Arbitrageur Factors

Jiaqi Zhang*

November 16, 2023

ABSTRACT

Five arbitrageur factors explain over 80% of the cross-sectional monthly return predictors found in the literature, outperforming existing factor models. Theoretically, stock price changes can be predicted by assuming that stocks are influenced by common, predictable sources of behavioral trading, and that arbitrageurs have a limited capacity for risk-taking. Empirically, factors are robustly chosen by maximizing the Sharpe ratio of the tangency portfolio. Beta loadings in this model identify sources of mispricing in asset pricing anomalies.

*Carroll School of Management, Boston College. zhangeq@bc.edu. I would like to specially thank Jefferey Pontiff for his guidance as my advisor. I would like to thank Rui Albuquerque, Samuel Hartzmark, Trevor Haynes, Hojoon Lee, Daniel Smith, David Solomon, Ziqi Yuan, and participants at the Boston College Finance PhD Seminar and the Third-year Paper Presentation for their helpful comments and suggestions. ChatGPT helps me improve the clarity and readability of the paper.

1 Introduction

Theoretically, assets in the market are not perfectly priced, because some returns are necessary as compensation to arbitrageurs for the risks associated with holding arbitrage positions (Grossman and Stiglitz 1980). Empirically, cross-sectional variation in asset returns are shown in a number of predictors in the literature. These so-called "anomalies" – long-short portfolios generating significantly positive returns over time – do not appear to arise from differences in systematic risk exposures. The fact that anomalies partially persist after publication challenges both the hypothesis that asset returns are determined by time-invariant risk compensations and the hypothesis that all anomalies result from data mining (McLean and Pontiff 2016).

In this paper, I propose a new framework linking the theory of arbitrageur risk taking with the empirical results of cross-sectional return predictability. It demonstrates that a few factors, reflecting arbitrageurs' required returns, explain nearly all the cross-sectional prediction power documented in the literature and outperform all existing factor asset pricing models. By assuming that the trading behaviors of less sophisticated investors are predictable, returns from trading against these behaviors are also predictable given arbitrageurs' risk taking capacity.

In traditional asset pricing models like the APT, the risk-taking capacity of arbitrageurs is considered unlimited. As soon as prices deviate from assets' fair value, arbitrageurs enter the market and drive prices back. Expected return variations are attributed solely to compensation for different risk sources, captured by a few risk factors. Instead, I assume that arbitrageurs have limited risk-taking capacity. Consequently, when an asset deviates from its fair value, it may remain mispriced as the mispricing provides returns to compensate arbitrageurs.

Additionally, I posit that the reasons for price deviations from fair value are systematic, stemming from behavioral biases in less sophisticated investors who do not properly maximize their risk-adjusted returns. Arbitrageurs then require a certain level of return to hold positions opposite to investors with each type of behavioral bias. Therefore, their expected returns are explained by a few factors, which I term "arbitrageur factors." These factors do not represent generic risk sources for all investors but are specifically relevant to the marginal investor, the arbitrageur, in this model.

Under the efficient market hypothesis, only new information moves prices. Price changes result from altered expectations of future cash flows, while the realization of these cash flows has no impact. However, in the context of limited arbitrage, it is not just information about behavioral trading volume that affects prices, but the predictable trading volume itself also influences them (Hartzmark and Solomon 2021). This concept illustrates the source of cross-sectional predictive power in asset returns. The assumption that positions influenced by systematical behavioral reasons are predictable leads to the implication that cross-sectional asset returns are also predictable and can be explained by a series of factor returns.

Returns from such predictable arbitrage activities should remain relatively constant, except in turbulent market conditions where significant changes occur. Therefore, I select these "arbitrageur factors" by maximizing the Sharpe ratio of their tangency portfolios, ensuring that they offer stable returns with limited correlation. This automated algorithm helps mitigate concerns of data mining. As long as a few factors exhibit high Sharpe ratios and are not highly correlated, they are likely effective proxies for arbitrage returns associated with different types of behavioral biases.

The five factors chosen in my main model represent the following types of behavioral biases respectively: under-reaction to recent news; trading influenced by attention or sentiment; mentally disconnect between dividends and capital returns; insufficient response to less salient information; and over-optimism in long-term expectations. These biases are well-documented in literature and account for different sources of predictable cross-sectional expected returns.

Out of 207 anomalies in monthly stock returns, 69 are significant at the 1% level in Gibbons, Ross, and Shanken (1989) (hereafter referred to as GRS) tests with the first two arbitrageur factors and the market factor, while 31 are significant in the five-factor model. In contrast, over 100 anomalies are significant in all existing models. Alternative groups of factors, chosen using the same algorithm, yield lower explanatory power but still outperform all benchmark models.

The arbitrageur factor model more effectively explains the returns of anomalies after their publication in academic journals, especially for those highly cited anomalies. As more arbitrageurs become aware of a pattern, their arbitrage return approaches the competitive equilibrium with no alphas. Additionally, this model effectively explains anomalies related to market patterns, valuations, professional forecasts, and financing activities, yet leaves an unexplained alpha on average for fundamental anomalies. These anomalies could indicate either unique sources of market inefficiency or systematic risk exposures other than market risk.

Beta loadings in the arbitrageur factor model offer clear interpretations. Each factor in the model represents a specific type of investor behavioral bias. Significant exposure to a factor suggests that the anomaly is from the corresponding trading activity. For example, positive exposure to the Under-reaction to News factor suggests that the long side of the portfolio contains assets with recent good news ignored by some investors. The limited risk-taking capacity of arbitrageurs explains the persistence of its predictive power.

This study is connected to a broad spectrum of literature. It builds upon the literature of behavioral biases in the stock market, detailed in Section 4.2, and on prior studies in asset pricing models, reviewed in Section 6.1.

Additionally, the study follows the concept that determine asset prices. Recent theoretical and empirical work demonstrates that variations in financial intermediary capital move asset prices (He and Krishnamurthy 2013; Brunnermeier and Sannikov 2014; He, Kelly, and Manela 2017). In the context of cross-sectional asset returns, Cho (2020) indicates that correlation in anomaly returns stem from shared exposure to arbitrage activities.

Furthermore, the study aligns with recent evidence suggesting that systematic risk factors may not be the main determinants of required returns. Analysis reveals that anomaly returns are primarily driven by cash flow news instead of discount rate news (Lochstoer and Tetlock 2020); long-term discount rates do not seem to contribute sufficiently to risk compensation (Keloharju, Linnainmaa, and Nyberg 2021); and surveys, along with mutual fund flow data, indicate that investors disregard their exposure to systematic risks in their investment decisions (Choi and Robertson 2020; Chinco, Hartzmark, and Sussman 2022; Ben-David, Li, Rossi, and Song 2022).

Recent studies employing statistical methods suggest that the a few factors are sufficient to encapsulate all information from asset return predictors (see, for example, Kelly, Pruitt, and Su 2019; Feng, Giglio, and Xiu 2020). This paper extends this notion, showing that even simple linear factor models can reach similar conclusions when the appropriate factors are chosen.

The remaining sections are organized as follows: Section 2 presents the theoretical framework of return predictability and factors. Section 3 introduces the data sources. Section 4 demonstrates the selection process for arbitrageur factors and the time series of factors. Section 5 discusses the empirical results regarding the cross-sectional explanatory power. Section 6 interprets the sources of mispricing from regression betas, and Section 7 concludes the paper.

2 Theoretical Framework

2.1 Motivation

In a world where all investors are rational and with the same information, they maximize their risk-adjusted returns and all trade assets only at their fair value. However, in reality, investors hold assets for various behavioral reasons. There is extensive literature on the behavioral biases of investors. For example, they may not be fully aware of information and underreact to news; they make decisions based on their attention or sentiment; they are more likely to notice salient information; they overestimate the likelihood of long-term success for growth firms; and some treat dividends as free money.

If the market were fully efficient, despite some investors being influenced by these biases, the price would still align with the rational level as implied by future cash flow and risk-adjusted discount rates. Under the assumptions of the Arbitrage Pricing Theory (APT) (Ross 1976), any deviation from this price would create an opportunity for a zero-cost, long-short portfolio to yield positive profits. As soon as assets are mispriced, arbitrageurs trade against these patterns, driving prices back to fair value.

In reality, however, investors' risk-taking capacity is limited. As Grossman and Stiglitz (1980) demonstrates, some positive arbitrage return is necessary to compensate arbitrageurs for their risk in arbitrage activities. This paper further shows that, with limited risk-taking capacity and predictable trading biases, future asset returns are predictable.

Figure 1 illustrates the idea in demand and supply curves following the framework by Shleifer (1986). Behavioral investors are on the supply side. They hold asset positions for behavioral reasons, affecting the number of shares available to other investors on the x-axis. Arbitrageurs are on the demand side, holding different positions for different prices on the y-axis and making profits. The market clearing condition determines the equilibrium asset price. For simplicity, I assume that the asset supply is fully inelastic, and without loss of generality, I show the case where the remaining share supply is negative here.

In a world with no limitations to arbitrage, the demand is fully elastic. Arbitrageurs can take an infinite amount of securities at the fair value. Consequently, regardless of the supply level, the equilibrium price is always at the fair value as shown in graph (a).

With limited risk-taking capacity of arbitrageurs, the elasticity of the demand curve is lower. They are only willing to take a limited number of positions with a certain rate of return. The farther the price deviates from the fair value, the higher the returns they earn and the larger the position they take. The downward slope of the demand curve results from higher rates of return compensating arbitrageurs for their exposure to idiosyncratic risks when they take larger positions. In the example in graph (b), they take a short position at equilibrium, and the asset should be overpriced. Future corrections of mispricing provide positive expected returns for them.

More interestingly, with predictable future behavioral investors' positions and limited risk-taking, price changes arise not only from the information about the positions but also from the trades themselves. This effect is defined as predictable price pressure by Hartzmark and Solomon (2021).

In traditional asset pricing models, all price changes beyond the discount rate are attributed to new information. The arrival of new information regarding cash flows and discount rates immediately moves prices, while the realization of cash flow does not change prices as long as there is no new information. However, under the assumption that arbitrageurs have limited risk-taking capacity, not only does the information about future behavioral investors' positions move prices, but also the realization of changes in behavioral investors' positions.

For example, in graph (b), if behavioral investors' positions are expected to be lower in a future period, the supply curve is expected to shift right. The level of overpricing would then be lower, and the return is predicted to be negative when the position changes. This result holds even if the behavioral investors' positions are known in advance, contradicting the common belief that price changes are all driven by new information.

In the remainder of this section, I formally demonstrate in a model that under certain assumptions, returns are predictable from previously known behavioral investors' positions and several factors capture the prediction power.

2.2 Model

Assume that there is a fair value for each asset i, $V_{i,t}$, which is determined by the discounted value of all future dividends:

$$V_{i,t} = E_t \left[m_{i,t+1}(V_{i,t+1} + D_{i,t+1}) \right], \tag{1}$$

where $D_{i,t+1}$ is the dividend payment at time $t+1$. For simplicity, assume that the payout ratio $d_i = \frac{D_{i,t+1}}{V_{i,t+1}+D_{i,t+1}}$ is constant over time.

I do not explicitly model the discount rates and news of dividends, but simply assume that the CAPM accounts for the evolution of the fundamental value and that the beta for each asset is constant:

$$\frac{V_{i,t+1} + D_{i,t+1}}{V_{i,t}} = 1 + r_{f,t} + \beta_i(r_{M,t+1} - r_{f,t}) + \epsilon^V_{i,t+1}, \tag{2}$$

where $r_{f,t}$ is the risk-free rate observed at time t, $r_{M,t+1}$ is the realized market return at time $t+1$, and $\epsilon^V_{i,t+1}$ denotes the total unexpected shock to the fair value, comprising both discount rate news and cash flow news.

The actual price of the asset, $P_{i,t}$, however, may deviate from the fair value. The realized return of the asset is then

$$r_{i,t+1} = \frac{P_{i,t+1} + D_{i,t+1}}{P_{i,t}}. \tag{3}$$

There are three types of agents in the economy: Passive investors, or CAPM investors, always hold the market portfolio, with their total position being time-varying. They can be rationalized if I assume that they only observe the fair value of assets, and the market portfolio is the tangency portfolio as in CAPM. However, for simplicity, I do not explicitly solve their problem.

Behavioral investors are not rational and their position is unrelated to expected returns. For asset i at time t, their total position is

$$\sum_{j=1}^{J} X_{i,t}^{(j)}, \tag{4}$$

where j represents different types of noise trading, and

The position of behavioral investors on each asset is expected to decay in each period, following

$$X_{i,t+1}^{(j)} = \delta_j X_{i,t}^{(j)} + \epsilon_{i,j,t+1}, \tag{5}$$

where δ_j is the decay rate of the corresponding type of behavior, assumed to be exogenous in the model. $\varepsilon_{i,j,t}$ is unexpected shocks on the position, which are assumed to be independent with mean zero. This setting ensures that behavioral investors' position on each asset is with mean zero in the long term, while they are always exposed to some assets from unexpected shocks.

A representative agent, "the arbitrageur," actively trades in the market and maximizes her utility. She has access to all information in the market, including the fair value of all assets and the positions from all types of noise trading. I assume that she has an exponential utility function (constant absolute risk aversion) $u(W) = -e^{-aW}$. Additionally, she does not want to be exposed to any market risk. For each dollar invested in asset i, she hedges the market risk with β_i dollar in the market portfolio. Denoting her position in asset i as $X_{i,t}$, the problem she solves is

$$\max_{X_{i,t}} E_t U \left(W_t(1 + r_{f,t}) + \sum_{i=1}^{N} X_{i,t} r_{i,t+1}^{\alpha} \right), \tag{6}$$

where

$$r_{i,t+1}^{\alpha} = r_{i,t+1} - r_{f,t} - \beta_i (r_{M,t+1} - r_{f,t}) \tag{7}$$

is the CAPM alpha.

