricing patterns suggests that analysts' forecasts reflect investors' expectations in the US market. This has led many researchers, including those referenced above, to leverage financial analysts' forecast data to explore general questions regarding belief formation.

Our results indicate that the US may be a special case in terms of how well analysts' forecasts represent investors' views. Do similar patterns hold in other markets? Addressing this question is important for understanding agents' belief formation, informational friction, and price efficiency in different market environments. Our results indicate that it would be premature to make claims about investors' expectations using analysts' forecasts outside the US market without appropriately accounting for the (mis)alignment between the two expectations.

Our findings offer practical valuable insights for investors, analysts, and financial policymakers in China and other similar markets. In particular, we show that investors' failure to incorporate analysts' information can lead to significant price inefficiencies. These price inefficiencies are (1) large, (2) persistent, and (3) more pronounced in larger firms with (4) better growth opportunities. The literature shows that asset pricing anomalies with these four characteristics are "real anomalies" because they have large economic implications (van Binsbergen and Opp, 2019). Our results imply a large amount of "money on the table" for investors, reveal the unmet demand for analysts' insights, and underscore an important source of information friction for policymakers. Our findings suggest that, all else equal, the overall welfare can improve if investors pay more attention to analysts' information, analysts more widely distribute their reports, or market designers promote the salience of analysts' forecasts in China.

Finally, our paper speaks to the relationship between anomaly replicability and market efficiency. We show that the same variable can predict abnormal returns with opposite signs in different markets. Instead of raising skepticism regarding the original findings on

[1] Analysts' forecasts also overreact (De Bondt and Thaler, 1990), fail to incorporate known predictors of stock returns (Engelberg et al., 2020; Guo et al., 2020; Jegadeesh et al., 2004), and are sticky (Bouchaud et al., 2019).

the ground of data mining, such discrepancies can provide deeper insights into the mechanisms underlying the initially documented anomalies. In particular, the differences in the growth forecast anomalies in the US and China markets suggest that analysts' biased forecasts per se are not enough to distort prices. The alignment between investors' beliefs and analysts' (biased) beliefs is a key determinant that varies significantly across market environments.

The remainder of this paper is structured as follows. Section 2 describes the data. Section 3 presents the results. Section 4 introduces a model. Section 5 provides a discussion. Section 6 concludes the paper.

2. Data

2.1. Sample construction

We use data from two sources: (1) monthly stock-level market data and firm-level financial data from the China Stock Market and Accounting Research (CSMAR) database and (2) daily analyst consensus forecast data from WIND.[2] Analysts typically forecast earnings for one year ahead (the current fiscal year, FY1), two years ahead (FY2), and three years ahead (FY3). Our sample period spans from January 2010 to December 2020. The starting time is limited by the availability of analyst forecast data. We begin with a sample consisting of all A-share stocks listed on the main boards of the Shanghai and Shenzhen exchanges as well as the board of the GEM (the Chinese counterpart of Nasdaq). We then apply standard filters to exclude observations in the banking and financial sectors, with ST or *ST status (which are typically small and illiquid), or listed for $<$12 months. Our final sample consists of 193,394 stock-month observations.

2.2. Variable definition

Our core variables are analysts' short- and long-horizon growth forecasts. At the end of each month t, we compute a stock i's short-horizon growth forecasts ($SFG_{i,t}$) as the difference between analysts' consensus earnings forecasts for the current fiscal year ($FE^{(1)}_{i,t}$) and the most recent actual earnings ($AE^{(0)}$), scaled by firms' most recent total assets ($TA_{i,t}$). In robustness tests, we show that our results are not sensitive to the choice of the scalar.

$$SFG_{i,t} = \frac{FE^{(1)}_{i,t} - AE^{(0)}_{i,t}}{TA_{i,t}} \tag{1}$$

Analysts in China do not provide forecasts equivalent to long-term growth forecasts (LTG) in the US, but they issue earnings forecasts up to three years ahead. We compute a counterpart of LTG — long-horizon growth forecast (LFG) — as the forward growth forecast from the end of the current fiscal year to three years ahead. This is calculated on a monthly basis as the consensus three-years-ahead earnings forecast minus the one-year-ahead forecast, scaled by total assets and annualized by multiplying it by one-half:

$$LFG_{i,t} = \frac{FE^{(3)}_{i,t} - FE^{(1)}_{i,t}}{TA_{i,t}} \times \frac{1}{2} \tag{2}$$

We validate our LFG measure by using the US data and demonstrating that LFG is highly correlated with LTG with a rank correlation of approximately 57%. In addition, we replicate the asset pricing results reported by La Porta (1996) and Bordalo et al. (2019) using both LFG and LTG and present them in Fig. A.1 in the Appendix. The top (bottom) LFG and LTG decile portfolios both earn monthly CAPM alphas of approximately -0.4% (0.2%), and the long-short differences at approximately 0.6% are statistically significant at the 5% level. These results suggest that our LFG largely captures the economic contents of LTG in the US market. Thus, the differences in the findings in China are not driven by our different long-horizon growth forecast measures.

We construct other analyst-related variables as follows. (1) Standardized unexpected earnings (SUE) is the actual earnings of the current year ($AE^{(1)}$) minus the analyst consensus forecast earnings ($FE^{(1)}$), scaled by total assets (TA). (2) Analyst coverage (ANUM) is the logarithm of the number of institutions that released earnings-per-share forecasts for the firm in the past 180 days. (3) Analysts' forecast dispersion (FDP) is the standard deviation of forecasts issued in the past 180 days, scaled by total assets (Diether et al., 2002). (4) Analyst's forecast revision (FREV) is defined as the sum of the monthly consensus earnings forecast revision in the past six months (Hawkins et al., 1984; Zhang et al., 2017).

Regarding other firm characteristics, we compute (1) a firm's short-horizon actual earnings growth (SAG) as the difference between the one-year-ahead actual earnings ($AE^{(1)}$) and the most recent actual earnings ($AE^{(0)}$), scaled by total assets (TA). (2) Long-horizon actual earnings growth (LAG) is calculated as the difference between the three-years-ahead actual earnings ($AE^{(3)}$) and one-year-ahead actual earnings ($AE^{(1)}$), scaled by the firm's most recent total assets (TA) and then multiplied by one-half. (3) We estimate the CAPM beta (BETA) of individual stocks using daily returns over the past 12 months (Fama and MacBeth, 1973). (4) Firm size (SIZE)

[2] The CSMAR database also provides analysts' consensus forecast data. However, we choose to use WIND for our main analysis because it computes the consensus using the same methodology as that used in IBES, assigning equal weights to all analysts. Conversely, CSMAR uses a more complex approach, considering factors such as the time since a forecast has been made and the analyst's past accuracy. Nevertheless, our results remain largely consistent when using data from CSMAR.

is the log of the stock's market capitalization. (5) Book-to-market ratio (BM) is the ratio of book equity (total shareholder equity minus the book value of preferred stocks) to market capitalization (Fama and French, 2015). (6) Profitability (ROE) is the firm's earnings divided by book equity. (7) Reversal (REV) is the stock's lagged one-month return, computed as the cumulative return over the past 20 trading days (Jegadeesh and Titman, 1993). (8) Turnover (TO) is the number of shares traded divided by the number of shares outstanding. (9) Illiquidity (ILLIQ) is a stock's Amihud (2002) illiquidity measure. (10) Idiosyncratic volatility (IVOL) is the standard deviation of the Fama-French three-factor model residual from the daily return regression in the previous month (Ang et al., 2006). To mitigate the impact of outliers, we winsorize all continuous variables at the 1st and 99th percentiles. We describe the detailed construction of these variables in Appendix A.

2.3. Summary statistics

Table 1 presents the summary statistics for stocks in each of the quintile portfolios formed on *SFG* (Panel A) and *LFG* (Panel B). Columns one and two in the two panels show that *SFG* and *LFG* are positively correlated, suggesting that analysts believe that a firm's earnings growth is positively autocorrelated. Both forecasts are positively correlated with the firm's current profitability (ROE), analyst coverage (ANUM), and forecast earnings revision (FREV). Columns three and four indicate that analysts' growth forecasts are informative about actual future earnings because *SFG* and *LFG* are positively associated with ex-post actual earnings growth (*SAG* and *LAG*). Meanwhile, growth forecasts are negatively associated with earnings surprise (*SUE*), indicating that optimistic forecasts tend to be overly optimistic. Table A.1 in the Appendix presents the summary statistics of these characteristics. Fig. A.2 plots the number of stocks covered by analysts over time.

3. Empirical results

3.1. Analysts' growth forecasts and expected returns

At the end of each month, we construct value-weighted quintile portfolios formed on *SFG* and *LFG* and hold the positions over the next month. Table 2 reports the portfolios' average returns (in excess of the risk-free rate) and factor model alphas. The last column reports the results for long-short portfolios that buy the stocks in the top forecast quintile and sell those in the bottom quintile.

Panel A reports the performance of the portfolios formed on *SFG*. The first row shows that the average monthly excess return increases monotonically from −0.10% to 1.07% from the bottom to the top quintile. The long-short portfolio earns a monthly value-weighted average return of 1.17% with a *t*-value of 2.68. The next three rows show that the results are similar after controlling for the exposures to common asset pricing factors. We consider three models: (1) the CAPM, (2) the China-four-factor (CH4) model (Liu et al., 2019) and (3) the Fama-French-Carhart six-factor (FF-C-6) model (Fama and French (2015) five-factor augmented with the momentum factor (Carhart, 1997)). The alphas of the long-short portfolios range from 0.76% to 1.67% and are all highly statistically significant.

Panel B shows that the same trading rule based on *LFG* yields even higher profits. Compared to the results in Panel A, the average return and the alphas under different pricing models all increase by approximately 50 basis points to between 0.97 and 2.47% per month and are highly significant.[3]

In Fig. 1, we plot the monthly FF-C-6 alphas and the 95% confidence intervals for the long-short portfolios formed on *SFG* or *LFG* over the 12 months after the portfolio construction.[4] The figure shows that the alphas on both portfolios are quite persistent, especially the *LFG* portfolio (Panel B). The *LFG* long-short portfolio alpha remains above 0.5% and statistically significant up to seven months after portfolio formation.

These results show that analysts' earnings growth forecasts significantly and positively predict abnormal returns. This predictive power is stronger for longer-horizon forecasts. These findings contrast with those regarding analysts' long-term growth forecasts *LTG* in the US market, but are consistent with the interpretation that Chinese investors largely neglect analysts' forecasts, and this neglect is more pronounced for longer-horizon forecasts.

