

The Intersection of Expected Returns

Austin Sobotka*

June 20, 2025

Abstract

A relatively small number of stocks plays a disproportionately large role in explaining the performance of 164 cross-sectional asset pricing anomalies. For instance, excluding the top 10% of stocks that are shared across the most anomaly portfolios for a given month reduces the average anomaly's return and alpha by approximately 40%. These stocks can be identified ex ante and used to form long-short portfolios that generate abnormal returns more than three times larger than that of the average anomaly portfolio. Consistent with prior research, I find evidence that biased investor expectations help explain the returns to these stocks, suggesting that a significant portion of the returns to the 164 anomalies can be attributed to mispricing. My results have implications for traditional asset pricing, behavioral finance, and for investors and practitioners in the factor investing space.

I thank Richard Sias, Scott Cederburg, Andrea Rossi, Chris Lamoureux, and seminar participants from the 2025 Eastern Finance Association annual meeting for their thoughtful comments and feedback.

*Eller College of Management, University of Arizona, McClelland Hall, Room 315R, P.O. Box 210108, Tucson, AZ 85721-0108, 205asobotka@arizona.edu

1 Introduction

In response to the exploding number of proposed cross-sectional return predictors, modern asset pricing research has shifted its focus away from identifying new anomalies and towards taming the so-called “factor zoo” (Cochrane (2011)). This shift in agenda has been driven by two main, not necessarily incompatible, intuitions. First, it could be the case that the large collection of proposed return predictors capture systematic risk factors. In this case, economic intuition suggests that they likely do not proxy for entirely distinct risk factors, but rather that they are each noisy versions of a relatively small collection of latent factors – that is, it seems implausible that the true return generating process is driven by hundreds of independent components. Alternatively, it could be the case that the collection of proposed return predictors capture mispricing. In this case, economic intuition once again suggests that they should be related in some way or another – if markets are at least reasonably efficient, then although the existence of a few anomalous return sources may be unsurprising, the proliferation of hundreds of distinct sources seems unlikely.

Building on the intuition that the large collection of cross-sectional return predictors are likely not entirely distinct from one another, I examine an underexplored source of potential commonality among the anomalies that comprise the factor zoo: the stocks that constitute them. Specifically, I examine stock overlap among a set of 164 anomaly portfolios. My tests yield four primary insights, each of which support the notion that a broad set of anomalies are more closely related than previously thought.

First, a small set of stocks appear simultaneously in the long (short) legs of anomaly portfolios in any given month. For example, in the final month of my sample, Dillard’s (ticker: DDS) is included in 36 of the 158 anomaly long legs with valid return observations for the month while Toughbuilt Industries (ticker: TBLT) is included in 55 of the 158 anomaly short legs. More generally, out of the top 10% of stocks appearing in the most anomaly portfolios in a given month, the average anomaly portfolio includes 60 of them. Moreover, every one of the 164 anomalies shares a meaningful amount of stocks with each of the other anomalies at some point in the time series.

Second, these “overlap” stocks contribute disproportionately to the performance of anomaly portfolios. For example, although the 10% of stocks that are included in the most anomaly portfolios in any given month account for only 17% of the stocks in the average anomaly portfolio, they account for approximately 40% of the average anomaly portfolio’s returns and alphas. That is, the overlap stocks earn alphas that are more than three times larger, on average, than the non-overlap stocks. The disproportionate contribution of overlap stocks to anomaly performance is relatively uniform across all of the 164 anomalies.

Third, a long-short portfolio constructed using the overlap stocks significantly outperforms the average anomaly portfolio. Over the period 1926-2022, the value-weighted portfolio that longs the 10% of stocks in

the most anomaly long legs and shorts the 10% of stocks in the most anomaly short legs earned an annualized CAPM alpha of 12.6%. In contrast, over the same period, the average value-weighted anomaly earned an annualized alpha of 3.82% while momentum (in the form of the Fama-French UMD factor), one of the most persistent and best performing anomalies, earned an annualized alpha of 10.2%.

