Decoding Anomalies through Alpha Dynamics

Shuhao Ren*

August 14, 2024

Abstract

This paper studies how alphas of the characteristic-sorted stock portfolios *evolve* over the months after the sorting date, which I refer to as "alpha dynamics". I develop new tests to examine the alpha dynamics predicted by economic theories. The results provide new insights relevant to assessing whether anomalies (1) are attributable to collective "data snooping" or are real, (2) can be attributed at least in part to mispricing, and (3) imply potential profits after considering trading costs. I study 205 published anomalies and find that t-tests of whether average alphas equal zero fail to detect many real anomalies, a problem that becomes more severe with higher t-statistic cutoffs. Further, the observed alpha dynamic pattern conforms to existing behavioral models rather than rational models for about sixty percent of characteristics, including net share issuance, idiosyncratic volatility, and momentum. Moreover, I show that after-cost profitability is significantly underestimated when alpha dynamics are not allowed for.

*WP Carey School of Business, Arizona State University. Email: sren14@asu.edu. I am deeply grateful to Hendrik Bessembinder for his support and guidance throughout my graduate studies, and to my other committee members, George Aragon, Seth Pruitt, and Sunil Wahal for their invaluable feedback. For helpful comments and suggestions, I also thank Fahiz Baba-Yara, Michael Barnett, Oliver Boguth, Amit Goyal, Shikun Ke, Fabio Moneta, Yuri Tserlukevich, as well as participants at the Financial Management Association (2023), Southwestern Finance Association (2024), Eastern Finance Association (2024), and ASU Brownbag Seminar series. I am responsible for any remaining errors and omissions.

1 Introduction

In recent decades, researchers have documented hundreds of apparent "anomalies" in stock returns.[1] This evidence has led to ongoing debates regarding key questions, including (1) Do the reported anomalies truly exist, or are they attributable to collective data snooping? (2) Do they arise due to mispricing, or are there rational expectations? (3) Is trading on the basis of the anomalies profitable after considering trading costs?[2]

The alpha related to a characteristic is typically estimated as the average excess return earned by a portfolio created by sorting on a characteristic and purchasing (selling) stocks with outcomes in one tail (the other tail) in the months after the portfolio sort. Most often, alpha is estimated using all available months. For instance, if stocks are sorted only once a year, it would be typical to estimate alphas using data for the twelve months after the sort. However, theories imply that alphas can vary across months, in which case this procedure estimates the average alpha across one to twelve months after the sort. I propose to instead examine alphas for each month after the portfolio sort, and to examine the *evolution* of alphas after sorting (henceforth, the *alpha dynamic*). Having done so, I explore the new insights that alpha dynamics provide.

Here is a concrete example. Using the accruals as the characteristic, researchers usually sort firms in June every year and hold characteristic-sorted portfolios from July to the following June. Then they calculate monthly returns in excess of a given factor model's predictions and average these returns over all months. This method is equivalent to first calculating the alphas for each month following the sorting period and then *averaging* the alphas (henceforth, the *alpha mean*) over the first twelve months after sorting.[3]

[1] If a firm characteristic's return predictability on future stock returns cannot be fully explained by an asset pricing model, it is considered an anomaly relative to that model.

[2] For (1), see e.g., Harvey, Liu, and Zhu (2016) and Hou, Xue, and Zhang (2020). For (2), see e.g., Engelberg, McLean, and Pontiff (2018) and McLean and Pontiff (2016). And for (3), see e.g., Chen and Velikov (2023) and Novy-Marx and Velikov (2016).

[3] Since alpha is an arithmetic average, the alpha mean is also an average. Additionally, when the alpha in one particular month after sorting is examined, the alpha mean is just that alpha itself. The term, "alpha mean" highlights that it is an average rather than a dynamic.

While studying the alpha mean is informative, it does not reveal how or if alphas evolve over the months after sorting. That is, the alpha earned (say) one month after sorting is blended with the alpha earned (say) six months after sorting. Are these alphas the same? Theoretically, should they be? Importantly, I propose to study how alphas vary in event time following the portfolio sorting date, a perspective that differs from prior work that examines whether alphas (or betas) vary across calendar time or vary around specified events.[4]

I examine 205 published characteristic-sorted anomalies. For each, I conduct portfolio sorts every month, form value-weighted long-short portfolios, and track the portfolios' monthly returns after the portfolio sort date.[5] I then study alpha estimates for returns measured in different numbers of months after the sort date. To do so, I develop new empirical tests that allow me to statistically assess whether alphas are constant in the months after sorting, or whether they exhibit certain patterns predicted by existing economic models.[6] Depending on the predicted pattern to be tested, I study alphas from the first twelve months up to nine years after the portfolio sort date.

I show that the consideration of alpha dynamics provides new and important insights into the three questions listed in the first paragraph of the paper. First, regarding the existence of anomalies, I show that tests that accommodate alpha dynamics can detect the existence of non-zero alphas even when tests focused on the mean alpha may not. For example, a build-up of mispricing followed by a correction could imply positive alphas for a period followed by negative alphas. In this case, anomalous mean returns exist, even while the mean alpha can be statistically indistinguishable from zero. I find that alpha-mean tests fail to detect many real anomalies. Further, the use of higher t-statistic cutoffs (e.g., to 3.0) to address p-hacking concerns (Harvey et al., 2016) leads to a higher rate of failure to detect real anomalies.[7]

[4]For calendar time, see e.g., Boguth, Carlson, Fisher, and Simutin (2011) and Lewellen and Nagel (2006). For specified events, e.g., McLean and Pontiff (2016) study publication dates and Engelberg et al. (2018) study corporate news and earnings announcement dates.

[5]Since portfolio sorts are conducted every month, results in this paper cannot be explained by seasonality (e.g., Heston & Sadka, 2008).

[6]Alphas are considered constant if the difference between any two adjacent alphas after sorting is statistically insignificant from zero.

[7]p-Hacking occurs when large t statistics result from searching for significant results among numerous

Second, I present results indicating that about 60% of the anomalies exhibit alpha dynamic patterns that conform to existing behavioral models rather than rational models. These results imply that the anomalies are at least partially due to mispricing. Examples of well-known categories include net share issuance, idiosyncratic volatility, and momentum.

Third, I develop a rule for the optimal holding period between entry and exit from a position that should be employed by a trader seeking to profit from an anomaly, showing that the holding period should be determined both by alpha dynamics and trading costs. I implement this rule and show that after-cost profitability may have been significantly underestimated in prior studies. The rule I develop achieves a statistically significant improvement in after-cost alpha as compared to that assumed in prior studies for about 20% of anomalies. Among these, the improvement averages about 0.3% per month.

To clarify why studying alpha dynamics can be used to detect anomalies, recognize that if a given anomaly does not exist, then the true alpha (i.e., the parameter) is zero in all months after a portfolio sort. This implies alpha is zero on average and does not differ across any subset of months after sorting. Therefore, either a statistically significant non-constant alpha dynamic or a non-zero alpha mean rejects the zero-alpha null hypothesis, implying that alpha-dynamic tests have the potential to detect some anomalies that alpha-mean tests do not.

Why would alphas be non-constant? Both rational and behavioral theories can imply non-constant alphas. For instance, Keloharju, Linnainmaa, and Nyberg (2021) show that several production-based models such as Berk, Green, and Naik (1999) can imply a monotone pattern of alphas after sorting because firm risks converge over time in those models. Further, changes in mispricing over time will be reflected as alphas; if the rate of such changes varies over time, then so will alphas. Models such as return extrapolation (e.g., Barberis, 2018) can imply ripple-like patterns, characterized by alternating increases and decreases in alphas over time.[8]

meaningless characteristics.

[8]Behavioral models usually assume a shock date when mispricing initially arises and study how alphas

Alphas must be defined relative to benchmark asset pricing models (Fama, 1970). Rather than adding to the voluminous literature seeking to assess the correct model, I focus on how alpha dynamics help to understand the anomalies implied by a *given* asset pricing model. While for simplicity, I primarily focus on anomalies implied by the CAPM, I also illustrate how the methodology can be applied to several canonical multi-factor models.[9]

Some behavioral models, e.g., return extrapolation (Barberis, 2018) and inattention (Duffie, 2010), imply a ripple pattern in alphas, which helps to illustrate how studying alpha dynamics can help to determine if anomalies are at least partially due to mispricing. In Barberis (2018), extrapolators overreact to both negative and positive returns, resulting in an alternating build-up and resolution of overpricing and underpricing over time. In contrast, existing rational explanations, to the best of my knowledge, do not imply such a pattern.[10]

I develop two new statistical tests to assess whether alphas are non-constant. As some models can imply monotone patterns of alphas, the first extends the monotonicity test of Paton and Timmermann (2010) to assess if alphas exhibit a monotonically increasing or decreasing pattern.[11] The second test generally distinguishes constant alphas from non-constant alphas. The intuition of the test is that if true alphas are constant, then the mean alpha estimate should be invariant to the timing of portfolio formation relative to the portfolio sort date. The rejection of this hypothesis implies that alphas are non-constant.[12]

or prices evolve after the shock date. In contrast, I study portfolio sorting dates when firms are sorted based on a specific firm characteristic. Empirically, studying sorting dates is a tradition (e.g., van Binsbergen, Boons, Opp, & Tamoni, 2023). Economically, in Appendix B, I discuss more how shock dates and sorting dates might be related.

[9]I use the CAPM as the baseline model because theories that explain the anomalies are mostly proposed relative to the CAPM (e.g., Zhang, 2005). Further, Jensen, Kelly, and Pedersen (2022) suggest the use of the CAPM to study the existence of anomalies.

[10]The ripple pattern can be interpreted as repeated drifts and reversals in the price. Therefore, it is related to the literature that uses price drift followed by a reversal as evidence of mispricing. While prior tests detect the pattern by examining *mean alphas*, I detect it by examining *alpha dynamics*. I further discuss the problems of prior tests and the benefits of my test in Section 2.3.

[11]Patton and Timmermann (2010) apply their method to examine monotone relations in the cross-section. In contrast, I examine monotone relations in the event time after the portfolio sort.

[12]All tests in this paper examining alpha dynamics have considered that exposures to the factors in the benchmark models could vary across different horizons after the sort. This is achieved by first estimating alphas and exposures separately for returns measured at different months after sorting. Then the estimated

To evaluate these hypotheses, I follow Hou et al. (2020) and construct value-weighted portfolios. Chen and Zimmermann (2021) identify the period studied by the original authors who documented each anomaly. Since additional data is now available, each calendar month can be assigned as before-sample, in-sample, or post-sample with respect to a given anomaly. I estimate average alphas for each anomaly and sample period, and I label those that fail the t test with a cutoff of 1.96 as suspicious anomalies. Among the remaining anomalies, I label those that fail the t test with a cutoff of 3.0 as Harvey et al. (2016, HLZ) suspicious anomalies. I find that, using the full sample period, 21% (19 of 92) of suspicious anomalies and 59% (24 of 41) of HLZ suspicious anomalies do exist, as they pass at least one of the two alpha-dynamic tests.[13] Notably, these results are robust in subperiods and when different benchmark models are used. For example, in the post-sample periods, these corresponding figures are 25% and 63%. Using the Fama and French (2015) five-factor model as the benchmark, the corresponding figures are 21% and 49%.

To assess potential ripple patterns, I study alphas over nine years after sorting. I consider this long horizon because ripple patterns are typically implied by the relevant models to manifest themselves after a delay. I then divide the nine years into five non-overlapping subperiods. To examine whether there is an increasing pattern in the first twelve months after sorting, I test the hypothesis that the alpha in the first month (the left edge of the subperiod) is the highest. A rejection implies that there is a higher alpha in later periods, i.e., an increasing pattern. Similarly, to examine whether there is a decreasing pattern, I test the hypothesis that the alpha in the twelfth month (the right edge) is the highest. I repeat this test over each of the subperiods. To aggregate outcomes over the subperiods, I apply multiple comparison corrections. In those cases where the tests detect both increasing and decreasing alphas, a ripple pattern is implied. Among the 132 anomalies showing evidence of non-zero alphas in the full sample period, 79, or sixty percent, exhibit a ripple pattern. Importantly,

alphas are jointly examined.

[13] While these saved suspicious anomalies usually have small alphas, they could be economically important as van Binsbergen and Opp (2019) show that small but persistent alphas imply more mispricing in the price.

this result also indicates that the alphas for many anomalies are persistent in the long run in a ripple pattern, which can include alternating periods of positive and negative alphas. For example, the alphas of momentum, accruals, net share issuance, and idiosyncratic volatility persist for at least nine years. Further, 41% of anomalies that exhibit the ripple pattern in the in-sample period do not exhibit it in the post-sample period, consistent with enhanced arbitrage after publication.

Turning to implementation issues, I estimate the optimal holding period that maximizes the after-cost alpha with expanding regressions.[14] To alleviate look-ahead bias and data mining concerns, I estimate optimal holding periods using known data but evaluate after-cost alpha based on out-of-sample returns. Relying on Monte Carlo simulations, I show that the improvement in after-cost alpha cannot be attributed to random variation alone.

On balance, the tests I introduce and the resulting empirical evidence show that the consideration of alpha dynamics provides new and important insights. While tests of whether mean alphas differ significantly from zero are useful, tests that accommodate alpha dynamics provide additional information relevant to the economic interpretation of characteristic-based anomaly returns.

2 Intuition

In this section, I use a simple model to provide economic intuition on why alpha dynamics provide new insights.

[14] While other elements, such as characteristic frequencies and cost mitigation techniques, may affect the after-cost profitability, the main focus of this analysis is to examine the role of alpha dynamics and trading costs in determining the optimal holding periods. Furthermore, while some studies (e.g., Gârleanu & Pedersen, 2013; Jensen, Kelly, Malamud, & Pedersen, 2022) investigate the optimal strategy *conditioned on* the anomaly generates positive after-cost alpha, this paper proposes a new rule to evaluate whether an anomaly generates positive after-cost alpha in the first place.

2.1 Motivation: Patterns of alpha dynamics

Before discussing the insights obtained from alpha dynamics, what patterns of alpha dynamics we should expect? Let j be the number of months that have passed after the portfolio sort, where $j = 0$ is the sorting month. Denote α_j as the true alpha in event time j: the average of returns in excess of the factor-model prediction in the population. I omit the subscripts for characteristic X and time t in α_j to ease exposition, but it can be different both across characteristics and across time. In all that follows, I focus on alpha with respect to the market factor.

2.1.1 True alphas are zero

The first situation is when true alpha is zero in every period after sorting. That is, $\alpha_j = 0$ for all j. As also shown in Figure 1a, within any subset of months after sorting, true alphas will be *constant* or there is no alpha dynamic.

2.1.2 Rational expectations

If true alphas do exist, one explanation is rational expectations. That is, the factor model used to assess anomalies may omit rational risk factors. The production-based asset pricing literature suggests that α_j can have a *non-constant* pattern. For example, in the models of Gomes, Kogan, and Zhang (2003) and Zhang (2005), production risk is mean-reverting. These models imply that α_j *decreases monotonically* as firms' risks converge over time.[15]

2.1.3 Mispricing

Alphas can also be due to mispricing. Mispricing can be an arbitrage opportunity that does not involve risk. It can also be collective and result from correlated errors in investor expectations (e.g., Barberis, Greenwood, Jin, & Shleifer, 2015; Stambaugh & Yuan, 2017).

[15] Keloharju et al. (2021) simulate the models of Berk et al. (1999), Gomes et al. (2003), Hackbarth and Johnson (2015), and Zhang (2005), and show that they all imply a monotonically decreasing pattern of alphas.

The changes in mispricing will be reflected as alphas. If the rate of changes is not constant over time, α_j will also have a *non-constant* pattern.

To explore potential alpha dynamics predicted by behavioral models, I further examine the mechanisms of return extrapolation (Barberis, 2018) and inattention (Duffie, 2010; Hendershott, Menkveld, Praz, & Seasholes, 2022). I calibrate and plot the alpha dynamic implied by each model in Figure 2, using the parameters from the original papers.

Model 1: Barberis (2018):[16] **return extrapolation**. There are two types of investors, extrapolators and arbitrageurs. Extrapolators' belief on future price change is a weighted average of past price changes. And arbitrageurs have bounded rationality in the way that they do not have a full understanding of extrapolator demand. Extrapolators push the price away from the fundamental price, while arbitrageurs drag the price back.

Model 2: Duffie (2010): **fixed periods of inattention**. When there is a supply shock, only a few investors (all attentive investors and part of inattentive investors) can absorb the shock. Therefore, there is a large price recession to compensate investors who absorb the shock. These investors then lay off the risk over time when other inattentive investors come to the market and the price reverts over time. The model has one class of inattentive investors and they fix their periods of inattention.

Model 3: Hendershott et al. (2022): **stochastic arrival of inattentive investors**. The model has market makers, attentive investors, and multiple classes of inattentive investors[17]. The inattentive investors arrive stochastically. Each class of investors has random private-value shocks each period. And the shocks between attentive and inattentive investors can be perfectly canceled off. However, only part of inattentive investors will adjust their portfolios after the shocks. Therefore, attentive investors are compensated with a price recession to absorb the shocks of inattentive investors and lay off the risk when other inattentive investors

[16] In Section 4.1 of the paper. The mechanism is similar to that in Hong and Stein (1999). Compared to the framework in Hong and Stein (1999), Barberis (2018) is more general as it models a well-known behavioral bias of return extrapolation that could persistently exist in the market.

[17] To provide the intuition of how inattention affects the pattern of alphas, I assume one class of inattentive investors when I simulate this model.

come to adjust their portfolios and price reverts over time.

These mechanisms suggest that alphas can have *monotonically decreasing, monotonically increasing, or ripple-like* patterns. All mechanisms exhibit a monotonically decreasing pattern of alphas within the first few periods after the initial shock. This implies that the rate of changes in mispricing decreases over time. Take the mechanism of Barberis (2018) in Figure 2a as an example. After a large positive cash flow shock in period 0, overpricing builds up. As the rate of build-up decreases over time, so does alpha.[18] Similarly, a monotonically increasing pattern in Barberis (2018) is due to an increasing rate of changes in mispricing.[19]

Furthermore, the mechanisms of Barberis (2018) and Duffie (2010) imply a ripple-like pattern, characterized by alternative increases and decreases in alphas. In Barberis (2018), the build-up and resolution of mispricing lead to opposite signs of alphas. Since extrapolators overreact to both positive and negative returns, the build-up and resolution of overpricing and underpricing occur alternatively over time, resulting in a ripple pattern. In Duffie (2010), the ripple pattern appears because some inattentive investors only adjust their portfolios every few periods. Therefore, prices surge above and plunge below their steady-state level over time due to imbalances in demand and supply.

2.2 Intuition for the tests on the existence of alphas

Section 2.1 shows that if α_j is always zero, then it follows that true alphas are also constant across j. However, non-zero alphas can be dynamic. Motivated by this, I proceed to design tests to examine the null hypothesis that α_j is constant in j.

[18]Arbitrageurs pull the price back to the fundamental price, while extrapolators reduce their opposite trades to those of arbitragers over time as they update their beliefs over time. Therefore, the build-up rate slows down over time and eventually reaches zero.

[19]For example, Figure 2a shows that alphas increase from period 2 to period 4 if we flip the sign of the negative alphas (same pattern from period 7 to period 9). This subperiod corresponds to the resolution of the overpricing. As arbitrageurs trade on the other side of extrapolators, the rate of resolution is sluggish at first. However, as extrapolators adjust their belief over time, the resolution rate increases.