The market clearing condition below ensures that the sum of positions of the arbitrageur and behavioral investors is a proportion of the market portfolio:

$$X_{i,t} = -\sum_{j=1}^{J} X_{i,t}^{(j)}. \tag{8}$$

The arbitrageur's problem, along with the market clearing condition, determines the equilibrium.

Proposition 1.

$$P_{i,t} = V_{i,t}(1 + \sum_{j=1}^{J} \frac{a\sigma_i^2}{1-(1-d_i)\delta_j} X_{i,t}^{(j)}), \tag{9}$$

where σ_i^2 *is the variance of* $\varepsilon_{i,t+1} = \epsilon_{i,t+1}^{i} + (1-d_i)\sum_{j=1}^{J}\epsilon_{i,j,t+1}$*, represents an equilibrium. In this equilibrium, the stock return in period* $t+1$ *is*

$$r_{i,t+1} = r_{f,t} + \beta_i(r_{M,t+1} - r_{f,t}) + \sum_{j=1}^{J} b_i^{(j)} X_{i,t}^{(j)} + \varepsilon_{i,t+1}, \tag{10}$$

where $b_i^{(j)} = -a\sigma_i^2$*.*

Proof: *See Appendix A.1.*

Equation (9) demonstrates that the level of mispricing is proportional to the position of each type of behavioral investor in the asset. With identical positions across noise trader types, the corresponding level of mispricing increases with a higher risk tolerance of arbitrageurs (a), greater total idiosyncratic risks in the asset (σ_i^2), more persistent behavioral trading (δ_j), and lower dividend payout ratios (d_i). The conclusion here mostly aligns with Pontiff (2006), while I also highlights persistence.

2.3 Factors

Empirically, the position of behavioral investors and its persistence are not observable, making it challenging to calibrate mispricing levels. In this paper, I do not discuss the levels of mispricing further but rather focus on the factor structure of returns. Equation (10) indicates that the return of an asset is a linear combination of the positions of behavioral investors of each type. This allows me to characterize the expected return of any single portfolio through linear regressions, provided there are tradable proxies for those types of behavioral investors' positions. In the rest of this subsection, I discuss the underlying assumptions for such empirical studies.

First, arbitrage portfolios constructed from fixed sorting rules maintain constant exposure to each type of noise trading volume. In other words, for any j and portfolio k with weight $w_{i,t}^{(k)}$ on asset i at time t,

$$\beta_k^{(j,M)} = \frac{\sum_{i=1}^{N} w_{i,t} b_i^{(j)} X_{i,t}^{(j)}}{\sum_{i=1}^{N} w_{i,t}^{M} b_i^{(j)} X_{i,t}^{(j)}} \tag{11}$$

remains unchanged over time t, where $w^M i, t$ is the weight of asset i in the market portfolio.

Similarly, such portfolios also exhibit constant exposure to market risk, meaning that

$$\beta_k^M = \sum_{i=1}^{N} w_{i,t} \beta_i \tag{12}$$

is also time-invariant.

Second, for each j, a self-funded arbitrage portfolio with weight $w_{i,t}^{(j)}$ has zero exposure to the market factor and all other types of noise $j' \neq j$, namely

$$\sum_{i=1}^{N} w_{i,t}^{(j)} = 0,\ \sum_{i=1}^{N} w_{i,t}^{(j)} \beta_i = 0,\ \sum_{i=1}^{N} w_{i,t}^{(j')} b_i^{(j')} X_{i,t}^{(j')} = 0,\ \sum_{i=1}^{N} w_{i,t}^{(j)} b_i^{(j)} X_{i,t}^{(j)} > 0, \tag{13}$$

and I denote its portfolio return as $F_{j,t}$ at time t.

Proposition 2. *With the aforementioned assumptions, the return of an arbitrage portfolio k constructed from fixed sorting rules can be expressed as*

$$r_{k,t+1} = r_{f,t} + \beta_k^M (r_{M,t+1} - r_{f,t}) + \sum_{j=1}^{J} \beta_k^{(j)} F_{j,t} + \varepsilon_{k,t+1}^P, \tag{14}$$

where β_k and $\beta_k^{(j)}$s are time-invariant and the noise term $\varepsilon_{k,t+1}^P$ has no serial correlation and a mean of zero.

Proof: *See Appendix A.2.*

Equation 14 explains the expected return of any single portfolio through linear regressions on the market factor and those factors defined. In the empirical section of this paper, I refer to these factors as 'arbitrageur factors', as each represents the required return of arbitrageurs on a specific type of investor behavior. I demonstrate that they explain the expected returns of most anomalies documented in the literature.

2.4 Unexpected shocks

In this model, I do not explicitly model changes in the risk appetite of arbitrageurs and the persistence of behavioral investors' position, instead assuming them to be constant. However, in reality, they may change over time. In this section, I study stock returns with unexpected changes in these parameters.

Suppose the risk appetite a changes from a_t to a_{t+1}, and the persistence of noise j changes from $\delta_{j,t}$ to $\delta_{j,t+1}$. The arbitrageur does not anticipate these changes and continues to assume that a and δ_js remain fixed. Under these circumstances, the equilibrium result in Equation (9) still applies:

$$P_{i,t} = V_{i,t}(1 + \sum_{j=1}^{J} \frac{a_t \sigma_i^2}{1-(1-d_i)\delta_{j,t}} X_{i,t}^{(j)}), \tag{15}$$

$$P_{i,t+1} = V_{i,t}(1 + \sum_{j=1}^{J} \frac{a_{t+1} \sigma_i^2}{1-(1-d_i)\delta_{j,t+1}} X_{i,t+1}^{(j)}), \tag{16}$$

Proposition 3. *With the above equilibrium prices, the return of an asset is*

$$r_{i,t+1} = r_{f,t} + \beta_i(r_{M,t+1} - r_{f,t}) + \sum_{j=1}^{J} b_{i,t+1}^{(j)} X_{i,t}^{(j)} + \varepsilon_{i,t+1}, \tag{17}$$

where

$$b_{i,t+1}^{(j)} = -a_{t+1}\sigma_i^2 + \frac{a_{t+1}\sigma_i^2}{1-(1-d_i)\delta_{j,t+1}} - \frac{a_t\sigma_i^2}{1-(1-d_i)\delta_{j,t}}. \tag{18}$$

Proof: *See Appendix A.3.*

Since the factor return at time j is proportional to $b^{(j)}$ by assumption, it is higher than the static casein one period when the arbitrageur becomes less risk-averse or when behavioral investors' position becomes less persistent in that period. Conversely, it is lower or even negative when the arbitrageur becomes more risk-averse or the noise becomes more persistent.

This explains why factors experience significant returns in some periods and crashes in others. These fluctuations may be driven by changes in the pattern of behavioral investors (like during bubble accumulation periods), or by shifts in the risk appetite of arbitrageurs (like during crises).

3 Data

The cross-sectional return data used in this study is sourced directly from the open-source cross-sectional asset pricing project conducted by Chen and Zimmermann (2022). This project offers monthly indicators and returns for 207 anomalies from prior literature on its website[1]. It closely replicates the data processing steps outlined in the corresponding original papers. These return series, derived from long-short arbitrage portfolios, are signed appropriately to ensure their expected returns are positive. The project demonstrates successful replication of almost all these anomalies within the sample periods described in the original studies. In this paper, I test whether these returns are explained over a sample period from 1973 to 2021, including the maximum available data for each anomaly within this timeframe.

I also choose factors directly from the same pool of arbitrage return series. This approach simplifies the selection process and addresses concerns about data mining in constructing factors. Specifically, for factor selection, I require candidate series to have non-missing data in at least 90% of the months from January 1973 to December 2021. This criterion retains 174 out of the 207 series from their dataset. I then calculate their sample mean returns and the covariance matrix. The market factor in the arbitrageur factor model, represented by the market excess return, is obtained from Kenneth R. French's Data Library.

Furthermore, I evaluate the performance of the arbitrageur factor model by comparing it with several established asset pricing models, including:

1. The three-factor model by Fama and French (1993) (FF3), and the five-factor model of Fama and French (2015) (FF5), both available on Kenneth R. French's Data Library.

2. The q^5 factor model by Hou, Mo, Xue, and Zhang (2021) (q^5), available on the Global q website.

[1] https://www.openassetpricing.com/

3. The mispricing factor model by Stambaugh and Yuan (2017) (M4), accessible on Robert F. Stambaugh's personal website, until December 2016.

4. The behavioral factor model by Daniel, Hirshleifer, and Sun (2020) (DHS), available on Kent D. Daniel's personal website, until December 2018.

5. The IPCA factors by Kelly, Pruitt, and Su (2019) (IPCA5), generated using replication codes from Seth Pruit's personal website, available until May 2014. Although not technically a linear factor model, these factors are included for comparison.

Unless specified differently, these data series span the entire sample period from January 1973 to December 2021.

Lastly, as an additional test, I include 300 bivariate sorted portfolios on size/value, size/investment, and size/profitability, obtained from Kenneth R. French's Data Library.

4 The Arbitrageur Factors

4.1 Factor selection

In this section, I show the process of selecting factors. Behavioral investors' positions are not directly observable, so it is not possible to directly verify if the selected factors indeed capture all types of behaviors in the market. Instead, I discuss in the subsequent subsection what types of behaviors these selected factors may represent as indirect verification.

The algorithm is based on the Sharpe ratio of the tangency portfolio of all selected factors, which includes the market factor inherently present in the model. Firstly, a high Sharpe ratio suggests that the realized return series contain a greater proportion of predictable returns relative to unpredictable market information, indicating that they are likely capturing predictable behaviors as in the model. Secondly, optimizing the tangency portfolio implicitly require the selected factors to demonstrate low correlation, suggesting that they likely originate from distinct types of behaviors.

Exploring all possible combinations of factors is not feasible due to computational constraints ($200^6 = 6.4 \times 10^{13}$ or 64 trillion). Therefore, I employ a greedy algorithm for the selection of arbitrageur factors. The market factor is included initially. With a set of factors already chosen in the model, I examine each anomaly return series as a potential additional factor. For each, I calculate the Sharpe ratio of the tangency portfolio, which is derived from the mean and covariance of returns within the sample period from January 1973 to December 2021. The return series that yields the highest Sharpe ratio upon incorporation into the model is then selected as the next factor.

Table 1 step-by-step displays the arbitrageur factors selected and the corresponding Sharpe ratios of the tangency portfolios. In the sample, the market portfolio has a Sharpe ratio of 0.494. When including the *AnnouncementReturn* anomaly, the tangency portfolio yields the highest Sharpe ratio of 2.558 among all potential ones; therefore, *AnnouncementReturn* is selected as the first factor. Subsequently, adding the *STReversal* anomaly to the two already chosen factors results in the highest Sharpe ratio of 4.203 compared to other potential tangency portfolios, leading to the selection of *STReversal* as the second factor. This procedure continues until the sixth factor is selected.

As shown in the table, the first two strategies—post-earnings-announcement drift, as measured by the return around the announcement date (Chan, Jegadeesh, and Lakonishok 1996), and the one-month short-term reversal (Jegadeesh 1990)—significantly enhance the Sharpe ratio of the tangency portfolio. The subsequent factors continue to increase this in-sample Sharpe ratio.

The sixth factor, representing the short interest of a stock, seems more indicative of limitations to arbitrage rather than investor behavior, which does not align with the model's assumptions. Consequently, the model excluding this factor—the 5-factor model—is adopted as the primary model in this study. Additionally, models with 2, 3, 4, and 6 factors are also tested for comparative purposes.

As demonstrated in the table, the first two strategies—the post-earnings-announcement drift, evaluated by the return around the announcement date (Chan, Jegadeesh, and Lakonishok 1996), and the one-month short-term reversal (Jegadeesh 1990)—make substantial contribution to the Sharpe ratio of the tangency portfolio, and the following factors further increases this in-sample Sharpe ratio. Since the sixth factor selected, the short interest of a stock, is not likely from investor behaviors but more likely a measure of limitation to arbitrage, the 5-factor model is deployed as the main model in this paper, while models with 2, 3, 4, 6 factors are also included for comparision.

To verify the robustness of the factor model and ensure its performance is not due to chance in factor selection, two alternative models are presented. Alternative model 1 is constructed by repeating the factor selection process to select five factors, excluding the six factors previously chosen. Alternative model 2 follows a similar process, excluding all 11 previously chosen factors. The factors and the Sharpe ratios of their respective tangency portfolios are detailed in the second part of Table 1. Although these models exhibit lower Sharpe ratios due to the exclusion of the least noisy predictors, Section 5.1 demonstrates that they still outperform existing models in explaining expected returns.

4.2 Types of behavior biases

In this subsection, I discuss the types of behavior biases that the five factors selected for the main model represent. These biases are well-documented in behavioral finance literature, though some have garnered more attention than others.

1. Under-reaction to News (NEW): The first factor selected is constructed from the market's return around earnings announcements predicting subsequent returns (Chan, Jegadeesh, and Lakonishok 1996). This post-earnings-announcement drift has been a long-standing subject in literature (Ball and Brown 1968; Foster, Olsen, and Shevlin 1984). The behavioral rationale for under-reaction may be attributed to the disposition effect, where investors are reluctant to realize their losses (Grinblatt and Han 2005, Frazzini 2006), or to limited investor attention to news (DellaVigna and Pollet 2009, Hirshleifer, Lim, and Teoh 2009, Ben-Rephael, Da, and Israelsen 2017).