Fourth, the superior performance of the overlap stocks appears to arise at least partially from biased expectations and mispricing: the average stock in the most anomaly long legs has an analyst return forecast error of -8.77% (indicating that analysts expect long leg stocks to perform much worse than they actually do), while the average stock in the most anomaly short legs has a return forecast error of +34.71% (indicating that analysts expect short leg stocks to perform much better than they actually do). That is, contrary to documented empirical patterns, analysts expect the stocks in the most anomaly short legs to significantly outperform the stocks in the most anomaly long legs. Consistent with the notion that analyst forecasts impact asset prices (or that they reflect the expectations of the marginal investor (e.g., Anarkulova et al. (2025))), the stocks in the most anomaly long legs experience significantly higher returns on earnings announcement days relative to the stocks in the most anomaly short legs (0.72% vs -0.10%). The asymmetric importance of the overlap stocks for anomaly returns paired with evidence that they are driven by biased expectations suggests that a large portion of the majority of all anomaly returns can be attributed to mispricing.¹

My results have implications for traditional asset pricing, behavioral finance, and for practitioners and investors in the factor investing space. First, recent evidence suggests that anomalies are not as independent as previously thought: the return space appears to be relatively low-dimensional, consistent with the hypothesis that a few latent factors describe returns rather than hundreds of unique ones. The evidence to this effect is largely statistically based, relying most commonly on some variation of principal component analysis (e.g., Kozak et al. (2018), Kelly et al. (2019), Lettau and Pelger (2020), Clarke (2022)). My results relate to this strand of literature by identifying a key dimension along which anomalies are similar and suggest that efforts to extract common information from the set of anomalies that comprise the factor zoo may be more effective if focused on the set of stocks that are most influential to them – the overlap stocks. Second, my results are related to the literature that examines the interactions between anomalies and, in particular, the complementarity of signals between anomalies (e.g., Novy-Marx (2013), Brandt et al. (2009), Lewellen (2015)). My results contribute to this strand of literature by providing evidence that the large set of cross-sectional return predictors identified in prior studies are noisy yet related proxies for a common signal, which helps explain why combining seemingly opposing anomalies, as in, e.g., Novy-Marx (2013), often yields superior

¹Estimates of the number of documented cross-sectional anomalies are as high as 452 (Hou et al. (2020)). Chen and Zimmermann (2021), however, point out that many of these anomalies are not constructed from distinct characteristics. Thus, although my set of 164 anomalies/characteristics is significantly lower than the proposed raw number of anomaly portfolios, the characteristics that I examine can be used to construct the vast majority of them.

performance. I propose a method to extract this signal that follows directly from my main analysis, avoids the need to estimate time-varying covariances or expected returns, and is simple to implement in practice.

My analysis is also related to the literature that attempts to relate anomaly returns to expectations errors and mispricing (e.g., Piotroski and So (2012), Lewellen (2015), Stambaugh et al. (2015), Stambaugh and Yuan (2017), McLean and Pontiff (2016), Engelberg et al. (2018), Engelberg et al. (2020)). For example, Piotroski and So (2012) provide evidence that the returns to the value anomaly are driven by the subset of stocks in each leg that are most likely to be mispriced and that a strategy that extracts just these stocks significantly outperforms a “naive” value strategy. Relatedly, McLean and Pontiff (2016) and Jacobs and Müller (2020) demonstrate that anomaly returns decay post-publication in a way that is consistent with mispricing. My methodology of identifying overlap stocks is closely related to that of Stambaugh et al. (2015) and Stambaugh and Yuan (2017), who combine 11 firm-level characteristics to form their “mispricing” factors and to that of Engelberg et al. (2018) and Engelberg et al. (2020), who examine the relationship between analyst expectations and 93 firm-level characteristics. My work builds on this strand of literature by focusing on the overlap of stocks across the majority of anomalies in the factor zoo, demonstrating that these stocks are disproportionately responsible for anomaly returns, and by providing corroborating evidence that the returns to these stocks are at least partially attributable to mispricing.

Finally, my work has direct implications for investors in practice. Due to implementation costs, investors rarely hold anomaly portfolios that correspond exactly to those explored in academic research and instead hold relatively more cost-effective portfolios that are constructed to mimic underlying anomaly portfolios (e.g., Cremers et al. (2022)). My main analysis demonstrates that exclusion of a relatively small number of stocks from the cross section drastically reduces anomaly alphas and suggests that investors seeking factor exposures may fail to capture the desirable features of anomalies if the portfolios that they use happen to exclude just a few key stocks in any given month.² On the other hand, my results suggest that the portfolio optimization problem for factor investors may be simpler than previously thought. That is, instead of choosing weights across a large set of individual anomaly portfolios, I provide evidence that investors may be able to capture many of the desirable features of anomalies by investing in a single, composite portfolio – what I will refer to throughout the rest of the paper as the “overlap portfolio.”