2.2.1 Mean alphas

With a specific combination of the number of skipped months after sorting (k) and holding period (h), the true mean alpha of characteristic-sorted portfolios is given by:

$$\alpha^{op}_{k,h} = \frac{1}{h} \sum_{j=k+1}^{k+h} \alpha_j \tag{1}$$

Here, I use the traditional overlapping portfolio approach (e.g., Jegadeesh & Titman, 1993). Therefore, in each month, there will be h portfolios that are sorted from $k+h-1$ to k months ago. Each portfolio has a weight of $\frac{1}{h}$. The portfolios are rebalanced monthly, so there is an overlap in returns when $h > 1$. To deal with this overlap, I use a calendar-time portfolio approach to calculate average monthly returns.

Lemma 1 *When α_j is constant in j, $\alpha^{op}_{k,h}$ is constant in k and h. And when α_j is non-constant in j, $\alpha^{op}_{k,h}$ is not constant in k and h.*

Lemma 1 follows directly from Equation 1. It suggests that testing whether α_j is constant is the same as testing whether the mean alpha $\alpha^{op}_{k,h}$ is constant in k and h. Therefore, in the next two sections, I develop two tests to examine whether $\alpha^{op}_{k,h}$ is constant in k and h.

Testing the null hypothesis of constant alphas is not easy. A test should consider multiple hypotheses testing adjustment and the alpha distribution may change at different horizons j. The two tests in the following two sections take these into account. Further, it matters whether investors can trade on those anomalies based on alpha dynamics. As I will show in the next two sections, results from the two tests will indicate how investors should optimally trade on the anomalies. Moreover, results from the monotonicity test will also provide evidence for the economic theories that predict monotone patterns.

2.2.2 The monotonicity test

As discussed in Section 2.1, economic models can imply monotone patterns of alphas. Motivated by theories, the first test examines whether alphas are monotonically increasing or decreasing within a subset of months after portfolio sorting. Specifically, the monotonicity test examines the following hypotheses:

$$\begin{aligned} &H0 : \alpha^{\text{op}}_{\text{0,h}} \text{ has a constant pattern in h.} \\ &H1 : \alpha^{\text{op}}_{\text{0,h}} \text{ has a strictly increasing or decreasing pattern in h.} \end{aligned} \tag{2}$$

Rejecting the null hypothesis implies the existence of non-zero alphas. Further, focusing on how $\alpha^{op}_{0,h}$ change in h instead of k is because we cannot directly observe when alphas disappear. For example, assume that true alphas strictly decrease over time and disappear after three months. Suppose that we examine alphas over the first twelve months after sorting in the test. In the first twelve months, true alphas decrease first and remain flat, which is not monotone. Therefore, we cannot identify monotone patterns of alphas by examining the patterns of $\alpha^{op}_{k,1}$ in k, where k ranges from zero to eleven. In contrast, examining patterns of $\alpha^{op}_{0,h}$ in h solves the problem. This is because $\alpha^{op}_{0,h}$ in h is a moving average of alphas from the first month to h.

2.2.3 The optimization test

Non-constant patterns are not limited to monotone patterns.[20] Therefore, the monotonicity test is inadequate, and I design the second test to identify a more general non-constant pattern of alphas.

According to Lemma 1, when α_j is constant in j, $\alpha^{op}_{k,h}$ is the same for all strategies based on k and h. In contrast, when α_j is non-constant, Let α_{j^*} be a maximum. To maximize the mean alpha, the problem is:

[20] One example is the ripple pattern discussed in Section 2.1.

$$\max_{k,h} \alpha^{op}_{k,h} = \max_{k,h} \frac{1}{h} \sum_{j=k+1}^{k+h} \alpha_j \tag{3}$$

Proposition 1 *The optimal strategy to obtain the highest* $\alpha^{op}_{k,h}$ *is* $k^* = j^* - 1$ *and* $h^* = 1$.[21]

Proof: See Appendix C.

When true alphas are non-constant, Proposition 1 shows that there is at least one strategy (the optimal strategy) that can generate a higher mean alpha than the average of the mean alphas of alternative strategies across k and h on average.

To examine whether alphas are constant in the first $k_{end} + 1$ months after sorting, I examine the strategies of ks from the set I, where k is within the range of $[0, k_{end}]$. Denote N_I as the number of ks in I. Specifically, the optimization test examines the following hypotheses:

$$\begin{aligned} &H0: \text{All strategies generate the same } \alpha^{\text{op}}_{\text{k,1}} \text{ for k } \in \text{ I.} \\ &H1: \text{There is at least one } \alpha^{\text{op}}_{\text{k,1}} \text{ for k } \in \text{ I that is higher than } \frac{1}{\text{N}_\text{I}} \sum_{\text{k} \in \text{I}} \alpha^{\text{op}}_{\text{k,1}} \end{aligned} \tag{4}$$

$\alpha^{op}_{k,1}$ can be interpreted as the true mean alphas at different horizons after the sort. According to the null hypothesis, the true mean alpha should be invariant across horizons. In contrast, the alternative hypothesis implies that the true mean alpha of at least one horizon differs from the others. Based on Proposition 1, to statistically test the null hypothesis, I examine whether an optimization strategy k^* from I generates a statistically significant higher alpha than that of the average of alphas of all strategies of k in I. To alleviate look-ahead bias and data mining concerns, I estimate k^* based on past known information before month t, form a portfolio at the end of month t based on the estimated k^*, and measure

[21] Further, this optimal strategy also maximizes the wealth of investors that have a long investment horizon. This is because this strategy is optimal in any month theoretically.

its returns in month $t+1$. Using the returns in $t+1$, I examine whether the optimization strategy generates a higher alpha with a t-stat hurdle than that of the average of alphas of all strategies of k in I. By using expanding regressions, the optimization test is conservative and mimics an investor's experience. More details for empirical design are in Section 4.1.3.

Furthermore, alpha-dynamic tests alleviate some of the potential problems associated with alpha-mean tests. First, as in the build-up and resolution of the mispricing example discussed in the Introduction, the choice of holding periods in portfolio construction often influences the results of alpha-mean tests and may lead to conflicting conclusions.[22] Given that the alpha dynamic is non-constant in that example, alpha-dynamic tests could serve to identify non-zero alphas. Moreover, alpha-dynamic tests study alphas at different horizons jointly, thus alleviating the data mining concerns in the choice of holding periods in alpha-mean tests. Additionally, failing alpha-mean tests could be due to low statistical power. This problem would be more significant for value-weighting portfolios and in the post-publication period due to typically small alphas. However, alpha-dynamic tests may not encounter this issue simultaneously, as their statistical power depends on the differences between alphas after sorting and the volatility of these differences.

Moreover, results from both the monotonicity test and the optimization test will indicate how investors should trade on an anomaly. The optimization test is designed to mimic investors' trading experience. As for the monotonicity test, suppose that there is a statistically significant decreasing pattern of alphas. Then the optimal strategy to trade on the anomaly is to take $k=0$ and $h=1$.

2.3 Intuition for detecting mispricing

Patterns of alpha dynamics provide important moments to evaluate asset pricing models. As discussed in Section 2.1, existing behavioral models such as return extrapolation (Barberis, 2018) and inattention (Duffie, 2010) can imply a ripple-like pattern in alphas. In contrast,

[22] See Bessembinder, Burt, and Hrdlicka (2022), Hasler (2022), and Hou et al. (2020).

as far as I am aware, existing rational models do not imply the ripple pattern.

As the ripple pattern is characterized by repeated drifts and reversals in prices, using the ripple pattern to detect mispricing is similar in spirit to the extant empirical studies that interpret price drift followed by a reversal as evidence of mispricing (e.g., Chan, 2003; Michaely, Thaler, & Womack, 1995; Tetlock, 2007, 2011). These studies use the drift-reversal pattern to detect mispricing because while rational models find it hard to explain momentum (Jegadeesh & Titman, 1993) followed by a reversal (De Bondt & Thaler, 1985), behavioral models (e.g., Barberis, Shleifer, & Vishny, 1998; Daniel, Hirshleifer, & Subrahmanyam, 1998; Hong & Stein, 1999) can explain them.

In detecting drift-reversal patterns, how does this paper differ from prior studies? Previous papers on the drift-reversal pattern typically conduct tests on *mean alphas* over different horizons. These tests evaluate the signs of mean alphas within non-overlapping subperiods after the portfolio sort date. A drift-reversal pattern is implied if both statistically significant positive and negative mean alphas are detected. Unlike these papers, my proposed test examines *alpha dynamics*. A drift-reversal pattern is implied if the test detects both statistically significant increasing and decreasing patterns of alphas. Thus, the two tests focus on different features of drift-reversal patterns.

However, alpha-mean tests might inadvertently fail to detect drift-reversal patterns for several reasons. First, by examining whether the signs of mean alphas change over time, these tests compare *mean alphas between subperiods*. Without prior knowledge of when prices will reverse, results are highly affected by how subperiods are specified. The true monthly alphas within a pre-specified subperiod could be a mix of positive and negative alphas. In this case, averaging these alphas may not result in a statistically significant negative or non-zero mean alpha, failing to detect a price reversal. Additionally, comparing the signs of mean alphas between subperiods ignores variations of alphas within each subperiod.

In comparison, the alpha-dynamic test has a stronger potential to detect alpha patterns. To detect increasing (or decreasing) patterns of alphas within a subperiod, the test simply assesses if the alpha in any month within the subperiod is higher by a significant margin as compared to the monthly alpha at the beginning (or end) of the subperiod. That is, this test compares *alphas across months within a given subperiod*. Therefore, the test captures more granular variations in alphas than alpha-mean tests and is less affected by the timing of price reversals due to pre-specified subperiods.

Empirically, I demonstrate the advantages of alpha-dynamic tests over alpha-mean tests in Section 4.2.2 and Appendix D. Alpha-dynamic tests detect about sixty percent of anomalies showing drift-reversal patterns, whereas alpha-mean tests detect only a few anomalies using the same subperiods.

2.4 Intuition for the impact on after-cost profitability

It does not challenge market efficiency if characteristic-sorted portfolios generate alphas on paper, but agents cannot trade profitably because of transaction costs. Therefore, many papers related to market efficiency include robust analyses of after-cost profitability (e.g., Bowles, Reed, Ringgenberg, & Thornock, 2023). To evaluate after-cost profitability, we need to choose an appropriate holding period. In this section, I examine the optimal holding period under different alpha dynamics.

2.4.1 After-cost alphas

The after-cost alpha is equal to the mean alpha minus the average rebalancing costs. Let c represent the population mean of monthly rebalancing costs (in percentage). Further, assume c is exogenous. c can also be different across characteristics X and across time t. The monthly true after-cost alpha for a combination of k and h is then given by:

$$\alpha_{k,h}^{ac} = \frac{1}{h}\sum_{j=k+1}^{k+h} \alpha_j - \frac{c}{h} = \alpha_{k,h}^{op} - \frac{c}{h} \tag{5}$$

Equation 5 shows that holding periods h affect both $\alpha_{k,h}^{op}$ and $\frac{c}{h}$ and k affects only $\alpha_{k,h}^{op}$.

To evaluate whether investors can generate positive after-cost alpha, the appropriate k and h should be the ones that maximize the $\alpha^{ac}_{k,h}$.[23] In the next section, I examine what are the optimal k^* and h^* with different alpha dynamics and trading costs.

2.4.2 Exponential decay in alphas

I model the behavior of alphas the same as that in Hendershott et al. (2022, HMPS). In HMPS, alphas are due to mispricing[24] and a shock occurs every month. Denote initial mispricing due to the shock as $-\delta_0$.[25] Denote the expected mispricing j months after the shock as P_j (in percentage). P_j is assumed to follow an Ornstein-Uhlenbeck (OU) process, characterized by exponential decay and reversion to zero over time. In equilibrium, the decay rate is the inattentiveness of inattentive investors, λ ($\lambda > 0$), and P_j can be expressed as:

$$P_j = -\delta_0 e^{-\lambda j} \qquad \text{for} \quad j \geq 0 \tag{6}$$

Let $\delta_0 > 0$.[26] Since alphas are changes in P_j, the evolution of alphas can be expressed as:

$$\alpha_j = P_j - P_{j-1} = \delta_0(1 - e^{-\lambda})e^{-\lambda(j-1)} \qquad \text{for} \quad j \geq 1 \tag{7}$$

Equation 7 indicates that α_j also reverts to zero over time.

2.4.3 Optimization

To maximize the after-cost alpha, the problem is:

[23] Empirically, I estimate optimal k^* and h^* based on past known information at month t to reduce look-ahead bias and data mining concerns. Then I form portfolios based on k^* and h^* at the end of month t, and measure their returns from $t + 1$ to $t + h^*$.

[24] The following results can also apply to alphas resulting from rational expectations if omitted risk premia follow the same behavior as in Eq 7.

[25] This is the $-pG_t$ in Equation 12 of HMPS. In HMPS, the absolute value of δ_0 decreases in the fraction of attentive investors (m_F in HMPS) and increases in inattentiveness.

[26] The sign of δ_0 does not affect the conclusion. We can flip the sign of long-short portfolios so that alphas are positive when $\delta_0 < 0$.

$$\max_{k,h} \alpha^{ac}_{k,h} = \max_{k,h} \frac{1}{h}\sum_{j=k+1}^{k+h} \alpha_j - \frac{c}{h} \\ = \max_{k,h} \frac{1}{h}\delta_0 e^{-\lambda k}(1-e^{-\lambda h}) - \frac{c}{h} \tag{8}$$

Since $\alpha^{ac}_{k,h}$ decreases in k, $k^* = 0$. Replace $k = 0$ and take FOC on h:

$$foc = \frac{\delta_0(\lambda h + 1)e^{-\lambda h} + c - \delta_0}{h^2} \tag{9}$$

Since alphas decay over time by assumption, if α_1 is less than c, $\alpha^{ac}_{k,h}$ will be negative for any combination of k and h. Therefore, let us assume that $\alpha_1 > c$. Then $\delta_0(1 - e^{-\lambda}) > c$ according to Equation 7 and $\delta_0 > c$. Since $\delta_0(\lambda h + 1)e^{-\lambda h}$ converges in h to zero and $\delta_0 > c$, foc cannot be always positive, and there are two cases:

Proposition 2 *Case 1: When* $\frac{\lambda+1}{e^{\lambda}} \leq 1 - \frac{c}{\delta_0}$, *the optimal holding period* $h^* = 1$.

Proof: See Appendix C.

The intuition is that when trading costs c are tiny relative to λ or δ_0, trading costs can be ignored, and the problem is similar to the one for the mean alpha. Therefore, similar to Proposition 1, the optimal holding period is $h^* = 1$.

Proposition 3 *Case 2: When* $\frac{\lambda+1}{e^{\lambda}} > 1 - \frac{c}{\delta_0}$, *the optimal holding period* h^* *satisfies* $\frac{\lambda h^*+1}{e^{\lambda h^*}} = 1 - \frac{c}{\delta_0}$. *And* h^* *(1) decreases in* δ_0, *(2) decreases in* λ, *and (3) increases in* c.

Proof: See Appendix C.

The closed-form solution does not exist when $\frac{\lambda+1}{e^{\lambda}} > 1 - \frac{c}{\delta_0}$,[27] but we can know how δ_0, λ, and c affect h^*. The intuition for (1) is that when initial mispricing δ_0 is larger, the decay in alphas will also be larger over time when everything else is the same. Then having a longer holding period will reduce the mean alpha and thereby reduce the after-cost alpha. The intuition behind (2) is similar. If the attentiveness or decay rate λ is greater, alphas decay

[27] It is a Lambert's W function.

more quickly. Finally, (3) is because when trading costs c are higher, a longer holding period can reduce rebalancing costs.

Overall, analysis indicates that when alphas decay exponentially over time, $k^* = 0$ and h^* depends on the initial mispricing level δ_0, rate of reversion of mispricing λ, and trading costs c. This implies that prior studies may significantly underestimate after-cost profitability by relying on ad-hoc holding periods or turnover rates for determining holding periods.

Is k^* always zero? The assumption that alphas decay monotonically is motivated by some of the existing models (e.g., HMPS). However, alphas may have other patterns. For example, when the alpha in the first period (α_1) is not the maximum among all possible α_j, Proposition 1 suggests that the optimal k^* to maximize the mean alpha will be greater than zero. Then the optimal k^* for after-cost alpha can also be different from zero.[28] Although such a scenario is possible, I do not find any model that implies such a pattern of alpha dynamics, making it difficult to model the behavior of alphas. Instead, I empirically examine whether k matters for after-cost profitability in Section 4.4.3.

3 Data

3.1 Characteristic-sorted portfolios

I collect 205 firm or return-based characteristics from the website.[29] Authors of Chen and Zimmermann (2021) create the website and kindly provide the data. The data is monthly and ranges from December 1925 to June 2022. Returns of the factors are from Kenneth French's website and the q-factor library.[30] I collect monthly stock-level data on returns (with and without dividends), prices, and numbers of shares outstanding for all US stocks traded on the NYSE, AMEX, and NASDAQ from the Center for Research in Security Prices (CRSP) database.

[28] For example, when trading costs are extremely low, the effect of trading costs can be ignored.
[29] https://www.openassetpricing.com/
[30] http://global-q.org/factors.html

For each characteristic, I sort stocks into decile groups at the end of each month. After each sort, I construct value-weighted long-short spread portfolios.[31], and track their monthly returns up to nine years. When a stock delists, I reinvest the amount of money in the stock (net of the delisting return) to the rest of the stocks in the portfolio with value weighting. Finally, as Chen and Zimmermann (2021) provide the sample period studied by the original authors, I can categorize each calendar month into before-sample, in-sample, and post-sample periods.

3.2 Trading cost measures

Stock-level trading cost measures are collected from Andrew Chen's website. Chen and Velikov (2023) use these cost measures to estimate the profit after trading costs for each characteristic. They estimate effective spreads as measures of trading costs and argue that their trading cost measures aim to measure the minimum amount by which prices would have been moved. Effective spreads are measured as twice the absolute difference between the midpoint of bid-ask spreads and the executed prices. The effective spread measures include one high-frequency (HF) measure and three low-frequency (LF) measures. Following Chen and Velikov (2023), I employ half HF effective spreads as trading costs whenever possible and take half of the average of the three LF spreads as trading costs when HF spreads are missing.

Upon merging the trading costs data with stock returns data, about 12% of the observations contain missing trading costs. Since I am interested in whether portfolio construction choices affect after-cost alphas, I need trading cost estimates for any stock in any month. Otherwise, the estimate of portfolio rebalancing costs is imprecise. Therefore, I fill the missing trading cost of a stock in month t in the following sequence: If the trading cost of the stock in month $t-1$ is available, I use it to fill in the trading cost for the stock in month

[31] Chen and Zimmermann (2021) sign the characteristics so that the average return of the spread portfolios is positive within the sample period of the original paper. I use the same signs as Chen and Zimmermann (2021).

t. If the trading cost in the previous month is not available, I sort stocks into deciles based on their firm sizes in month t. I use the average of the trading costs of the size decile the stock is in month t to fill in the missing trading cost. Last, if the firm size of the stock is missing in month t, I use the average of the trading costs of all stocks in month t to fill in the missing trading cost.

In Appendix F, I plot how effective spreads vary across time and stocks in Figure F1. I also plot the portfolio-level distribution of rebalancing costs and turnover in Figure F2 and Figure F3. These figures show that trading costs vary significantly across stocks, portfolios, and time.