2. Attention-Induced Trading (ATT): The second factor identified is the one-month short-term reversal. One possible channel is that less sophisticated traders purchase stocks that capture their attention, leading to short-term price increases and subsequent reversals (Da, Engelberg, and Gao 2011, Barber, Huang, Odean, and Schwarz 2022).

3. chasing dividends (DIV): The third factor is derived from the returns of stocks predicted to issue dividends (Hartzmark and Solomon 2013). Investors tend to perceive dividends as a distinct component of their income (Hartzmark and Solomon 2019, Bräuer, Hackethal, and Hanspal 2022), which leads to different behaviors around dividend payments.

4. Slow Response to Less Salient Information (SAL): The fourth factor is based on the premise that returns of large firms predict the returns of other firms in the same industry

(Hou 2007). This pattern of predictability in the stock market is mirrored in findings related to economic links (Cohen and Frazzini 2008), supplier and customer relationships (Menzly and Ozbas 2010), and conglomerates (Cohen and Lou (2012)). These scenarios involve information indirectly related to the firm, which investors find more challenging to process. Bordalo, Gennaioli, and Shleifer (2022) discuss how salience influences decision making.

5. Bias in Long-Term Expectations (EXP): The fifth factor, net equity financing (Bradshaw, Richardson, and Sloan 2006), represents the prediction power from the difference between share issuance and share repurchase. Firms time the market, issuing shares when they detect over-optimism in their prices and repurchase shares otherwise, and such decisions reveal more accurate information of long-term valuation. Misinterpretation of conditional probabilities (Bordalo, Gennaioli, Porta, and Shleifer 2019) and the fading memory of past events (Nagel and Xu 2022) contribute to biases in long-term expectations.

4.3 Time series

Figure 2 presents the time series of log cumulative returns for the five factors selected in the main model, along with the tangency portfolio. The Attention-Induced Trading (ATT) factor exhibits the highest expected return and volatility, contrasting with the Chasing Dividend (DIV) factor, which shows lower but more consistent returns. This disparity illustrates that some investor behaviors, like chasing dividends, are relatively stable over time, whereas behaviors influenced by attention can be highly volatile. In the theoretical framework discussed in Section 2.4, this suggests that the DIV factor represents one type of investor behavior with relatively constant persistence, while the persistence investor behavior captured by the ATT factor fluctuates more frequently.

Generally, the returns of these factors are stable during normal market conditions but exhibit significant returns and reversals in times of crisis. This pattern aligns with the interpretation that investor and arbitrageur behaviors undergo changes during crises.

Notably, the cumulative return line of the tangency portfolio indicates a turning point around 2003, coinciding with the end of the dot-com bubble crash. The average return of the portfolio is observed to be lower after this period. Possible explanations for this shift include increased market efficiency, from reduced behavioral trading volume or heightened risk-taking by arbitrageurs. Alternatively, it could be that these factors, derived from literature, no longer capture the predictive powers in the market as effectively, particularly in more recent times.

5 Empirical Tests

5.1 207 anomalies

Table 2 exhibits whether various factor models can explain the 207 anomaly return series shared by Chen and Zimmermann (2022). The first part presents the performance of the arbitrageur factor models with varying numbers of factors, alongside the market factor. The second part compares these results with benchmark models, including the performance of t-tests on the anomaly returns without any model, and the FF3, FF5, q^5, M4, DHS, and IPCA models, as discussed in Section 3. The following parts presents the outcomes for alternative models, excluding each of the factors or excluding all factors from the main model as described in Section 4.1 and detailed in Table 1.

It includes the number of anomalies significant in t-tests on alphas, average standardized alphas, and average R^2 values, along with the GRS test results. The "standardized alpha" adjusts the alpha value for volatility, defined as the alpha estimate divided by the sample standard deviation of the anomaly return series, and then multiplied by $\sqrt{12}$. This metric makes it comparable to annual Sharpe ratios. For the "no model" row, this column directly represents the annualized Sharpe ratio. The use of standardized alpha, instead of original alpha estimates, ensures that the mean alpha statistic is not overly influenced by anomalies with high volatility.

This test primarily focuses the expected returns of portfolios, rather than on the total variation in returns, to reveal whether the anomalies documented in the literature offer additional predictive power beyond the selected arbitrageur factors. If most of the anomalies are found to be statistically insignificant, it would support the argument that the chosen arbitrageur factors capture the bulk of predictive power inherent in all documented anomalies, even if some variation in returns is not explained. Both the t-test on alpha and the GRS test are aligned with this objective.

In the model incorporating only the first two arbitrageur factors—Under-reaction to

News (NEW) and Attention-Induced Trading (ATT)—the alphas of 98 anomalies are significant at $|t| > 2$, and 55 at $|t| > 3$. These numbers are notably lower compared to those obtained using benchmark models. The GRS tests also reflect a similar relationship in the number of significant anomalies, and the average standardized alpha of the anomalies is smaller than those yielded by benchmark models. According to the theoretical framework of this paper, this superior explanatory power indicates that a substantial portion of the predictive power on asset returns can be attributed to these two major market behaviors and the limited risk-taking capacity of arbitrageurs in counteracting their price impacts.

The inclusion of three additional factors—Chasing Dividends (DIV), Salience of Information (SAL), and Biased Long-term Expectations (EXP)—leads to more anomalies being explained. In this case, only 31 anomalies remain significant at the 1% level in GRS tests, demonstrating the added explanatory power of these extra behavioral trading sources.

Regarding benchmark models, the limited explanatory capacity of the FF3 model is expected, as most anomaly studies already include FF3 in their analysis, and anomalies explained by FF3 are less likely to be reported. The FF5 model shows marginal improvements in explanatory power. The q^5 factors account for more anomalies, while the M4 and DHS models, grounded in behavioral biases and mispricing concepts, exhibit similar performance to the q factors, as noted by Hou, Mo, Xue, and Zhang (2019). Despite the IPCA factors achieving high R-square values, indicative of a strong ability to explain variations in asset returns, they fall short in explaining expected portfolio returns in linear regressions.

The average R^2 value for the two-factor arbitrageur model stands at 13.56%, substantially lower than that of benchmark models. Even with the five-factor model, the R^2 value only matches those of benchmark models. These low R^2 values suggest potential presence of various risk sources in realized returns not captured by the arbitrageur factors, although they may not necessarily offer risk compensation.

Factor models excluding each of the five main factors still demonstrate strong performance in explaining the anomalies. Models without either the Under-reaction to News (NEW) factor or the Attention-Induced Trading (ATT) factor do not match the efficacy of the complete five-factor model, underscoring the significance of these two behavioral aspects. However, they still surpass all existing models documented in the literature in terms of explanatory power on anomalies. Omitting the Chasing Dividend (DIV) factor does not significantly impact the model's explanatory power, as this behavior is distinct and not commonly associated with many anomalies.

Furthermore, the two alternative factor models, encompassing three or five factors and excluding all factors from the main model (as discussed in Section 4.1), also works well. While they explain fewer anomalies than the main model, the number of anomalies explained is still larger than those in all existing factor models in the literature. This outcome reinforces the robustness of the factor selection process and suggests that the superior explanatory power of the main model is not merely due to chance.

5.2 After publication

While the majority of anomalies documented in the literature are accounted for by the five factors, the persistence of 31 anomalies as significant even at the 1% level warrants further investigation. McLean and Pontiff (2016) suggest that if anomaly returns are due to mispricing, the influx of arbitrageurs into the market post-publication, capitalizing on these patterns, should lead to diminished but still positive returns. This partial attenuation is because arbitrageurs require some level of positive returns to offset their risks and costs in trading.

I test post-publication returns in the framework of the arbitrageur factor model. If this model's factors encompass all sources of arbitrageurs' required returns and arbitrageurs are fully cognizant of the anomaly, post-publication alphas should be zero, indicating a fully competitive arbitrage market. By contrast, pre-publication, some positive returns might be present due to the lack of universal awareness among arbitrageurs and the economic rent enjoyed by those capitalizing on their knowledge advantage. This scenario could explain why some anomalies exhibit significant alphas in the full sample.

Table 3 presents these test results. I rerun all tests from the previous subsection, but only for samples starting from the year each anomaly was first published in an academic journal. The findings with the five-factor arbitrageur model suggest that the hypothesis of $\alpha = 0$ post-publication is only rejected for a few anomalies. Specifically, only six anomalies are significant at $|t| > 3$ in linear regressions, and 12 at the 1% level in GRS tests.

In all existing factor models, the number of significant anomalies post-publication is even higher than the number of significant ones in t-tests on anomaly returns without any asset pricing model. These models partially explain the predictive power before publication but fail to account for the explanatory power that persists after publication.

5.3 Discussion of remained prediction power

A few anomalies remain significant in the post-publication sample. Additionally, the average alpha of these anomalies is still positive post-publication. These observations indicate that the arbitrageur factor model does not entirely capture the cross-sectional predictive power. In this subsection, I explore two possible reasons for this.

Firstly, some anomalies may not have garnered the attention of all arbitrageurs even after publication, leading to a deviation from a fully competitive equilibrium. If this hypothesis is correct, we should generally observe lower alphas for anomalies that have been more widely cited.

Figure 3 shows the results of a linear regression of post-publication standardized alphas on the logarithm of citation numbers, sourced from Google Scholar as of April 17th, 2023. A significantly negative correlation is evident. The estimated coefficient implies that well-known anomalies (with around 10,000 citations, such as size, governance, and cash flow to price) tend to have an alpha close to zero post-publication, suggesting that arbitrage activity around these anomalies is highly competitive.

Secondly, it's plausible that the five factors in the model do not encompass all sources of predictive power in asset prices. There might be other systematic behavioral biases or systematic risks beyond market risk that are not accounted for. To investigate this, I categorize the anomalies into six groups, as shown in Table 4. This table presents the average standardized alpha values for each group under various models. It also includes the significance levels of t-tests comparing the average standardized alphas in the model with the average Sharpe ratios of the raw anomaly returns (as listed under "no model"). Here, one star indicates a significance level of $p < 0.05$, and two stars signify $p < 0.01$."

The assignment to the six groups is based on how the anomalies are derived:

1) *Fundamental:* solely from state variables that are not related to the market, such as

accounting data or patent data.

2) *Change:* from changes in fundamental variables.

3) *Market:* solely from past market prices, volumes, and shareholders.

4) *Valuation:* from both fundamental and market variables.

5) *Forecast:* from outside professionals' forecasts, such as earnings forecasts, analyst recommendations, and credit ratings.

6) *Financing:* from companies' financing activities, such as dividends, repurchases, issues, IPOs, takeovers, spinoffs, and exchange switches.

This classification broadly follows the approach used by McLean and Pontiff (2016), with a modification: I further divide their 'Event' group into three distinct groups—*Change*, *Forecast*, and *Financing*—each representing different types of events.

As indicated in Table 4, the five-factor model significantly explains part of the alphas in four out of the six groups post-publication. The average standardized alpha values for anomalies related to market patterns and forecasts are nearly zero, suggesting that these groups are largely accounted for by the chosen factors. However, the model shows limited explanatory power for anomalies related to fundamental changes and those arising from changes in fundamentals.

This pattern provides suggestive evidence that there may be other forms of risk compensation not captured by the model, since fundamental anomalies are more likely to reflect intrinsic risks in firms. However, this conclusion is speculative, and the framework in this paper does not allow for a definitive test of whether there are unrecognized sources of risk beyond market risk.

In Section 6.3, I detail all the anomalies that remain unexplained and explore another potential explanation: the absence of short selling costs in the model. This omission might prevent prices from reaching equilibrium within the framework I propose.

6 Interpretation of Anomalies

6.1 Factor models

In this section, I demonstrate how alphas and beta loadings within the arbitrageur factor model explain cross-sectional return predictors documented in the literature through identifying the sources of mispricing these predictors represent. I begin with an analysis of predictors that have been integrated as factors in established factor models, including the five-factor model by Fama and French (2015) (FF5) augmented with the momentum factor, the q^5 model by Hou, Mo, Xue, and Zhang (2021), the four-factor mispricing model by Stambaugh and Yuan (2017) (M4), and the short- and long-horizon factor model by Daniel, Hirshleifer, and Sun (2020) (DHS). These predictors, chosen as factors, have garnered significant attention from both researchers and practitioners.

Table 5 presents the alphas and beta loadings for these factors. The first column displays the annual Sharpe ratios of returns along with the t-values for tests asserting that the expected returns are non-zero. The second column shows the anomaly alphas, standardized by dividing by the sample standard deviation and multiplying by $\sqrt{12}$, as previously discussed in Section 5.1. Subsequent columns feature the standardized beta values, representing the portion of returns explained by each corresponding factor. These beta values are standardized in the same manner as the alphas:

$$\beta_{i,std} = \frac{\sqrt{12}\beta_i \bar{F}_i}{\sigma}, \tag{19}$$

where $\bar{F}_i$ is the sample mean of the factor return and σ is the sample standard deviation of the anomaly return.

This method allows for a clear demonstration of how the Sharpe ratio of an anomaly return is decomposed through standardized alpha and beta values. Essentially, the Sharpe ratio of the raw anomaly return is equal to the sum of the standardized alpha and the standardized betas for all factors. The numbers in parentheses represent the corresponding t-values from the regressions, testing whether the respective alpha and betas significantly differ from zero.

As indicated in the table, none of the factors within the FF5, q^5, M4, and DHS models display alphas with t-values exceeding 3, suggesting that existing factors are substantially explained by the arbitrageur factor model. Below, I delve into the sources of mispricing that account for the predictive power inherent in these factors.