The rest of this paper is organized as follows. Section 2 describes my data, how I form anomaly portfolios, identify the stocks that are most common across anomaly portfolios, and test whether these stocks play an important role in anomaly portfolios. Section 3 examines the degree to which anomalies share stocks, presents my main results, and explores the performance of long-short portfolios formed using overlap stocks. Section 4 explores some of the implications for investors in the factor investing space. Section 5 examines whether

²Relatedly, my results help explain why anomaly performance is so sensitive to portfolio construction choices.

biased expectations and mispricing can help explain the returns of overlap stocks. Section 6 concludes.

2 Data and Methods

I gather monthly firm-level return data from CRSP, data on 209 monthly firm-level characteristics from Chen and Zimmermann (2021), and Fama-French three and four factor data from Ken French’s website. From CRSP, I keep only regular share codes, stocks that trade on the NYSE, NASDAQ, or AMEX, observations with nonzero market values of equity, stocks with prices above \$1 (i.e., exclude penny stocks), and non-financials (i.e., exclude SIC codes 6000-7000). Following Banz (1981), I compute size as the natural log of market capitalization. From the 209 characteristics in Chen and Zimmermann (2021), I keep only continuous characteristics (such that I can form decile portfolios) and characteristics that have sufficient data to form at least twenty years of monthly portfolios. After filtering I am left with 164 unique firm-level characteristics.

Chen and Zimmermann (2021) provide characteristics that are available as of month end. Although theoretically accessible for trading at the start of the following month, it is unlikely – especially before the the advent of the SEC’s EDGAR filing system in the 1990s – that investors had such timely access to this information. As such, and as is standard in the literature (e.g., Chen and Zimmermann (2021)), I lag all characteristics by one month.³ Thus, for example, the inclusion of a firm in the value portfolio at the beginning of month t is determined by its book-to-market ratio at the end of month $t - 2$. I winsorize each of the 164 characteristics at the 1% and 99% levels and, following Chen and Zimmermann (2021), sign them so that expected returns are increasing in the characteristic.^{4,5} For example, small stocks, high book-to-market stocks, and high lag 12-month return stocks are signed such that they appear in the 10th decile of size, value, and momentum sorts, respectively. Since data for each of the characteristics are not always available over the entire sample period, the number of anomaly portfolios varies by month and generally increases over time. For example, characteristics requiring accounting data only enter my sample post 1961, corresponding to the beginning of Compustat data, whereas characteristics that rely on analyst forecasts only begin to enter my sample post 1970, corresponding to the beginning of IBES data. A comprehensive list of the 164 characteristics, their descriptions, and their sample periods can be found in the appendix.

I first compute monthly firm-level decile ranks for each characteristic. I then compute the number (both gross and net) of extreme deciles that a stock falls into each month. For example, if a stock has a high lag

³It is possible, pre EDGAR, that the average investor would not have had access to this information even one month after it was initially available (e.g., Ivkovich et al. (2021)). Regardless, it is likely that sophisticated, highly capitalized investors would have. I also note that my results are robust to not lagging by an extra month.

⁴I winsorize because it is standard in the literature to do so. The winsorization process does not affect my results.

⁵Chen and Zimmermann (2021) sign characteristics so that expected returns are increasing in them for the sample period of the original paper in which they were presented as return predictors. Some of the long-short anomaly portfolios have full sample average returns (and alphas) that are negative. I do not resign these portfolios, reflecting the fact that a real-time investor would have used the signs in Chen and Zimmermann (2021) if they were trying to exploit the predictors following their publications.

12-month return (i.e., is included in the long leg of the momentum strategy), a low book-to-market ratio (i.e., is included in the short leg of the value strategy), and is small (i.e., is included in the long leg of the size strategy), where short and long legs refer to the bottom and top decile for that characteristic, then the stock's gross short leg (long leg) inclusion for the month is 1 (2) while its net short leg (long leg) inclusion is -1 (1).⁶ More formally, a stock's net long inclusion in month t is given by,

$$Net_Long_{i,t} = \sum_{i=1}^j Long_Leg_{i,j,t} - \sum_{i=1}^j Short_Leg_{i,j,t} \quad (1)$$

where $Long_Leg_{i,j,t}$ is an indicator variable equal to 1 if stock i is in the long leg of anomaly j in month t . Similarly, $Short_Leg_{i,j,t}$ is an indicator variable equal to 1 if stock i is in the short leg of anomaly j in month t . A stock's net short inclusion is just the inverse of its net long inclusion ($Net_Short_{i,t} = \sum_{i=1}^j Short_{i,j,t} - \sum_{i=1}^j Long_{i,j,t}$) and its gross long (short) inclusion is $\sum_{i=1}^j Long_{i,j,t}$ ($\sum_{i=1}^j Short_{i,j,t}$). After calculating each stock's monthly anomaly inclusion measures, I determine monthly percentile (decile) ranks of short-leg and long-leg inclusion (both gross and net) for each stock.