We further examine the asset pricing implications of *SFG* and *LFG* using Fama and MacBeth (1973) regressions. Specifically, each month, we estimate cross-sectional regressions of excess stock returns on lagged values of *SFG*, *LFG*, and control variables:

$$Ret_{i,t} = \alpha_{t-1} + \gamma_{t-1} FG_{i,t-1} + \boldsymbol{\phi}_{t-1}^{'} \boldsymbol{X}_{i,t-1} + \epsilon_{i,t}, \tag{3}$$

where $\text{Ret}_{i,t}$ is the monthly excess return on stock *i* in month *t*; $FG_{i,t-1}$ is analysts' growth forecast (i.e., *SFG* or *LFG*); $\boldsymbol{X}_{i,t-1}$ is a set of lagged firm-level controls.

Table 3 reports the time-series averages of γ and Newey and West (1987) adjusted *t*-values in parentheses. In Panel A columns one and two, we report univariate regression results for *SFG* and *LFG*. The coefficients for both variables are all positive and significant: 8.92 for *SFG* and 18.31 for *LFG*, with *t*-values of 3.26 and 3.11. In column three, we simultaneously include *SFG* and *LFG* on the right-hand-side. The coefficients of both variables remain positive and significant, but the magnitude of the coefficient on *SFG* decreases by half. This result suggests that, although *SFG* and *LFG* contain independent information, the predictive power mostly stems from

[3] These alphas seem exceedingly large relative to the typical anomaly in the US market, but are of similar magnitudes to those in Liu et al. (2019) and other China-focused studies. These studies attribute the larger alphas in China to short-sale constrain and uninformed trading.

[4] The numerical values are shown in Table A.2.

Table 1
Descriptive statistics.

	SFG (%)	LFG (%)	SAG (%)	LAG (%)	SUE (%)	SIZE	BETA	BM	ROE	ANUM	FREV (%)
Panel A: SFG portfolios (monthly average $N = 1515$)											
1	−0.69	1.40	−1.64	0.48	−1.29	15.99	1.08	0.60	3.29	1.50	−1.72
2	0.83	1.38	−0.13	0.23	−1.07	15.93	1.08	0.51	4.61	1.63	−0.89
3	1.69	2.08	0.32	0.31	−1.41	15.82	1.07	0.38	5.66	1.76	−0.72
4	2.89	2.98	1.02	0.32	−1.85	15.84	1.06	0.30	6.59	1.87	−0.67
5	7.67	4.80	3.95	0.49	−3.21	15.97	1.04	0.23	8.44	1.83	−0.62
H–L	8.36	3.40	5.58	0.01	−1.92	−0.01	−0.04	−0.37	5.14	0.33	1.11
Panel B: LFG portfolios (monthly average $N = 1345$)											
1	1.14	0.40	0.16	0.10	−0.92	16.11	1.08	0.69	4.28	1.53	−0.92
2	1.25	1.12	0.11	0.17	−1.15	15.96	1.09	0.50	4.76	1.70	−1.01
3	1.86	1.89	0.35	0.20	−1.54	15.87	1.08	0.37	5.69	1.81	−0.92
4	2.58	2.98	0.75	0.31	−1.90	15.86	1.06	0.29	6.55	1.91	−0.89
5	5.08	6.25	2.42	1.05	−2.70	16.05	1.03	0.18	8.47	2.01	−0.87
H–L	3.94	5.85	2.26	0.95	−1.77	−0.06	−0.05	−0.50	4.19	0.48	0.06

This table presents the average stock characteristics for quintile portfolios formed on *SFG* (Panel A) and *LFG* (Panel B). The characteristics are short-horizon actual growth (*SAG*), long-horizon actual growth (*LAG*), standardized unexpected earnings (*SUE*), size (SIZE), market beta (BETA), book-to-market equity (BM), return on equity (ROE), number of analysts tracking (ANUM), and analysts' forecast revision (FREV). Detailed construction of these variables is described in Appendix A. All of the characteristics are cross-sectionally winsorized at the 1st and 99th percentiles. The sample period spans from January 2010 to December 2020.

Table 2
Univariate portfolio analysis.

Panel A. Portfolios formed on *SFG*						
	1 (L)	2	3	4	5 (H)	H − L
Excess return	−0.10	0.37	0.68	1.07	1.07	1.17
	(−0.16)	(0.58)	(1.01)	(1.50)	(1.51)	(2.68)
CAPM-α	−0.38	0.09	0.37	0.76	0.77	1.15
	(−2.64)	(0.60)	(3.00)	(2.98)	(2.90)	(3.23)
CH4-α	−0.51	−0.14	0.29	1.06	1.16	1.67
	(−2.58)	(−0.87)	(1.80)	(2.67)	(3.70)	(3.54)
FFC6-α	−0.27	−0.03	0.08	0.36	0.49	0.76
	(−1.88)	(−0.32)	(0.68)	(2.43)	(3.55)	(3.62)
Panel B. Portfolio formed on *LFG*						
	1 (L)	2	3	4	5 (H)	H − L
Excess return	−0.17	0.25	0.70	0.90	1.46	1.63
	(−0.28)	(0.38)	(0.98)	(1.29)	(1.97)	(2.94)
CAPM-α	−0.43	−0.05	0.38	0.60	1.17	1.60
	(−1.83)	(−0.50)	(2.26)	(2.30)	(3.37)	(2.92)
CH4-α	−0.80	−0.13	0.50	0.91	1.66	2.47
	(−2.88)	(−0.93)	(2.27)	(3.20)	(3.65)	(3.48)
FFC6-α	−0.30	−0.05	0.08	0.31	0.68	0.97
	(−3.12)	(−0.42)	(0.88)	(1.93)	(3.83)	(4.13)

This table presents the monthly value-weighted average excess returns (returns minus the one-year savings interest rates) and alphas on quintile portfolios formed on *SFG* (Panel A) and *LFG* (Panel B). The portfolios are rebalanced monthly. We use three pricing models: the CAPM, China-four-factor model (Liu et al., 2019, CH4), and Fama-French-Carhart six-factor model (Fama and French (2015) five-factor augmented with the momentum factor (Carhart, 1997)). Standard errors are Newey and West (1987) adjusted with 12 lags. *t*-values are reported in paratheses.

investors' neglect of the information about longer-horizon growth.

In column four, we include well-known stock return predictors as controls: market beta (BETA), market capitalization (SIZE), book-to-market ratio (BM), profitability (ROE), short-term reversal (REV), share turnover (TO), idiosyncratic volatility (IVOL), and Amihud (2002) illiquidity (ILLIQ). We also control for various analyst-related variables, including analysts' forecast revision (FREV) and forecast dispersion (FDP). The coefficients on these control variables align with those in the prior literature and shown in Table A.3. More importantly, including these controls does not affect the coefficients on *SFG* and *LFG*.

Liu et al. (2019) shows that the valuation of the smallest 30% firms in China is contaminated by their "shell value." This could be a

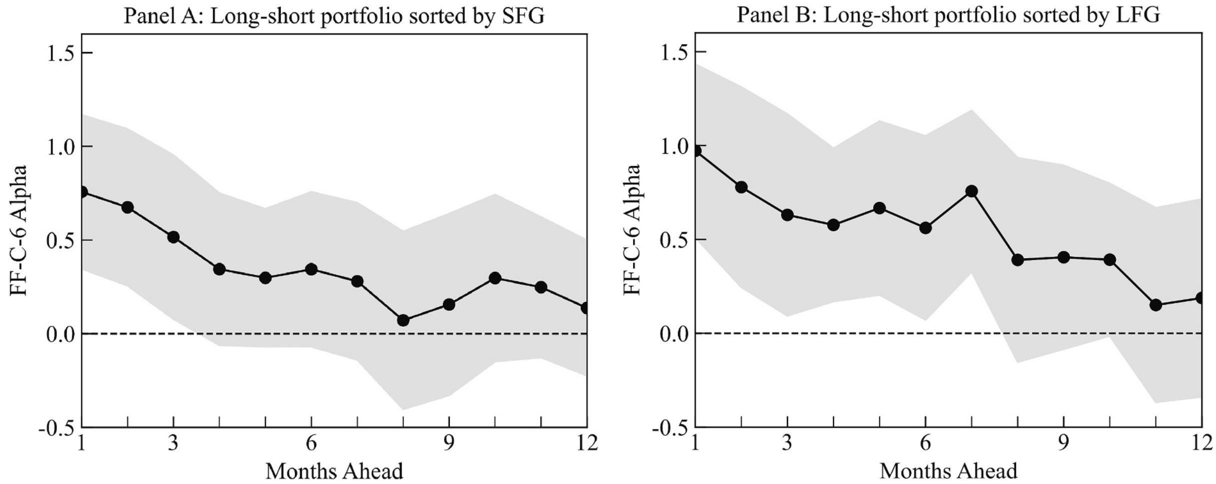

Fig. 1. Monthly alphas of *SFG* and *LFG* long-short portfolios. This figure plots the monthly value-weighted FF-C-6 alphas in the 12 months after portfolio formation and 95% confidence intervals for long-short portfolios formed on *SFG* and *LFG*. Standard errors are Newey and West (1987) adjusted with 12 lags.

Table 3
Fama-MacBeth regression.

	Full sample				Excluding micro-caps			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SFG	8.92		4.03	4.52	9.86		3.34	3.34
	(3.26)		(2.15)	(2.74)	(2.98)		(1.57)	(1.95)
LFG		18.31	15.42	16.01		19.88	17.52	17.42
		(3.11)	(2.61)	(5.14)		(2.88)	(2.58)	(4.27)
Controls	NO	NO	NO	YES	NO	NO	NO	YES
N	174,669	174,669	174,669	172,669	126,633	126,633	126,633	126,633
Adj.R^2	0.65%	1.60%	1.83%	11.03%	0.91%	2.11%	2.43%	12.12%

This table presents the time-series averages of the slope coefficients and the associated *t*-values from the Fama-MacBeth regressions of monthly excess stock returns on *SFG* and *LFG* and control variables: $Ret_{i,t} = \alpha_{0,t-1} + \gamma_{t-1}FG_{i,t-1} + \boldsymbol{\phi}'_{t-1}\boldsymbol{X}_{i,t-1} + \epsilon_{i,t}$, where $Ret_{i,t}$ is the monthly excess return for stock *i* observed in month *t*; $FG_{i,t-1}$ is the analysts' growth forecast, *SFG* or *LFG*; $\boldsymbol{X}_{i,t-1}$ is a set of firm-level controls, including market beta (BETA), market capitalization (SIZE), book-to-market ratio (BM), profitability (ROE), short-term reversal (REV), share turnover (TO), idiosyncratic volatility (IVOL), Amihud (2002) illiquidity (ILLIQ), analysts' forecast revision (FREV), and analysts' forecast dispersion (FDP). The coefficients on these control variables are listed in Table A.3. The *t*-values are computed using standard errors that are Newey-West adjusted with 12 lags. The last two rows show the sample size and the average adjusted R-squared.

concern as a firm's shell value may not be related to its earnings, leading to a misalignment between earnings forecasts and market valuation. We test whether our results still hold after removing the micro-cap stocks and report the results in columns five to eight.[5] The results remain virtually unchanged and, if anything, become slightly stronger.