Three of the four models (FF5, q^5, and M4) incorporate a size factor, which, in the arbitrageur factor model, is explained by Information Salience (SAL) and Chasing Dividends (DIV). The rationale for SAL is straightforward: information about smaller firms is more challenging to analyze and process. DIV's influence may be attributed to dividend chasing behavior correlating with other time-invariant investor preferences for certain stocks. Stocks are more influenced by these behaviors when their total shares outstanding are fewer and their liquidity is lower.

Additionally, three models (q^5, M4, and DHS) include factors primarily explained by Under-reaction to News (NEW). These are profitability in the q^5 model, performance in the M4 model, and PEAD in the DHS model. All these factors capture recent earnings announcement information and the market's slow response to it. The profitability factor in the FF5 model, which uses a different sorting method than the q^5 model, is not explained by NEW. Instead, the FF5 model includes an additional momentum factor to account for the delayed response to recent information. This momentum factor is also partly explained by Chasing Dividends (DIV), as the pattern of chasing past winners is similarly stable.

Several factors reflect biases in Long-term Expectation (EXP), such as value, profitability, and investment in the FF5 model; investment, profitability, and expected growth in the q^5 model; management and performance in the M4 model; and financing in the DHS model. These factors generally measure persistent overvaluation of assets. For instance, stocks with low book-to-market ratios might be overvalued due to overly optimistic growth expectations, leading to lower subsequent returns. Among these, value and investment in the FF5 and q^5 models, investment and expected growth in the q^5 model, management in the M4 model, and financing in the DHS model are also partially attributed to lower Information Salience (SAL).

Finally, the alternative size and management factors in the M4 model, the value and investment factors in the FF5 model, and both factors in the DHS model, also encapsulate some predictive power stemming from short-term misvaluation driven by Attention-Induced Trading (ATT).

6.2 Most cited anomalies

In Table 6, I present the two most cited anomalies in each group as defined in Section 5.3, excluding those already covered in Table 5. Alongside alphas and betas, I also include their post-publication Sharpe ratios and the standardized post-publication alpha, calculated from the sample starting from the year each anomaly was published. The beta loadings help to identify which type of investor behavior is associated with each anomaly.

Fundamental anomalies, book leverage (Fama and French 1992) and governance (Gompers, Ishii, and Metrick 2003), do not show significant positive correlation with any factors and are negatively correlated with the Expectation factor (EXP). This is because they are considered high-quality firms, where investors may have overly optimistic expectations. Unexplained alphas may indicate other sources of mispricing or risk..

Two anomalies related to changes in fundamentals—past revenue growth (Lakonishok, Shleifer, and Vishny 1994) and asset growth (Cooper, Gulen, and Schill 2008)—highlight that firms with higher past revenue growth have higher expected returns, while those with higher past year asset growth have lower subsequent returns. Their loadings on the Attention-Induced Trading factor (ATT) suggest that firms with the highest asset growth and lowest revenue growth attract excessive attention, leading to overvaluation and subsequent correction. The Salience factor (SAL) loading indicates that firms with significant asset changes are difficult to analyze.

The anomaly that illiquid stocks yield higher returns (Amihud 2002) is mainly attributed to Attention-Induced Trading (ATT) and Chasing Dividend (DIV) factors, as less liquid stocks are more impacted by these types of trading volumes. Long-run reversal (De Bondt and Thaler 1985) stems from misvaluation in both Attention-Induced Trading (ATT) and under-study of less salient information (SAL).

For firm valuation measures, cash flow to price (Lakonishok, Shleifer, and Vishny 1994) and earnings to price (Basu 1977) primarily capture mispricing from biases in Long-term Expectation (EXP). The earnings to price ratio is also positively correlated with the reversal of Attention-Induced Trading (ATT) as unsophisticated investors tend to overreact to accruals.

Anomalies related to analysts' forecasts, such as revision (Chan, Jegadeesh, and Lakonishok 1996) and dispersion (Diether, Malloy, and Scherbina 2002), capture both Underreaction to News (NEW) and biases in Long-term Expectation (EXP). They are negatively correlated with the Information Salience factor (SAL), as analysts disseminate recent information, thereby increasing its salience.

Finally, two anomalies from financing activities—lower expected returns for firms 6-36 months post-IPO (Ritter 1991) and higher expected returns for firms with more share repurchases (Ikenberry, Lakonishok, and Vermaelen 1995)—both stem from biases in Long-term Expectation (EXP). This indicates a tendency for firms to opt for equity financing when they believe their valuation is inflated.

6.3 Unexplained anomalies

In Table 7, I showcase the 11 anomalies that remain unexplained in the post-publication sample, aside from book leverage which is included in Table 6. The columns follow the same format as those in Table 6.

Three of these anomalies are linked to limits of arbitrage beyond idiosyncratic risks. Two of them relate to short selling fees, while the third indirectly measures the cost of short selling through option prices. Since the asymmetric cost between longing and short selling is not included in the model, anomalies capturing it should not be explained.

The majority of the remaining unexplained anomalies are fundamentally oriented. They encompass tangibility, leverage, debt issuance, accruals, dividend yield, the concept of the efficient frontier, and the impact of switching exchanges. These anomalies could stem from unique behavioral biases, systematic risks other than market risk, or a lack of sufficient competition among arbitrageurs. Many of these anomalies are tied to intrinsic properties of firms, indicating potential sources of systematic risks, albeit different from the risk factors traditionally documented in the literature. However, this paper does not aim to delve deeper into these aspects.

7 Conclusion

In this study, I introduce an alternative theoretical framework for factor asset pricing models that diverges from the traditional "no arbitrage" assumption. Instead, I acknowledge the existence of arbitrage returns, demonstrating that they are subject to an equilibrium condition. In this model, arbitrageurs are willing to hold only a limited number of shares based on their expected returns. Consequently, returns must be higher when arbitrageurs counterbalance a larger volume of trades stemming from behavioral biases. This leads to the hypothesis that systematic and predictable positions of behavioral investors result in predictable returns, which can be encapsulated by a few factors.

Empirically, I use a robust factor selection process based on the Sharpe ratios of the tangency portfolio. From this process emerges a five-factor model, comprising Underreaction to News, Attention-induced Trading, Chasing Dividends, Information Salience, and Long-term Expectation. This model successfully explains the majority of anomalies identified in the literature. Additionally, the beta loadings within this model provide a clear and intuitive explanation of the sources of mispricing for each anomaly.

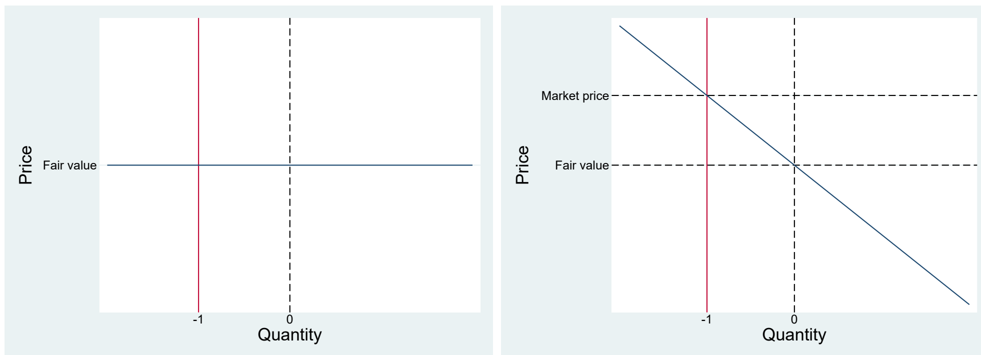
This framework effectively explains stock return predictability and aligns closely with empirical findings. It does not insist that common investors consider risk factors in their investment decisions, and it accommodates the existence of asset mispricing. A few factors are able to explain almost all the predictability because of correlation in sources of mispricing.

While this paper concentrates on asset returns and does not explore the underlying determinants of mispricing, such as the persistence of behavioral trading and the risk tolerance of arbitrageurs, it establishes a foundation for future research. Future studies could build upon this framework to estimate the level of mispricing and explore its implications, particularly in the context of costly equity finance and its impact on firm decision-making.

Figures and Tables

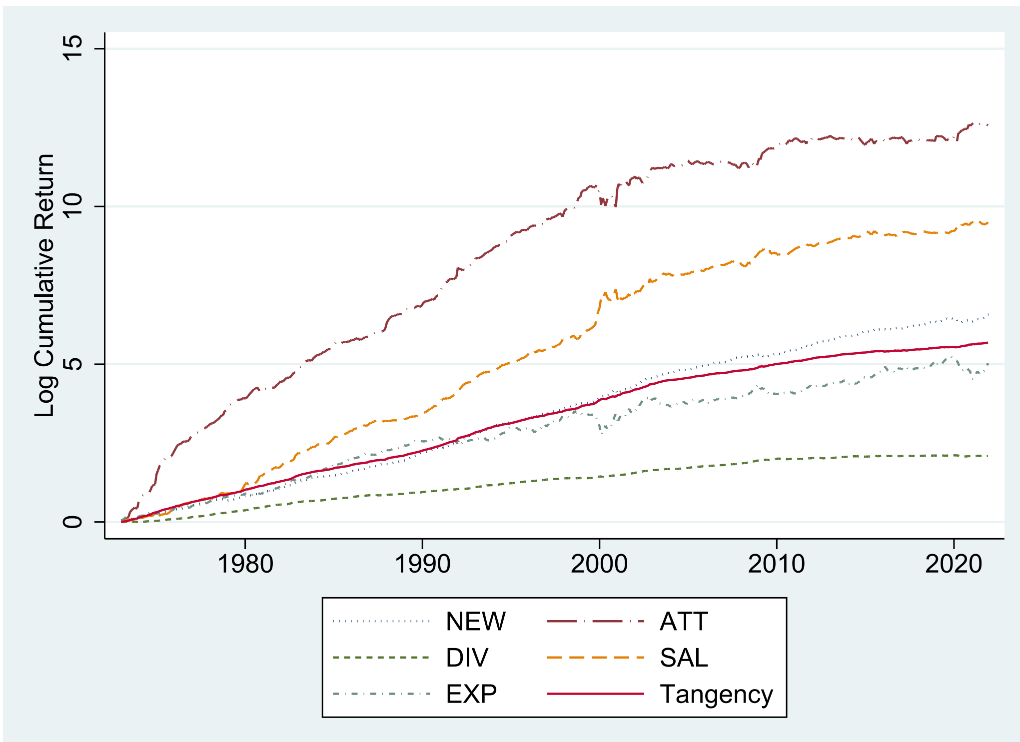
Figure 1: Demand and Supply Curves under Different Assumptions

(a) Traditional asset pricing (b) With limited arbitrageur risk-taking



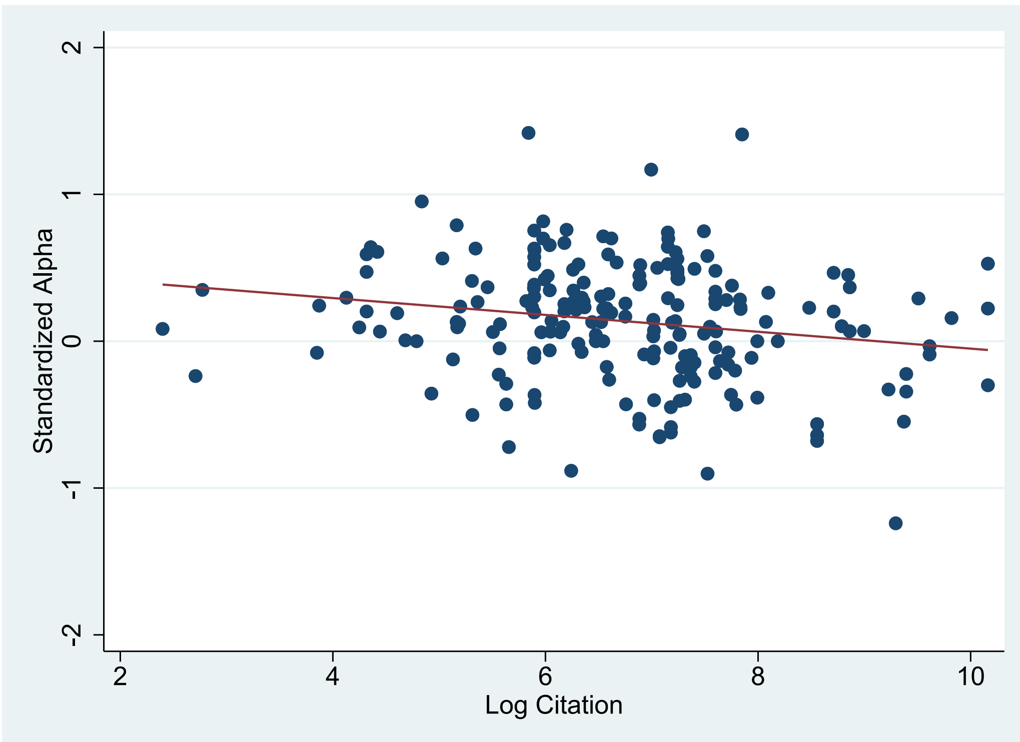
Notes: These graphs intuitively show the equilibrium prices with a positive behavioral investor position under two different assumptions. The x-axis is the shares available to arbitrageurs, opposite with the shares that behavioral investors hold. The y-axis is the price. The supply curve is assumed to be inelastic. The price is at the fair price if the demand curve of arbitrageurs is fully elastic. The price is higher than the fair price if the demand curve is not, when the risk-taking capacity of arbitrageurs is limited.

Figure 2: Time Series of Arbitrageur Factors



Notes: This figure shows the time series of log cumulative returns of the five arbitrageur factors in the main model: Under-reaction to news (NEW), Attention-induced trading (ATT), Chasing Dividend (DIV), Information Salience (SAL), and Long-term Expectation (EXP). They are from data shared by Chen and Zimmermann (2022).