4 Empirical design and results

4.1 Existence

4.1.1 Motivating evidence

To examine the existence of anomalies, the traditional method is to conduct t tests under the null hypothesis of a zero mean alpha. As a motivation for the alpha dynamics tests, I conduct t tests on 205 characteristic-sorted portfolios to show that t tests are insufficient to determine the existence of anomalies.

Similar to Hou et al. (2020), for each characteristic, I construct value-weighted long-short decile portfolios. Portfolios are then held for one month and rebalanced. Within each sample period (full-sample, before-sample, in-sample, and post-sample periods), I conduct t tests on the CAPM alpha, and I label anomalies that fail the t test with a cutoff of 1.96 as suspicious anomalies. Moreover, in light of potential p-hacking, Harvey et al. (2016) recommend increasing the t-cutoff to 3.0. I label the anomalies that pass the t test with a cutoff of 1.96 but fail with a cutoff of 3.0 as HLZ suspicious anomalies. The remaining anomalies are labeled as robust.

In the full sample period, results show that there are 92 suspicious, 41 HLZ suspicious, and 72 robust anomalies. The failure rate based on t tests is similar to Harvey et al. (2016) and Hou et al. (2020). Further, some researchers also worry that the portfolio construction method adopted by the original authors may be subject to data mining (Hasler, 2022). Overall, these studies argue that many anomalies may not truly exist, and collective "data snooping" might be a significant issue.

In contrast, Chen and Zimmermann (2021) show that all 205 anomalies exist if we use the same portfolio construction method adopted by the original authors.[32]

To conclude, it is still uncertain whether certain anomalies exist or not based solely on the t tests on the mean alpha. Motivated by this, I proceed to examine whether alpha dynamics can provide new insights relevant to assessing whether anomalies are attributable to collective data snooping or are real.

4.1.2 The monotonicity test

I start by testing Hypotheses 2 to examine the existence of monotone patterns in alphas after portfolio sorting within each sample period. I measure the returns of eleven characteristic-sorted portfolios with a fixed number of months skipped after sorting ($k = 0$) and different holding periods ($h = 1, 6, 12, 18, ..., 60$). In Figure 3a, I illustrate the relationship between calendar time t, k, and h. Choosing a subset of holding periods rather than considering all possible holding periods is to maintain statistical power. Further, the six-month gap between adjacent portfolios allows for relatively large differences in mean alphas, which enhances statistical power.

To assess whether the CAPM alphas across the eleven portfolios exhibit a statistically significant monotone pattern, I extend the monotonicity test proposed by Patton and Timmermann (2010, PT). While PT use the test to examine monotone relations in the cross-

[32] The construction decisions include the holding period, listing exchanges, the weighting scheme, etc. Bessembinder et al. (2022) document that about half of the original papers use equal weight to construct portfolios.

section of stocks, I use it to examine monotone relations in the event time following portfolio sorting. The monotonicity test also allows betas to vary across the eleven portfolios, as CAPM alphas are separately estimated for each portfolio.

Denote the CAPM alphas of the eleven portfolios as $\alpha^{op}_{1,t}$, $\alpha^{op}_{2,t}$, ..., $\alpha^{op}_{11,t}$. Let $\Delta_i = E[\alpha^{op}_{i,t}] - E[\alpha^{op}_{i+1,t}]$ for $i = 1, ..., 11$, where E is an expectation operator. Let $\Delta \equiv [\Delta_1, ..., \Delta_{11}]'$. I examine four tests proposed by PT. The first is to test for a strictly monotone relation (MR test) in Δ. The null is a constant or weakly decreasing pattern and the alternative is a strictly increasing pattern.

$$\begin{aligned} H_0 &: \Delta \leq 0 \\ H_1 &: \Delta > 0 \end{aligned} \tag{10}$$

The alternative hypothesis is the one that I would like to test. Therefore, I flip the long and short portfolios when investigating whether alphas decrease over time and do not flip when investigating whether alphas increase over time.

If statistical power is low, the MR test may fail to detect a monotone pattern for these anomalies. To address the issue of low statistical power, PT propose the next two tests. They examine whether at least some parts of the pattern of Δ are strictly positive (Up test) or negative (Down test). The Up and Down tests both have a null hypothesis of a constant pattern. Therefore, they are less restricted than the MR test and potentially detect any small deviation from a constant pattern. Specifically, for the Up test,

$$\begin{aligned} H_0 &: \Delta = 0 \\ H_1^+ &: \sum_{i=1}^{11} |\Delta_i| \mathbb{1}\{\Delta_i > 0\} > 0 \end{aligned} \tag{11}$$

where the indicator $\mathbb{1}\{\Delta_i > 0\}$ is one if $\Delta_i > 0$. And for the Down test,

$$
\begin{aligned}
&H_0 : \Delta = 0 \\
&H_1^- : \sum_{i=1}^{11} |\Delta_i| \mathbb{1}\{\Delta_i < 0\} > 0
\end{aligned} \tag{12}
$$

where the indicator $\mathbb{1}\{\Delta_i < 0\}$ is one if $\Delta_i < 0$. Up and Down tests apply a non-parametric method. The distribution of the statistics in the tests is estimated from 1,000 bootstrapping replications.

The last test is Bonferroni bound. This test is more conservative as discussed by PT. It analyzes whether the minimum t-statistic on estimated $\Delta_i,\ i = 1, ..., 11$, falls below the critical value derived from a bound on the probability of a Type I error.

Finally, under a 5% significance level, I examine whether the statistics in either the MR test, the Up test, the Down test, or the Bonferroni bound test are significant.

4.1.3 The optimization test

I continue to test Hypotheses 4 and examine whether alphas are constant or not within a subset of months after sorting within each sample period. The first step is to measure the returns of the optimization strategy. Motivated by Proposition 1, the holding period is set to one month ($h = 1$). To alleviate data mining concerns and look-ahead bias, I conduct a conditional analysis by running expanding regressions. The first 60 months are burn-out months. At the end of any month $t \geq 60$, I search for the optimal number of months to skip (k^*) that maximizes the mean alpha from the set of $\{0, 12, 24, ..., 60\}$ and is based on the information already known at t.[33] I form a spread portfolio based on k^* at the end of month t. And I calculate the return of the portfolio in $t + 1$. This process is repeated until the end of the sample. The returns in $t + 1$ represent the out-of-sample returns of the strategy, and I use them to estimate the CAPM alpha. In Figure 3b, I present the timeline for the estimation of k^*, formation of the portfolio, and measurement of the returns. This timeline

[33]The set of k ensures that the same subset of months is considered as in the monotonicity test. Further, as discussed in the previous section, this reduces the concerns of data mining.

is also used in all optimization strategies of the paper.

Next, I measure the returns of a benchmark strategy. The returns are used to estimate the mean alpha of the strategies across the ks from the set $\{0, 12, 24, ..., 60\}$, as described in Hypotheses 4. At the end of any month $t \geq 60$, I form a portfolio by equally weighting the portfolios based on $k = 0, 12, 24, ..., 60$. I then calculate the return of this portfolio in month $t + 1$. This process is repeated until the end of the sample.

Finally, I compare the alpha of the optimization strategy to that of the benchmark strategy. In each sample period, I run the following regression:

$$r_{opt,t} - r_{b,t} = a_{opt} + b_m r_{m,t} + \epsilon_{opt,t} \tag{13}$$

where $r_{opt,t}$, $r_{b,t}$, and $r_{m,t}$ are the returns of the optimization strategy, the benchmark strategy, and the market. Similar to the monotonicity test, this test also allows the betas to the market factor to vary at different horizons. b_m is the difference between the beta of the optimization strategy and that of the benchmark strategy.

Under a 5% significance, if a_{opt} is positive and statistically significant, the optimization strategy generates a higher alpha than the average of the alphas of alternative strategies in k. This rejects the null hypothesis and indicates the existence of non-zero alphas.

The challenge then lies in how to infer the statistical significance of a_{opt}. Throughout the paper, all regressions require at least 20 observations. However, statistical inference may still be biased when the sample size is small. Furthermore, errors ($\epsilon_{opt,t}$) may have heteroskedasticity and autocorrelation issues.

To address these concerns, I use a bootstrapping approach to estimate the p-value of a_{opt}. Let $\{r_{i,t} \mid t = 1, ..., T;\ i = opt,\, b,\, m\}$ be the actual returns recorded for the optimization strategy, the benchmark strategy, and the market over T months for a sample period. I first use the stationary bootstrap of Politis and Romano (1994) to randomly draw (with replacement) a new sample of returns $\{\tilde{r}^{(b)}_{i,\tau} \mid \tau(1), ..., \tau(T);\ i = opt,\, b,\, m\}$. Here, $\tau(t)$ is the new time index, randomly drawn from the actual data $\{1, ..., T\}$. $\tau(t)$ is common across i to preserve cross-sectional dependencies in returns. The bootstrap replication number, denoted as b, ranges from 1 to 2,000. Furthermore, to account for time series dependencies, returns data are drawn in blocks, I choose the average block length to be 10 months. Within each bootstrapping replication, I estimate Equation 13 and obtain an estimate of a_{opt}. Finally, I obtain a distribution of a_{opt}, and I calculate whether a_{opt} is greater than 0 in 95% of 2,000 replications.

4.1.4 The CAPM

Results for the CAPM alpha are shown in Table 2. In the full sample period, out of 92 suspicious and 41 HLZ suspicious anomalies, 19 (21%) suspicious and 24 (59%) HLZ suspicious anomalies reject the null hypothesis of constant alphas in either the monotonicity test or the optimization test. Results indicate that these anomalies generate non-zero alphas after sorting. Further, results indicate that alpha-dynamics tests can detect the existence of non-zero alphas when alpha-mean tests do not. Moreover, results indicate that raising the t cutoff from 1.96 to 3.0 leads to a higher rate of failure to detect real anomalies.

Results in Panel D of Table 2 provide information about the statistical power of alpha-dynamic tests. In the full-sample period, 82% of robust anomalies pass alpha-dynamic tests. This indicates that alpha-dynamic tests do not fully subsume alpha-mean tests. It would be more informative to consider both alpha-dynamic tests and alpha-mean tests.

Additionally, in the before-sample, in-sample, and post-sample periods, the numbers of suspicious anomalies that exhibit statistically significant non-constant alphas are 17, 28, and 37. Results indicate that the tests are useful across subperiods. Importantly, alpha-dynamic tests are more useful in the post-sample periods.

Among all 205 anomalies, the total number of anomalies that present a statistically significant non-constant pattern of alphas are 93 and 75 in the in-sample and post-sample periods, which is consistent with McLean and Pontiff (2016). Since anomalies that are due to mispricing are likely to be arbitraged away after publication, fewer anomalies should generate non-zero alphas in the post-sample period.

Furthermore, the monotonicity test and the optimization test complement each other well. In each subperiod, both tests can detect anomalies that may have true alphas while the other test cannot.

4.1.5 Multi-factor models

I use the CAPM as an illustration. It is also interesting to examine alpha dynamics relative to canonical multi-factor models. In this section, I examine alphas relative to Fama and French (1993, FF3) three-factor model, a four-factor model including the factors in the FF3 and momentum (FF3+MOM), Fama and French (2015, FF5) five-factor model, and Hou, Xue, and Zhang (2015, HXZ) model. For each model, I conduct the same alpha-dynamic tests (both the monotonicity test and the optimization test).

Results are shown in Table 3. The classifications of suspicious, HLZ suspicious, and robust anomalies are determined by the results of the t-test on the mean alpha relative to each model, following the same approach as with the CAPM. In the full sample period, the total number of suspicious anomalies exhibiting statistically significant non-constant alphas for the CAPM, FF3, FF3+MOM, FF5, and HXZ are 19, 23, 24, 23, and 34, respectively. These corresponding figures are also similar in the in-sample and post-sample periods. The results indicate that alpha-dynamic tests remain useful for these canonical multi-factor models. Furthermore, while the corresponding figures for robust anomalies decrease with bigger models as bigger models digest more anomalies, those figures for suspicious and HLZ suspicious anomalies do not decrease. The results indicate that the performance of those bigger models was overstated.

4.2 Mispricing or rational expectations

If alphas exist, they could be due to mispricing or rational expectations. In this section, I focus on the anomalies in my sample that show statistically significant non-zero alphas based on either t tests on mean alphas (with a cutoff of 1.96) or alpha-dynamic tests. In total, I examine 132, 65, 128, and 92 anomalies in the full-sample, before-sample, in-sample, and after-sample periods.

4.2.1 The ripple test

To identify the ripple pattern in alphas described in Section 2.3, I design a new test that extends the optimization test. Specifically, I study alphas over nine years after the sorting date. I consider this long horizon because the relevant models often imply that ripple patterns will manifest after a delay and prior papers show that mispricing can persist for many years (Cho & Polk, 2023; van Binsbergen et al., 2023).

I divide the nine years into five non-overlapping subperiods ([1,13], [13,37], [37,61], [61,85], and [85,109]). For each subperiod, I measure the returns of an optimization strategy using expanding regressions as in the optimization test. I examine whether the optimization strategy generates a statistically significant higher alpha than those at the start and end of the subperiod. By comparing the alpha of the optimization strategy with that at the start, I test the null hypothesis that the alpha at the start is the highest. A rejection implies that there is a higher alpha in later periods, indicating an increasing pattern within the subperiod. Similarly, I test the null hypothesis that the alpha at the end is the highest to examine whether there is a decreasing pattern.

To provide more intuition, denote a subperiod as $[k_{start}{+}1,\ k_{end}{+}1]$. At the end of each month t, the optimal strategy searches for the optimal k^* from the range $[k_{start}, k_{end}]$ that generates the highest alpha.[34] By running expanding regressions over time, this search is based on known information before month t. Then at the end of month t, a long-short portfolio is constructed based on k^* and the return in month $t+1$ is calculated.

To estimate the CAPM alphas at the start and end of the subperiod ($\alpha_{k_{start}+1}$ and

[34] In the first subperiod, k is searched from the set of 0, 3, 6, 9, 12. For the rest of the subperiods, k is searched from the smallest possible value to the largest possible value with a step of 6. For example, in the second subperiod, k is searched from the set of 12, 18, 24, 30, 36.

$\alpha_{k_{end}+1}$), I measure the returns of two benchmark strategies: one that always sets $k = k_{start}$ and $h = 1$, and another that always sets $k = k_{end}$ and $h = 1$. Last, to assess whether the alpha of the optimization strategy generates a statistically significant higher alpha than those at the start and end of the subperiod, I run the following regression:

$$r_{opt,t} - r_{edge,t} = \phi_0 + \phi_1 r_{m,t} + \epsilon_t \tag{14}$$

where $r_{opt,t}$ represents the return of the optimization strategy and $r_{edge,t}$ represents the return of either of the two benchmark strategies. $r_{m,t}$ is the market return and ϵ_t is the residual. Under a 5% significance level, if ϕ_0 is statistically significant and positive, it implies that the optimization strategy generates a higher CAPM alpha.

If the optimal strategy generates a statistically significant higher CAPM alpha than that on the left edge of the subperiod ($\alpha_{k_{start}+1}$), it implies that $\alpha_{k_{start}+1}$ is not the highest within the subperiod. This indicates that there is at least one increasing pattern of alphas within the subperiod [k_{start}+1, k_{end}+1]. With the same logic, if the optimal strategy generates a statistically significant higher CAPM alpha than that on the right edge of the subperiod ($\alpha_{k_{end}+1}$), then $\alpha_{k_{end}+1}$ is not the highest. This indicates that there is at least one decreasing pattern within the subperiod. Figure 4 presents this intuition with examples.

This empirical design separates the whole period into five subperiods. This is because the alpha in the first month after sorting ($k = 0$) is the highest for most anomalies as shown in Figure F4. When the alpha in the first month is highest, I cannot detect any increasing pattern if I do not divide the whole period into subperiods. Further, while having more subperiods may theoretically provide a better recovery of the pattern of alphas, empirically, we cannot use a large number of subperiods due to reduced statistical power after adjusting for multiple testing. Moreover, after sorting, alphas usually decay faster at first, then decay slower. As a result, I set the range of the first subperiod to twelve months and the subsequent subperiods to two years.

Finally, since I examine five subperiods and compare two differences in CAPM alphas in each subperiod, there are ten hypotheses in total. To correct for multiple hypothesis testing, I adjust the p values by the Benjamini-Hochberg method to ensure that the expected proportion of false discovery rate is no greater than 5%. The Benjamini-Hochberg method is also employed by Harvey et al. (2016) and Keloharju et al. (2021). Under a 5% significance level, if there are both increasing and decreasing patterns of alphas within the nine-year period after sorting, it implies that the ripple pattern exists.

Figure 5 gives an example. The ripple test is designed to detect a statistically significant decrease in alphas in the first subperiod [1,13], no pattern in [13,37],[35] an increase in [37,61], both increase and decrease in [61,85], and a decrease in [85,109].

4.2.2 Results

In Table 4, I present the patterns of several well-known anomalies. The table shows that there is a statistically significant ripple pattern for accruals ($Accruals$), idiosyncratic volatility ($IdioRisk$), momentum ($Mom12m$), and net share issuance ($SharIss1Y$) in the full sample period.

Comparing the results with prior studies, book-to-market (B/M) shows a statistically significant increasing pattern in the first twelve months after sorting. The increasing pattern is consistent with the finding in Giglio, Kelly, and Kozak (2023). This indicates the robustness of the ripple test. In contrast to their study, while they estimate portfolios' term structure, they have not formally tested the statistical significance of any pattern in alphas, which is the focus of this paper.

In Table 5, I present all the anomalies in my sample that exhibit a statistically significant ripple pattern in different sample periods. The total number of anomalies that show this pattern in the full-sample, before-sample, in-sample, and post-sample periods are 79 (60%),

[35] We can design another test to examine the hypotheses that alphas at the start or end of a subperiod is the lowest. This test should detect both statistically significant increase and decrease in alphas in the subperiod [13,37]. Theoretically, testing these hypotheses can provide us with a better identification of increasing and decreasing patterns of alphas. However, empirically, testing these hypotheses may reduce statistical power after adjusting for multiple testing.

27 (41%), 69 (54%), and 41 (45%). The results indicate that this test is useful in identifying a large number of anomalies that can be attributed at least in part to mispricing. Further, the results show that many other anomalies in the categories of idiosyncratic volatility, net share issuance, and momentum also exhibit a statistically significant ripple pattern. Moreover, many anomalies such as accruals ($Accruals$) and net share issuance ($ShareIss1Y$) exhibit a statistically significant ripple pattern in the in-sample period, but not in the post-sample period. One potential explanation is that arbitragers correct their mispricing following the publication of anomalies (McLean & Pontiff, 2016).

Furthermore, Results in Table 4 and 5 also indicate that many anomalies generate persistent alphas. For example, Table 4 shows that the CAPM alphas of accruals, idiosyncratic volatility, momentum, and net share issuance persist for at least nine years after sorting.

Moreover, to compare with the tests on mean alphas, I examine the performance of alpha-mean tests in detecting the pattern of price drifts followed by reversals In Appendix D. Results in Table D1 show that alpha-mean tests have much lower power.

4.3 After-cost profitability

If an anomaly does not have non-zero alphas, it is relatively meaningless to study its after-cost profitability. Therefore, I focus on the anomalies that show statistically significant non-zero alphas based on either t tests on mean alphas (with a cutoff of 1.96) or alpha-dynamic tests in different sample periods.