We conduct a batch of robustness tests and present the results in Table 4. The results indicate that our findings are stable across time (Panel A), not driven by low-price stocks (Panel B), not sensitive to the choice of the scaler when measuring the forecasts (Panel C), and continue to hold after industry-adjusting the forecasts (Panel D).

While the *SFG* strategy's profitability has experienced some decay after 2016, as indicated by the smaller regression coefficient (4.37) in the post-2016 period, the *LFG* strategy remains highly profitable across the two sample periods. This result suggests that even though investors may have become more attentive to analysts' short-horizon forecasts over time, longer-horizon forecasts still escape investors' attention. In Fig. 2, we plot the value of a one-dollar investment in the *LFG* long-short portfolio at the beginning of our sample period. The figure shows that the strategy performs consistently, and the one-dollar investment turns into about six dollars by the end of the period.

Thomas et al. (2023) find that stock price momentum (Jegadeesh and Titman, 1993) exists outside the month of February in China. Their results suggest that the stock return seasonal reversal in February can mask an underreaction-type anomaly such as price momentum and ours. In robustness tests, we find that our portfolio earns a sizable average return of 4% in February, suggesting our results

[5] We also apply weighted least square regressions, where weights are given by the stock's market capital- ization, and the results are similar.

Table 4
Robustness: Fama-MacBeth regressions.

			(1)	(2)	(3)
Panel A: Subperiods	$\leq$ 2015	*SFG*	11.53		7.29
			(5.67)		(3.46)
		LFG		8.41	6.27
				(6.06)	(4.65)
	$\geq$ 2016	*SFG*	4.37		1.25
			(2.50)		(0.63)
		LFG		10.47	10.06
				(4.01)	(3.75)
Panel B: Exclude low-priced stocks	Price $\geq$ 5	*SFG*	8.24		4.29
			(5.13)		(2.68)
		LFG		9.33	8.06
				(7.05)	(5.65)
Panel C: Different scaler	$\mid AE^{(0)} \mid$	*SFG*	7.28		3.22
			(3.84)		(1.16)
		LFG		6.41	5.64
				(3.28)	(2.26)
	ME	*SFG*	9.54		7.99
			(2.85)		(2.67)
		LFG		14.78	12.65
				(4.88)	(4.69)
Panel D: Industry-adjusted		*SFG*	4.67		2.99
			(2.62)		(1.76)
		LFG		4.98	3.96
				(4.22)	(3.50)

The table presents the robustness test results with Fama-MacBeth regressions of excess return on *SFG*, *LFG* and the full set of controls as in Table 3. Panel A divides the sample into two halves. Panel B removes stocks with prices below 5 RMB. Panel C redefines *SFG* and *LFG* by using the absolute value of the previous actual earnings or the market value of equity as the scaling variable. Panel D adjusts the *SFG* and *LFG* by subtracting the cross-sectional industry averages. We report the average slopes and *t*-values computed using Newey and West (1987) adjusted standard errors with 12 lags.

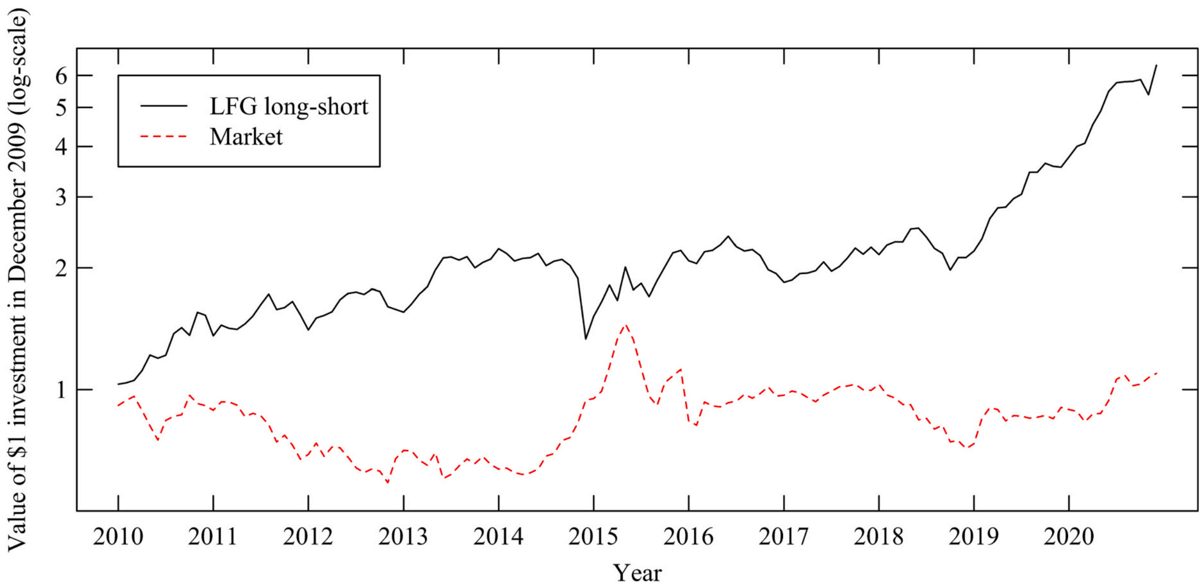

Fig. 2. Performance of *LFG* long-short portfolio. This figure plots the value of investing $1 in a portfolio that buys stocks in the top *LFG* quintile and sells stocks in the bottom quintile at the end of December 2009 (solid line). The red dashed line shows the value of a $1 investment (borrowed at the risk-free rate) in the market portfolio. The sample period spans from January 2010 to December 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are less affected by the February reversal.

3.2. Fundamental information and biases in analysts' forecasts

Fig. 3 visualizes our core evidence on the decoupling of analysts' forecasts and investors' expectations. We plot the average market-adjusted returns (Panel A), the average earnings surprises relative to analysts' consensus forecasts (Panel B), and the average consensus forecast revisions (Panel C) by quintiles of analysts' long-horizon earnings growth forecasts (*LFG*). To ensure the comparability of the three panels, we measure the returns, earnings surprises, and forecast revisions around firms' annual earnings announcement months.

Panel A shows that the average abnormal stock return monotonically increases with *LFG*. In contrast, Panel B shows a strong negative association between earnings surprise and *LFG*. That is, firms with higher growth forecasts suffer from larger negative earnings surprises.[6] Panel C shows the same negative association between growth forecasts and analysts' forecast revisions. The sharp contrast between the pattern in Panel A and those in Panels B and C dramatizes our main point that the market and analysts have divergent perspectives — what seems to be a pleasant surprise for the investors of high *LFG* firms appears to be a big disappointment for the analysts who cover them.

The patterns in the return space (Panel A) and those in the forecast space (Panels B and C) are difficult to reconcile if one assumes analysts' expectations are good proxies for investors' expectations in China. Our proposed interpretation of these results is that investors fail to price in the valuable information in analysts' forecasts, but at the same time, these forecasts are predictably biased.

If analysts' forecasts are both informative and biased, rational investors would need to filter out the biases when incorporating these forecasts into their beliefs. The optimal filtering rule depends on the structure of the bias. We estimate a possible and parsimonious bias structure in which the bias is proportional to the objective expected growth (proxied by the ex-post realized growth). This model is motivated by the findings by La Porta (1996) and Bordalo et al. (2019), which show that analysts' high growth forecasts tend to be too high and vice versa.

Our estimation approach is straightforward. We regress actual earnings growth on the growth forecasts over the same horizon. We include time and industry fixed effects to absorb the impacts of unobserved time-varying macro shocks and persistent industry shocks on our estimates.[7] Our specification is as follows:

$$AG_{i,t} = \alpha + \beta FG_{i,t} + \text{Year}_t + \text{Industry}_i + \epsilon_{i,t}, \tag{4}$$

where $AG_{i,t}$ denotes the actual earnings growth at the short and long horizons, $AG \in \{SAG, LAG\}$; $FG_{i,t}$ is the growth forecast, $FG \in \{SFG, LFG\}$. $Year_t$ and $Industry_i$ are the year and industry fixed effects. $\epsilon_{i,t}$ is the error term. If analysts' growth forecasts contain information about future earnings growth, we expect β to be positive. If extreme forecasts are also on average too extreme, we expect that $0 < \beta < 1$.

Panel A of Table 5 presents our results. As expected, the coefficients on *SFG* and *LFG* are significantly positive, but below one. Column one shows that the coefficient on *SFG* is quite high, at 0.90, which means that a one-unit increase in the growth forecast is, on average, associated with a 0.9-unit increase in the actual growth. This large coefficient, coupled with the high R^2 of 65.93%, suggests that the growth forecast for the current fiscal year is highly informative. Column two shows that the long-horizon growth forecast, *LFG*, also predicts short-horizon growth, with a coefficient of 0.41. This association, however, appears to stem entirely from the correlation between *SFG* and *LFG* as suggested by the small and negative coefficient on *LFG* in column three.

Columns four through six repeat the exercise but replace the dependent variable by *LAG*, the actual earnings growth from one- to three-years-ahead. This horizon corresponds to the forecast horizon of *LFG*. The near-zero coefficient in column four suggests that short-horizon forecasts are not informative about longer-horizon growth. The long-horizon growth forecasts, on the other hand, are informative about *LAG*, as indicated by the highly significant coefficients of 0.29 and 0.37 in columns five and six. However, these coefficients are also substantially below one — a one-unit increase in the growth forecast is only associated with about one-third of a unit increase in the actual growth over the period. This result is consistent with existing findings in the US market (Bordalo et al., 2019) and suggests that analysts' forecasts in China are subject to similar predictable distortions. Our estimates suggest that a way to filter out such distortions is to "discount" the short-horizon growth forecasts by a factor of about 0.90 and the long-horizon forecast by a factor of roughly 0.29.

In Panel B, we regress subsequent earnings surprise and forecast revision on lagged analysts' forecasts to show that the biases in analysts' forecasts are indeed strongly predictable by *SFG* and *LFG* (as shown in Fig. 3). The results show that firms with optimistic forecasts are more likely to have disappointing earnings in the eyes of analysts, as indicated by the significantly negative coefficients in columns one through three. These firms are also more likely to receive downward forecast revisions as shown in columns four through six. These results are consistent with those in Panel A that extreme forecasts are, on average, too extreme.

[6] The average earnings surprises are negative in all *LFG* quintiles. This pattern suggests that Chinese analysts are overwhelmingly overoptimistic about firms' earnings, which is consistent with prior studies such as Xu et al. (2012).