Figure 3: Correlation between numbers of citation and standardized alpha values of anomalies



$$\alpha_{std} = 0.523^{**} - 0.057^{**} \, Log\ Citation$$
$$(3.71) \quad (-2.80)$$

Notes: This figure shows linear regressions of the 207 anomalies. Standardized 5-arbitrageur-factor alphas are on the y-axis, defined as alphas in the arbitrageur factor model divided by standard deviation of the corresponding anomaly return multiplied by $\sqrt{12}$. Log citation numbers are on the x-axis. Citation numbers are from Google Scholar on April 17th, 2023. Regressions are in the post-publication sample starting from the year when the corresponding anomaly was published. t values are reported in parentheses. * $p < 0.05$, ** $p < 0.01$.

Table 1: Factor Selection

Factor Model	Sharpe
(0)RMRF	0.494
Main Models	
(1)RMRF+*AnnouncementReturn*	2.558
(2)RMRF+*AnnouncementReturn*+*STReversal*	4.203
(3)RMRF+*AnnouncementReturn*+*STReversal*+*DivSeason*	4.579
(4)RMRF+*AnnouncementReturn*+*STReversal*+*DivSeason*+*IndRetBig*	4.821
(5)RMRF+*AnnouncementReturn*+*STReversal*+*DivSeason*+*IndRetBig*+*NetEquityFinance*	5.192
(6)RMRF+...+*ShortInterest*	5.463
Alternative Models	
(A1)RMRF+*DelFinLiab*+*DivYieldST*+*NumEarnIncrease*+*Frontier*+*EarningsConsistency*	3.558
(A2)RMRF+*VolumeTrend*+*AnalystRevision*+*BM*+*ChTax*+*Mom12m*	3.154

Notes: This table shows the factors selected in the models as described in Section 4.1 and annual Sharpe ratios of tangency portfolios for each group of factors. Sharpe ratios of tangency portfolios are calculated through estimated expected returns and covariance matrix of anomalies in the sample. The sample period is from January 1973 to December 2021. "Main Models" shows factors selected one by one through the greedy algorithm by maximizing the Sharpe ratio in each step, from the first one to the twelfth. "Alternative Models" show other groups of five factors that exclude previously selected factors. Full anomaly names and citation are in the appendix.

Table 2: Explanation power of factor models on 207 anomalies

	α				GRS test	
	$n(\lvert t\rvert > 2)$	$n(\lvert t\rvert > 3)$	mean(α_{std})	mean(R^2)	$n(p < 0.05)$	$n(p < 0.01)$
Arbitraguer Factors (2)	98	55	0.414	0.1356	98	69
Arbitraguer Factors (3)	89	44	0.352	0.1433	93	62
Arbitraguer Factors (4)	100	60	0.291	0.1744	101	76
Arbitrageur Factors (5)	**50**	**22**	0.201	0.2704	**55**	**31**
Arbitraguer Factors (6)	62	25	**0.128**	0.2898	68	37
Benchmark models						
No model	163	105	0.551			
Fama-French 3	169	136	0.574	0.2221	170	148
Fama-French 5	156	123	0.513	0.2953	157	135
Hou-Mo-Xue-Zhang q^5	130	80	0.411	0.2622	131	108
Stambaugh-Yuan M4	129	88	0.440	0.2551	129	101
Daniel-Hirshleifer-Sun	142	105	0.474	0.1966	144	116
Kelly-Pruit-Su IPCA5	171	154	0.719	**0.3787**	172	160
Excluding one factor						
(5) -*AnnouncementReturn* (NEW)	101	64	0.311	0.2570	105	78
(5) -*STReversal* (ATT)	115	79	0.350	0.2530	116	95
(5) -*DivSeason* (DIV)	65	26	0.256	0.2650	67	35
(5) -*IndRetBig* (SAL)	89	37	0.312	0.2557	91	55
Alternative Factor Models (see Table 1 for specific factor names)						
(A1) 3 Factors	84	43	0.334	0.1580	89	62
(A1) 5 Factors	86	51	0.255	0.2556	89	63
(A2) 3 Factors	120	70	0.377	0.2180	121	86
(A2) 5 Factors	95	54	0.291	0.2960	95	66

Notes: This table shows the time-series regression results and GRS test results of 207 anomaly returns from Chen and Zimmermann (2022) using various asset pricing models. Arbitrageur factor models and alternative factor choices are shown in Table 1. Benchmark models are discussed in Section 3. The sample period is from January 1973 to December 2021 unless otherwise noted in Section 3. Columns 1-2 show statistics of t values of alphas in 207 time-series regressions. Column 3 shows the average standardized alphas in those regressions, defined as alphas in the arbitrageur factor model divided by standard deviation of the corresponding anomaly return multiplied by $\sqrt{12}$, and the annual Sharpe ratio of anomalies in the row of "no model". Column 4 shows the average R^2 values of those regressions. Columns 5-6 show the number of significant GRS tests on each of the 207 anomalies. The smallest count numbers in each column, the smallest $|\alpha_{std}|$, and the largest R^2 are marked in bold.

Table 3: Explanation power after publication of the anomalies

	α				GRS test	
	$n(\|t\| > 2)$	$n(\|t\| > 3)$	mean(α_{std})	mean(R^2)	$n(p < 0.05)$	$n(p < 0.01)$
Arbitraguer Factors (2)	44	11	0.272	0.1626	44	21
Arbitrageur Factors (5)	**29**	**6**	**0.223**	0.3119	**32**	**12**
Benchmark models						
No model	57	19	0.316			
Fama-French 3	93	48	0.382	0.2550	94	70
Fama-French 5	69	30	0.339	0.3291	71	49
Hou-Mo-Xue-Zhang q^5	68	22	0.303	0.2997	69	34
Stambaugh-Yuan M4	73	25	0.323	0.3237	71	35
Daniel-Hirshleifer-Sun	62	26	0.302	0.2449	62	42
Kelly-Pruit-Su IPCA5	108	60	0.441	**0.4661**	108	78

Notes: This table shows the time-series regression results and GRS test results of 207 anomaly returns from Chen and Zimmermann (2022) using various asset pricing models. Arbitrageur factor models and alternative factor choices are shown in Table 1. Benchmark models are discussed in Section 3. The sample period is from January 1973 to December 2021 unless otherwise noted in Section 3. Tests in this table use the sample in or after the year when the anomaly is published. Columns 1-2 show statistics of t values of alphas in 207 time-series regressions. Column 3 shows the average standardized alphas in those regressions, defined as alphas in the arbitrageur factor model divided by standard deviation of the corresponding anomaly return multiplied by $\sqrt{12}$, and the annual Sharpe ratio of anomalies in the row of "no model". Column 4 shows the average R^2 values of those regressions. Columns 5-6 show the number of significant GRS tests on each of the 207 anomalies. The smallest count numbers in each column, the smallest $|\alpha_{std}|$, and the largest R^2 are marked in bold.

Table 4: Post-publication explanation power on six groups of anomalies

	Fundamental	Change	Market	Valuation	Forecast	Financing
Average post-publication α_{std} in different models						
Arbitraguer Factors (2)	0.209	0.298	0.098**	0.373	-0.013**	0.420
Arbitraguer Factors (5)	0.203	0.226	0.020**	0.220*	0.057**	0.259**
Other models						
No model	0.190	0.264	0.349	0.383	0.307	0.482
Fama-French 3	0.165	0.324	0.438	0.411	0.462	0.609
Fama-French 5	0.177	0.290	0.405	0.366	0.363	0.466
Hou-Mo-Xue-Zhang q^5	0.126	0.258	0.411	0.375	0.265	0.411
Stambaugh-Yuan M4	0.149	0.303	0.368	0.468	0.289	0.448
Daniel-Hirshleifer-Sun	0.182	0.276	0.364	0.346	0.251	0.381*
Kelly-Pruit-Su IPCA5	0.100*	0.432	0.429	0.517	0.592	0.782
Average full-sample Sharpe	0.701	0.547	0.694	0.385	0.546	0.500
n(Anomalies)	43	36	69	18	20	21

Notes: This table shows the average post-publication standardized α values in each group using various asset pricing models, and the annual Sharpe ratio of anomalies in the row of "no model". Arbitrageur factor models are shown in Table 1. Benchmark models are discussed in Section 3. The sample period is from January 1973 to December 2021 unless otherwise noted in Section 3. Standardized alphas are defined as alphas in the arbitrageur factor model divided by standard deviation of the corresponding anomaly return multiplied by $\sqrt{12}$. Tests in this table use the sample in or after the year when the anomaly is published. "Fundamental" includes anomalies resulting from fundamental states, "Change" includes anomalies resulting from changes in fundamentals, "Market" includes anomalies resulting from only market data, "Valuation" includes anomalies resulting from both fundamental and market data, "Forecast" includes anomalies resulting from analysts and credit ratings, and "Financing" includes anomalies resulting from firm financing activities. Stars indicate the significance level of t tests where the null hypothesis is the average standardized alpha value in the model is equal to the average Sharpe ratio of anomalies. $*p < 0.05$, $**p < 0.01$.

Table 5: Alphas and beta loadings of factors in existing models

	Sharpe	α	β_{NEW}	β_{ATT}	β_{DIV}	β_{SAL}	β_{EXP}
Fama and French (2015) 5 factors and the momentum factor							
SMB	0.221	-0.206	0.059	0.112	0.264**	0.150*	-0.224**
(Size)	(1.550)	(-0.909)	(0.527)	(1.703)	(3.066)	(2.338)	(-6.405)
HML	0.317*	0.077	-0.397**	0.213**	-0.054	0.152*	0.387**
(Value)	(2.216)	(0.350)	(-3.631)	(3.341)	(-0.651)	(2.442)	(11.410)
RMW	0.440**	-0.071	-0.103	0.080	0.071	-0.056	0.523**
(Profitability)	(3.079)	(-0.367)	(-1.083)	(1.427)	(0.970)	(-1.041)	(17.664)
CMA	0.529**	0.060	-0.068	0.179*	-0.044	0.310**	0.240**
(Investment)	(3.702)	(0.272)	(-0.622)	(2.805)	(-0.532)	(4.969)	(7.074)
Mom	0.483**	-0.387	0.935**	-0.216**	0.306**	-0.068	-0.053
(Momentum)	(3.379)	(-1.802)	(8.776)	(-3.465)	(3.760)	(-1.118)	(-1.605)
Hou, Mo, Xue, and Zhang (2021) q^5 factors							
I/A	0.606**	0.002	-0.083	0.186*	0.037	0.294**	0.292**
(Investment)	(4.245)	(0.010)	(-0.758)	(2.917)	(0.442)	(4.728)	(8.604)
Roe	0.692**	-0.181	0.575**	-0.045	0.124	-0.085	0.313**
(Profitability)	(4.841)	(-0.871)	(5.584)	(-0.748)	(1.573)	(-1.444)	(9.804)
Eg	1.345**	0.538*	0.312**	0.117	0.036	0.184**	0.292**
(Expected Growth)	(9.414)	(2.622)	(3.060)	(1.974)	(0.468)	(3.180)	(9.242)
Stambaugh and Yuan (2017) M4 factors							
SMB	0.521**	-0.083	0.041	0.176*	0.291*	0.171*	-0.132**
(Size)	(3.453)	(-0.307)	(0.306)	(2.249)	(2.767)	(2.284)	(-3.095)
MGMT	0.794**	0.150	-0.106	0.193**	-0.029	0.249**	0.483**
(Management)	(5.264)	(0.742)	(-1.056)	(3.290)	(-0.366)	(4.440)	(15.029)
PERF	0.568**	-0.558*	0.877**	-0.120	0.288**	0.007	0.128**
(Performance)	(3.770)	(-2.314)	(7.333)	(-1.718)	(3.057)	(0.104)	(3.348)
Daniel, Hirshleifer, and Sun (2020) 3 factors							
PEAD	1.108**	-0.605**	1.679**	0.184**	-0.082	0.113*	-0.137**
(Announcement)	(7.754)	(-3.173)	(17.721)	(3.319)	(-1.132)	(2.092)	(-4.670)
FIN	0.618**	0.165	-0.231*	0.182**	-0.070	0.135**	0.567**
(Financing)	(4.325)	(1.061)	(-2.982)	(4.030)	(-1.178)	(3.063)	(23.577)

Notes: This table shows results of regressing factors in other factor models in literature on the five arbitrageur factors. The five factor selected for the arbitrageur factor model are discussed in Section 4.2. NEW factor is under-reaction to news measured by post-earnings-announcement drift. ATT factor is attention-induced trading measured by one-month return reversal. DIV factor is from dividend seasonality. SAL factor is information salience measured by lag of return for small firms. EXP factor is biases in long term expectation measured by net equity finance. The sample period is from January 1973 to December 2021 unless otherwise noted in Section 3. The first column shows Sharpe ratio of factors returns in the full sample. Following columns show standardized alphas and betas. $\alpha_{std} = \sqrt{12}\alpha/\sigma$, and $\beta_{i,std} = \sqrt{12}\beta_i \bar{F}_i/\sigma$, where σ is the standard deviation of the corresponding anomaly return, and $\bar{F}_i$ is the sample mean of the corresponding factor return. t values for alphas and betas are reported in parentheses. * $|t| > 2$, ** $|t| > 3$.