4.3.1 Optimization strategy: A new rule

I develop a new rule for the optimal holding period h. Similar to the procedure of the optimization strategy for mean alphas, I conduct conditional analysis by running expanding regressions over time. Motivated by the discussion in Section 2.4, I first restrict the number of months skipped $k = 0$. At the end of any month t, the optimal holding period h^* is searched from the set of $\{1, 3, 6, 9, 12\}$ to maximize the after-cost alpha and is based on the information already known at t. This set is chosen to enable a comparison with prior studies, as prior studies mainly use $h = 1$ and $h = 12$ (Chen & Zimmermann, 2021). Then I construct spread portfolios based on h^* with the overlapping portfolio approach at the end of month t, and record the monthly after-cost returns of the portfolios between $t + 1$ and $t + h^*$. To reduce the computing time in calculating the rebalancing costs, I estimate the optimal h^* every five years. That is, I use the same h^* in the following five years after an estimation.

4.3.2 Benchmark strategies: Prior rules

To compare with prior methods, I first consider a benchmark strategy that always sets $k = 0$ and $h = 12$ (henceforth, H12). Since the optimization strategy searches h^* from $\{1, 3, 6, 9, 12\}$ and H12 already uses the longest holding period in this set, trading costs cannot be further reduced by h^*. Therefore, any superior performance from the optimization strategy should only come from the impact of alpha dynamics on the alpha mean. That is, comparing the performance of the optimization strategy with that of H12 provides direct evidence of the importance of considering alpha dynamics when evaluating after-cost profitability.

I also consider three other benchmark strategies used in prior papers. The second benchmark strategy always sets $k = 0$ and $h = 1$ (henceforth, H1). $h = 1$ is also the most frequently used holding period in the literature (Chen & Zimmermann, 2021). The third benchmark strategy determines the holding period by the turnover rate rule in Novy-Marx and Velikov (2016, henceforth, the NV method). The rule is that if each of the long and short sides, on average, turns over less than once a year, a one-month holding period is used. In other cases, a twelve-month holding period is used. Finally, the fourth benchmark strategy sets the holding period the same as that in the original papers (Chen & Velikov, 2023, henceforth, the CV method).[36]

[36] In the original papers, holding periods were distributed as follows: one month (104 anomalies), three months (7 anomalies), six months (2 anomalies), twelve months (91 anomalies), and thirty-six months (1 anomaly).

4.3.3 Performance metrics

To compare the performance between the optimization strategy and benchmark strategies, I consider two performance metrics. For the first one, I run the following regression to examine whether the optimization strategy generates a statistically significant higher after-cost alpha than the benchmark strategy.

$$r_{s,t} - r_{b,t} = a + b_m r_{m,t} + \epsilon_{s,t} \tag{15}$$

where $r_{s,t}$ represents the after-cost returns of the optimization strategy, $r_{b,t}$ represents the after-cost returns of a benchmark strategy, and $r_{m,t}$ represents the after-cost market return. If a is positive and significant at the 5% significance level, the optimization strategy generates a higher after-cost CAPM alpha than the benchmark strategy. This metric is also of interest to an investor who only invests in the anomaly.

Additionally, investors may be interested in how to optimally trade both the anomaly and the market factor to maximize the Sharpe ratio. If $r_{s,t}$ is exposed to additional risk factors that $r_{b,t}$ is not exposed to, investors should combine $r_{s,t}$ with the market factor. This is because allocating some positive weight to the optimization strategy can improve the investment opportunity of the investors already trading the benchmark strategy and the market factor. The second performance metric is obtained from the regression of:

$$r_{s,t} = a_s + b_s r_{b,t} + b_m r_{m,t} + \epsilon_{s,t} \tag{16}$$

The optimization strategy is considered to outperform a benchmark strategy if a_s is positive and significant at a 5% significance level.

4.3.4 Results

In Table 6, I present the number of anomalies for which the optimization strategy outperforms the benchmark strategies. In Panel A, I show the results when H12 is the benchmark strategy. The optimization strategy generates a statistically significant higher after-cost alpha (a) for 31, 6, 30, and 12 anomalies in the full-sample, before-sample, in-sample, and post-sample periods, respectively. Further, the returns of the optimization strategy cannot be fully explained by the returns of H12 and the market factor ($a_s > 0$) for 26, 5, 28, and 10 anomalies, respectively. In Panel B, I show the results when H1 is the benchmark strategy. Although fewer anomalies (16) generate a statistically significant higher after-cost alpha in the full sample period, more anomalies (34) cannot be explained by H1 and the market factor. These results indicate that the optimal holding period varies across anomalies.

In Panel C and D, I show the results when the benchmark strategies are NV and CV, respectively. Results are similar to those for H1 and H12. For economic magnitudes, the average improvement in monthly after-cost alpha (a) is 0.30% when NV is the benchmark and 0.28% when CV is the benchmark in the full-sample period.

Furthermore, the average improvement is stronger in the post-sample period compared to the in-sample period. For example, the average improvement in monthly after-cost alpha is 0.46% when NV is the benchmark and 0.44% when CV is the benchmark. This might be surprising considering that trading costs have been much smaller in recent years as shown in Figure F1. However, alphas are also smaller after publication (McLean & Pontiff, 2016). This means that trading costs are still large relative to alphas for several anomalies and cannot be ignored (Proposition 3). Therefore, considering an optimal holding period to evaluate after-cost profitability remains important in the post-sample period.

Overall, the results indicate that prior studies may have significantly underestimated the after-cost profitability of anomalies.

4.4 Robustness

4.4.1 Random variation in mean alphas

This section examines whether the statistically significant non-constant alphas present in Table 2 can be solely attributed to random variation. In Appendix F, Figure F4 shows the number of anomalies that generate the highest CAPM alpha at different k values in different sample periods. Unlike the optimization test, which conducts conditional analysis with expanding regressions, these results are based on unconditional analysis. This figure shows that $k = 0$ generates the highest alphas for most of the anomalies, indicating alphas decay over time for the majority of the anomalies. If alpha dynamics are random, the distribution should be evenly distributed across different k values. Results imply that the results in Table 2 cannot be explained by random variation alone.

4.4.2 Random variation in after-cost alphas

Can the results for after-cost profitability in Table 6 be explained by random variation alone? To examine this question, I conduct simulations following a similar procedure as in Bessembinder, Burt, and Hrdlicka (2021). In Appendix E, I describe how I conduct the simulations. The basic idea is to assume a data-generating process that no strategy based on k and h can outperform H1 (a strategy that always sets $k = 0$ and $h = 1$) based on the statistics a and a_s in Equations 15 and 16. Then I conduct the same optimization strategies with simulated returns and investigate how many anomalies can outperform H1 in each simulation. I report the results in Figure E1. The results indicate that random variation alone cannot explain the results. For example, under a t-cutoff of 1.96, 2.00, 2.50, 3.00, 3.50, and 4.00, the maximum numbers of anomalies that have positive and statistically significant a from 2,000 simulations are 10, 10, 5, 4, 2, 1 in the full-sample period. In contrast, these numbers are 16, 15, 10, 10, 7, and 7 from the actual data when the optimization strategy searches for h only.

4.4.3 Does k affect after-cost profitability?

As discussed in Section 2.4, k may also affect after-cost profitability. To investigate the impact of k on after-cost profitability, I remove the constraint of $k = 0$ in the optimization strategy outlined in Section 4.3 and conduct an optimization strategy that searches for both

k and h to maximize after-cost alphas. Results are shown in Table F1. I compare the results to those in Panel A and B in Table 6 when the optimization strategy only searches for h. Under different performance metrics and different benchmark strategies, the number of anomalies that outperform the benchmark strategies is always similar or even lower when the optimization strategy searches for both k and h. Results imply that k does not have a significant impact on after-cost profitability.

4.4.4 Information ratio

Alphas have not considered risk. Studies such as Barillas and Shanken (2017) suggest that information ratio (IR) is also important, where IR equals alpha divided by the volatility of the residual. It is important because the optimal factor construction method should maximize the information ratio relative to a benchmark model. The intuition is that the maximum obtainable Sharpe ratio of the new model that includes the new factor is equal to the sum of the maximum obtainable Sharpe ratio of the old model and the information ratio of the new factor. In this section, I examine the impact of considering the dynamics of information ratio in the construction of the size and book-to-market factors in the three-factor model of Fama and French (1993, FF3).

FF3 considers a holding period of twelve months (H12). H12 is also the traditional method used for constructing factors (e.g., Hou et al., 2015). Therefore, I take H12 as the benchmark strategy and examine whether H12 is optimal for the size and book-to-market factors based on the IR relative to the CAPM. For book-to-market, I investigate both the version that uses the market equity in December of the prior year ($BMdec$) and the version that uses the latest market equity (BM). I conduct a similar optimization strategy as before. The difference is that, in the optimization strategy, I search for k from $\{0, 3, 6, 9, 12\}$ and set $h = 1$ to maximize IR instead of alphas. Then I compare the IR of the optimization strategy with that of H12.

Results are shown in Table F2. For both size and BM, there is a significant increase in

IR compared to H12. When considering the optimization strategy, the IR almost triples in the full sample period. For example, the IR for the optimization strategy is 0.118 for size and 0.077 for BM. In contrast, these values for H12 are 0.042 and 0.0.027, respectively.

Holding periods h may also affect the IR. Since the overlapping portfolio approach is used, a longer holding period means holding multiple portfolios each month. This leads to a diversification effect among the portfolios. To investigate the effect of h, I conduct optimization strategies that search for h only or search for both k and h. Results in Table F2 imply that h does not have an impact on IR since these strategies do not outperform the strategy that searches for k only.

A comprehensive analysis of different factors and factor models is beyond the scope of this paper. However, results for the construction of size and book-to-market factors in FF3 indicate that other dynamics, such as the dynamic of IR, could also be useful. This poses additional challenges to standardized portfolio construction methods in prior studies that largely overlook the dynamics of these metrics.

5 Conclusion

While prior studies on anomalies focus on the average of alphas, this paper studies how alphas evolve over time after sorting. The paper shows that studying alpha dynamics helps us better understand anomalies.

I introduce new tests to examine alpha dynamics predicted by economic theories. Results show that alpha dynamics provide important insights relevant to evaluating whether anomalies (1) exist, (2) can be attributed at least in part to mispricing, and (3) generate positive profit after considering trading costs. First, alpha-mean tests have several problems, and relying solely on alpha-mean tests fails to detect many real anomalies. Alpha-dynamic tests alleviate these problems and help better detect real anomalies. Furthermore, a large proportion of anomalies exhibit statistically significant ripple patterns that conform to existing behavioral models rather than rational models. Moreover, determining the holding period based on both alpha dynamics and trading costs significantly improves after-cost profitability. Therefore, this indicates that this rule is more appropriate to evaluate after-cost profitability. Overall, results indicate that more published anomalies are real and more are profitable than previously thought, and a large proportion of anomalies can be attributed at least in part to mispricing.

These results have important implications for academics seeking to understand anomalies, firm managers estimating discount rates, and investors considering asset allocations and trading strategies.

References

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of behavioral economics: Applications and foundations 1* (pp. 79–175, Vol. 1). Elsevier.

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2015). X-capm: An extrapolative capital asset pricing model. *Journal of Financial Economics*, *115*(1), 1–24.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343.

Barillas, F., & Shanken, J. (2017). Which alpha? *The Review of Financial Studies*, *30*(4), 1316–1338.

Berk, J. B., Green, R. C., & Naik, V. (1999). Optimal investment, growth options, and security returns. *The Journal of Finance*, *54*(5), 1553–1607.

Bessembinder, H., Burt, A., & Hrdlicka, C. M. (2021). Time series variation in the factor zoo. *Aaron Paul and Hrdlicka, Christopher M., Time Series Variation in the Factor Zoo (December 22, 2021).*

Bessembinder, H., Burt, A., & Hrdlicka, C. M. (2022). Factor returns and out-of-sample alphas: Factor construction matters. *Aaron Paul and Hrdlicka, Christopher M., Factor Returns and Out-of-Sample Alphas: Factor Construction Matters (November 19, 2022).*

Boguth, O., Carlson, M., Fisher, A., & Simutin, M. (2011). Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics*, *102*(2), 363–389.

Bowles, B., Reed, A. V., Ringgenberg, M. C., & Thornock, J. R. (2023). Anomaly time. *Available at SSRN 3069026.*

Chan, W. S. (2003). Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of financial economics*, *70*(2), 223–260.

Chen, A. Y., & Velikov, M. (2023). Zeroing in on the expected returns of anomalies. *Journal of Financial and Quantitative Analysis*, *58*(3), 968–1004.

Chen, A. Y., & Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming.*

Cho, T., & Polk, C. (2023). Putting the price in asset pricing. *Journal of Finance forthcoming.*

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, *53*(6), 1839–1885.

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

Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *The Journal of Finance*, *65*(4), 1237–1267.

Engelberg, J., McLean, R. D., & Pontiff, J. (2018). Anomalies and news. *The Journal of Finance*, *73*(5), 1971–2001.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

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

Gârleanu, N., & Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, *68*(6), 2309–2340.

Giglio, S., Kelly, B. T., & Kozak, S. (2023). *Equity term structures without dividend strips data* (tech. rep.). National Bureau of Economic Research.

Gomes, J., Kogan, L., & Zhang, L. (2003). Equilibrium cross section of returns. *Journal of Political Economy*, *111*(4), 693–732.

Hackbarth, D., & Johnson, T. (2015). Real options and risk dynamics. *The Review of Economic Studies*, *82*(4), 1449–1482.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hasler, M. (2022). Looking under the hood of data-mining. *Available at SSRN 4279944*.

Hendershott, T., Menkveld, A. J., Praz, R., & Seasholes, M. (2022). Asset price dynamics with limited attention. *The Review of Financial Studies*, *35*(2), 962–1008.

Heston, S. L., & Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, *87*(2), 418–445.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, *54*(6), 2143–2184.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91.

Jensen, T. I., Kelly, B., & Pedersen, L. H. (2022). Is there a replication crisis in finance? *The Journal of Finance*.

Jensen, T. I., Kelly, B. T., Malamud, S., & Pedersen, L. H. (2022). Machine learning and the implementable efficient frontier. *Available at SSRN 4187217*.

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

Lewellen, J., & Nagel, S. (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics*, *82*(2), 289–314.

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

Michaely, R., Thaler, R. H., & Womack, K. L. (1995). Price reactions to dividend initiations and omissions: Overreaction or drift? *the Journal of Finance*, *50*(2), 573–608.

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, *29*(1), 104–147.

Patton, A. J., & Timmermann, A. (2010). Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics*, *98*(3), 605–625.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313.

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

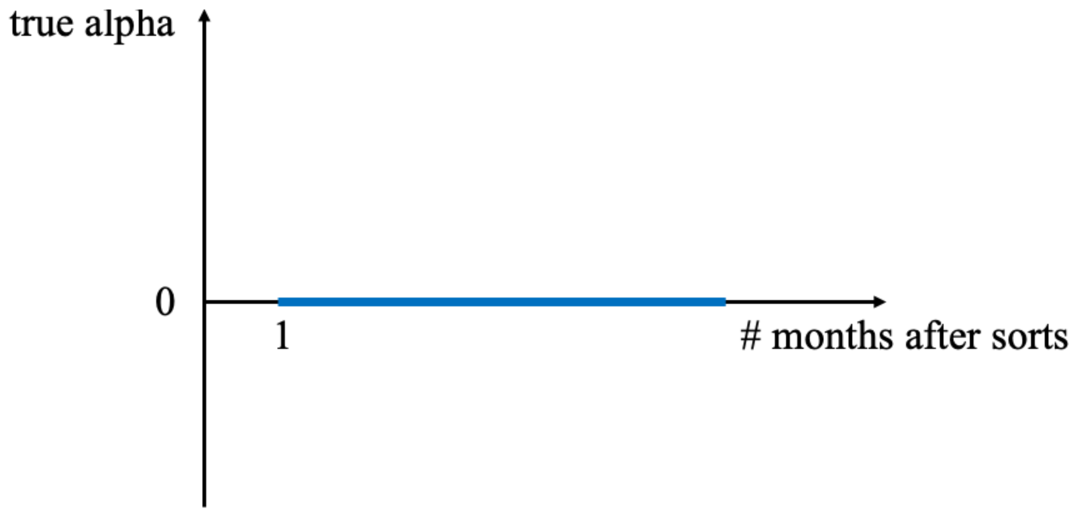
Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies*, *24*(5), 1481–1512.

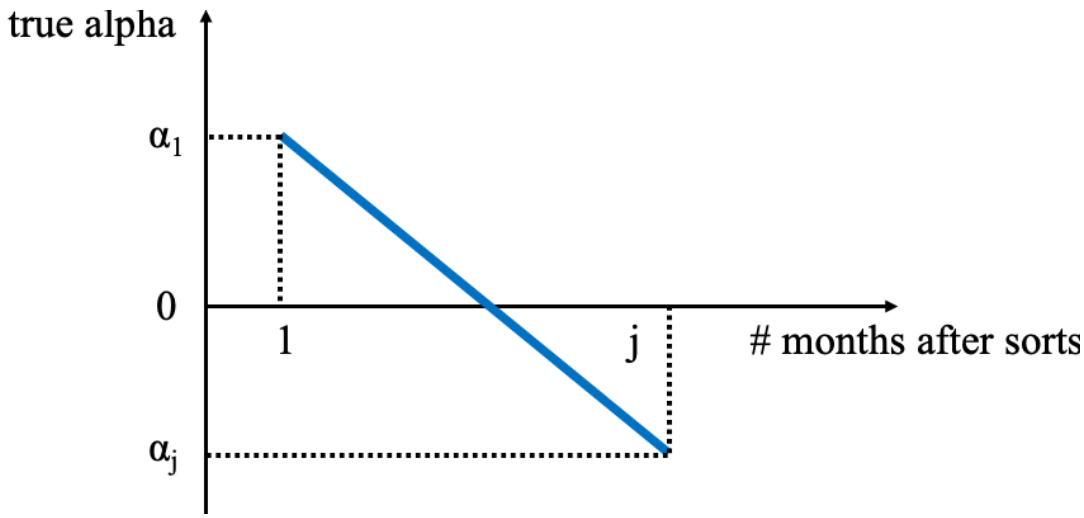
van Binsbergen, J. H., Boons, M., Opp, C. C., & Tamoni, A. (2023). Dynamic asset (mis) pricing: Build-up versus resolution anomalies. *Journal of Financial Economics*, *147*(2), 406–431.

van Binsbergen, J. H., & Opp, C. C. (2019). Real anomalies. *The Journal of Finance*, *74*(4), 1659–1706.

Zhang, L. (2005). The value premium. *The Journal of Finance*, *60*(1), 67–103.