My general approach is to form three sets of portfolios: anomaly portfolios, anomaly portfolios that exclude the overlap stocks, and portfolios consisting only of the overlap stocks. I begin by forming standard zero net investment anomaly portfolios that long (decile 10) and short (decile 1) the extreme deciles of a characteristic and that rebalance monthly. Because I construct these portfolios using the entire cleaned universe of stocks, I denote them "unfiltered portfolios." Second, using the same extreme deciles, I reform anomaly portfolios while excluding stocks above the N th monthly percentile of net inclusion.⁷ I denote these "filtered portfolios." For example, the 90th percentile filtered portfolios are constructed after excluding the 10% of stocks with the greatest long leg overlap and the 10% of stocks with the greatest short leg overlap for the month. Importantly, I do not re-rank stocks with respect to their characteristics after excluding the N th percentile overlap stocks. For example, if a stock is in the 10th decile of B/M pre-filtering, it will also be in the 10th decile post-filtering (supposing it is not one of the stocks that is filtered out). I do, however, reweight stocks post-filtering so that the filtered anomaly portfolios are still zero net investment.⁸ Third, I construct a long-short portfolio that uses only the overlap stocks, which I denote the "overlap portfolio." Specifically, the N th percentile overlap portfolio longs stocks above the N th percentile of long leg inclusion and shorts stocks above the N th percentile of short leg inclusion. I construct and report results for both equal-weighted

⁶As noted above, the top decile is always defined as the long leg. For example, the top decile (long leg) of value consists of high book to market stocks while the top decile of size (long leg) consists of small stocks.

⁷I focus on net overlap stocks for much of the paper, but the results remain qualitatively similar when filtered based on gross overlap.

⁸For instance, if an equal-weighted unfiltered anomaly portfolio long leg consisted of 100 stocks, each stock would receive a weight of 1/100. If 5 stocks were filtered, the equal-weighted filtered long leg would consist of 95 stocks, each with a weight of 1/95.

and value-weighted versions of all portfolios. As robustness checks, I repeat my analysis using quintile sorts, using various liquidity filters (no liquidity filter, price greater than \$5, NYSE only, and stocks with total market capitalization greater than the NYSE 20th percentile), using portfolios that rebalance yearly, and using alternative sample periods. All results are robust to these alternative specifications.

Figure 1 provides an illustrative example of how I form my unfiltered and filtered anomaly portfolios. The data is a fictional snapshot of the cross section for a given month and assumes that there were only four stocks in each of the extreme deciles of B/M for the month. Stocks A, B, C, and D belong to the short leg of the value portfolio (decile 1) for the month while stocks E, F, G, and H belong to the long leg (decile 10). The column labeled ‘Unf. Weight’ reports the weight in each stock required to form an equal-weighted unfiltered value portfolio. ‘Long Percentile’ reports the (fictional) percentile rank of net anomaly long-leg inclusions for the month while ‘Short Percentile’ reports the percentile rank of net anomaly short-leg inclusions. For example, out of the eight stocks in the table, stock F was in the most long legs (net) for the month, and so has the highest long percentile ranking. Similarly, stock C was included in the least long legs (net), so has the lowest long percentile ranking and, since net rankings are symmetric, has the highest short percentile ranking. The column labeled ‘Filt. Weight’ reports the weight in each stock required to form the equal-weighted 90th percentile filtered value portfolio. Since stocks B, C, E, and F are filtered at the 90th percentile, they are not included in the filtered portfolio. The filtered short leg thus takes an equal weight between stocks A and D while the filtered long leg takes an equal weight between stocks G and H. Since they are filtered at the 90th percentile, stocks B, C, E, and F will be excluded not only from the filtered value portfolio but also from all other 90th percentile filtered anomaly portfolios. Instead, they will belong to the set of stocks used to form the 90th percentile overlap portfolio, with stocks B and C belonging to the short leg and stocks E and F belonging to the long leg.