[7] Excluding these fixed effects do not materially change the results.

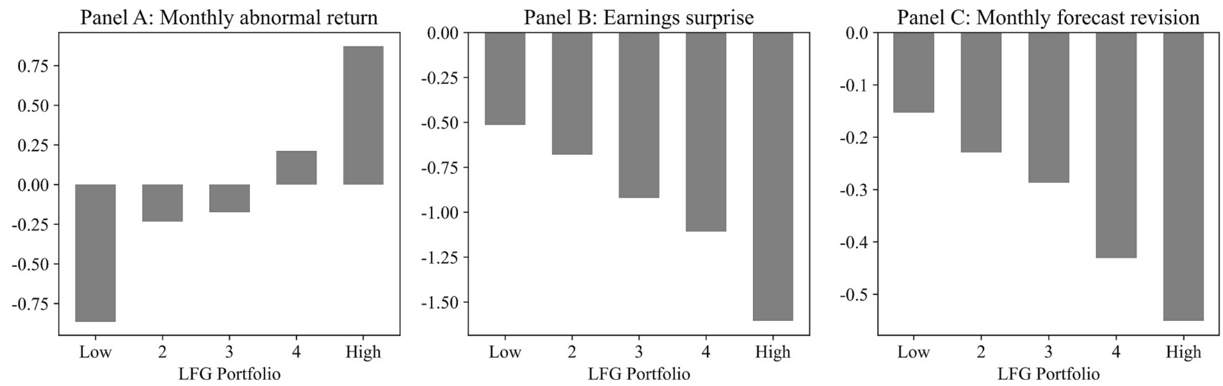

Fig. 3. Growth forecasts, stock returns, earnings surprises, and forecast revisions. This figure plots the average market-adjusted returns (Panel A), earnings surprises (Panel B), and forecast revisions (Panel C). All of the values are measured during annual earning announcement months. Earnings surprises are calculated as actual earnings minus analysts' consensus (mean) forecasts, scaled by lagged total assets. Forecast revisions are computed as changes in analysts' forecasts from before to after announcements, scaled by lagged total assets. Stocks are sorted monthly into quintiles by analysts' long-horizon growth forecasts, *LFG*, defined in Section 2. All values are in percentages.

Table 5
Information and predictable biases in analysts' growth forecasts.

Panel A. Predict actual earnings growth

	SAG			*LAG*		
	(1)	(2)	(3)	(4)	(5)	(6)
SFG	0.90		0.96	−0.03		−0.12
	(50.90)		(54.51)	(−0.65)		(−4.11)
LFG		0.41	−0.18		0.29	0.37
		(12.19)	(−10.18)		(6.83)	(9.49)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
N	14,751	14,751	14,751	14,751	14,751	14,751
Adj. R^2	65.93%	6.00%	66.88%	0.04%	2.59%	3.45%

Panel B. Predict earnings surprises and forecast revisions

	SAG			*LAG*		
	(1)	(2)	(3)	(4)	(5)	(6)
SFG	−0.09		−0.05	−0.67		0.75
	(−4.81)		(−2.68)	(−1.22)		(1.00)
LFG		−0.17	−0.14		−4.61	−5.08
		(−7.97)	(−7.62)		(−5.78)	(−5.18)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
N	14,751	14,751	14,751	13,785	13,785	13,785
Adj. R^2	1.88%	3.23%	3.71%	0.05%	1.10%	1.15%

This table reports the coefficients and associated *t*-values from the panel regressions of the actual earnings growth (Panel A), earnings surprises (Panel B), and forecast revisions (Panel B) on analysts' growth forecasts using yearly data from 2010 to 2020. The sample includes firm-year observations around annual earnings announcements. The specification in Panel A regression is:$AG_{i,t} = \beta_{1,t} + \beta_{2,t}FG_{i,t} + Industry_i + Year_t + \epsilon_{i,t}$, where $AG_{i,t}$ is the actual earnings growth in the current fiscal year ($SAG_{i,t}$) (columns (1)–(3)) or actual average earnings growth in the second and third years ($LAG_{i,t}$) (columns (4)–(6)). $FG_{i,t}$is analysts' growth forecasts, $SFG_{i,t}$ or $LFG_{i,t}$, measured before the annual earnings announcements. In Panel B, the dependent variables are standardized unexpected earnings (*SUE*) (columns (1)–(3)) and forecast revisions (columns (4)–(6)). We define forecast revisions as the difference between consensus earnings forecasts for $FY(1)$ in annual announcement month t, $FE^{(1)}$, and analysts' $FY(2)$ forecasts in previous month $FE^{(2)}_{i,t-1}$, scaled by total assets. All specifications include year fixed effects and industry fixed effects. Standard errors are two-way clustered by year and industry.

3.3. Mechanisms

3.3.1. Ex-post earnings growth

In this subsection, we further investigate the source of the return predictability by analysts' forecasts. Earnings announcements provide a powerful setting for identifying potential drivers of market surprises, as changes in beliefs surrounding earnings results can strongly influence prices during this period. We focus on the 11-day window around earnings announcements, which is wider than usual due to evidence of information leakage prior to earnings announcements in China (Xu, 2021). We define the cumulative abnormal return (*CAR*) as the stock return in excess of the market return from $t = -5$ to $t = 5$ relative to earnings announcements.[8] We then regress CAR on *SFG* and *LFG*, and subsequently include additional variables to identify the underlying mechanisms.

Table 6 presents our results. In columns one and two, the coefficients on the *SFG* and *LFG* are 4.69 (t-value = 1.93) and 13.80 (t-value = 5.71), confirming that the asset pricing results in the previous sections. In column three, we again observe that the predictive power of *SFG* is largely subsumed by *LFG*. In terms of economic significance, the coefficient on *LFG* implies that the average CAR of the top *LFG* quintile portfolio exceeds that of the bottom quintile portfolio by approximately 80.7($\approx 5.85 \times 13.8$) basis points in the 11-day window (or 18.5% annually). In column four, we include contemporaneous earnings surprise relative to analysts' consensus forecast (*SUE*) as a control variable. We find that *SUE* does not explain the predictive power of *LFG*, which is consistent with *LFG* capturing information beyond the current period earnings.

We test our main proposed mechanisms in columns five to seven by including the ex-post realized earnings growth, *SAG* and *LAG*, on the right-hand-side. *SAG*, like *SUE*, is known during the earnings announcement. On the other hand, *LAG* is unknown during the announcement as it is the average actual earnings growth starting from the current new fiscal year to two years ahead. If the predictive power of the forecasts stems from investors neglecting the informative signals in the forecasts, controlling for the ex-post realization of the earnings growth should largely "explain away" the predictive power of the forecasts. Here, we rely on the assumption that the ex-post earnings growth, *LAG*, is a good proxy for the ex-ante informative signal contained in *LFG*.

The results show that after including the actual growth as explanatory variables, the coefficient on *SFG* turns negative (column five) and the coefficient on *LFG* becomes insignificant (column seven). Column six shows that *SAG* only explains a small portion of the predictive power of *LFG*, which is consistent with the information contained in *LFG* being about the longer horizons. The results in Table 6 are consistent with our proposed mechanism that investors neglect the signals (the parts that correlate with future realized growth) in analysts' forecasts.

3.3.2. Forecast informativeness

We explore a source of heterogeneity — forecast informativeness — in this subsection. If investors' neglect of analysts' forecasts is the mechanism that drives our results, we expect the association between analysts' forecasts and subsequent stock returns to be stronger when these forecasts are (ex-ante) more informative. Consider an extreme scenario in which analyst forecasts are purely random noise; even if investors ignore such forecasts, there would be no impact on stock returns. In contrast, if analysts' forecasts perfectly predict realized earnings growth and investors ignore these forecasts, then these forecasts will predict large abnormal returns going forward. We further develop this argument with a model in the next section. To measure forecast informativeness, we employ three proxies.

(i) Past forecast accuracy (FACC), which is defined as the negative absolute value of the percentage forecast error for the last earnings:

$$FACC_{i,t} = -\frac{\left|FE^{(1)}_{i,t-1} - AE^{(0)}_{i,t}\right|}{\left|AE^{(0)}_{i,t}\right|}, \tag{5}$$

where $FE^{(1)}_{i,t-1}$ is the consensus one-year-ahead forecast measured in the previous fiscal year t–1. $AE^{(0)}_{i,t}$ is the corresponding actual earnings.

(ii) Analyst coverage (ANUM), which is defined as the log of the number of analysts covering the stock in the last 180 days.
(iii) Institutional ownership ratio (INST), which is computed from firms' most recent quarterly financial statements and institutional holdings data. We download this variable directly from CSMAR.

We expect analysts' consensus forecasts to be more informative if they have been more accurate in the past, if more analysts are producing forecasts (so the consensus is an average across more individual forecasts), and if institutional ownership of the stock is higher (so that analysts have stronger incentives to produce more accurate forecasts).

We test the effects of forecast informativeness using a portfolio approach. At the end of each month, we sort stocks into three groups by the 30th and 70th percentiles of informativeness, and into quintiles by *SFG* or *LFG*. In Table 7, we present the FF-C-6 value-weighted alphas for the long-short portfolios that buy stocks with optimistic forecasts and sell those with pessimistic forecasts within each forecast informativeness group. The H − L columns report the alphas of the return differences between the "High" and "Low" long-

[8] We identify earnings announcement dates as the dates when the firms file their financial statements to the SEC in China.

Table 6
Earnings announcement returns.

	$CAR[-5,5]$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SFG	4.69		1.14	2.66	−17.90		
	(1.93)		(0.42)	(1.01)	(−4.49)		
LFG		13.80	13.10	17.55		9.45	3.50
		(5.71)	(5.04)	(6.26)		(4.23)	(1.29)
SUE				31.84			
				(8.11)			
SAG					24.90	10.53	13.01
					(7.13)	(5.13)	(6.62)
LAG							8.43
							(5.81)
Industry FE	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES
N	14,751	14,751	14,751	14,751	14,751	14,751	14,751
Adj. R^2	0.04%	0.15%	0.15%	0.82%	0.51%	0.38%	1.06%

This table reports the coefficients and associated *t*-values from the panel regressions of the earning announcement period cumulative abnormal return (CAR) on *SFG, LFG, SUE, SAG* and *LAG* as defined in previous tables. The dependent variable, $CAR[5,+5]_{i,t}$, is calculated as the stock return minus the market return in the 2 weeks around the announcement for stock *i* in year *t*. All specifications include year fixed effects and industry fixed effects. Standard errors are two-way clustered by year and industry.

Table 7
Forecast informativeness: bivariate portfolio analysis.