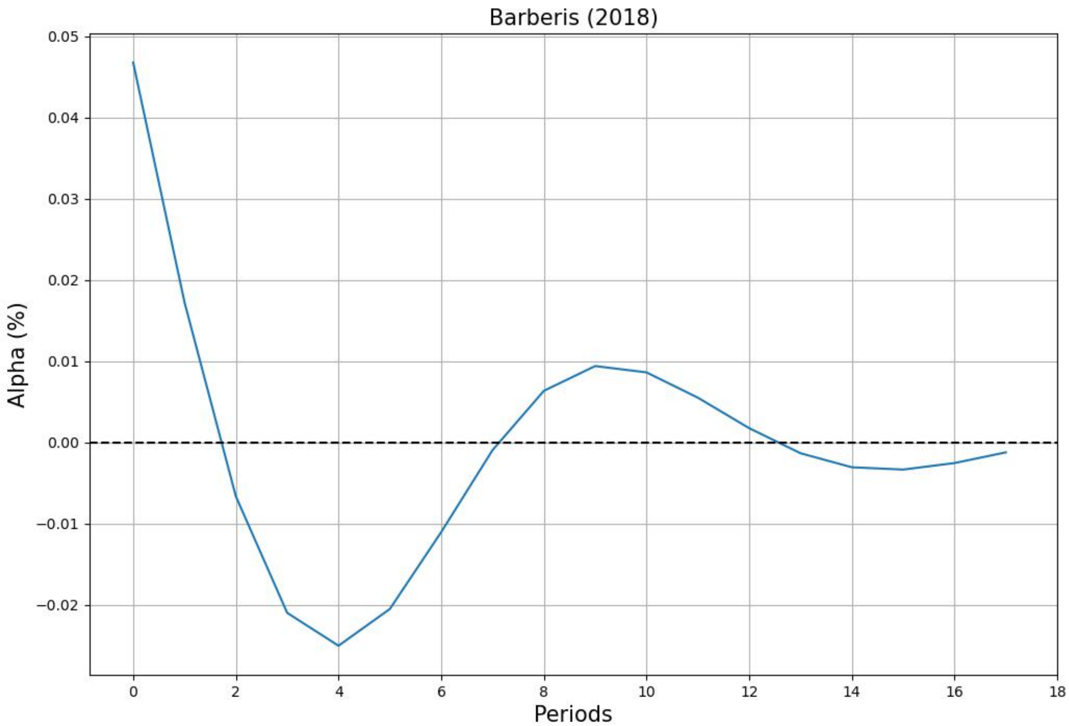
(a) The alpha dynamic when alphas do not exist



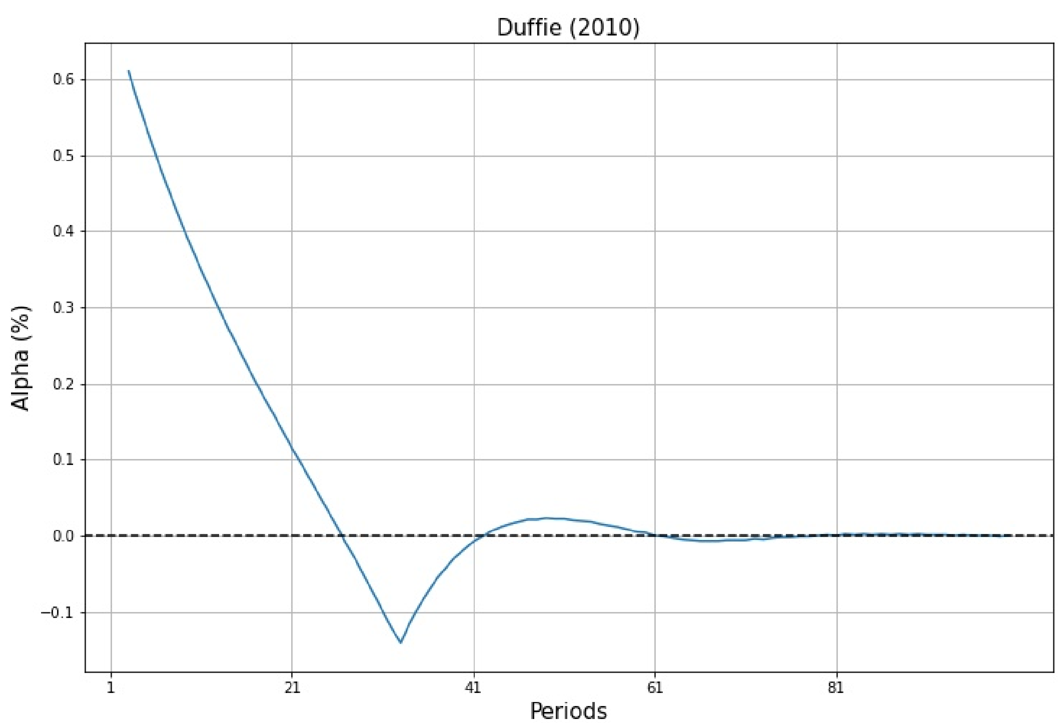
(b) An example of non-constant alpha dynamic

Figure 1: Patterns of alpha dynamics

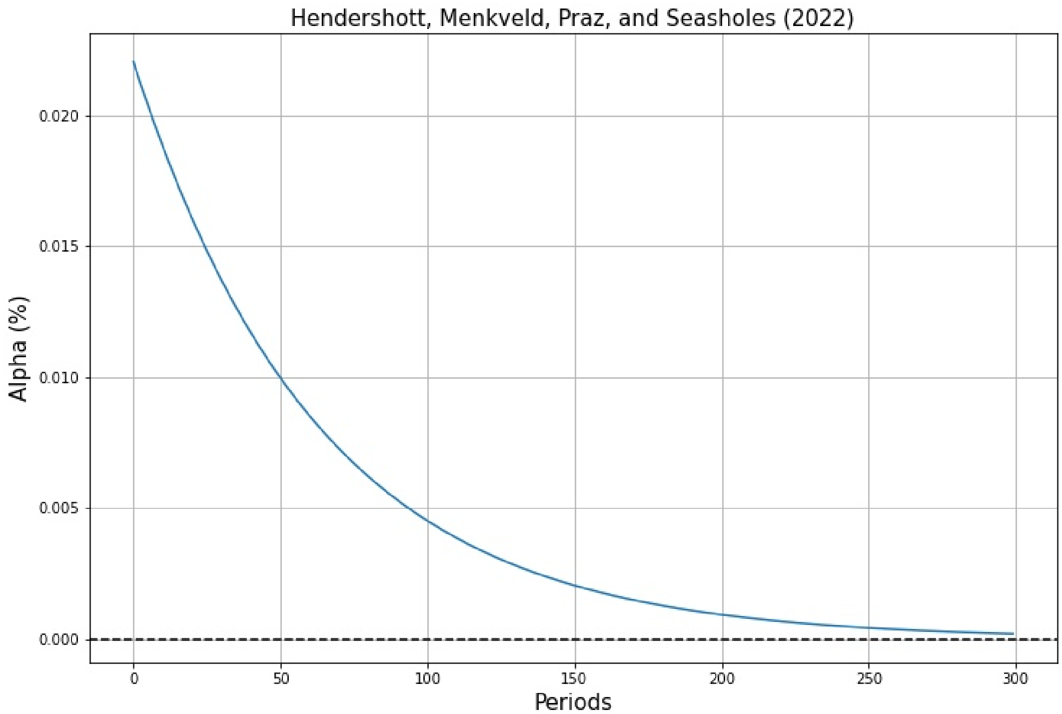
This figure shows the patterns of alpha dynamics. The y axis represents true alphas. x axis represents the number of months after the portfolio sorting period. The true alpha represents the population mean. Figure (a) shows how alphas after sorting evolve under the null hypothesis that alphas do not exist. And Figure (b) shows an example of how true alphas after sorting evolve under a non-constant alpha dynamic.



(a) The alpha dynamic of Barberis (2018)



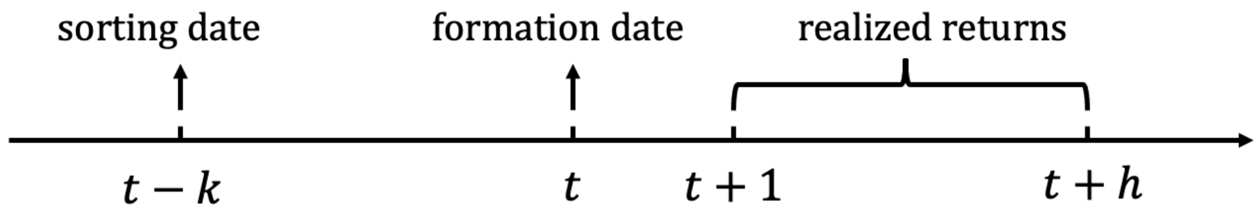
(b) The alpha dynamic of Duffie (2010)



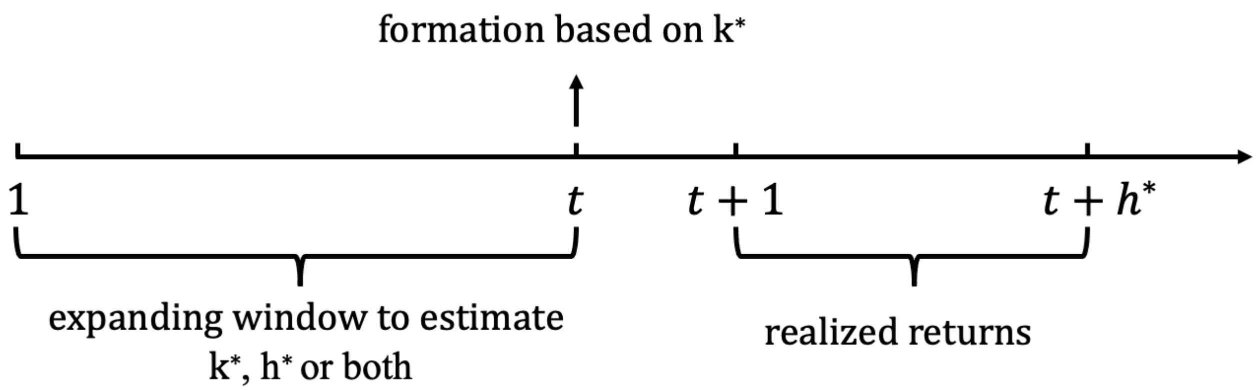
(c) The alpha dynamic of Hendershott, Menkveld, Praz, and Seasholes (2022)

Figure 2: Dynamics of alphas in Barberis (2018), Duffie (2010), and Hendershott, Menkveld, Praz, and Seasholes (2022)

The figures show the patterns of alpha dynamics implied by the models of Barberis (2018), Duffie (2010), and Hendershott, Menkveld, Praz, and Seasholes (2022). The y axis represents true alphas. x axis represents the number of periods after the initial shocking dates when mispricing arises. Alphas equal changes of mispricing over time.



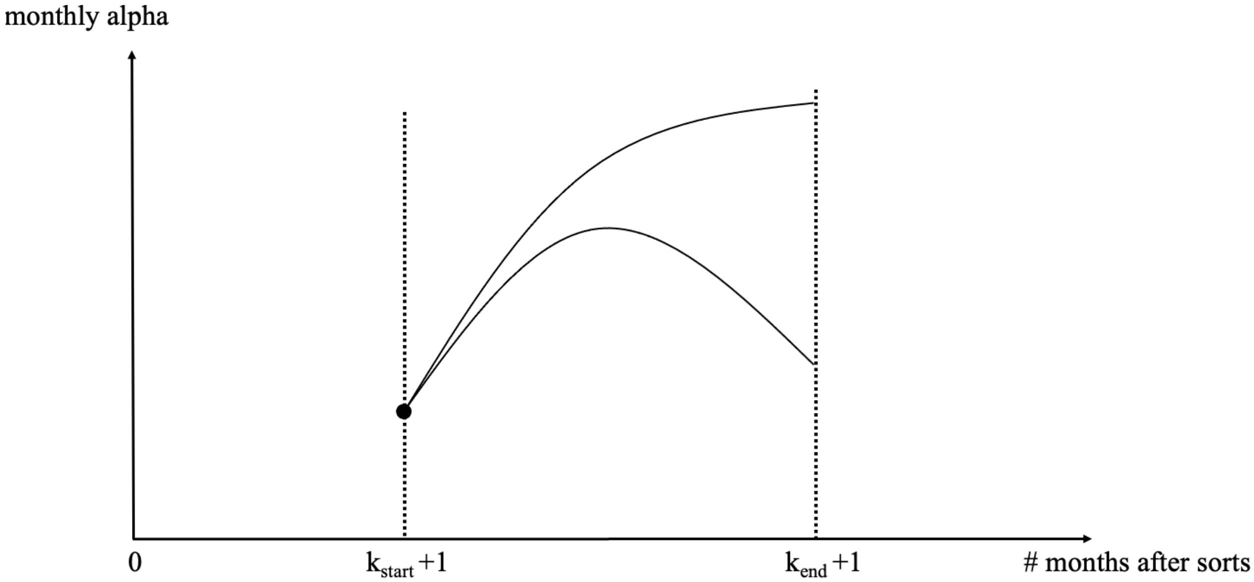
(a) Relationship between t, k, and h



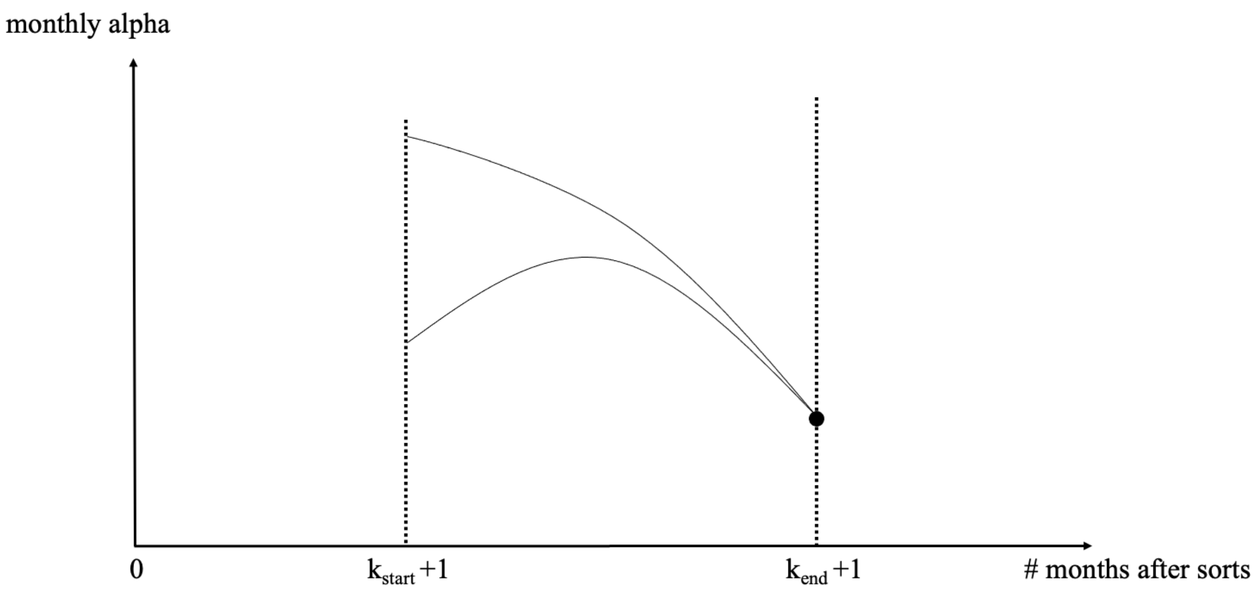
(b) Timeline for estimation, portfolio formation, and measurement of returns

Figure 3: Timelines

Figure 3a shows the relationship between calendar time t, the number of months skipped after sorting k and holding period h. Figure 3b shows the timeline for estimation, portfolio formation, and measurement of returns. When searching only for the optimal k^*, $h = 1$. Additionally, when searching for the optimal h^*, I measure returns from $t+1$ and $t+h^*$ using the overlapping portfolio approach. All optimization strategies in the paper use this timeline.



(a) Examples of alpha dynamics when the optimization strategy generates a higher alpha than that at the start of a subperiod



(b) Examples of alpha dynamics when the optimization strategy generates a higher alpha than that at the end of a subperiod

Figure 4: Potential patterns within a subperiod

The figures show the potential patterns of alpha dynamics within a subperiod of months after portfolio sorting by comparing the alpha of the optimization strategy with those at the start and end of the subperiod. The y axis represents true alphas. x axis represents the number of months after the sorting date. The top figure shows examples of patterns of alpha dynamics within the subperiod $[k_{start}+1, k_{end}+1]$ when the optimization strategy generates a higher alpha than $\alpha_{k_{start}+1}$. The bottom figure shows examples of patterns of alpha dynamics within the subperiod $[k_{start}+1, k_{end}+1]$ when the optimization strategy generates a higher alpha than $\alpha_{k_{end}+1}$.

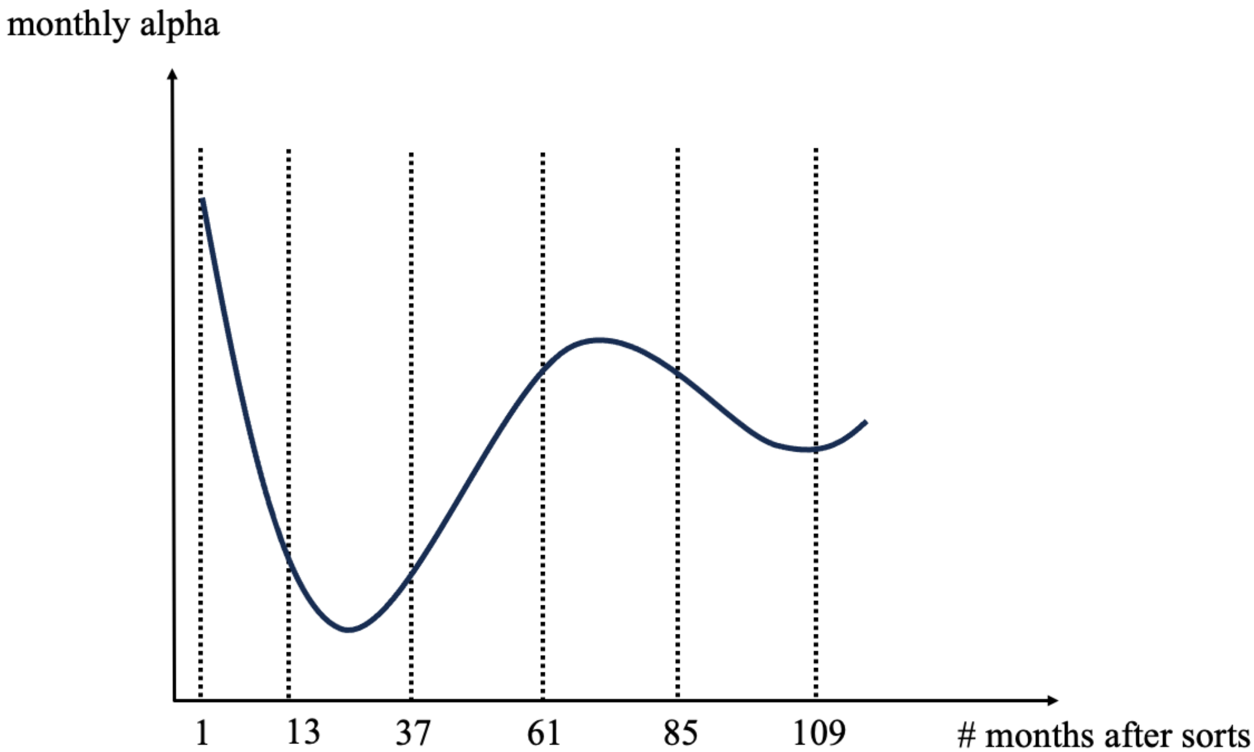


Figure 5: Examining the ripple pattern in the alpha dynamic

The figure provides an example to illustrate how I identify the ripple pattern in the alpha dynamic. The y axis represents monthly alphas, while the x axis represents the number of months after the sorting date. I examine five consecutive, non-overlapping subperiods after sorting: [1,13], [13,37], [37,61], [61,85], and [85,109]. Within each subperiod, I analyze whether there is an increasing pattern, a decreasing pattern, or both in the alpha dynamic. A ripple pattern occurs when the alpha dynamic exhibits both increasing and decreasing patterns over the nine years (109 months) following sorting.