Table 6: Alphas and beta loadings of some most cited anomalies

	α	β_{NEW}	β_{ATT}	β_{DIV}	β_{SAL}	β_{EXP}	PP Sharpe	PP α
Fundamental								
BookLeverage	0.848**	0.105	-0.074	-0.087	0.019	-0.642**	0.130	0.648**
	(5.255)	(1.315)	(-1.583)	(-1.430)	(0.411)	(-25.826)	(0.913)	(3.067)
Governance	0.041	0.016	-0.194	0.116	-0.087	-0.175**	-0.144	-1.336
	(0.089)	(0.063)	(-1.688)	(0.594)	(-0.640)	(-3.788)	(-0.582)	(-1.938)
Change								
RevGrowth	-0.058	0.018	0.223**	0.057	0.130	0.079*	0.367*	0.361
	(-0.235)	(0.146)	(3.132)	(0.610)	(1.879)	(2.082)	(2.572)	(1.392)
AssetGrowth	0.124	0.147	0.351**	0.203*	0.503**	-0.238**	0.984**	0.743*
	(0.579)	(1.382)	(5.637)	(2.495)	(8.303)	(-7.183)	(6.885)	(2.646)
Market								
Illiquidity	-0.566*	0.317*	0.239**	0.358**	0.169*	0.002	0.306*	0.194
	(-2.605)	(2.939)	(3.785)	(4.348)	(2.751)	(0.048)	(2.144)	(1.168)
LRReversal	-0.252	-0.160	0.428**	0.117	0.517**	-0.221**	0.381*	-0.135
	(-1.210)	(-1.545)	(7.089)	(1.484)	(8.793)	(-6.878)	(2.669)	(-0.617)
Valuation								
CashFlow	0.519*	-0.273*	-0.124*	-0.096	-0.217**	0.435**	0.273	-0.108
	(2.534)	(-2.682)	(-2.084)	(-1.238)	(-3.758)	(13.769)	(1.909)	(-0.414)
EarningsToPrice	0.358	-0.613**	0.204*	0.099	0.100	0.196**	0.317*	0.234
	(1.518)	(-5.235)	(2.989)	(1.107)	(1.502)	(5.402)	(2.219)	(1.010)
Forecast								
Revision	-0.170	0.705**	-0.047	0.028	-0.226**	0.139**	0.431*	-0.411
	(-0.728)	(5.887)	(-0.751)	(0.289)	(-3.360)	(3.877)	(2.904)	(-1.268)
Dispersion	-0.051	0.350**	-0.083	-0.023	-0.150**	0.425**	0.342*	0.023
	(-0.324)	(4.361)	(-1.978)	(-0.356)	(-3.315)	(17.753)	(2.316)	(0.109)
Financing								
IPO	0.367	-0.259*	0.016	-0.146	0.111	0.451**	0.464**	0.320
	(1.832)	(-2.511)	(0.286)	(-1.770)	(1.917)	(14.698)	(3.173)	(1.455)
ShareRepurchase	-0.046	0.076	-0.059	-0.178*	-0.089*	0.662**	0.394*	-0.282
	(-0.383)	(1.264)	(-1.692)	(-3.881)	(-2.619)	(35.574)	(2.760)	(-1.852)

Notes: This table shows results of regressing two of the most cited anomaly return series in each group in literature on the five arbitrageur factors. Groups of anomalies are discussed in Section 5.3. The five factor selected for the arbitrageur factor model are discussed in Section 4.2. NEW factor is under-reaction to news measured by post-earnings-announcement drift. ATT factor is attention-induced trading measured by one-month return reversal. DIV factor is from dividend seasonality. SAL factor is information salience measured by lag of return for small firms. EXP factor is biases in long term expectation measured by net equity finance. The sample period is from January 1973 to December 2021. The first six columns show standardized alphas and betas. $\alpha_{std} = \sqrt{12}\alpha/\sigma$, and $\beta_{i,std} = \sqrt{12}\beta_i \bar{F}_i/\sigma$, where σ is the standard deviation of the corresponding anomaly return, and $\bar{F}_i$ is the sample mean of the corresponding factor return. The last two columns show the Sharpe ratio of the anomaly return and standadized regression alphas in the sampkle from the year when the anomaly was published. t values are reported in parentheses. * $|t| > 2$, ** $|t| > 3$.

Table 7: Unexplained anomalies

	α	β_{NEW}	β_{ATT}	β_{DIV}	β_{SAL}	β_{EXP}	PP Sharpe	PP α
Limit to arbitrage								
IOShortInterest	1.330**	-0.351*	-0.182*	-0.113	-0.039	-0.067	0.573**	1.759**
	(5.123)	(-2.609)	(-2.811)	(-1.101)	(-0.514)	(-1.744)	(3.707)	(3.788)
ShortInterest	0.922**	0.268*	0.033	0.047	-0.090	-0.036	0.879**	1.241**
	(4.450)	(2.597)	(0.557)	(0.595)	(-1.531)	(-1.122)	(6.149)	(4.026)
SmileSlope	1.138**	0.128	0.364**	0.197*	0.133	0.081*	2.014**	1.204**
	(4.224)	(0.977)	(7.273)	(2.220)	(1.970)	(2.780)	(10.069)	(4.151)
Insufficient notice, other behavioral biases, or other risk compensation								
Tangibility	0.914**	0.114	-0.140*	0.060	0.130*	-0.606**	0.398*	0.351
	(5.260)	(1.316)	(-2.778)	(0.913)	(2.646)	(-22.616)	(2.785)	(1.337)
NetDebtPrice	0.841**	0.259*	-0.072	-0.036	0.049	-0.419**	0.519**	0.939*
	(3.834)	(2.375)	(-1.135)	(-0.436)	(0.794)	(-12.372)	(3.635)	(2.657)
DebtIssuance	0.793**	0.218*	-0.028	0.254**	0.125*	-0.456**	0.733**	0.809*
	(3.955)	(2.192)	(-0.475)	(3.340)	(2.213)	(-14.765)	(5.134)	(2.859)
Accruals	0.969**	0.050	-0.107	0.052	-0.042	-0.191**	0.699**	0.178
	(3.951)	(0.413)	(-1.501)	(0.558)	(-0.599)	(-5.051)	(4.896)	(0.670)
PctAcc	0.795**	-0.243*	-0.110	-0.052	0.035	0.217**	0.611**	-0.082
	(3.345)	(-2.060)	(-1.598)	(-0.572)	(0.522)	(5.926)	(4.279)	(-0.250)
DivYieldST	0.996**	-0.369**	0.190**	0.528**	0.076	0.044	1.280**	0.881**
	(4.565)	(-3.406)	(3.000)	(6.379)	(1.234)	(1.300)	(8.962)	(3.729)
Frontier	1.015**	-0.334*	0.333**	-0.048	0.343**	-0.097*	1.053**	0.978*
	(4.495)	(-2.981)	(5.086)	(-0.564)	(5.378)	(-2.793)	(7.374)	(2.856)
ExchSwitch	0.959**	-0.156	0.075	-0.137	0.209**	-0.154**	0.642**	0.922*
	(4.090)	(-1.341)	(1.100)	(-1.537)	(3.161)	(-4.254)	(4.492)	(2.783)

Notes: This table shows results of regressing all unexplained anomalies ($p < 0.01$ in GRS tests in the post-publication sample) in literature on the five arbitrageur factors. The five factor selected for the arbitrageur factor model are discussed in Section 4.2. NEW factor is under-reaction to news measured by post-earnings-announcement drift. ATT factor is attention-induced trading measured by one-month return reversal. DIV factor is from dividend seasonality. SAL factor is information salience measured by lag of return for small firms. EXP factor is biases in long term expectation measured by net equity finance. The sample period is from January 1973 to December 2021. The first six columns show standardized alphas and betas. $\alpha_{std} = \sqrt{12}\alpha/\sigma$, and $\beta_{i,std} = \sqrt{12}\beta_i \bar{F}_i/\sigma$, where σ is the standard deviation of the corresponding anomaly return, and $\bar{F}_i$ is the sample mean of the corresponding factor return. The last two columns show the Sharpe ratio of the anomaly return and standadized regression alphas in the sampkle from the year when the anomaly was published. t values are reported in parentheses. * $|t| > 2$, ** $|t| > 3$.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 159–178.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz (2022). Attention-induced trading and returns: Evidence from robinhood users. *The Journal of Finance 77*(6), 3141–3190.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance 32*(3), 663–682.

Ben-David, I., J. Li, A. Rossi, and Y. Song (2022). What do mutual fund investors really care about? *The Review of Financial Studies 35*(4), 1723–1774.

Ben-Rephael, A., Z. Da, and R. D. Israelsen (2017). It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies 30*(9), 3009–3047.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2019). Diagnostic expectations and stock returns. *The Journal of Finance 74*(6), 2839–2874.

Bordalo, P., N. Gennaioli, and A. Shleifer (2022). Salience. *Annual Review of Economics 14*, 521–544.

Bradshaw, M. T., S. A. Richardson, and R. G. Sloan (2006). The relation between corporate financing activities, analysts' forecasts and stock returns. *Journal of Accounting and Economics 42*(1-2), 53–85.

Bräuer, K., A. Hackethal, and T. Hanspal (2022). Consuming dividends. *The Review of Financial Studies 35*(10), 4802–4857.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*(2), 379–421.

Chan, L. K., N. Jegadeesh, and J. Lakonishok (1996). Momentum strategies. *The Journal of Finance 51*(5), 1681–1713.

Chen, A. Y. and T. Zimmermann (2022). Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming*.

Chinco, A., S. M. Hartzmark, and A. B. Sussman (2022). A new test of risk factor relevance. *The Journal of Finance 77*(4), 2183–2238.

Cho, T. (2020). Turning alphas into betas: Arbitrage and endogenous risk. *Journal of Financial Economics 137*(2), 550–570.

Choi, J. J. and A. Z. Robertson (2020). What matters to individual investors? evidence from the horse's mouth. *The Journal of Finance 75*(4), 1965–2020.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *The Journal of Finance 63*(4), 1977–2011.

Cohen, L. and D. Lou (2012). Complicated firms. *Journal of financial economics 104*(2), 383–400.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *the Journal of Finance 63*(4), 1609–1651.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short-and long-horizon behavioral factors. *The Review of Financial Studies 33*(4), 1673–1736.

De Bondt, W. F. and R. Thaler (1985). Does the stock market overreact? *The Journal of Finance 40*(3), 793–805.

DellaVigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *The Journal of Finance 64*(2), 709–749.

Diether, K. B., C. J. Malloy, and A. Scherbina (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance 57*(5), 2113–2141.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *The Journal of Finance 47*(2), 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*(1), 3–56.

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

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Foster, G., C. Olsen, and T. Shevlin (1984). Earnings releases, anomalies, and the behavior of security returns. *Accounting Review*, 574–603.

Frazzini, A. (2006). The disposition effect and underreaction to news. *The Journal of Finance 61*(4), 2017–2046.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121–1152.

Gompers, P., J. Ishii, and A. Metrick (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics 118*(1), 107–156.

Grinblatt, M. and B. Han (2005). Prospect theory, mental accounting, and momentum. *Journal of Financial Economics 78*(2), 311–339.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American Economic Review 70*(3), 393–408.

Hartzmark, S. M. and D. H. Solomon (2013). The dividend month premium. *Journal of Financial Economics 109*(3), 640–660.

Hartzmark, S. M. and D. H. Solomon (2019). The dividend disconnect. *The Journal of Finance 74*(5), 2153–2199.

Hartzmark, S. M. and D. H. Solomon (2021). Predictable price pressure. *Available at SSRN 3853096*.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *American Economic Review 103*(2), 732–770.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance 64*(5), 2289–2325.

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *The Review of Financial Studies 20*(4), 1113–1138.

Hou, K., H. Mo, C. Xue, and L. Zhang (2019). Which factors? *Review of Finance 23*(1), 1–35.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021). An augmented q-factor model with expected growth. *Review of Finance 25*(1), 1–41.

Ikenberry, D., J. Lakonishok, and T. Vermaelen (1995). Market underreaction to open market share repurchases. *Journal of Financial Economics 39*(2-3), 181–208.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance 45*(3), 881–898.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Keloharju, M., J. T. Linnainmaa, and P. Nyberg (2021). Long-term discount rates do not vary across firms. *Journal of Financial Economics 141*(3), 946–967.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance 49*(5), 1541–1578.

Lochstoer, L. A. and P. C. Tetlock (2020). What drives anomaly returns? *The Journal of Finance 75*(3), 1417–1455.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *The Journal of Finance 71*(1), 5–32.

Menzly, L. and O. Ozbas (2010). Market segmentation and cross-predictability of returns. *The Journal of Finance 65*(4), 1555–1580.

Nagel, S. and Z. Xu (2022). Asset pricing with fading memory. *The Review of Financial Studies 35*(5), 2190–2245.

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics 42*(1-2), 35–52.

Ritter, J. R. (1991). The long-run performance of initial public offerings. *The Journal of Finance 46*(1), 3–27.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Shleifer, A. (1986). Do demand curves for stocks slope down? *The Journal of Finance 41*(3), 579–590.

Stambaugh, R. F. and Y. Yuan (2017). Mispricing factors. *The Review of Financial Studies 30*(4), 1270–1315.