	LS portfolios on *SFG*			LS portfolios on *LFG*		
Informativeness	Low	High	H − L	Low	High	H − L
FACC	0.23	1.20	0.97	0.15	1.36	1.21
	(0.91)	(3.66)	(2.96)	(0.51)	(3.54)	(3.45)
ANUM	0.20	1.06	0.86	0.06	1.31	1.26
	(1.11)	(3.54)	(2.80)	(0.27)	(4.63)	(3.90)
INST	0.34	1.25	0.91	0.10	1.38	1.28
	(1.10)	(5.04)	(2.35)	(0.16)	(3.97)	(2.20)

At the end of each month, we sort stocks into three groups by three measures of forecast informativeness: past forecast accuracy (FACC), analyst coverage (ANUM), or institutional ownership (INST) by the 30th and 70th percentiles. We then independently sort stocks into quintiles by *SFG* or *LFG*. This table reports the FFC6-alphas (*t*-values in parentheses) on the long-short portfolios that buy stocks with high forecasts (i.e., *SFG* or *LFG*) and sell those with low forecasts among stocks with low or high forecast informativeness. The H − L columns report the alphas of the return differences between the long-short portfolios that contain the high and low forecast informativeness stocks.

short portfolios.

The results in Table 7 indicate that for stocks with low forecast informativeness, the alphas on both the *SFG* and *LFG* strategies are insignificant. The magnitudes are small and range from 0.06 to 0.34% per month. In contrast, both strategies prove highly profitable when applied to stocks with high forecast informativeness. The alphas range from 1.06 to 1.38%, all of which are statistically significant. Furthermore, the differences between the alphas on the high and low informativeness groups across all three measures (ranging from 0.86 to 1.51) are large and highly significant. These results further support our proposed mechanism that investors largely neglect analysts' informative forecasts.

3.3.3. Public information, private information, and bias

Prior research, such as So (2013) and van Binsbergen et al. (2023), has proposed statistical models to compute the rational earnings forecast benchmarks. They show that analysts' deviations from these benchmarks likely reflect predictable biases which investors fail to filter. Consequently, trading rules that bet on the correction of these biases are profitable.

This pattern may also be flipped in China. Conceptually, analysts' forecasts contain four parts: (1) public information, (2) private information, (3) bias, and (4) noise.[9] We capture the public information component using statistical benchmarks and compute analysts' deviations from these benchmarks. These deviations reflect the sum of analysts' private signals and biases. The asset pricing implications of these forecast deviations depend on the relative amount of unpriced private information and unpriced biases. For example, if investors take analysts' forecasts at face value, given that these forecasts are often too extreme, one should observe that

[9] See de Silva and Thesmar (2021) for a fully developed framework.

analysts' deviation from statistical benchmarks predicts subsequent returns negatively, as in the US market. On the other extreme, if investors completely ignore analysts' forecasts, then neither the private information nor the biases are reflected in prices. In this case, analysts' deviations should positively predict subsequent returns.

We follow the methodologies of So (2013) and van Binsbergen et al. (2023) to construct two statistical benchmarks. We recursively train an OLS and a random forest model to predict firms' realized earnings using firm characteristics. The OLS model is:

$$\begin{aligned} E_{i,t} \quad &= \beta_{0,n} + \beta_{1,n}E^{+}_{i,t-n} + \beta_{2,n}NEGE_{i,t-n} + \beta_{3,n}ACC^{+}_{i,t-n} + \beta_{4,n}ACC^{-}_{i,t-n} + \beta_{5,n}AG_{i,t-n} \\ &+\beta_{6,n}\ \text{Dividend}_{i,t-n} + \beta_{7,n}DD_{i,t-n} + \beta_{8,n}BM_{\text{i},t-n} + \epsilon_{i,t}, \text{for } n \in \{1,2,3\}, \end{aligned} \tag{6}$$

where the dependent variable is a firm's earnings scaled by the total asset (E_i). On the right-hand-side are the lagged firm characteristics in year $t-n$: earnings scaled by total asset when earnings are positive and zero otherwise (E^+_i); a dummy variable that equals one if earnings are negative ($NEGE_i$); positive and negative accruals scaled by the total asset (ACC^+_i, ACC^-_i) where accruals equal the operating profit minus operating cash flow; the percent change in total assets (AG_i); dividends scaled by the total assets ($Dividend_i$); a dummy variable that equals one if the firm pays no dividend (DD_i); book-to-market ratio (BM_i). After obtaining the coefficients,[10] we use the model to predict the one-year-ahead, two-years-ahead, and three-years-ahead earnings using firms' latest financial data.

We train the random forest model using a similar procedure, but the number of variables expands to 28 plus 12 month-of-the-year dummies and 27 industry dummies. A complete list of these variables can be found in Table A.5 in the Appendix.

Table 8 reports the model performance. The first column shows that the average Pearson correlation between one-year-ahead OLS earnings forecasts and the actual earnings is quite high, at 0.74. The random forest model performs even better, with a correlation of 0.80. The correlation between analysts' forecasts and the actual earnings is 0.75. As expected, these correlations decrease with the forecast horizon because outcomes in more distant futures are less predictable. The model-based forecasts are highly correlated with analysts' forecasts (column two) and are statistically unbiased while analysts' forecasts exhibit a significant optimism bias (columns three through five).

With these model-based benchmarks, we compute analysts' deviation from them, the short-horizon forecast deviation (*SFD*) and long-horizon deviation (*LFD*) as follows:

$$\begin{aligned} &SFD_{i,t} = FE^{(1)}_{ana,i,t} - FE^{(1)}_{bm,i,t}, \\ &LFD_{i,t} = \left(FE^{(3)}_{ana,i,t} - FE^{(1)}_{ana,i,t}\right) - \left(FE^{(3)}_{bm,i,t} - FE^{(1)}_{bm,i,t}\right), \end{aligned} \tag{7}$$

where $FE^{(h)}_{ana,i,t}$ is analysts' consensus earnings forecast for stock i in month t at horizon h, scaled by total assets and $FE^{(h)}_{bm,i,t}$ is the associated statistical benchmark using the OLS or the random forest model. We then examine the cross-sectional relationship between these deviations and stock returns using Fama and MacBeth (1973) regressions.[11]

Table 9 presents the results. Columns one through four show that analysts' deviations from the benchmarks *positively* predict returns, especially for the long-horizon forecasts. These results, again, contrast with those in the US market but are consistent with analysts' forecasts in China containing valuable private information which investors appear to neglect.

4. Expectation formation with analysts: a model

Two channels potentially explain why analysts' information is not fully incorporated into prices: (1) investors may be inattentive to analysts' forecasts due to information friction, and (2) even when investors observe analysts' forecasts, they "shrink" the forecasts too much due to distrust. These two channels have different policy implications. For example, if investors' underweighting of analysts' forecasts arises from inattention, measures to increase the salience of analysts' forecasts will improve market efficiency. If investors distrust analysts' forecasts due to skepticism about analysts' incentives, restoring financial institutions' reputations through incentive reforms can improve efficiency. Prior research typically does not attempt to disentangle these two mechanisms. In this section, we introduce a model that accounts for both channels and discuss its implications.

4.1. Baseline case

The model has infinite periods, N firms, one investor, and one analyst. Earnings follow a partially forecastable random walk. The expected earnings of firm i at time $t+1$, $\pi_{i,t+1}$, is

$$\mathbb{E}\left[\pi_{i,t+1}\right] = \pi_{i,t} + \nu_{i,t+1} + \epsilon_{i,t+1}, \tag{8}$$

where $\mathbb{E}[]$ is the mathematical expectation operator. $\nu_{i,t+1} \sim N\left(0,\sigma^2_\nu\right)$ is a random earnings innovation from time t to $t+1$ that the investor and the analyst try to forecast. $\epsilon_{i,t+1} \sim N\left(0,\sigma^2_\varepsilon\right)$ is an unforecastable noise. The analyst produces a forecast $\widehat{\nu}_{A,i,t+1}$ and the

[10] Our baseline results use a three-year look-back window for training. Using different look-back windows does not materially change the results. Table A.4 reports the time-series average coefficients of the models.

[11] Results from portfolio regressions are similar so are not reported to conserve space.

Table 8
Earnings forecasts via statistical models.

		$corr(FE, AE)$	$corr(FE, FE)$	Mean	Mean error	t-value
	FE_{OLS}	0.74	0.79	5.58	−0.11	−1.09
$FY(1)$	FE_{RF}	0.80	0.93	5.99	0.21	1.59
	FE_{ana}	0.75		7.59	1.91	20.60
	FE_{OLS}	0.49	0.70	5.90	−0.29	−0.77
$FY(2)$	FE_{RF}	0.58	0.91	6.60	0.32	1.08
	FE_{ana}	0.57		9.91	3.72	15.96
	FE_{OLS}	0.36	0.63	6.36	−0.29	−0.42
$FY(3)$	FE_{RF}	0.40	0.86	7.04	0.23	0.45
	FE_{ana}	0.39		12.36	5.71	12.13

The first column shows the average Pearson correlations between earnings forecasts (model-based and analysts' forecasts) and actual earnings. Column two shows the average Pearson correlations between model-based forecasts (FE_{OLS} or FE_{RF}) and analysts' forecasts (FE_{ana}). The last three columns present the time-series averages of the mean forecasts, the mean forecast errors (forecasts minus actual earnings), and the corresponding Newey and West (1987) adjusted t-value.

Table 9
Forecast deviation and stock returns.

	OLS			Random forest		
	(1)	(2)	(3)	(4)	(5)	(6)
SFD	4.45		5.24	5.15		6.03
	(1.02)		(1.13)	(1.07)		(1.07)
LFD		10.13	8.26		6.61	6.43
		(2.98)	(2.16)		(3.53)	(3.54)
SFG_{bm}			4.04			5.30
			(1.13)			(1.69)
LFG_{bm}			−10.61			−1.14
			(−1.12)			(−0.40)
N	132,200	132,200	132,200	132,200	132,200	132,200
Adj.R^2	0.25%	0.98%	2.96%	0.32%	0.63%	1.61%

This table presents the time-series averages of the slope coefficients and the associated t-values from Fama-MacBeth regressions of excess stock returns on analysts' forecast bias predicted by statistical models. SFD_{OLS} and LFD_{OLS} are the short- and long-horizon biases implied by the OLS model. SFD_{RF} and LFD_{RF} are the biases implied by the random forest model. The t-values are computed using standard errors that are Newey-West adjusted with 12 lags.

investor generates an own forecast $\widehat{v}_{I,i,t+1}$.