Table 1: A summary of the contributions of alpha dynamics to anomaly explanations

Explanations	New Information from Alpha Dynamics
Existence	Alpha dynamics can detect non-zero alphas when the alpha mean may fail to do so.
Mispricing versus Rational Expectations	Existing behavioral models can imply a ripple-like pattern of alphas. Such a pattern is not implied in any existing rational models. Therefore, the existence of a statistically significant ripple pattern is consistent with anomalies being at least partially due to mispricing.
After-cost Profitability	Holding periods should be determined by jointly considering alpha dynamics and trading costs.

Table 2: alpha-dynamic tests on the existence of anomalies (or non-zero alphas)

This table presents the number of anomalies that pass alpha-dynamic tests under the 5% significance level. Alphas are relative to the CAPM. Within a sample period and with a holding period of one month, I label those that fail the t test with a cutoff of 1.96 as suspicious anomalies. Among the remaining anomalies, I label those that fail the t test with a cutoff of 3.0 as HLZ suspicious anomalies, while the rest as robust anomalies. Panel A shows the results for all anomalies. Panel B shows the results for suspicious anomalies. Panel C shows the results for HLZ suspicious anomalies. And Panel D shows the results for robust anomalies (72 anomalies). $Mono$ shows the results from the monotonicity test. *Opt* shows the results from the optimization test. $Total$ presents the number of anomalies that either passes the monotonicity test or the optimization test. The last two columns ("#" and "%") compare the results of alpha-dynamic tests with those of t tests. "#" shows the number of anomalies within a category and "%" shows the percentage of anomalies passing alpha-dynamic tests within a category.

Panel A: all anomalies					
Period	Mono	Opt	Total		
Full-sample	77	75	102		
Before-sample	28	23	39		
In-sample	68	72	93		
Post-sample	52	45	75		
Panel B: suspicious anomalies					
Period	Mono	Opt	Total	#	%
Full-sample	13	8	19	92	21%
Before-sample	14	8	17	137	12%
In-sample	16	19	28	105	27%
Post-sample	23	18	37	150	25%
Panel C: HLZ suspicious anomalies					
Period	Mono	Opt	Total	#	%
Full-sample	15	15	24	41	59%
Before-sample	7	9	14	32	44%
In-sample	7	7	12	35	34%
Post-sample	21	18	27	43	63%
Panel D: robust anomalies					
Period	Mono	Opt	Total	#	%
Full-sample	49	52	59	72	82%
Before-sample	7	6	8	16	50%
In-sample	45	46	53	65	82%
Post-sample	8	9	11	12	92%

Table 3: alpha-dynamic tests with different benchmark models

This table presents the total number of anomalies that pass alpha-dynamic tests relative to different asset pricing models. Columns 2 to 6 separately examine the alphas relative to the CAPM, Fama and French (1993, FF3) three-factor model, a four-factor model including the factors in the FF3 and momentum (FF3+MOM), Fama and French (2015, FF5) five-factor model, and Hou, Xue, and Zhang (2015, HXZ) model. Within a sample period and with a holding period of one month, I label those that fail the t test with a cutoff of 1.96 as suspicious anomalies. Among the remaining anomalies, I label those that fail the t test with a cutoff of 3.0 as HLZ suspicious anomalies, while the rest as robust anomalies. Panel A shows for results for all anomalies. Panel B shows the results for suspicious anomalies. Panel C shows the results for HLZ suspicious anomalies. Panel D shows the results for robust anomalies.

Panel A: all anomalies					
Period	CAPM	FF3	FF3+MOM	FF5	HXZ
Full-sample	102	96	85	77	65
Before-sample	39	40	32	23	14
In-sample	93	90	80	66	60
Post-sample	75	67	68	54	51
Panel B: suspicious anomalies					
Full-sample	19	23	24	23	34
Before-sample	17	14	12	11	10
In-sample	28	29	29	23	32
Post-sample	37	31	33	34	36
Panel C: HLZ suspicious anomalies					
Full-sample	24	14	22	19	17
Before-sample	14	15	11	8	3
In-sample	12	11	16	13	12
Post-sample	27	20	23	11	10
Panel D: robust anomalies					
Full-sample	59	59	39	35	14
Before-sample	8	11	9	4	1
In-sample	53	50	35	30	16
Post-sample	11	16	12	9	5

Table 4: Alpha dynamic patterns of well-known anomalies

This table presents the patterns of alphas of a few well-known anomalies in the first nine years of months after portfolio sorting, accruals, idiosyncratic volatility, momentum, net share issuance, and book-to-market. Start and End are the start and end number of months away from the portfolio sorting period. For example, the first row examines the alpha dynamic pattern within the first twelve months after the portfolio sorting period. *increasing*, *decreasing*, and *both* are statistically significant patterns under a 5% significance level. *increase* means there is at least an increasing pattern within the subperiod. *decrease* means there is at least a decreasing pattern within the subperiod. And *both* means there are both increasing and decreasing patterns within the subperiod. *none* means there is no statistically significant pattern within the subperiod. To aggregate outcomes over the five subperiods, I adjust the p values by the Benjamini-Hochberg method to correct for multiple hypothesis testing.

Start	End	Accruals	Idio Vol	Mom12m	Net Share Issuance	B/M
1	13	decrease	decrease	decrease	none	increase
13	37	decrease	both	both	decrease	none
37	61	both	both	both	both	none
61	85	both	both	both	both	none
85	109	decrease	both	both	both	none

Table 5: Anomalies that exhibit the ripple pattern in alpha dynamics.

This table presents the anomalies that exhibit a statistically significant ripple pattern in alpha dynamics under a 5% false discovery rate. I consider the anomalies that pass either the t test or alpha-dynamic tests in each sample period. In total, I examine 132, 65, 128, and 92 anomalies in the full-sample, before-sample, in-sample, and post-sample periods. The ripple pattern is a pattern that features both increasing and decreasing patterns in the whole alpha dynamic. Detailed descriptions of anomaly acronyms can be found in Appendix A.

Period	Total Number	Acronym
Full-sample	79	'AM' 'AbnormalAccruals' 'Accruals' 'AnalystRevision' 'AnnouncementReturn' 'AssetGrowth' 'CBOperProf' 'CF' 'ChAssetTurnover' 'ChEQ' 'ChInv' 'ChInvIA' 'ChNWC' 'CompEquIss' 'Coskewness' 'CustomerMomentum' 'DelCOA' 'DelEqu' 'DelFINL' 'DelNetFin' 'EntMult' 'EquityDuration' 'FEPS' 'FirmAgeMom' 'ForecastDispersion' 'GP' 'GrSaleToGrInv' 'High52' 'IO_ShortInterest' 'IdioRisk' 'IdioVol3F' 'IdioVolAHT' 'Illiquidity' 'IndMom' 'IndRetBig' 'IntMom' 'InvGrowth' 'InvestPPEInv' 'Leverage' 'MaxRet' 'Mom12m' 'Mom12mOffSeason' 'Mom6m' 'Mom6mJunk' 'MomOffSeason06YrPlus' 'MomRev' 'MomSeason06YrPlus' 'MomSeason11YrPlus' 'MomVol' 'NOA' 'NetDebtFinance' 'NetEquityFinance' 'NetPayoutYield' 'OPLeverage' 'OScore' 'OperProf' 'OperProfRD' 'OrgCap' 'PS' 'RD' 'REV6' 'RIO_Volatility' 'ResidualMomentum' 'SP' 'ShareIss1Y' 'Size' 'SmileSlope' 'VolMkt' 'VolSD' 'XFIN' 'betaVIX' 'dNoa' 'grcapx' 'grcapx3y' 'roaq' 'sfe' 'std_turn' 'zerotrade' 'zerotradeAlt1' 'zerotradeAlt12'
Before-sample	27	'ChInv' 'ChNWC' 'CompEquIss' 'DelCOA' 'FirmAgeMom' 'GrSaleToGrInv' 'High52' 'IdioRisk' 'IdioVol3F' 'IndMom' 'IndRetBig' 'InvestPPEInv' 'MaxRet' 'Mom12m' 'Mom12mOffSeason' 'Mom6m' 'MomOffSeason06YrPlus' 'MomSeason' 'MomSeasonShort' 'MomVol' 'NetEquityFinance' 'ShareIss1Y' 'VolMkt' 'VolSD' 'std_turn' 'zerotrade'
In-sample	69	'AM' 'AbnormalAccruals' 'Accruals' 'AnnouncementReturn' 'AssetGrowth' 'BMdec' 'CBOperProf' 'ChAssetTurnover' 'ChEQ' 'ChInv' 'CompositeDebtIssuance' 'DelCOA' 'DelEqu' 'DelFINL' 'DelNetFin' 'EntMult' 'EquityDuration' 'FEPS' 'FirmAgeMom' 'GP' 'GrSaleToGrInv' 'High52' 'IdioRisk' 'IdioVol3F' 'IdioVolAHT' 'IndRetBig' 'IntMom' 'InvGrowth' 'InvestPPEInv' 'MaxRet' 'Mom12m' 'Mom12mOffSeason' 'Mom6m' 'Mom6mJunk' 'MomOffSeason' 'MomRev' 'MomSeason11YrPlus' 'MomSeason16YrPlus' 'MomSeasonShort' 'MomVol' 'NOA' 'NetEquityFinance' 'NetPayoutYield' 'OPLeverage' 'OperProf' 'OperProfRD' 'OrgCap' 'ProbInformedTrading' 'REV6' 'RIO_MB' 'RIO_Volatility' 'ResidualMomentum' 'ShareIss1Y' 'ShareIss5Y' 'Size' 'VolMkt' 'VolSD' 'XFIN' 'betaVIX' 'dNoa' 'roaq' 'sfe' 'std_turn' 'tang' 'zerotrade' 'zerotradeAlt1' 'zerotradeAlt12'

Period	Total Number	Acronym
Post-sample	41	'AnalystRevision' 'AnnouncementReturn' 'ChInv' 'ChInvIA' 'CompEquIss' 'Coskewness' 'EarningsForecastDisparity' 'EquityDuration' 'FEPS' 'FirmAgeMom' 'ForecastDispersion' 'High52' 'IO_ShortInterest' 'IdioRisk' 'IdioVolAHT' 'IndRetBig' 'MaxRet' 'Mom12m' 'Mom12mOffSeason' 'Mom6m' 'MomOffSeason06YrPlus' 'MomSeason11YrPlus' 'MomVol' 'NetEquityFinance' 'NetPayoutYield' 'OperProf' 'OperProfRD' 'PS' 'RD' 'REV6' 'SP' 'Size' 'SmileSlope' 'Tax' 'VolMkt' 'VolSD' 'XFIN' 'grcapx3y' 'roaq' 'std_turn'

Table 6: Impact of alpha dynamics on after-cost profitability

The table presents the number of anomalies for which the optimization strategy outperforms the benchmark strategies and the magnitude of the improvement. I only consider anomalies that pass either the t test or alpha-dynamic tests in each sample period. In total, I examine 132, 65, 128, and 92 anomalies in the full-sample, before-sample, in-sample, and post-sample periods. In Panel A, the benchmark strategy always takes a holding period of twelve months (H12). In Panel B, the benchmark strategy always takes a holding period of one month (H1). In Panel C, the benchmark strategy determines the holding period based on the turnover rate rule of Novy-Marx and Velikov (2016) (NV). And in Panel D, the benchmark strategy takes the holding period that is used in the original paper (CV). The performance metrics are a and a_s. They are defined in Equation 15 and 16. $a > 0$ implies that the optimization strategy generates a higher after-cost alpha than a benchmark strategy. And $a_s > 0$ implies that the returns of the optimization strategy cannot be completely explained by a benchmark strategy and the market factor. a and a_s are expressed in monthly percentages.

Panel A: Benchmark is H12

	a						a_s					
Period	N	Mean	STD	25%	50%	75%	N	Mean	STD	25%	50%	75%
Full-sample	31	0.25	0.19	0.12	0.20	0.28	26	0.27	0.2	0.12	0.19	0.29
Before-sample	6	0.44	0.18	0.30	0.38	0.60	5	0.50	0.22	0.32	0.44	0.66
In-sample	30	0.36	0.26	0.19	0.26	0.47	28	0.37	0.27	0.20	0.28	0.43
Post-sample	12	0.42	0.28	0.22	0.34	0.56	10	0.45	0.28	0.27	0.32	0.72

Panel B: Benchmark is H1

	a						a_s					
Period	N	Mean	STD	25%	50%	75%	N	Mean	STD	25%	50%	75%
Full-sample	16	0.36	0.25	0.22	0.24	0.44	34	0.27	0.19	0.13	0.23	0.43
Before-sample	4	0.59	0.41	0.29	0.49	0.79	9	0.37	0.34	0.10	0.37	0.45
In-sample	19	0.39	0.22	0.24	0.37	0.50	29	0.40	0.29	0.15	0.28	0.63
Post-sample	6	0.47	0.41	0.15	0.31	0.75	8	0.34	0.11	0.27	0.33	0.41

Panel C: Benchmark is NV

	a						a_s					
Period	N	Mean	STD	25%	50%	75%	N	Mean	STD	25%	50%	75%
Full-sample	26	0.30	0.21	0.16	0.23	0.32	37	0.30	0.17	0.18	0.27	0.40
Before-sample	7	0.51	0.33	0.28	0.32	0.66	9	0.51	0.28	0.32	0.41	0.75
In-sample	27	0.34	0.20	0.18	0.30	0.46	31	0.41	0.28	0.20	0.30	0.55
Post-sample	7	0.46	0.37	0.24	0.34	0.62	10	0.32	0.12	0.26	0.31	0.39

Table 6: Impact of alpha dynamics on after-cost profitability

Panel D: Benchmark is CV												
	a						a_s					
Period	N	Mean	STD	25%	50%	75%	N	Mean	STD	25%	50%	75%
Full-sample	21	0.28	0.20	0.16	0.23	0.30	30	0.26	0.13	0.16	0.26	0.30
Before-sample	5	0.36	0.17	0.26	0.31	0.32	6	0.40	0.21	0.29	0.36	0.44
In-sample	23	0.31	0.18	0.17	0.27	0.42	29	0.34	0.22	0.19	0.28	0.45
Post-sample	8	0.44	0.39	0.12	0.29	0.71	7	0.35	0.18	0.27	0.34	0.45

A Appendix: Descriptions of anomalies

Acronym	Authors	Year	Description
AbnormalAccruals	Xie	2001	Abnormal Accruals
AbnormalAccrualsPercent	Hafzalla, Lundholm, Van Winkle	2011	Percent Abnormal Accruals
AccrualQuality	Francis, LaFond, Olsson, Schipper	2005	Accrual Quality
AccrualQualityJune	Francis, LaFond, Olsson, Schipper	2005	Accrual Quality in June
Accruals	Sloan	1996	Accruals
Activism2	Cremers and Nair	2005	Active shareholders
AdExp	Chan, Lakonishok and Sougiannis	2001	Advertising Expense
AMq	Fama and French	1992	Total assets to market (quarterly)
AnalystValue	Frankel and Lee	1998	Analyst Value
AnnouncementReturn	Chan, Jegadeesh and Lakonishok	1996	Earnings announcement return
AOP	Frankel and Lee	1998	Analyst Optimism
AssetGrowth	Cooper, Gulen and Schill	2008	Asset growth
AssetLiquidityMarket	Ortiz-Molina and Phillips	2014	Asset liquidity over market
AssetLiquidityMarketQuart	Ortiz-Molina and Phillips	2014	Asset liquidity over market (qtrly)
AssetTurnover	Soliman	2008	Asset Turnover
Beta	Fama and MacBeth	1973	CAPM beta
BetaBDLeverage	Adrian, Etula and Muir	2014	Broker-Dealer Leverage Beta
betaCC	Acharya and Pedersen	2005	Illiquidity-illiquidity beta (beta2i)
betaCR	Acharya and Pedersen	2005	Illiquidity-market return beta (beta4i)
BetaDimson	Dimson	1979	Dimson Beta
BetaFP	Frazzini and Pedersen	2014	Frazzini-Pedersen Beta
betaNet	Acharya and Pedersen	2005	Net liquidity beta (betanet,p)
betaRR	Acharya and Pedersen	2005	Return-market return illiquidity beta
BetaTailRisk	Kelly and Jiang	2014	Tail risk beta
BidAskTAQ	Hou and Loh	2016	Bid-ask spread (TAQ)
BM	Rosenberg, Reid, and Lanstein	1985	Book to market using most recent ME
BMdec	Fama and French	1992	Book to market using December ME
BMq	Rosenberg, Reid, and Lanstein	1985	Book to market (quarterly)
BookLeverageQuarterly	Fama and French	1992	Book leverage (quarterly)
BrandCapital	Belo, Lin and Vitorino	2014	Brand capital to assets
CapTurnover	Haugen and Baker	1996	Capital turnover
CapTurnover_q	Haugen and Baker	1996	Capital turnover (quarterly)
Cash	Palazzo	2012	Cash to assets
cashdebt	Ou and Penman	1989	CF to debt
CBOperProf	Ball et al.	2016	Cash-based operating profitability
CBOperProfLagAT_q	Ball et al.	2016	Cash-based oper prof lagged assets qtrly
CF	Lakonishok, Shleifer, Vishny	1994	Cash flow to market
CFq	Lakonishok, Shleifer, Vishny	1994	Cash flow to market quarterly
ChangeInRecommendation	Jegadeesh et al.	2004	Change in recommendation

ChAssetTurnover	Soliman	2008	Change in Asset Turnover
ChEQ	Lockwood and Prombutr	2010	Growth in book equity
ChInv	Thomas and Zhang	2002	Inventory Growth
ChInvIA	Abarbanell and Bushee	1998	Change in capital inv (ind adj)
ChNAnalyst	Scherbina	2008	Decline in Analyst Coverage
ChNCOA	Soliman	2008	Change in Noncurrent Operating Assets
ChNCOL	Soliman	2008	Change in Noncurrent Operating Liab
ChNNCOA	Soliman	2008	Change in Net Noncurrent Op Assets
ChPM	Soliman	2008	Change in Profit Margin
ConsNegRet	Watkins	2003	Consistently negative return
ConsRecomm	Barber et al.	2002	Consensus Recommendation
ConvDebt	Valta	2016	Convertible debt indicator
Coskewness	Harvey and Siddique	2000	Coskewness
CredRatDG	Dichev and Piotroski	2001	Credit Rating Downgrade
currat	Ou and Penman	1989	Current Ratio
CustomerMomentum	Cohen and Frazzini	2008	Customer momentum
DelayAcct	Callen, Khan and Lu	2013	Accounting component of price delay
DelayNonAcct	Callen, Khan and Lu	2013	Non-accounting component of price delay
DelBreadth	Chen, Hong and Stein	2002	Breadth of ownership
DelSTI	Richardson et al.	2005	Change in short-term investment
depr	Holthausen and Larcker	1992	Depreciation to PPE
DivInit	Michaely, Thaler and Womack	1995	Dividend Initiation
DivOmit	Michaely, Thaler and Womack	1995	Dividend Omission
DivSeason	Hartzmark and Salomon	2013	Dividend seasonality
DivYield	Naranjo, Nimalendran, Ryngaert	1998	Dividend yield for small stocks
DivYieldAnn	Naranjo, Nimalendran, Ryngaert	1998	Last year's dividends over price
DivYieldST	Litzenberger and Ramaswamy	1979	Predicted div yield next month
dNoa	Hirshleifer, Hou, Teoh, Zhang	2004	change in net operating assets
DolVol	Brennan, Chordia, Subra	1998	Past trading volume
EarningsConsistency	Alwathainani	2009	Earnings consistency
EarningsForecastDisparity	Da and Warachka	2011	Long-vs-short EPS forecasts
EarningsPredictability	Francis, LaFond, Olsson, Schipper	2004	Earnings Predictability
EarningsSmoothness	Francis, LaFond, Olsson, Schipper	2004	Earnings Smoothness
EarningsSurprise	Foster, Olsen and Shevlin	1984	Earnings Surprise
EarningsValueRelevance	Francis, LaFond, Olsson, Schipper	2004	Value relevance of earnings
EarnSupBig	Hou	2007	Earnings surprise of big firms
EBM	Penman, Richardson and Tuna	2007	Enterprise component of BM
EBM_q	Penman, Richardson and Tuna	2007	Enterprise component of BM
EP	Basu	1977	Earnings-to-Price Ratio
EPq	Basu	1977	Earnings-to-Price Ratio
ExchSwitch	Dharan and Ikenberry	1995	Exchange Switch
FEPS	Cen, Wei, and Zhang	2006	Analyst earnings per share
fgr5yrLag	La Porta	1996	Long-term EPS forecast