3 Main Results

3.1 Portfolio Overlap

I begin my analysis by examining whether, and the degree to which, anomalies share stocks. I first examine the number of anomaly portfolios each month that include at least N stocks that are above the 90th percentile of net overlap. For example, if in a given month the sample consists of 100 anomalies and 400 overlap stocks (200 long leg overlap stocks and 200 short leg overlap stocks), I count the number of anomalies, out of 100, that include at least N of the 400 overlap stocks. Second, I examine the average number of anomaly long (short) legs that the average stock in the N th percentile of long leg (short leg) overlap is included in each

month. Continuing with the example above, I count (separately) the number of anomaly long legs, out of 100, that the 200 long leg overlap stocks are included in for the month and the number of anomaly short legs that the 200 short leg overlap stocks are included in. Although the first two metrics capture overlap in equal-weighted portfolios they may not be appropriate for value-weighted portfolios. As such, I also examine the average percent of value-weighted anomaly portfolio market capitalization accounted for by stocks above the N th percentile of portfolio overlap. Continuing once more with the example above, for each of the 100 value-weighted anomaly long (short) legs, I compute the fraction of anomaly portfolio weight accounted for by the 200 long (short) leg overlap stocks.

Figure 2 plots the number of anomaly portfolios with valid return observations each month as well as the number of anomaly portfolios that include at least N stocks that are above the 90th percentile of net overlap. Effectively all anomaly portfolios include at least 10 of these stocks each month, the vast majority include at least 25, and, in the latter half of my sample, very many include at least 50. For example, in the last month of my sample, 155 (98%) of the 158 anomalies with valid returns for the month include at least 10 of the 90th percentile net overlap stocks, 149 (94%) include at least 25, and 141 (89%) include at least 50. Although it is not shown in Figure 2, every anomaly portfolio includes at least one of these stocks each month, ensuring that all anomaly portfolios are affected every month, to some extent, by filtering at the 90th percentile level.

Figure 3 shows that the average stock in the 10th decile (90th percentile) of gross monthly long portfolio overlap is present in approximately 20% of anomaly long legs each month and that the average stock in the 10th decile of gross short overlap is present in approximately 25% of anomaly short legs each month. When I calculate the total percent of anomalies that a stock in the 10th decile of gross overlap is included in, rather than anomaly long and short leg inclusions separately, I find, in untabulated results, that the average overlap stock is included in approximately 35% of anomalies each month. Coverage of the average 10th decile net overlap stock is slightly less than that of the average gross overlap stock since net overlap is necessarily less than or equal to gross overlap.

Table 1 reports the average percent of anomaly portfolio market capitalization that the N th percentile gross overlap stocks account for each month.⁹ These stocks account for a nontrivial portion of anomaly portfolios. For example, the stocks above the 90th percentile of gross overlap account for an average of 21.57% of the market capitalization of the average value-weighted anomaly in any given month. In terms of raw inclusion, the average anomaly portfolio in my sample includes 105 90th percentile gross overlap stocks each month. Further, the average total percent of anomaly market capitalization accounted for by the N th percentile gross overlap stocks is always greater than the percent of stocks excluded from the investable

⁹For each anomaly portfolio, each month, I compute the percent of the anomaly's market capitalization that is accounted for by the N th percentile gross overlap stocks that are included in the anomaly portfolio that month. Then, for each anomaly, I take the time-series average and report statistics of these averages.

universe. For instance, the 99th percentile overlap stocks account for 2% of the investable universe (the top 1% of long leg inclusion stocks and the top 1% of short leg inclusion stocks) but account for approximately 3% of the average anomaly’s market capitalization. This is consistent with later results and indicates that overlap stocks tend to be larger than the average stock in anomaly portfolios. Panel B of Table 1 indicates that much of this effect comes from the stocks in the short legs of the overlap portfolios. It is also worth noting that every one of the 164 anomaly portfolios that I examine has exposure to the overlap stocks at some point in time (i.e., the column labeled ‘Min’ in Table 1 is always greater than 0).