In this baseline scenario, we assume that both forecasts are unbiased. After observing the analyst's forecast, the investor forms the final forecast by weighing $\widehat{v}_{I,i,t+1}$ against $\widehat{v}_{A,i,t+1}$:

$$\mathbb{F}\left[v_{i,t+1}\right] = \beta_I \widehat{v}_{I,i,t+1} + \beta_A \widehat{v}_{A,i,t+1}, \tag{9}$$

where $\mathbb{F}[]$ is the investor's subjective forecast operator. β_I and β_A are the weights. The investor chooses their values by solving an optimization problem. For example, if the investor aims to minimize the mean squared error of the forecast, then the problem becomes estimating a multiple regression model. Namely, the investor minimizes $\mathbb{E}\left[\left(\pi_{i,t+1} - \mathbb{F}\left[\pi_{i,t+1}\right]\right)^2\right]$ by choosing the weights β_I and β_A.

In reality, the investor may not be able to obtain the optimal weights β_I^* and β_A^* due to inefficiencies such as inattention, informational frictions, and various forms of cognitive limitations or behavioral biases. We denote the subjective weights that the investor assigns to the two forecasts as β_I^S and β_A^S. The rational and subjective forecasts are written as:

$$\mathbb{E}\left[v_{i,t+1}\right] = \beta_I^* \widehat{v}_{I,i,t+1} + \beta_A^* \widehat{v}_{A,i,t+1}, \tag{10}$$

$$\mathbb{F}\left[v_{i,t+1}\right] = \beta_I^S \widehat{v}_{I,i,t+1} + \beta_A^S \widehat{v}_{A,i,t+1}. \tag{11}$$

The differences between the optimal and subjective weights, $\beta_I^* - \beta_I^S$ and $\beta_A^* - \beta_A^S$ reflect how much the investor under- or overweights the two forecasts. Taking the difference between the two equations, we have the expected market earnings surprise as:

$$\mathbb{E}\left[v_{i,t+1} - \mathbb{F}\left[v_{i,t+1}\right]\right] = \left(\beta_I^* - \beta_I^S\right)\widehat{v}_{I,i,t+1} + \left(\beta_A^* - \beta_A^S\right)\widehat{v}_{A,i,t+1}. \tag{12}$$

This equation implies that market surprises are predictable by the forecasts if the subjective weights deviate from the optimal weights. For example, if the investor under-weights the analyst's forecasts, or $\beta_A^* - \beta_A^S > 0$, then the analyst's forecasts should positively predict earnings surprises.

4.2. Attention or trust?

In reality, analysts' forecasts are predictably biased. Therefore, investors may need to de-bias these forecasts before incorporating them into their final forecasts. This is similar to a situation where a friend offers suggestions on an issue. To determine whether to follow the suggestions, you would first evaluate whether the friend is biased and, if so, to what extent and in which direction. In this case, you are perfectly aware of the suggestions, so it is not a matter of inattention. However, suboptimal integration of this information can still occur due to excessive trust or distrust toward the suggestions. Thus, *attention* and *trust* are two separate elements, which we jointly model here.

We assume a simple linear de-biasing rule in our model:

$$\overline{v}_A^* = \alpha^* + \zeta^* \widehat{v}_A \quad , \tag{13}$$

$$\overline{v}_A^S = \alpha^S + \zeta^S \widehat{v}_A \quad , \tag{14}$$

where $\overline{v}_A^*$ and are $\overline{v}_A^S$ the optimally and subjectively de-biased analyst's forecasts. α^* optimally corrects for the analyst's unconditional bias. α^S is the subjective correction term for the unconditional bias; ζ^* and ζ^S optimally and subjectively correct for the bias proportional to $\widehat{v}_A$. Existing research suggests that $\alpha^* < 0$ and $0 < \zeta^* < 1$, which means that investors should shift the forecasts down and shrink them toward zero. The rational and subjective earnings growth expectations are equal to:

$$\mathbb{E}\left[v_{i,t+1}\right] = \beta_I^* \widehat{v}_{I,i,t+1} + \beta_A^* \overline{v}_{A,i,t+1}^* , \tag{15}$$

$$\mathbb{F}\left[v_{i,t+1}\right] = \beta_I^S \widehat{v}_{I,i,t+1} + \beta_A^S \overline{v}_{A,i,t+1}^S . \tag{16}$$

The market earnings surprise is now written as:

$$\mathbb{E}\left[v_{i,t+1} - \mathbb{F}\left[v_{i,t+1}\right]\right] = \left(\beta_A^* \alpha^* - \beta_A^S \alpha^S\right) + \left(\beta_I^* - \beta_I^S\right)\widehat{v}_{I,i,t+1} + \left(\beta_A^* \zeta^* - \beta_A^S \zeta^S\right)\widehat{v}_{A,i,t+1}. \tag{17}$$

As is clear from this equation, the earnings surprise to investors depends on the optimality of both the subjective attention weights ($\beta_I^S$and β_A^S) and the de-biasing rule (α^S and ζ^S).

4.3. Implications

Disentangling the inattention and distrust channels is challenging since we cannot observe investors' internal forecast $\widehat{v}_I$ or their final forecast ($\mathbb{F}[v_i]$). Therefore, it is not feasible to undertake structural estimation of this model. We briefly explore the model's reduced-form implications in the cross-section and link them to our empirical results.

If investors are inattentive to analysts' forecasts (i.e., β_A^S is too small), we expect that analysts' forecasts will more strongly and positively predict subsequent stock returns when these forecasts are more informative. This is because the optimal weight on analysts' forecasts, β_A^*, increases with the signal-to-noise ratio of analysts' forecasts. Therefore, all else equal, $\beta_A^* - \beta_A^S$ increases with the informativeness of analysts' forecasts.

In contrast, if investors are fully aware of analysts' forecasts but overly discount them at least initially, we expect the more "trustworthy" forecasts to exhibit diminished ability to predict subsequent returns over subsequent sample periods. This is because, under the assumption of perfect attention, investors can learn from the informativeness of analysts' forecasts. Consequently, forecasts that have been more accurate in the past should receive less discounting in the future, thus reducing their predictability over time. Such learning, however, cannot take place when investors are inattentive to the forecasts to begin with.

We find no support for an adaptive learning story. In particular, the results in Table 4 show that the predictive power of analysts' forecasts persists in the second half of the sample. Forecasts that are more informative ex-ante, which should be more trustworthy in the cross-section, more strongly predict subsequent returns, and more importantly, this pattern also persists in the second half of the sample (see Table A.6). These results suggest that information friction likely plays a more important role than investors' distrust in the price inefficiencies we document.

5. Discussion

We briefly discuss three interpretation issues regarding our results.

5.1. Why do investors in China overlook analyst information?

This phenomenon can be largely attributed to the composition of China's stock market investor base in China. As of the end of 2015, retail investors held 88% of all free-floating shares (see Liu et al., 2019). These retail investors may lack easy access to analyst reports and may have limited financial literacy, as per survey evidence. In addition, statistics show that retail investors in China tend to trade frequently, indicating a greater focus on speculation than on value investing and less willingness to process information from analyst reports (Han and Shi, 2022; Han and Li, 2017). Furthermore, institutional investors may encounter market frictions, such as short-sale constraints, which limit their ability to fully utilize valuable information. Collectively, these factors can explain why some information, such as analysts' forecasts, which is deemed valuable in other markets, is not fully incorporated by investors in China.

5.2. Omitted risk factors

Although we cannot completely rule out the possibility of analysts' forecasts aligning with unobserved risks, we contend that a risk story is unlikely to be the primary mechanism behind our results. Our tests on the mechanism (Table 6) demonstrate that ex-post earnings growth largely explains the predictive power of *SFG* and *LFG*. Therefore, any unknown risk factor must align with ex-post realized growth. Furthermore, our results regarding forecast informativeness suggest that this unknown risk factor affects only a subset of firms with higher analyst coverage, more predictable earnings, and higher institutional ownership, despite these firms typically being perceived as less risky by conventional standards. Moreover, analysts' long-horizon growth forecasts have been shown to negatively predict returns in the US market. Hence, a risk-based explanation must account for this inter-market discrepancy.

5.3. Replicability of the LTG anomaly

The negative relationship between long-term growth (*LTG*) and subsequent stock returns has been criticized for being fragile, model-specific, or confined to small stocks in the US (Bordalo et al., 2019; Hou et al., 2020). In Fig. A.1, we replicate this finding and confirm that the high-minus-low return spreads in the US are only significant in equal-weighted portfolios, which echoes concerns raised by previous studies. Da and Warachka (2011) show that the disparity between *LTG* and short- horizon growth (after industry adjustment) is a more robust negative predictor of returns. Bordalo et al. (2019) also provide a discussion of the robustness of the *LTG* anomaly.

Our framework provides a novel explanation for the difficulties in replicating the *LTG* anomaly. In particular, investors' beliefs need to align excessively with analysts' forecasts for the negative association between *LTG* and stock returns to emerge in the data. Such excessive alignment may not naturally occur in certain market environments, such as those where investors and analysts do not think alike or analysts' information is difficult to acquire, limiting the robustness of the *LTG* anomaly.

6. Conclusion

In this study, we examine the implications of analysts' earnings growth forecasts on asset pricing in China. Our findings diverge from those of previous studies conducted in the US market. Specifically, we find that analysts' growth forecasts have a *positive* predictive power for stock returns in China. Our results suggest that investors' expectations do not align with analysts' forecasts in China. Thus, biases in analysts' forecasts do not appear to distort prices. This is not because investors are efficient in filtering out these biases; rather, it is because these forecasts are largely overlooked from the outset.

Investigating the precise determinants of the alignment between investors' and analysts' beliefs across various markets holds great potential. Although we argue that the composition of market participants plays an important role, other factors, including information barriers, financial literacy, and (dis)trust in financial institutions, may also exert significant effects.

Our findings underscore the importance of comprehending variations in market environments in out-of-sample asset pricing examinations across different markets. The absence of specific anomalies in China, such as the price momentum in its original form, as found in Jegadeesh and Titman (1993), may not necessarily indicate a more efficient market. When faced with results like ours that diverge from existing findings, instead of concluding that the original anomalies fail to hold out-of-sample or that the markets without the anomalies are more efficient, identifying the causes of the disparity can provide deeper insights into the mechanisms underlying the anomalies.