A Model Solution

A.1 Proof of Proposition 1

Proof. Guess that

$$P_{i,t} = V_{i,t}(1 + \sum_{j=1}^{J} k_i^{(j)} X_{i,t}^{(j)}). \tag{20}$$

Then

$$\begin{aligned}
1 + r_{i,t+1} &= \frac{P_{i,t+1} + D_{i,t+1}}{P_{i,t}} && (21)\\
&= \frac{V_{i,t+1}(1 + \sum_{j=1}^{J} k_i^{(j)} X_{i,t+1}^{(j)}) + D_{i,t+1}}{V_{i,t}(1 + \sum_{j=1}^{J} k_i^{(j)} X_{i,t}^{(j)})} && (22)\\
&= \frac{V_{i,t+1} + D_{i,t+1}}{V_{i,t}} \frac{1 + (1 - d_i)\left(\sum_{j=1}^{J} k_i^{(j)} X_{i,t+1}^{(j)}\right)}{1 + \sum_{j=1}^{J} k_i^{(j)} X_{i,t}^{(j)}} && (23)\\
&= (1 + R_{i,t} + \epsilon_{i,t+1}^{V}) \frac{1 + (1 - d_i)\left(\sum_{j=1}^{J} k_i^{(j)} (\delta_j X_{i,t}^{(j)} + \epsilon_{i,j,t+1})\right)}{1 + \sum_{j=1}^{J} k_i^{(j)} X_{i,t}^{(j)}} && (24)\\
&= 1 + R_{i,t} + \epsilon_{i,t+1}^{V} + \sum_{j=1}^{J} \left[((1 - d_i)\delta_j - 1) k_i^{(j)} X_{i,t}^{(j)} + (1 - d_i)\epsilon_{i,j,t+1}\right] && (25)\\
&= 1 + r_{f,t} + \beta_i(r_{M,t+1} - r_{f,t}) + \sum_{j=1}^{J} b_i^{(j)} X_{i,t}^{(j)} + \varepsilon_{i,t+1}, && (26)
\end{aligned}$$

where

$$\varepsilon_{i,t+1} = \epsilon_{i,t+1}^{V} + (1 - d_i) \sum_{j=1}^{J} \epsilon_{i,j,t+1} \tag{27}$$

is the total idiosyncratic risk with variance σ_i^2, and

$$b_i^{(j)} = ((1 - d_i)\delta_j - 1) k_i^{(j)} \tag{28}$$

is the loading of expected return on a certain noise trading level.

Note that Equation (25) is a linear approximation, which holds when the expected returns are close to zero.

The total CAPM alpha that the arbitrageur optimizes $\sum_{i=1}^{N} X_{i,t}(r_{i,t+1}^{\alpha})$ is

$$\sum_{i=1}^{N} \sum_{j=1}^{J} b_i^{(j)} X_{i,t}^{(j)} X_{i,t} + \sum_{i=1}^{N} \varepsilon_{i,t+1} X_{i,t}. \tag{29}$$

Her first-order condition on $X_{i,t}$ is

$$\frac{\partial(-\log - E_t U(W_{t+1}))}{\partial X_{i,t}} = \frac{\partial(aEW_{t+1} - \frac{1}{2}a^2 Var(W_{t+1}))}{\partial X_{i,t}} \tag{30}$$

$$= a\sum_{j=1}^{J} b_i^{(j)} X_{i,t}^{(j)} - a^2\sigma_i^2 X_{i,t} = 0, \tag{31}$$

and the solution is

$$X_{i,t} = \sum_{j=1}^{J} \frac{b_i^{(j)}}{a\sigma_i^2} X_{i,t}^{(j)}. \tag{32}$$

Comparing it with the market clear condition in Equation (8) gives

$$b_i^{(j)} = -a\sigma_i^2, \tag{33}$$

$$k_i^{(j)} = \frac{a\sigma_i^2}{1-(1-d_i)\delta_j}. \tag{34}$$

This shows that the guess in Equation (20) is one possible equilibrium. In this equilibrium,

$$P_{i,t} = V_{i,t}(1 + \sum_{j=1}^{J} \frac{a\sigma_i^2}{1-(1-d_i)\delta_j} X_{i,t}^{(j)}), \tag{35}$$

and

$$r_{i,t+1} = r_{f,t} + \beta_i(r_{M,t+1} - r_{f,t}) + \sum_{j=1}^{J} b_i^{(j)} X_{i,t}^{(j)} + \varepsilon_{i,t+1}. \tag{36}$$

□

A.2 Proof of proposition 2

Proof. The return of factor j is

$$F_{j,t} = \sum_{i=1}^{N} w_{i,t}^{(j)} r_{f,t} + \sum_{i=1}^{N} w_{i,t}^{(j)} \beta_i(r_{M,t+1} - r_{f,t}) + \sum_{i=1}^{N}\sum_{j'=1}^{J} w_{i,t}^{(j)} b_i^{(j')} X_{i,t}^{(j')} + \sum_{i=1}^{N} w_{i,t}^{(j)} \varepsilon_{i,t+1} \tag{37}$$

$$= \sum_{i=1}^{N} w_{i,t}^{(j)} b_i^{(j)} X_{i,t}^{(j)} + \sum_{i=1}^{N} w_{i,t}^{(j)} \varepsilon_{i,t+1} \tag{38}$$

From the assumption in Equation 11, let the portfolio k to be factor j, we have that

$$\beta_j^{(j,M)} = \frac{\sum_{i=1}^{N} w_{i,t}^{(j)} b_i^{(j)} X_{i,t}^{(j)}}{\sum_{i=1}^{N} w_{i,t}^{M} b_i^{(j)} X_{i,t}^{(j)}} \tag{39}$$

is also time-invariant. Define

$$\beta_k^{(j)} = \frac{\beta_k^{(j,M)}}{\beta_j^{(j,M)}} = \frac{\sum_{i=1}^N w_{i,t}^{(k)} b_i^{(j)} X_{i,t}^{(j)}}{\sum_{i=1}^N w_{i,t}^{(j)} b_i^{(j)} X_{i,t}^{(j)}}. \tag{40}$$

Under this notation,

$$r_{k,t+1} = \sum_{i=1}^N w_{i,t}^{(k)} r_{f,t} + \sum_{i=1}^N w_{i,t}^{(k)} \beta_i (r_{M,t+1} - r_{f,t}) + \sum_{i=1}^N \sum_{j'=1}^J w_{i,t}^{(k)} b_i^{(j)} X_{i,t}^{(j)} + \sum_{i=1}^N w_{i,t}^{(k)} \varepsilon_{i,t+1} \tag{41}$$

$$= r_{f,t} + \beta_k (r_{M,t+1} - r_{f,t}) + \sum_{j=1}^J \beta_k^{(j)} (\sum_{i=1}^N w_{i,t}^{(j)} b_i^{(j)} X_{i,t}^{(j)}) + \sum_{i=1}^N w_{i,t}^{(k)} \varepsilon_{i,t+1} \tag{42}$$

$$= r_{f,t} + \beta_k (r_{M,t+1} - r_{f,t}) + \sum_{j=1}^J \beta_k^{(j)} (F_{j,t} - \sum_{i=1}^N w_{i,t}^{(j)} \varepsilon_{i,t+1}) + \sum_{i=1}^N w_{i,t}^{(k)} \varepsilon_{i,t+1} \tag{43}$$

$$= r_{f,t} + \beta_k (r_{M,t+1} - r_{f,t}) + \sum_{j=1}^J \beta_k^{(j)} F_{j,t} + \varepsilon_{k,t+1}^P. \tag{44}$$

where $\varepsilon_{k,t+1}^P = \sum_{i=1}^N (w_{i,t}^{(k)} - \beta_k^{(j)} w_{i,t}^{(j)}) \varepsilon_{i,t+1}$ is not autocorrelated and has mean zero. □

A.3 Proof of proposition 3

Proof. Denote

$$k_{i,t}^{(j)} = \frac{a_t \sigma_i^2}{1 - (1 - d_i)\delta_{j,t}}, k_{i,t+1}^{(j)} = \frac{a_{t+1} \sigma_i^2}{1 - (1 - d_i)\delta_{j,t+1}}. \tag{45}$$

Then

$$
\begin{aligned}
1 + r_{i,t+1} &= \frac{P_{i,t+1} + D_{i,t+1}}{P_{i,t}} && (46)\\
&= \frac{V_{i,t+1}(1 + \sum_{j=1}^{J} k_{i,t+1}^{(j)} X_{i,t+1}^{(j)}) + D_{i,t+1}}{V_{i,t}(1 + \sum_{j=1}^{J} k_{i,t}^{(j)} X_{i,t}^{(j)})} && (47)\\
&= \frac{V_{i,t+1} + D_{i,t+1}}{V_{i,t}} \frac{1 + (1 - d_i)\left(\sum_{j=1}^{J} k_{i,t+1}^{(j)} X_{i,t+1}^{(j)}\right)}{1 + \sum_{j=1}^{J} k_{i,t}^{(j)} X_{i,t}^{(j)}} && (48)\\
&= (1 + R_{i,t} + \epsilon_{i,t+1}^{V}) \frac{1 + (1 - d_i)\left(\sum_{j=1}^{J} k_{i,t+1}^{(j)} (\delta_j X_{i,t}^{(j)} + \epsilon_{i,j,t+1})\right)}{1 + \sum_{j=1}^{J} k_{i,t}^{(j)} X_{i,t}^{(j)}} && (49)\\
&= 1 + R_{i,t} + \epsilon_{i,t+1}^{V} + \sum_{j=1}^{J} \left[((1 - d_i)\delta_{j,t+1} k_{i,t+1}^{(j)} - k_{i,t}^{(j)}) X_{i,t}^{(j)} + (1 - d_i)\epsilon_{i,j,t+1}\right] && (50)\\
&= 1 + r_{f,t} + \beta_i (r_{M,t+1} - r_{f,t}) + \sum_{j=1}^{J} b_{i,t+1}^{(j)} X_{i,t}^{(j)} + \varepsilon_{i,t+1}, && (51)
\end{aligned}
$$

where

$$
\begin{aligned}
b_{i,t+1}^{(j)} &= (1 - d_i)\delta_{j,t+1} \frac{a_{t+1}\sigma_i^2}{1 - (1 - d_i)\delta_{j,t+1}} - \frac{a_t \sigma_i^2}{1 - (1 - d_i)\delta_{j,t}} && (52)\\
&= -a_{t+1}\sigma_i^2 + \frac{a_{t+1}\sigma_i^2}{1 - (1 - d_i)\delta_{j,t+1}} - \frac{a_t \sigma_i^2}{1 - (1 - d_i)\delta_{j,t}}, && (53)
\end{aligned}
$$

and

$$
\varepsilon_{i,t+1} = \epsilon_{i,t+1}^{V} + (1 - d_i) \sum_{j=1}^{J} \epsilon_{i,j,t+1}. \qquad (54)
$$

□

B Anomalies

This section displays all 207 anomalies from the open-source asset pricing website by Chen and Zimmermann (2022), which were used as both the pool of candidate factors and the test sample in this study. Anomalies are categorized into groups as defined in Section 5.3. Acronyms, which are used in Table 1, are included for easy reference, along with the names of the authors and the publication years. Brief descriptions provided by Chen and Zimmermann (2022) are also given. Additionally, the table provides the citation counts for each anomaly, as sourced from Google Scholar on April 17th, 2023.

Acronym	Authors	Publication	Description	Citation
Fundamental				
AbnormalAccruals	Xie	2001	Abnormal Accruals	1794
Accruals	Sloan	1996	Accruals	6962
AdExp	Chan, Lakonishok and Sougiannis	2001	Advertising Expense	2527
BookLeverage	Fama and French	1992	Book leverage (annual)	25911
BrandInvest	Belo, Lin and Vitorino	2014	Brand capital investment	202
Cash	Palazzo	2012	Cash to assets	356
CashProd	Chandrashekar and Rao	2009	Cash Productivity	48
CBOperProf	Ball et al.	2016	Cash-based operating profitability	395
CitationsRD	Hirschleifer, Hsu and Li	2013	Citations to RD expenses	894
DelBreadth	Chen, Hong and Stein	2002	Breadth of ownership	1500
DelDRC	Prakash and Sinha	2012	Deferred Revenue	70
EarningsConsistency	Alwathainani	2009	Earnings consistency	47
ExclExp	Doyle, Lundholm and Soliman	2003	Excluded Expenses	626
FirmAge	Barry and Brown	1984	Firm age based on CRSP	856
FR	Franzoni and Marin	2006	Pension Funding Status	286
Governance	Gompers, Ishii and Metrick	2003	Governance Index	10884
GP	Novy-Marx	2013	gross profits / total assets	2332
Herf	Hou and Robinson	2006	Industry concentration (sales)	1113
HerfAsset	Hou and Robinson	2006	Industry concentration (assets)	1113
HerfBE	Hou and Robinson	2006	Industry concentration (equity)	1113
Investment	Titman, Wei and Xie	2004	Investment to revenue	2009
MS	Mohanram	2005	Mohanram G-score	477
NOA	Hirshleifer et al.	2004	Net Operating Assets	984
OperProf	Fama and French	2006	operating profits / book equity	1327
OperProfRD	Ball et al.	2016	Operating profitability R&D adjusted	395
OPLeverage	Novy-Marx	2010	Operating leverage	412
OrderBacklog	Rajgopal, Shevlin, Venkatachalam	2003	Order backlog	260
OrgCap	Eisfeldt and Papanikolaou	2013	Organizational capital	855