FirmAge	Barry and Brown	1984	Firm age based on CRSP
FR	Franzoni and Marin	2006	Pension Funding Status
FRbook	Franzoni and Marin	2006	Pension Funding Status
Frontier	Nguyen and Swanson	2009	Efficient frontier index
Governance	Gompers, Ishii and Metrick	2003	Governance Index
GP	Novy-Marx	2013	gross profits / total assets
GPlag	Novy-Marx	2013	gross profits / total assets
GrAdExp	Lou	2014	Growth in advertising expenses
GrSaleToGrOverhead	Abarbanell and Bushee	1998	Sales growth over overhead growth
GrSaleToGrReceivables	Abarbanell and Bushee	1998	Change in sales vs change in receiv
Herf	Hou and Robinson	2006	Industry concentration (sales)
HerfAsset	Hou and Robinson	2006	Industry concentration (assets)
HerfBE	Hou and Robinson	2006	Industry concentration (equity)
High52	George and Hwang	2004	52 week high
IdioVol3F	Ang et al.	2006	Idiosyncratic risk (3 factor)
IdioVolAHT	Ali, Hwang, and Trombley	2003	Idiosyncratic risk (AHT)
Illiquidity	Amihud	2002	Amihud's illiquidity
IndIPO	Ritter	1991	Initial Public Offerings
IndMom	Grinblatt and Moskowitz	1999	Industry Momentum
IndRetBig	Hou	2007	Industry return of big firms
IntanCFP	Daniel and Titman	2006	Intangible return using CFtoP
IntanEP	Daniel and Titman	2006	Intangible return using EP
IntanSP	Daniel and Titman	2006	Intangible return using Sale2P
IntrinsicValue	Frankel and Lee	1998	Intrinsic or historical value
invest	Chen and Zhang	2010	Capex and Inventory Change
Investment	Titman, Wei and Xie	2004	Investment to revenue
IO_ShortInterest	Asquith Pathak and Ritter	2005	Inst own among high short interest
KZ	Lamont, Polk and Saa-Requejo	2001	Kaplan Zingales index
LaborforceEfficiency	Abarbanell and Bushee	1998	Laborforce efficiency
Leverage_q	Bhandari	1988	Market leverage quarterly
MaxRet	Bali, Cakici, and Whitelaw	2010	Maximum return over month
MeanRankRevGrowth	Lakonishok, Shleifer, Vishny	1994	Revenue Growth Rank
Mom12m	Jegadeesh and Titman	1993	Momentum (12 month)
Mom12mOffSeason	Heston and Sadka	2008	Momentum without the seasonal part
Mom6mJunk	Avramov et al	2007	Junk Stock Momentum
MomOffSeason	Heston and Sadka	2008	Off season long-term reversal
MomOffSeason06YrPlus	Heston and Sadka	2008	Off season reversal years 6 to 10
MomOffSeason11YrPlus	Heston and Sadka	2008	Off season reversal years 11 to 15
MomOffSeason16YrPlus	Heston and Sadka	2008	Off season reversal years 16 to 20
MomRev	Chan and Ko	2006	Momentum and LT Reversal
MomSeason	Heston and Sadka	2008	Return seasonality years 2 to 5
MomSeasonShort	Heston and Sadka	2008	Return seasonality last year
MomVol	Lee and Swaminathan	2000	Momentum in high volume stocks

MRreversal	De Bondt and Thaler	1985	Medium-run reversal
nanalyst	Elgers, Lo and Pfeiffer	2001	Number of analysts
NetDebtFinance	Bradshaw, Richardson, Sloan	2006	Net debt financing
NetDebtPrice	Penman, Richardson and Tuna	2007	Net debt to price
NetDebtPrice_q	Penman, Richardson and Tuna	2007	Net debt to price
NetEquityFinance	Bradshaw, Richardson, Sloan	2006	Net equity financing
NetPayoutYield	Boudoukh et al.	2007	Net Payout Yield
NetPayoutYield_q	Boudoukh et al.	2007	Net Payout Yield quarterly
NOA	Hirshleifer et al.	2004	Net Operating Assets
NumEarnIncrease	Loh and Warachka	2012	Earnings streak length
OperProf	Fama and French	2006	operating profits / book equity
OperProfLag	Fama and French	2006	operating profits / book equity
OperProfRDLagAT	Ball et al.	2016	Oper prof R&D adj lagged assets
OPLeverage	Novy-Marx	2010	Operating leverage
OptionVolume1	Johnson and So	2012	Option to stock volume
OrderBacklog	Rajgopal, Shevlin, Venkatachalam	2003	Order backlog
OrgCap	Eisfeldt and Papanikolaou	2013	Organizational capital
OScore	Dichev	1998	O Score
PatentsRD	Hirschleifer, Hsu and Li	2013	Patents to RD expenses
pchcurrat	Ou and Penman	1989	Change in Current Ratio
pchdepr	Holthausen and Larcker	1992	Change in depreciation to PPE
pchgm_pchsale	Abarbanell and Bushee	1998	Change in gross margin vs sales
pchquick	Ou and Penman	1989	Change in quick ratio
PM	Soliman	2008	Profit Margin
PM_q	Soliman	2008	Profit Margin
Price	Blume and Husic	1972	Price
PriceDelayRsq	Hou and Moskowitz	2005	Price delay r square
PriceDelaySlope	Hou and Moskowitz	2005	Price delay coeff
PriceDelayTstat	Hou and Moskowitz	2005	Price delay SE adjusted
ProbInformedTrading	Easley, Hvidkjaer and O'Hara	2002	Probability of Informed Trading
PS	Piotroski	2000	Piotroski F-score
PS_q	Piotroski	2000	Piotroski F-score
quick	Ou and Penman	1989	Quick ratio
RD_q	Chan, Lakonishok and Sougiannis	2001	R&D over market cap quarterly
rd_sale	Chan, Lakonishok and Sougiannis	2001	R&D to sales
RDAbility	Cohen, Diether and Malloy	2013	R&D ability
RDcap	Li	2011	R&D capital-to-assets
RDIPO	Gou, Lev and Shi	2006	IPO and no R&D spending
realestate	Tuzel	2010	Real estate holdings
ResidualMomentum	Blitz, Huij and Martens	2011	Momentum based on FF3 residuals
ResidualMomentum6m	Blitz, Huij and Martens	2011	6 month residual momentum
retConglomerate	Cohen and Lou	2012	Conglomerate return
RetNOA	Soliman	2008	Return on Net Operating Assets

RetNOA_q	Soliman	2008	Return on Net Operating Assets
ReturnSkew	Bali, Engle and Murray	2015	Return skewness
ReturnSkew3F	Bali, Engle and Murray	2015	Idiosyncratic skewness (3F model)
ReturnSkewQF	Bali, Engle and Murray	2015	Idiosyncratic skewness (Q model)
REV6	Chan, Jegadeesh and Lakonishok	1996	Earnings forecast revisions
RevenueSurprise	Jegadeesh and Livnat	2006	Revenue Surprise
RIO_Turnover	Nagel	2005	Inst Own and Turnover
RIO_Volatility	Nagel	2005	Inst Own and Idio Vol
RoE	Haugen and Baker	1996	net income / book equity
roic	Brown and Rowe	2007	Return on invested capital
salerec	Ou and Penman	1989	Sales to receivables
secured	Valta	2016	Secured debt
securedind	Valta	2016	Secured debt indicator
sfe	Elgers, Lo and Pfeiffer	2001	Earnings Forecast to price
ShareIss1Y	Pontiff and Woodgate	2008	Share issuance (1 year)
ShareIss5Y	Daniel and Titman	2006	Share issuance (5 year)
ShareRepurchase	Ikenberry, Lakonishok, Vermaelen	1995	Share repurchases
ShortInterest	Dechow et al.	2001	Short Interest
sinAlgo	Hong and Kacperczyk	2009	Sin Stock (selection criteria)
sinOrig	Hong and Kacperczyk	2009	Sin Stock (original list)
Size	Banz	1981	Size
skew1	Xing, Zhang and Zhao	2010	Volatility smirk near the money
Spinoff	Cusatis, Miles and Woolridge	1993	Spinoffs
std_turn	Chordia, Subra, Anshuman	2001	Share turnover volatility
STreversal	Jegadeesh	1989	Short term reversal
tang	Hahn and Lee	2009	Tangibility
tang_q	Hahn and Lee	2009	Tangibility quarterly
Tax_q	Lev and Nissim	2004	Taxable income to income (qtrly)
UpRecomm	Barber et al.	2002	Up Forecast
VarCF	Haugen and Baker	1996	Cash-flow to price variance
VolMkt	Haugen and Baker	1996	Volume to market equity
VolSD	Chordia, Subra, Anshuman	2001	Volume Variance
WW	Whited and Wu	2006	Whited-Wu index
WW_Q	Whited and Wu	2006	Whited-Wu index
zerotrade	Liu	2006	Days with zero trades
zerotradeAlt1	Liu	2006	Days with zero trades
ZScore_q	Dichev	1998	Altman Z-Score quarterly

B Appendix: Sorting Dates and Shock Dates

The behavioral models described in Section 2.1.3 all assume a shock date when mispricing initially arises and study how alphas or prices evolve after the shock date. In contrast, I study how alphas evolve after the sorting dates, which correspond to the dates when firms are sorted based on a specific characteristic. In this section, I discuss how sorting dates are related to shock dates.

First, while the exact relationship between shock dates and sorting dates may not be directly observed, these models can still imply a non-constant alpha dynamic after sorting dates. It is because, in those models, the alpha dynamic is non-constant over any subset of periods before mispricing is completely resolved as shown in Figure 2. If a characteristic is associated with remaining mispricing at sorting dates ($j = 0$), then the alpha dynamic after sorting α_j can be non-constant according to these models.

Further, we may also be able to predict the pattern of the alpha dynamic based on the models. That is, where does $j = 0$ fall in these models? First, in the models that explain anomalies endogenously like Model 1 (Barberis, 2018), there is a clear linkage between shock dates and sorting dates. Further, in exogenous models that do not explain anomalies, exogenous investment opportunities (mispricing) appear on shock dates. For example, Model 2 (Duffie, 2010) and Model 3 (Hendershott et al., 2022) describe how prices evolve when some rational traders do not trade on investment opportunities immediately. Since mispricing is exogenous, we can interpret trading on the anomaly characteristics as investment opportunities and shock dates as sorting dates. That is, $j = 0$ should be the shock date. For instance, if some traders do not adjust their characteristic-sorted portfolios based on firm size immediately at the end of each month, alpha dynamics after sorting dates may display a non-constant pattern based on Models 2 and 3.

C Appendix: Proofs

Proof of Proposition 1. $\alpha^{op}_{k,h} = \frac{1}{h}\sum_{j=k+1}^{k+h} \alpha_j \leq \frac{1}{h} h\alpha_{j^*} = \alpha^{op}_{j^*-1,1}$. ■

Proof of Proposition 2. Since, $foc = \frac{\delta_0(\lambda h+1)e^{-\lambda h}+c-\delta_0}{h^2}$, $foc(h=1) = \delta 0(\lambda+1)e^{-\lambda}+c-\delta_0$. Then $foc(h=1) \leq 0 \Leftrightarrow \frac{\lambda+1}{e^\lambda} \leq 1-\frac{c}{\delta_0}$.

As $\delta_0(\lambda h+1)e^{-\lambda h}$ decreases in h, $foc < 0$ when $h > 1$ if $\frac{\lambda+1}{e^\lambda} \leq 1-\frac{c}{\delta_0}$. Therefore, after-cost alpha $\alpha^{ac}_{k,h}$ strictly decreases in h and $h^* = 1$ when $\frac{\lambda+1}{e^\lambda} \leq 1-\frac{c}{\delta_0}$. ■

Proof of Proposition 3. Take $foc = 0$:

$$\frac{\lambda h^* + 1}{e^{\lambda h^*}} = 1 - \frac{c}{\delta_0} \tag{17}$$

Since $\delta_0(\lambda h+1)e^{-\lambda h}$ decreases in h, $foc > 0$ when $h < h^*$ and $foc < 0$ when $h > h^*$. Therefore, h^* maximizes $\alpha^{ac}_{k,h}$ and solves the problem. ■

D Appendix: Performance of alpha-mean tests in detecting drift-reversal patterns

In this section, I examine the performance of tests on mean alphas to detect the pattern of a price drift followed by a reversal. Consistent with the alpha-dynamic test, I examine five subperiods after the sorting date: [1,13], [13,37], [37,61], [61,85], and [85,109]. Additionally, I only consider anomalies that pass either the t test or alpha-dynamic tests in each sample period. In total, I examine 132, 65, 128, and 92 anomalies in the full-sample, before-sample, in-sample, and post-sample periods.

To estimate the average of the alphas for all the months within a subperiod, I measure the returns of a strategy. To estimate the average of the alphas for subperiod [1,13], I form a portfolio at the end of each month t by equally weighting the portfolios based on the numbers of months skipped after sorting $k = 0, 1, 2, ..., 12$. I calculate the return of this portfolio in month $t + 1$. This process is repeated until the end of the sample. I then estimate the CAPM alpha of this strategy and its p value. Finally, I repeat this test over each of the other subperiods to estimate the mean CAPM alphas and their p values for each subperiod.

Since I examine five subperiods, there are five hypotheses in total. To adjust for multiple hypothesis testing, I use the Benjamini-Hochberg method to ensure that the expected proportion of false discovery rate is no greater than 5%. Under a 5% significance level, if there are both positive and negative mean alphas within the nine years after sorting, it indicates that the drift-reversal pattern exists.

Results are shown in Table D1. In the full-sample, before-sample, in-sample, and post-sample periods, the numbers of anomalies that exhibit statistically significant drift-reversal patterns are 2, 1, 7, and 0, separately. In comparison, those figures are 79, 27, 69, and 41 for alpha-dynamic tests as shown in Table 5. Therefore, results indicate that alpha-dynamic tests have much higher power than alpha-mean tests in detecting drift-reversal patterns.

Table D1: Anomalies that exhibit the drift-reversal pattern under alpha-mean tests

Period	Total Number	Acronym
Full-sample	6	'FirmAgeMom' 'Mom12mOffSeason' 'Mom6mJunk' 'CustomerMomentum' 'IndRetBig' 'MomVol'
Before-sample	1	'FirmAgeMom'
In-sample	7	'FirmAgeMom' 'Mom6m' 'Mom12mOffSeason' 'Mom6mJunk' 'IndRetBig' 'MomVol' 'Mom12m'
Post-sample	0	

E Appendix: Simulated performance under random variation

In Section 4.3, I show that the benchmark strategy that always takes the number of months skipped $k = 0$ and holding period $h = 1$ (H1) is outperformed by an optimization strategy for many anomalies when after-cost alphas are compared.

One possible concern is whether the improvement in after-cost alphas can be obtained with random variation. This section is to investigate this concern with simulations. The null hypothesis is that H1 is optimal and no strategy can outperform it based on after-cost alphas.

The null hypothesis is examined as follows. I conduct 2,000 simulations. In each simulation, I first simulate a time series of after-cost market returns calibrated to the sample mean and standard deviation of after-cost market returns over the sample period. The sample period is between January 1936 and December 2021.

For each characteristic, I estimate a_d, b_d, and the residual volatility by regressing the actual returns of H1 on the actual after-cost market excess returns (not the simulated returns):

$$r_{b,t} = a_d + b_d r_{m,t} + \epsilon_{b,t} \tag{18}$$

I then generate a simulated time series of returns of the benchmark strategy $r_{b,t}$ for each characteristic with simulated market returns. The date range of the simulated returns for a characteristic match the actual returns of the characteristic-sorted portfolios.

Next, within each characteristic, I estimate a_s, b_s, b_m and the residual volatility from the following regression:

$$r_{s,t} = a_s + b_s r_{b,t} + b_m r_{m,t} + \epsilon_{s,t} \tag{19}$$

$r_{s,t}$ are after-cost returns of a strategy that restricts $k = 0$ and takes different $h = 1, 3, 6, 9, 12$ throughout the sample. Returns of each $r_{s,t}$ and $r_{b,t}$ have the same length.

After estimating $\hat{b}_s$, $\hat{b}_m$, and the volatility of $\hat{\epsilon}_{s,t}$ with the actual data, I create simulated time series of returns for the strategies based on $\hat{b}_s$, $\hat{b}_m$, estimated residual volatility, and simulated returns of $r_{b,t}$ and $r_{m,t}$:

$$r_{s,t} = \hat{b}_s r_{b,t} + \hat{b}_m r_{m,t} + \hat{\epsilon}_{s,t} \tag{20}$$

That is, I demean the intercept, a_s. This is to generate a data-generating process that neither strategy outperforms H1. In the meantime, the correlation structure as well as other moments are preserved. Let us denote this data-generating process as DGP_{as}.