Appendix

A. Variable definitions

- Short-horizon growth forecasts (SFG). For each month *t*, we assess a stock's short- horizon growth forecasts (*SFG*) as the difference between analyst consensus forecast earnings for the current fiscal year ($FE^{(1)}$) and the most recent earnings ($AE^{(0)}$), scaled by firms' most recent total assets (*TA*):

$$SFG_{i,t} = \frac{FE_{i,t}^{(1)} - AE_{i,t}^{(0)}}{TA_{i,t}}. \tag{A.1}$$

- Long-horizon growth forecast (LFG). We calculate long-horizon growth forecast (*LFG*) as the difference the consensus three-year-ahead earnings forecast and the one-year-ahead forecast, scaled by total assets and annualized by multiplying it by one-half:

$$LFG_{i,t} = \frac{FE_{i,t}^{(3)} - FE_{i,t}^{(1)}}{TA_{i,t}} \times \frac{1}{2}. \tag{A.2}$$

- Short-horizon actual earnings' growth (SAG). We measure a stock's short-horizon actual earnings' growth (*SAG*) as the difference between the one-year-ahead actual earnings ($AE^{(1)}$) and the most recent actual earnings ($AE^{(0)}$):

$$SAG_{i,t} = \frac{AE_{i,t}^{(1)} - AE_{i,t}^{(0)}}{TA_{i,t}}. \tag{A.3}$$

- Long-horizon actual earnings' growth (LAG). We measure a stock's long-horizon actual earnings' growth (*LAG*) as the difference between three-years-ahead actual earn- ings ($AE^{(3)}$) and one-year-ahead actual earnings ($AE^{(1)}$), scaled by 2 times the firm's most recent total assets (*TA*):

$$LAG_{i,t} = \frac{AE_{i,t}^{(3)} - AE_{i,t}^{(1)}}{TA_{i,t}} \times \frac{1}{2}. \tag{A.4}$$

- Standardized unexpected earnings (SUE). We calculate standardized unexpected earnings (*SUE*) as the actual earnings of the current year ($AE^{(1)}$) minus the analyst consensus forecast earnings ($FE^{(1)}$), standardized by firms' most recent total assets (*TA*).
- Analyst coverage (ANUM). The logarithm of the number of institutions that have released EPS forecasts for the firm in the past 180 days.
- Analysts' forecasts revision (FREV). Following Hawkins et al. (1984) and Zhang et al. (2017), analysts' forecast revision (FREV) is defined as the sum of the monthly consensus earnings forecast revision in the past six months.
- Analysts' growth forecasts dispersion (FDP). Following Diether et al. (2002), analysts' earnings forecasts dispersion (FDP) is defined as the standard deviation of analyst earnings forecast by different institutions in the past 180 days and divided by the total assets.
- Market Beta (BETA). Following Fama and MacBeth (1973), we estimate the market beta of individual stocks (BETA) using daily returns over the prior 12 months.
- Size (SIZE). The log value of stock's market capitalization (CSMAR item Msmvttl).
- Book-to-market ratio (BM). Following Fama and French (2015), book equity equals total shareholder equity (CSMAR item A003000000) minus the book value of preferred stocks (CSMAR item A003112101). A stock's BM is the ratio of book equity to market capitalization.
- Reversal (REV). Following Jegadeesh and Titman (1993), We define Reversal (REV) as the cumulative return over the past month.
- Profitability (ROE). We define Profitability (ROE) as the ratio of a firm's earnings (CSMAR item B002000000) to book equity (CSMAR item A003000000).
- Turnover (TO). We define Turnover (TO) as the ratio of shares traded at month t to shares outstanding.
- Illiquidity (ILLIQ). Following Amihud (2002), a stock's illiquidity measure for day t is calculated as $ILL_t = |\mathrm{ret}_t| \times 10^6/\mathrm{volume}_t$, where $|\mathrm{ret}_t|$ is the stock's absolute return on day t, and volumet is the stock's dollar trading volume on day t.
- Idiosyncratic volatility (IVOL). Following Ang et al. (2006), we calculate idiosyncratic volatility (IVOL) as the standard deviation of the Fama-French three-factor model residuals from daily return regression in the previous month.
- Institutional ownership ratio (INST). We define INST as the proportion of total shares held by institutional investors.

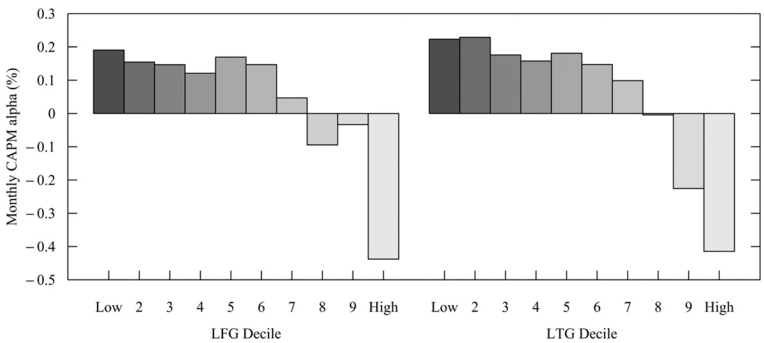

Fig. A.1. Validating *LFG* in the U.S. data. This figure plots the monthly CAPM alphas on the decile portfolios formed on *LFT* (left panel) and *LTG* (right panel). We report equal-weighted results to be comparable to Bordalo et al. (2019). The sample period is from July 1985 to June 2019. *LFG* is as defined in Section 2.

Fig. A.2. Number of stocks covered by analysts over time. This figure plots the number of stocks covered by at least one analyst from 2010 to 2021.

Table A.1
Descriptive statistics.

Variable	Obs	Mean	Std.	P25	P50	P75
Ret	1517	1.066	11.020	−5.329	−0.426	5.809
SFG	1515	2.483	3.699	0.620	1.667	3.267
LFG	1345	2.524	2.373	0.936	1.882	3.323
SAG	1516	0.710	4.507	−0.743	0.557	2.138
LAG	1384	0.356	4.614	−1.004	0.293	1.832
SUE	1515	−1.762	3.593	−2.456	−0.894	−0.043
SIZE	1516	1.068	0.226	0.922	1.067	1.212
BM	1517	15.908	0.885	15.254	15.757	16.411
BETA	1517	0.405	0.288	0.205	0.329	0.515
ROE	1517	5.691	5.428	2.437	4.760	7.882
REV	1517	0.010	0.097	−0.053	−0.004	0.058
TO	1517	0.444	0.372	0.199	0.336	0.564
FREV	1506	−0.918	1.858	−1.419	−0.442	0.016
ANUM	1511	1.710	0.778	1.053	1.663	2.343
FDP	1511	0.008	0.010	0.001	0.005	0.010

This table presents the time-series average descriptive statistics of stock characteristics: number of observations (Obs), mean, standard deviation (Std.), 25th percentile (P25), median and 75th percentile (P75) of the main variables: stock return (Ret), analysts' short-horizon growth forecasts (*SFG*), analysts' long-horizon growth forecasts (*LFG*), short-horizon actual growth (*SAG*), long-horizon actual growth (*LAG*), standardized unexpected earnings (*SUE*), size (SIZE), book to market ratio (BM), market beta (BETA), return on equity (ROE), short term reversal (REV), turnover (TO), analysts' forecasts revision (FREV), analysts coverage number (ANUM), analysts' earnings forecasts dispersion (FDP). All the characteristics are cross-sectionally winsorized at the 1st and 99th percentiles. The sample period is from January 2010 to December 2020.

Table A.2
Alphas in the 12 months after portfolio formation.

	SFG						*LFG*					
	1	2	3	4	5	H − L	1	2	3	4	5	H − L
$t+1$	−0.27	−0.03	0.08	0.36	0.49	0.76	−0.30	−0.05	0.08	0.31	0.68	0.97
	(−1.88)	(−0.32)	(0.68)	(2.43)	(3.55)	(3.62)	(−3.12)	(−0.42)	(0.88)	(1.93)	(3.83)	(4.13)
$t+2$	−0.22	−0.04	0.06	0.31	0.45	0.67	−0.17	−0.05	0.10	0.34	0.61	0.78
	(−1.64)	(−0.35)	(0.37)	(2.23)	(3.11)	(3.17)	(−1.51)	(−0.55)	(1.26)	(2.52)	(3.27)	(2.86)
$t+3$	−0.20	0.08	0.18	0.32	0.32	0.52	−0.14	−0.04	0.12	0.22	0.49	0.63
	(−1.33)	(0.55)	(1.22)	(2.17)	(2.24)	(2.31)	(−1.37)	(−0.49)	(1.16)	(1.56)	(2.50)	(2.30)
$t+4$	−0.10	−0.05	0.09	0.29	0.24	0.34	−0.12	0.08	0.20	0.32	0.45	0.58
	(−0.80)	(−0.39)	(0.73)	(1.80)	(1.64)	(1.66)	(−1.60)	(0.82)	(1.59)	(2.06)	(2.55)	(2.78)
$t+5$	−0.07	−0.12	0.18	0.13	0.23	0.30	−0.23	0.18	0.13	0.25	0.44	0.67
	(−0.51)	(−1.21)	(1.47)	(0.90)	(1.64)	(1.59)	(−2.17)	(1.93)	(1.00)	(1.47)	(2.53)	(2.83)
$t+6$	−0.06	0.02	0.04	0.22	0.28	0.34	−0.14	0.13	0.12	0.32	0.42	0.56
	(−0.44)	(0.14)	(0.29)	(1.90)	(1.62)	(1.63)	(−1.43)	(1.00)	(1.17)	(1.70)	(2.38)	(2.25)
$t+7$	−0.01	−0.02	0.13	0.20	0.27	0.28	−0.31	0.29	0.19	0.30	0.45	0.76
	(−0.08)	(−0.18)	(0.85)	(1.76)	(1.38)	(1.31)	(−3.15)	(2.20)	(1.77)	(1.68)	(2.69)	(3.46)
$t+8$	0.08	−0.07	0.23	0.07	0.15	0.07	−0.13	0.17	0.22	0.40	0.26	0.39
	(0.63)	(−0.64)	(1.72)	(0.49)	(0.73)	(0.29)	(−1.11)	(1.99)	(2.18)	(2.14)	(1.34)	(1.41)
$t+9$	0.01	−0.02	0.36	0.05	0.17	0.16	−0.12	0.09	0.33	0.36	0.29	0.41
	(0.10)	(−0.22)	(2.43)	(0.27)	(0.81)	(0.63)	(−1.07)	(0.74)	(2.21)	(1.67)	(1.49)	(1.62)
$t+10$	−0.13	0.08	0.49	0.09	0.17	0.30	−0.18	0.26	0.17	0.37	0.21	0.39
	(−0.84)	(0.61)	(3.14)	(0.52)	(0.89)	(1.31)	(−2.00)	(2.63)	(1.24)	(1.86)	(1.39)	(1.90)
$t+11$	−0.02	−0.06	0.54	0.12	0.23	0.25	−0.08	−0.03	0.31	0.37	0.07	0.15
	(−0.10)	(−0.49)	(2.98)	(0.71)	(1.09)	(1.30)	(−0.64)	(−0.24)	(2.43)	(1.90)	(0.38)	(0.57)

(continued on next page)

Table A.2 (*continued*)

	SFG						*LFG*					
	1	2	3	4	5	H − L	1	2	3	4	5	H − L
$t+12$	0.02	0.02	0.39	0.06	0.16	0.14	−0.11	−0.02	0.17	0.34	0.08	0.19
	(0.14)	(0.13)	(2.62)	(0.34)	(0.81)	(0.74)	(−1.02)	(−0.14)	(1.02)	(1.52)	(0.39)	(0.70)

This table presents the monthly value-weighted FF-C-6 alphas on the long-short portfolios formed on *SFG* or *LFG* in the twelve months after portfolio formation. *t*-values are computed with Newey and West (1987) adjusted standard errors with 12 lags. The H − L columns contain the values plotted in Fig. 1.