OScore	Dichev	1998	O Score	1410
PatentsRD	Hirschleifer, Hsu and Li	2013	Patents to RD expenses	894
PctAcc	Hafzalla, Lundholm, Van Winkle	2011	Percent Operating Accruals	175
PctTotAcc	Hafzalla, Lundholm, Van Winkle	2011	Percent Total Accruals	175
RDAbility	Cohen, Diether and Malloy	2013	R&D ability	427
RDcap	Li	2011	R&D capital-to-assets	523
RDS	Landsman et al.	2011	Real dirty surplus	62
realestate	Tuzel	2010	Real estate holdings	209
roaq	Balakrishnan, Bartov and Faurel	2010	Return on assets (qtrly)	263
RoE	Haugen and Baker	1996	net income / book equity	1422
sinAlgo	Hong and Kacperczyk	2009	Sin Stock (selection criteria)	2514
tang	Hahn and Lee	2009	Tangibility	181
Tax	Lev and Nissim	2004	Taxable income to income	750
TotalAccruals	Richardson et al.	2005	Total accruals	1996
Change				
AssetGrowth	Cooper, Gulen and Schill	2008	Asset growth	1791
ChAssetTurnover	Soliman	2008	Change in Asset Turnover	718
ChEQ	Lockwood and Prombutr	2010	Growth in book equity	78
ChInv	Thomas and Zhang	2002	Inventory Growth	749
ChInvIA	Abarbanell and Bushee	1998	Change in capital inv (ind adj)	976
ChNNCOA	Soliman	2008	Change in Net Noncurrent Op Assets	718
ChNWC	Soliman	2008	Change in Net Working Capital	718
ChTax	Thomas and Zhang	2011	Change in Taxes	153
DelCOA	Richardson et al.	2005	Change in current operating assets	1996
DelCOL	Richardson et al.	2005	Change in current operating liabilities	1996
DelEqu	Richardson et al.	2005	Change in equity to assets	1996
DelFINL	Richardson et al.	2005	Change in financial liabilities	1996
DelLTI	Richardson et al.	2005	Change in long-term investment	1996
DelNetFin	Richardson et al.	2005	Change in net financial assets	1996
dNoa	Hirshleifer, Hou, Teoh, Zhang	2004	change in net operating assets	984
EarningsStreak	Loh and Warachka	2012	Earnings surprise streak	75

EarningsSurprise	Foster, Olsen and Shevlin	1984	Earnings Surprise	1458
GrAdExp	Lou	2014	Growth in advertising expenses	464
grcapx	Anderson and Garcia-Feijoo	2006	Change in capex (two years)	525
grcapx3y	Anderson and Garcia-Feijoo	2006	Change in capex (three years)	525
GrLTNOA	Fairfield, Whisenant and Yohn	2003	Growth in long term operating assets	1021
GrSaleToGrInv	Abarbanell and Bushee	1998	Sales growth over inventory growth	976
GrSaleToGrOverhead	Abarbanell and Bushee	1998	Sales growth over overhead growth	976
hire	Bazdresch, Belo and Lin	2014	Employment growth	423
IntanBM	Daniel and Titman	2006	Intangible return using BM	1396
IntanCFP	Daniel and Titman	2006	Intangible return using CFtoP	1396
IntanEP	Daniel and Titman	2006	Intangible return using EP	1396
IntanSP	Daniel and Titman	2006	Intangible return using Sale2P	1396
InvestPPEInv	Lyandres, Sun and Zhang	2008	change in ppe and inv/assets	578
InvGrowth	Belo and Lin	2012	Inventory Growth	179
MeanRankRevGrowth	Lakonishok, Shleifer, Vishny	1994	Revenue Growth Rank	7062
NumEarnIncrease	Loh and Warachka	2012	Earnings streak length	75
OrderBacklogChg	Baik and Ahn	2007	Change in order backlog	16
PS	Piotroski	2000	Piotroski F-score	1637
RevenueSurprise	Jegadeesh and Livnat	2006	Revenue Surprise	400
SurpriseRD	Eberhart, Maxwell and Siddique	2004	Unexpected R&D increase	1155
Market				
Activism1	Cremers and Nair	2005	Takeover vulnerability	1854
Activism2	Cremers and Nair	2005	Active shareholders	1851
AnnouncementReturn	Chan, Jegadeesh and Lakonishok	1996	Earnings announcement return	2961
Beta	Fama and MacBeth	1973	CAPM beta	18405
BetaFP	Frazzini and Pedersen	2014	Frazzini-Pedersen Beta	2403
BetaLiquidityPS	Pastor and Stambaugh	2003	Pastor-Stambaugh liquidity beta	6551
BetaTailRisk	Kelly and Jiang	2014	Tail risk beta	735
betaVIX	Ang et al.	2006	Systematic volatility	5196
BidAskSpread	Amihud and Mendelsohn	1986	Bid-ask spread	8081
CoskewACX	Ang, Chen and Xing	2006	Coskewness using daily returns	1373

Coskewness	Harvey and Siddique	2000	Coskewness	3211
CustomerMomentum	Cohen and Frazzini	2008	Customer momentum	1369
DolVol	Brennan, Chordia, Subra	1998	Past trading volume	2083
FirmAgeMom	Zhang	2004	Firm Age - Momentum	2213
High52	George and Hwang	2004	52 week high	975
IdioRisk	Ang et al.	2006	Idiosyncratic risk	5196
IdioVol3F	Ang et al.	2006	Idiosyncratic risk (3 factor)	5196
IdioVolAHT	Ali, Hwang, and Trombley	2003	Idiosyncratic risk (AHT)	203
Illiquidity	Amihud	2002	Amihud's illiquidity	11753
IndMom	Grinblatt and Moskowitz	1999	Industry Momentum	2255
IndRetBig	Hou	2007	Industry return of big firms	648
IntMom	Novy-Marx	2012	Intermediate Momentum	567
IO_ShortInterest	Asquith Pathak and Ritter	2005	Inst own among high short interest	1091
iomom_cust	Menzly and Ozbas	2010	Customers momentum	550
iomom_supp	Menzly and Ozbas	2010	Suppliers momentum	550
LRreversal	De Bondt and Thaler	1985	Long-run reversal	12021
MaxRet	Bali, Cakici, and Whitelaw	2010	Maximum return over month	1500
Mom12m	Jegadeesh and Titman	1993	Momentum (12 month)	14970
Mom12mOffSeason	Heston and Sadka	2008	Momentum without the seasonal part	363
Mom6m	Jegadeesh and Titman	1993	Momentum (6 month)	14970
Mom6mJunk	Avramov et al	2007	Junk Stock Momentum	514
MomOffSeason	Heston and Sadka	2008	Off season long-term reversal	363
MomOffSeason06YrPlus	Heston and Sadka	2008	Off season reversal years 6 to 10	363
MomOffSeason11YrPlus	Heston and Sadka	2008	Off season reversal years 11 to 15	363
MomOffSeason16YrPlus	Heston and Sadka	2008	Off season reversal years 16 to 20	363
MomRev	Chan and Ko	2006	Momentum and LT Reversal	15
MomSeason	Heston and Sadka	2008	Return seasonality years 2 to 5	363
MomSeason06YrPlus	Heston and Sadka	2008	Return seasonality years 6 to 10	363
MomSeason11YrPlus	Heston and Sadka	2008	Return seasonality years 11 to 15	363
MomSeason16YrPlus	Heston and Sadka	2008	Return seasonality years 16 to 20	363
MomSeasonShort	Heston and Sadka	2008	Return seasonality last year	363
MomVol	Lee and Swaminathan	2000	Momentum in high volume stocks	2251

MRreversal	De Bondt and Thaler	1985	Medium-run reversal	12021
OptionVolume1	Johnson and So	2012	Option to stock volume	364
OptionVolume2	Johnson and So	2012	Option volume to average	365
Price	Blume and Husic	1972	Price	169
PriceDelayRsq	Hou and Moskowitz	2005	Price delay r square	1120
PriceDelaySlope	Hou and Moskowitz	2005	Price delay coeff	1121
PriceDelayTstat	Hou and Moskowitz	2005	Price delay SE adjusted	1121
ProbInformedTrading	Easley, Hvidkjaer and O'Hara	2002	Probability of Informed Trading	2566
ResidualMomentum	Blitz, Huij and Martens	2011	Momentum based on FF3 residuals	246
retConglomerate	Cohen and Lou	2012	Conglomerate return	492
ReturnSkew	Bali, Engle and Murray	2015	Return skewness	279
ReturnSkew3F	Bali, Engle and Murray	2015	Idiosyncratic skewness (3F model)	279
RIO_Turnover	Nagel	2005	Inst Own and Turnover	1276
RIO_Volatility	Nagel	2005	Inst Own and Idio Vol	1276
ShareVol	Datar, Naik and Radcliffe	1998	Share Volume	1896
ShortInterest	Dechow et al.	2001	Short Interest	13480
Size	Banz	1981	Size	10169
skew1	Xing, Zhang and Zhao	2010	Volatility smirk near the money	788
SmileSlope	Yan	2011	Put volatility minus call volatility	344
std_turn	Chordia, Subra, Anshuman	2001	Share turnover volatility	1181
STreversal	Jegadeesh	1989	Short term reversal	3590
TrendFactor	Han, Zhou, Zhu	2016	Trend Factor	138
VolMkt	Haugen and Baker	1996	Volume to market equity	1426
VolSD	Chordia, Subra, Anshuman	2001	Volume Variance	1181
VolumeTrend	Haugen and Baker	1996	Volume Trend	1426
zerotrade	Liu	2006	Days with zero trades	1313
zerotradeAlt1	Liu	2006	Days with zero trades	1313
zerotradeAlt12	Liu	2006	Days with zero trades	1313
Valuation				
AccrualsBM	Bartov and Kim	2004	Book-to-market and accruals	75
AM	Fama and French	1992	Total assets to market	25911

BM	Rosenberg, Reid, and Lanstein	1985	Book to market using most recent ME	3282
BMdec	Fama and French	1992	Book to market using December ME	25911
BPEBM	Penman, Richardson and Tuna	2007	Leverage component of BM	482
CF	Lakonishok, Shleifer, Vishny	1994	Cash flow to market	7062
cfp	Desai, Rajgopal, Venkatachalam	2004	Operating Cash flows to price	583
DivYieldST	Litzenberger and Ramaswamy	1979	Predicted div yield next month	728
EBM	Penman, Richardson and Tuna	2007	Enterprise component of BM	482
EntMult	Loughran and Wellman	2011	Enterprise Multiple	176
EP	Basu	1977	Earnings-to-Price Ratio	4823
EquityDuration	Dechow, Sloan and Soliman	2004	Equity Duration	388
Frontier	Nguyen and Swanson	2009	Efficient frontier index	126
Leverage	Bhandari	1988	Market leverage	2312
NetDebtPrice	Penman, Richardson and Tuna	2007	Net debt to price	482
RD	Chan, Lakonishok and Sougiannis	2001	R&D over market cap	2530
RIO_MB	Nagel	2005	Inst Own and Market to Book	1278
SP	Barbee, Mukherji and Raines	1996	Sales-to-price	337
VarCF	Haugen and Baker	1996	Cash-flow to price variance	1426
Forecast				
AnalystRevision	Hawkins, Chamberlin, Daniel	1984	EPS forecast revision	108
AnalystValue	Frankel and Lee	1998	Analyst Value	1583
AOP	Frankel and Lee	1998	Analyst Optimism	1583
ChangeInRecommendation	Jegadeesh et al.	2004	Change in recommendation	1281
ChForecastAccrual	Barth and Hutton	2004	Change in Forecast and Accrual	362
ChNAnalyst	Scherbina	2008	Decline in Analyst Coverage	100
ConsRecomm	Barber et al.	2002	Consensus Recommendation	420
CredRatDG	Dichev and Piotroski	2001	Credit Rating Downgrade	861
DownRecomm	Barber et al.	2002	Down forecast EPS	420
EarningsForecastDisparity	Da and Warachka	2011	Long-vs-short EPS forecasts	85
EarnSupBig	Hou	2007	Earnings surprise of big firms	648
FEPS	Cen, Wei, and Zhang	2006	Analyst earnings per share	11
fgr5yrLag	La Porta	1996	Long-term EPS forecast	1308

ForecastDispersion	Diether, Malloy and Scherbina	2002	EPS Forecast Dispersion	2430
PredictedFE	Frankel and Lee	1998	Predicted Analyst forecast error	1583
Recomm_ShortInterest	Drake, Rees and Swanson	2011	Analyst Recommendations and Short-Interest	213
REV6	Chan, Jegadeesh and Lakonishok	1996	Earnings forecast revisions	2961
RIO_Disp	Nagel	2005	Inst Own and Forecast Dispersion	1276
sfe	Elgers, Lo and Pfeiffer	2001	Earnings Forecast to price	262
UpRecomm	Barber et al.	2002	Up Forecast	420
Financing				
AgeIPO	Ritter	1991	IPO and age	6071
CompEquIss	Daniel and Titman	2006	Composite equity issuance	1396
CompositeDebtIssuance	Lyandres, Sun and Zhang	2008	Composite debt issuance	578
ConvDebt	Valta	2016	Convertible debt indicator	83
DebtIssuance	Spiess and Affleck-Graves	1999	Debt Issuance	538
DivInit	Michaely, Thaler and Womack	1995	Dividend Initiation	1639
DivOmit	Michaely, Thaler and Womack	1995	Dividend Omission	1639
DivSeason	Hartzmark and Salomon	2013	Dividend seasonality	120
ExchSwitch	Dharan and Ikenberry	1995	Exchange Switch	364
IndIPO	Ritter	1991	Initial Public Offerings	6071
NetDebtFinance	Bradshaw, Richardson, Sloan	2006	Net debt financing	695
NetEquityFinance	Bradshaw, Richardson, Sloan	2006	Net equity financing	695
NetPayoutYield	Boudoukh et al.	2007	Net Payout Yield	681
PayoutYield	Boudoukh et al.	2007	Payout Yield	681
RDIPO	Gou, Lev and Shi	2006	IPO and no R&D spending	234
ShareIss1Y	Pontiff and Woodgate	2008	Share issuance (1 year)	729
ShareIss5Y	Daniel and Titman	2006	Share issuance (5 year)	1396
ShareRepurchase	Ikenberry, Lakonishok, Vermaelen	1995	Share repurchases	2803
Spinoff	Cusatis, Miles and Woolridge	1993	Spinoffs	568
XFIN	Bradshaw, Richardson, Sloan	2006	Net external financing	695