In Section 4.3.3, I run another regression that examines whether a strategy generates a higher alpha than a benchmark strategy:

$$r_{s,t} - r_{b,t} = a + b_m r_{m,t} + \epsilon_{s,t} \tag{21}$$

If I use DGP_{as} to generate simulated returns of the other strategies, a problem is that a in Eq 21 may not be zero. As I have demeaned the intercept, a_s, $r_{s,t} - r_{b,t} = a = (\hat{b}_s - 1)r_{b,t}$. That is, a will not be zero with DGP_{as} if $\hat{b}_s - 1 \neq 0$ and the mean of $r_{b,t}$ is not zero. To address this issue, I also simulate returns of the other strategies based on the following DGP, which I denote as DGP_a:

$$r_{s,t} - r_{b,t} = \hat{b}_m r_{m,t} + \hat{\epsilon}_{s,t} \tag{22}$$

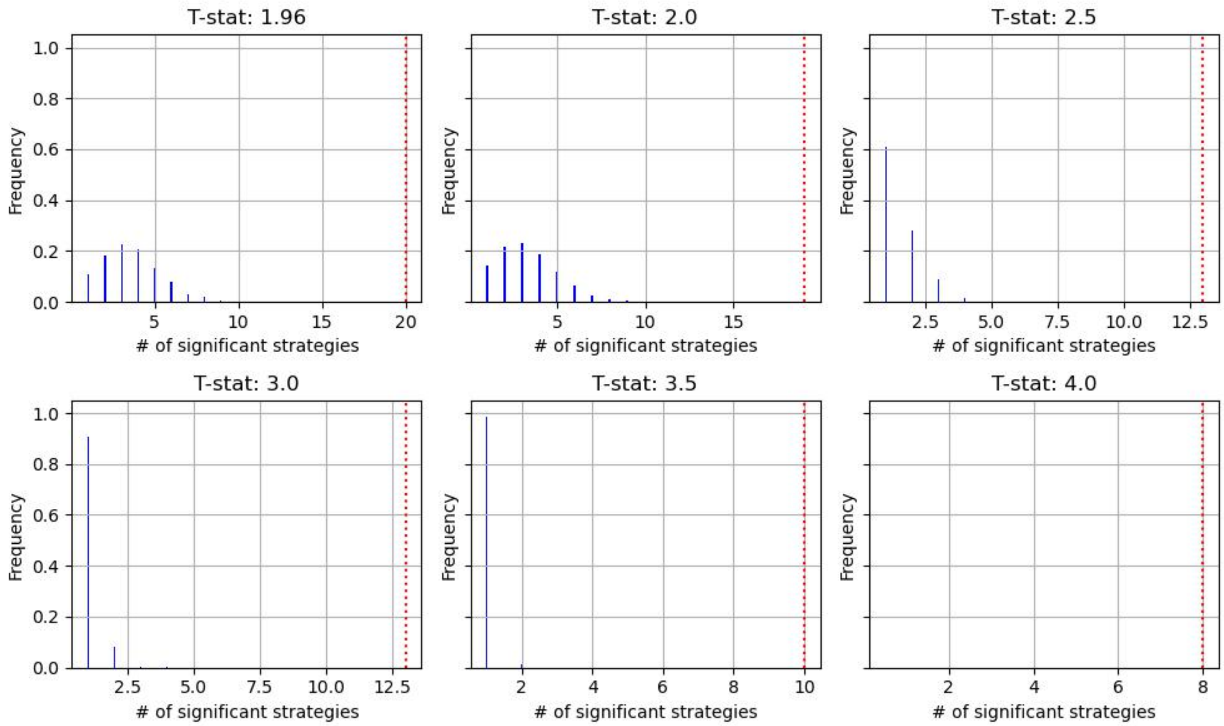
Here, $\hat{b}_m$ and $\hat{\epsilon}_{s,t}$ is generated based on estimated b_m and the volatility of $\epsilon_{s,t}$ in Equation 21 with the actual data.

Then I use the same procedure as described in Section 4.3 to obtain the return series of the optimization strategy with simulated returns.

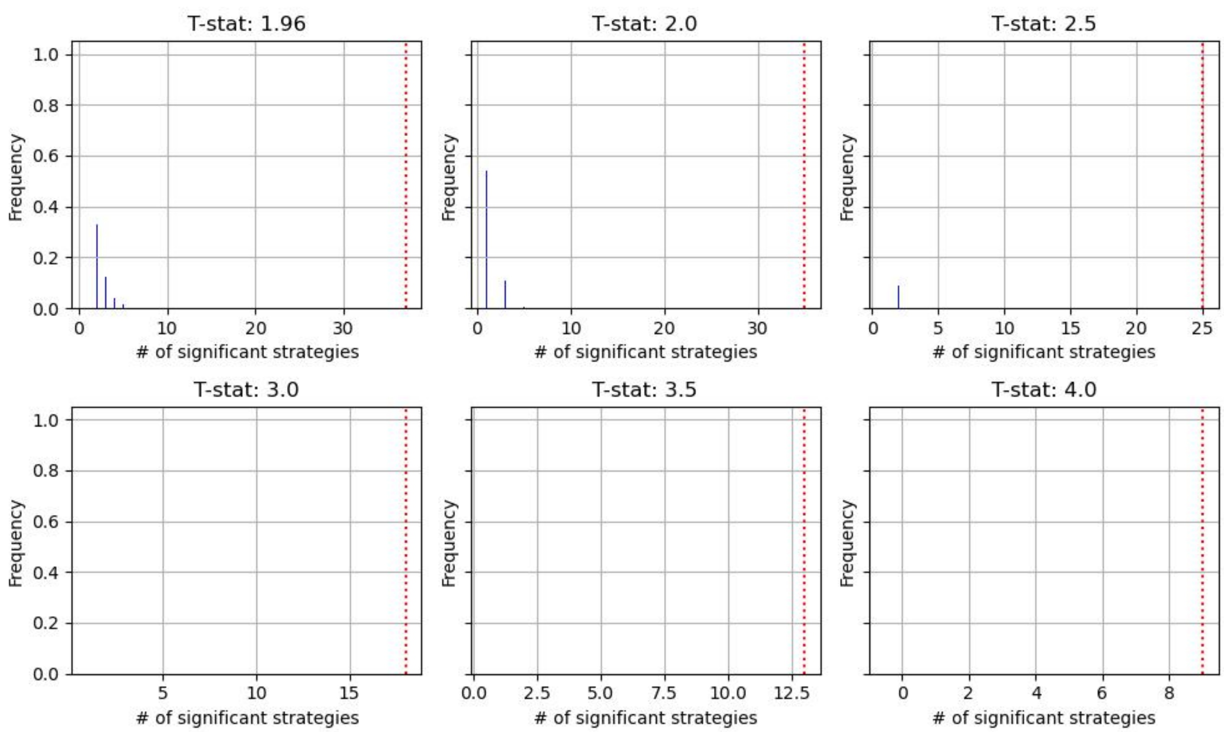
When the optimization strategy uses the simulated returns from DGP_{as}, I run regression Equation 19 to examine whether the optimization strategy outperforms H1. And when the optimization strategy uses the simulated returns from DGP_a, I run regression Equation 21 to examine whether the optimization strategy outperforms H1.

A strategy outperforms H1 if a_s or a is statistically significant and positive when the cutoff of t-statistics ranges from 1.96 to 4.00. Then I obtain the counts of characteristics for which the optimization strategy outperforms the benchmark strategy from the regressions under different t-statistic cutoffs. The simulation is repeated 2,000 times so that I obtain a distribution of the counts under different cutoffs.

Results are shown in Figure E1. The red vertical line is the number of characteristics for which the optimization strategy outperforms H1 with the actual data. In subfigure (a), a is the performance metric. And in subfigure (b), a_s is the performance metric. Under t-statistic cutoffs of 1.96, 2.00, 2.50, 3.00, 3.50, and 4.00, the number of characteristics for which the optimization strategy outperforms H1 is always beyond the maximum both when a_s is considered and when a is considered. Therefore, the null hypothesis that H1 is optimal can be rejected.



(a) Peformance metric is a



(b) Peformance metric is a_s

Figure E1: Simulated distributions for the number of strategies that outperform the benchmark strategy

F Appendix: Additional empirical evidence

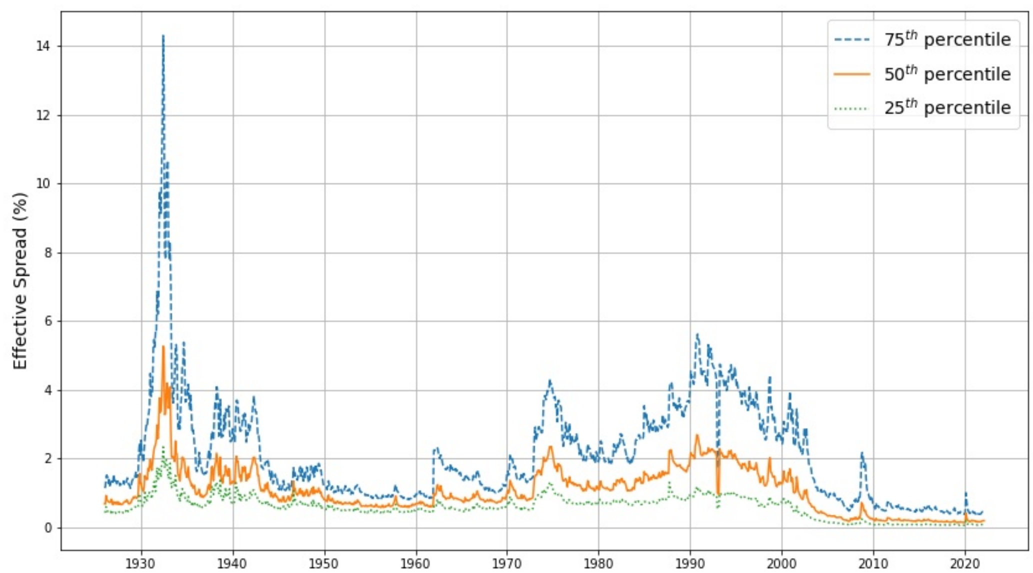


Figure F1: Monthly stock effective spreads over time

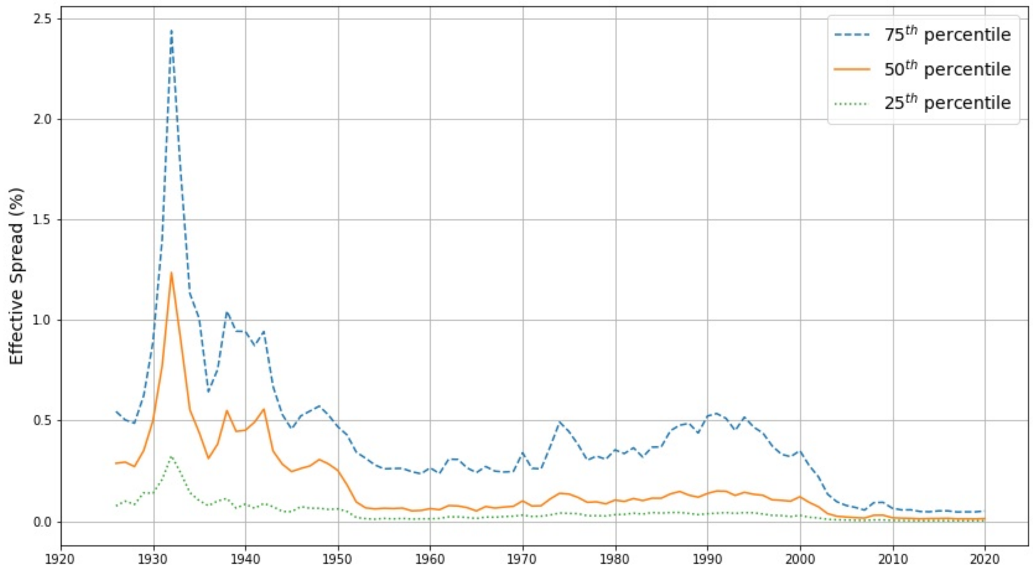


Figure F2: Monthly portfolio rebalancing costs over time

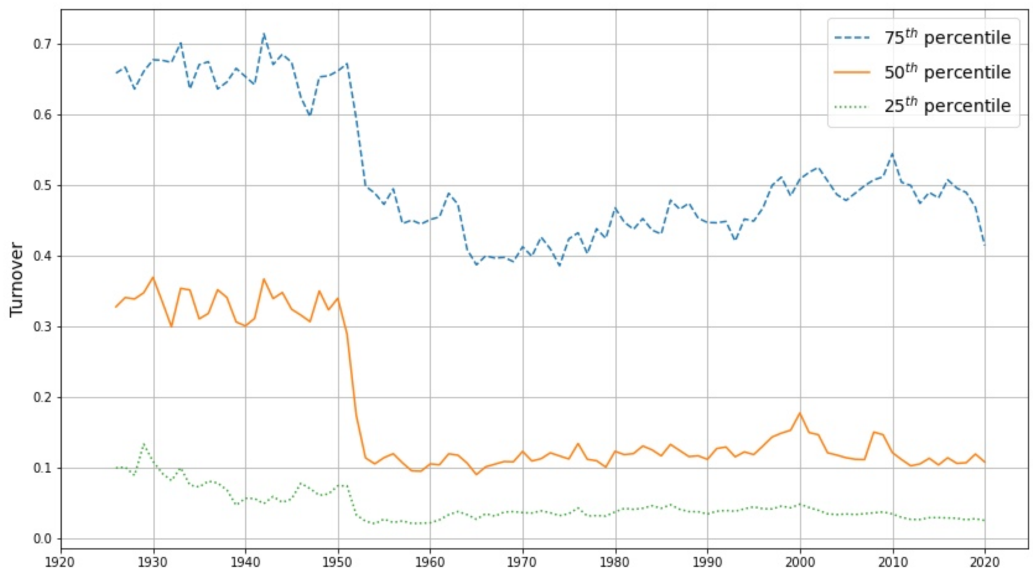


Figure F3: Monthly portfolio turnover rate over time

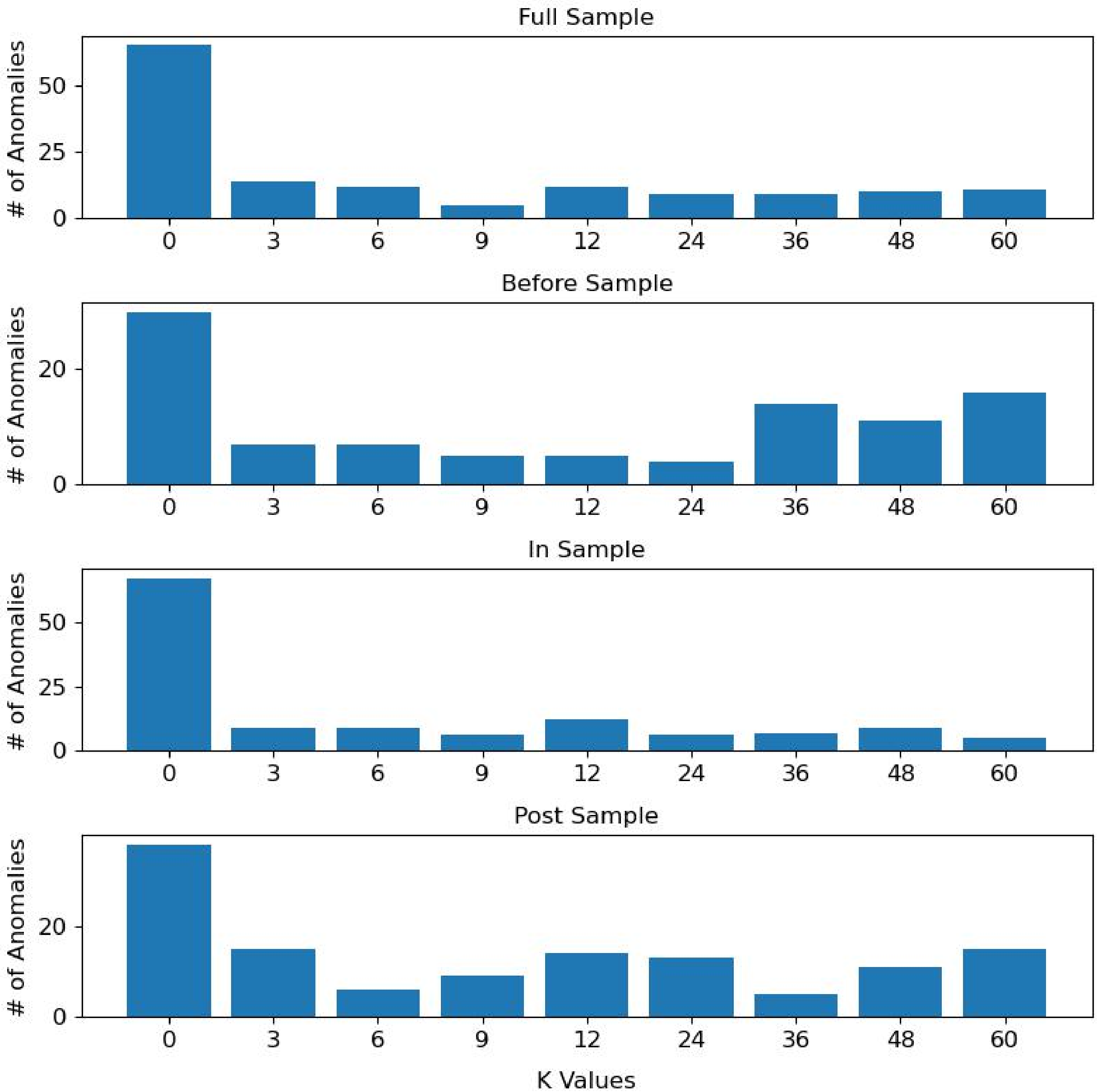


Figure F4: Number of anomalies with the highest alpha at different k values

The figure plots the number of anomalies with the highest CAPM alpha at different k values under an unconditional analysis. I study 205 published anomalies. k is the number of skipped months after a portfolio sort. The holding period is one month. The y axis is the number of anomalies, while the x axis are the k values. From top to bottom, I plot the distributions in the full-sample, before-sample, in-sample, and post-sample periods.

Table F1: Does k have an impact on after-cost profitability?

This table presents the number of anomalies for which the optimization strategy outperforms the default strategies. I compare results for two optimization strategies. The first searches for h only and the second searches for both k and h. The benchmark strategies take either $k = 0$ and $h = 12$ (H12) or $k = 0$ and $h = 1$ (H1) based on after-cost alphas. k is the number of skipped months and h is the holding period. I only consider anomalies that pass either the t test or alpha-dynamic tests in each sample period. In total, I examine 132, 65, 128, and 92 anomalies in the full-sample, before-sample, in-sample, and post-sample periods. Panel A shows results when k is restricted to zero and only optimal h is searched for. Panel B shows results when both optimal k and h are searched for. The performance metrics are a and a_s as in Equation 15 and 16. a examines whether the optimization strategy generates a higher after-cost alpha than a default strategy. And a_s examines whether the returns of the optimization strategy can be completely explained by a default strategy and the market factor.

Panel A: searching for h only				
	Default Strategy			
	H12		H1	
Period	a	a_s	a	a_s
Full-sample	30	26	16	33
Before-sample	6	5	4	8
In-sample	30	28	19	29
Post-sample	12	10	6	8
Panel B: searching for both k and h				
	Default Strategy			
	H12		H1	
Period	a	a_s	a	a_s
Full-sample	24	24	15	29
Before-sample	3	5	4	9
In-sample	27	30	19	24
Post-sample	20	14	7	9

Table F2: Optimization on the information ratio

This table compares the results between the optimization strategy and the benchmark strategy that takes $k = 0$ and $h = 12$ (H12) based on the information ratio (IR). k is the number of skipped months and h is the holding period. I consider *Size* and *Book-to-market* factors. BM uses the latest market equity to construct the book-to-market. And $BMdec$ uses the market equity in December of the prior year to construct book-to-market. k is searched from $\{0, 3, 6, 9, 12\}$ and h is searched from $\{1, 3, 6, 9, 12\}$ to maximize IR over time. k^* searches for k only and restricts $h = 1$. h^* searches for h only and restricts $k = 0$. And k^*h^* searches for both k and h. T is the sample length.

Period	T	k^*	h^*	k^*h^*	H12
Panel A: Size					
Full-sample	1070	0.118	0.118	0.118	0.042
In-sample	514	0.075	0.075	0.075	0.038
Post-sample	556	0.101	0.101	0.101	0.027
Panel B: BM					
Full-sample	618	0.077	0.028	0.075	0.027
Before-sample	53	0.002	-0.009	0.003	-0.007
In-sample	132	0.112	0.086	0.112	0.086
Post-sample	433	0.025	-0.018	0.023	-0.017
Panel C: BMdec					
Full-sample	739	0.069	0.064	0.064	0.065
Before-sample	48	0.045	0.039	0.032	0.052
In-sample	330	0.103	0.091	0.099	0.093
Post-sample	361	-0.019	-0.016	-0.019	-0.019