Table A.3
Fama–MacBeth regression.

	Full sample				Excluding micro-caps			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SFG	8.92		4.03	4.52	9.86		3.34	3.34
	(3.26)		(2.15)	(2.74)	(2.98)		(1.57)	(1.95)
LFG		18.31	15.42	16.01		19.88	17.52	17.42
		(3.11)	(2.61)	(5.14)		(2.88)	(2.58)	(4.27)
SIZE				−0.49				−0.33
				(−1.96)				(−1.72)
BM				0.29				0.29
				(0.55)				(0.57)
BETA				0.56				0.38
				(0.93)				(0.67)
ROE				0.04				0.03
				(1.84)				(1.58)
REV				−2.63				−2.50
				(−2.82)				(−2.28)
TO				−0.88				−1.27
				(−3.23)				(−4.11)
IVOL				−4.46				−1.71
				(−1.80)				(−0.65)
ILLIQ				16.81				8.57
				(3.46)				(1.60)
FREV				0.13				0.16
				(4.23)				(5.70)
FDP				−10.38				−7.52
				(−2.07)				(−1.38)
N	174,669	174,669	174,669	172,669	126,633	126,633	126,633	126,633
Adj.R^2	0.65%	1.60%	1.83%	11.03%	0.91%	2.11%	2.43%	12.12%

The table is the full version of Table 3.

Table A.4
OLS model coefficients.

	FY(1)		*FY*(2)		*FY*(3)	
	Avg.coef	*Avg.t*	*Avg.coef*	*Avg.t*	*Avg.coef*	*Avg.t*
E^+	0.944	43.660	0.925	33.35	0.937	23.412
NEGE	0.009	1.312	0.014	1.850	0.017	1.667
ACC^+	−0.063	−3.433	−0.079	−3.028	−0.049	−1.282
ACC^-	−0.066	−2.442	−0.095	−2.823	−0.062	−1.243
AG	0.005	1.262	0.001	−0.254	−0.016	−2.919
Dividend	−0.276	−0.931	0.490	0.857	0.880	1.061
DD	−0.002	−1.072	−0.004	−1.324	−0.007	−2.069
BM	−0.033	−9.449	−0.058	−9.747	−0.070	−8.314
Const	0.014	5.029	0.030	8.188	0.041	8.087
Avg.R^2	55.792		31.471		21.754	

The table reports time series average coefficients and *t*-value of earnings forecast model 6 using yearly data from 2010 to 2020. In June of each calendar year *t*, we estimate the following OLS regression.

$$E_{i,t} = \beta_{0,n} + \beta_{1,n}E^+_{i,t-n} + \beta_{2,n}NEGE_{i,t-n} + \beta_{3,n}ACC^+_{i,t-n} + \beta_{4,n}ACC^-_{i,t-n} + \beta_{5,n}AG_{i,t-n} + \beta_{6,n}Dividend_{i,t-n} + \beta_{7,n}DD_{i,t-n} + \beta_{8,n}BM_{i,t-n} + \epsilon_{i,t},$$

for $n \in \{1,2,3\}$. The dependent variable $E_{i,t}$ is a firm's earnings scaled by the total asset in year t. Independent variables are lagged firm characteristics in year $t-n$: earnings scaled by total asset when earnings are positive and zero otherwise (E_i^+), a dummy variable that equals one if earnings are negative ($NEGE_i$), positive and negative accruals scaled by the total asset (ACC_i^+, ACC_i^-) where accruals equal the operating profit minus operating cash flow, the percent change in total assets (AG_i), dividends scaled by the total asset ($Dividend_i$), a binary variable indicating zero dividends (DD_i), and book-to-market ratio (BM_i).

Table A.5
Random forest regressions setup.

Variable	Definition
account_receivable_turnover_rate	Net Credit Sales / Average Accounts Receivable
bm	Book Equity / Market Equity
cash_ratio	Total Cash and Cash Equivalents / Current Liabilities
current_asset_turnover	Operating Revenue / Current Assets
current_ratio	Current Assets / Current Liabilities
debt_to_asset_ratio	Total Debt / Total Assets
debt_to_equity_ratio	Total Debt / Total Equity
dividend_yield	Dividend Yield
ev_to_ebitda	Enterprise Value Multiple
fixed_asset_ratio	Fixed Assets / Total Assets
fixed_asset_turnover	Operating Revenue / Fixed Assets
intangible_asset_ratio	Intangible Assets / Total Assets
inventory_turnover	Operating Revenue / Intangible Assets
long_term_debt_to_working_capital	Non-current Liabilities / Working Capital
net_profit_margin	Net Profit / Operating Revenue
ocf_to_interest_bearing_debt	Operating Cash Flow / Interest Bearing Debt
operating_revenue_growth	Operating Revenue Growth
pcf	Price / Cash flow
pe	Price / Net Profit
ps	Price / Sales
roa	Return on Assets
roe	Return on Equity
total_asset_growth	Total Asset Growth
total_asset_turnover	Operating Revenue / Total Assets
me	Market Capitalization
analyst_forecast	Analyst Forecast Earnings / Total Assets
anum	Analyst Coverage
fdp	Analyst Forecast Dispersion
Month of the year	
Industry Dummy	
Hyper-parameters	Number of Trees: 200, Maximum Depth: 5

The table presents the forecasting variables and hyper-parameters in the random forest model. Number of trees is the number of decision trees in the forest. Maximum Depth is the maximum number of splits that each decision tree can use. We vary the hyper-parameters Number of Trees from 100 to 2000 and Maximum Depth from 4 to 8. The results are similar.

Table A.6
LFG and forecast informativeness before and after 2015.

	2010 to 2015			2016 to 2020		
Informativeness	Low	High	H–L	Low	High	H–L
FACC	−0.41	0.46	0.87	−0.10	1.70	1.80
	(−2.37)	(1.61)	(2.79)	(−0.18)	(4.53)	(3.42)
ANUM	0.05	0.80	0.75	0.37	1.38	1.01
	(0.15)	(2.78)	(1.96)	(1.16)	(3.03)	(1.52)
INST	−0.53	0.41	0.94	0.95	1.60	0.65
	(−0.92)	(2.48)	(1.60)	(2.06)	(3.77)	(1.86)

This table reports the *LFG* results as in Table 7 in two subperiods. At the end of each month, we sort stocks into three groups by three measures of forecast informativeness: past forecast accuracy (FACC), analyst coverage (ANUM), or institutional ownership (INST) by the 30th and 70th percentiles. We then independently sort stocks into quintiles by *LFG*. This table reports the FF-C-6 alphas (and *t*-values in parentheses) on the long-short portfolios that buy stocks with high forecasts (*LFG*) and sell those with low forecasts among stocks with low or high forecast informativeness. The H − L columns report the alphas of the return differences between the long-short portfolios that contain the high- and low-forecast informativeness stocks.

References

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. J. Financ. Mark. 5, 31–56.
Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. J. Financ. 61, 259–299.
Bordalo, P., Gennaioli, N., La Porta, R., Shleifer, A., 2019. Diagnostic expectations and stock returns. J. Financ. 74, 2839–2874.
Bouchaud, J.-P., Krüger, P., Landier, A., Thesmar, D., 2019. Sticky expectations and the profitability anomaly. J. Financ. 74, 639–674.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Financ. 52, 57–82.
Da, Z., Warachka, M., 2011. The disparity between long-term and short-term forecasted earnings growth. J. Financ. Econ. 100, 424–442.
De Bondt, W.F.M., Thaler, R.H., 1990. Do security analysts overreact? Am. Econ. Rev. 80, 52–57.
de Silva, T., Thesmar, D., 2021. Noise in Expectations: Evidence from Analyst Forecasts. Working Paper..
Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. J. Financ. 57, 2113–2141.

Engelberg, J., McLean, R.D., Pontiff, J., 2020. Analysts and anomalies. J. Account. Econ. 69, 1–13.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81, 607–636.
Guo, L., Li, F.W., Wei, K.J., 2020. Security analysts and capital market anomalies. J. Financ. Econ. 137, 204–230.
Han, X., Li, Y., 2017. Can investor sentiment be a momentum time-series predictor? Evidence from China. J. Empir. Financ. 42, 212–239.
Han, C., Shi, Y., 2022. Chinese stock anomalies and investor sentiment. Pac. Basin Financ. J. 73, 101739.
Hawkins, E.H., Chamberlin, S.C., Daniel, W.E., 1984. Earnings expectations and security prices. Financ. Anal. J. 40, 24–74.
Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. Rev. Financ. Stud. 33, 2019–2133.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. J. Financ. 48, 65–91.
Jegadeesh, N., Kim, J., Krische, S.D., Lee, C.M., 2004. Analyzing the analysts: when do recommendations add value? J. Financ. 59, 1083–1124.
La Porta, R., 1996. Expectations and the cross-section of stock returns. J. Financ. 51, 1715–1742.
Liu, J., Stambaugh, R.F., Yuan, Y., 2019. Size and value in China. J. Financ. Econ. 134, 48–69.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Richardson, S., Teoh, S.H., Wysocki, P.D., 2004. The walk-down to beatable analyst forecasts: the role of equity issuance and insider trading incentives. Contemp. Account. Res. 21, 885–924.
So, E.C., 2013. A new approach to predicting analyst forecast errors: do investors overweight analyst forecasts? J. Financ. Econ. 108, 615–640.
Thomas, G.J., Hwang, C.-Y., Li, Y., 2023. February, Share Turnover, and Momentum in China. Working Paper.
van Binsbergen, J.H., Opp, C.C., 2019. Real anomalies. J. Financ. 74, 1659–1706.
van Binsbergen, J.H., Han, X., Lopez-Lira, A., 2023. Man versus machine learning: the term structure of earnings expectations and conditional biases. Rev. Financ. Stud. 36, 2361–2396.
Xu, X., 2021. Earnings information disclosure, investor behavior, and market insider trading (translated from Mandarin). Econ. Theory Bus. Manag. 41, 52–64.
Xu, N., Jiang, X., Yi, Z., Xu, X., 2012. Conflicts of interest, analyst optimism, and stock price crash risk (translated from Mandarin). Econ. Res. J. 47, 127–140.
Zhang, R., Wang, R., Wang, S., 2017. Analysts' revisions, fundamental analysis, and future stock returns (translated from Mandarin). J. Financ. Res. 38, 156–174.