3.2 Anomaly Performance

The results in the previous section suggest that the stocks above the N th percentile of anomaly portfolio inclusion in a given month are not just selecting stocks from a few anomaly portfolios – these stocks have a meaningful overlap between all 164 of the anomalies that I examine. Although the overlap stocks are present, to varying degrees, in all anomaly portfolios, it may be that these stocks are largely inconsequential for the performance of the majority of anomalies. For example, it could be the case that the overlap portfolio only ever selects stocks that are important for the performance of value-type anomalies and selects stocks that, while they are present in them, contribute very little to the returns of momentum-type anomalies. To examine whether these stocks are generally important for anomaly portfolio returns, I compute performance metrics for anomaly portfolios formed using all available stocks (“unfiltered portfolios”) and for portfolios formed after excluding stocks in the N th percentile of monthly net portfolio overlap (“filtered portfolios”).¹⁰ Although it is common in the literature to test whether a portfolio of stocks prices the cross section, I focus on the effects of excluding stocks, rather than on the ability of the overlap portfolio to price anomalies, because I am interested in (a) whether or not anomalies share many common stocks and (b) whether those stocks are the ones that are “important” for anomaly performance in general. The ability of a factor to price an anomaly says nothing about whether that factor shares stocks with the anomaly. Similarly, the fact that a factor shares stocks with an anomaly says nothing about its ability to price the anomaly (or whether the shared stocks are disproportionately important for the anomaly). To see the first point it is sufficient to note that a factor need not even be traded to price the cross section (e.g., it was long thought that consumption should work as a standalone factor). To see the second point, it is sufficient to note that the market portfolio does not price CAPM anomalies.

The results in Table 2 suggest that the stocks that are included in the most anomaly portfolios for a given month have a disproportionately large impact on anomaly performance in general. Consider the first column of Panel A; the average equal-weighted unfiltered anomaly returns 4.86% per year. Excluding just 10% of

¹⁰Recall that I retain the original (unfiltered) characteristic decile ranks when forming the filtered portfolios.

stocks from the cross section each month (the top 5% of net long overlap and the top 5% of net short overlap) reduces this to 2.81% (a 42% reduction). Turning to Panel B, an average of just 17% of anomaly portfolio stocks (corresponding to an average of just 73 stocks) each month account for over 40% of the average returns and 37% of the average alphas of the 164 value-weighted anomaly portfolios. For comparison, this is similar in magnitude to the reductions in the momentum portfolio alphas documented in Boguth et al. (2011) after controlling for time-varying betas and to the average reduction in (absolute) alphas documented in Kozak et al. (2018) after controlling for the first five principal components of the return space. Filtering out the top 20% of stocks included in the most anomaly portfolios (the top 10% included in the most long legs and the top 10% included in the most short legs) reduces average value-weighted anomaly returns and alphas by 57% and 53%, respectively. Further, the filtered portfolios still contain a relatively large number of stocks – well above the threshold beyond which we typically consider a portfolio as well-diversified – such that if the excluded stocks were not asymmetrically consequential (or, equivalently, if all of the anomaly stocks were equally consequential) we should expect to see no meaningful, systematic change in portfolio performance.¹¹ That is, the results in Table 2 clearly demonstrate that the overlap stocks greatly outperform the average anomaly stock.

Although filtering does not entirely remove alphas, it is likely that the remaining abnormal returns would not survive transaction costs. For example, Barroso and Detzel (2021) estimate average annual trading costs of 2.6% for a set of 8 well-known anomalies while Chen and Velikov (2017) estimate average annual trading costs of 3.6% for a broad set of 120 anomalies. Even after employing cost-mitigation strategies, Chen and Velikov (2017) estimate that their best performing anomaly portfolios would only have earned 2.40% annually net of trading costs (post-publication). The results in Table 2 suggest, then, that nearly all of the attainable alphas of even the best performing anomaly portfolios are accounted for entirely by the overlap stocks.

Table 3 shows that the stocks that are present in the most anomaly portfolios for a given month have also accounted for the vast majority of the cumulative returns to anomaly portfolios. For example, excluding an average of just 50 stocks each month reduces the full-sample average equal-weighted cumulative return by 50%. Excluding 250 reduces it by 80%. Given that there are an average of 2,498 stocks per month in my baseline sample, excluding 250 stocks still leaves an average of 2,248 stocks from which to form anomaly portfolios in any given month.

Tables 2 and 3 provide evidence that the stocks in the overlap portfolio account for a significant portion of the average anomaly portfolio’s performance. However, the impact may not be consistent across all anomalies.

¹¹It is important to note that I reweight the filtered portfolios so that they are still zero net investment strategies. Since the alpha of a portfolio is equal to the weighted sum of the alphas of the stocks that comprise it, extracting the overlap stocks without reweighting the remaining ones would mechanically reduce alphas (supposing that all of the stocks earn approximately the same alphas), since the resulting portfolio would represent a de-leveraged version of the original. That is, all else equal, a portfolio with a cumulative weight less than one will have a lower alpha than a portfolio with a cumulative weight of one.

That is, it might be that a handful of anomalies experience large negative changes after excluding the overlap stocks while the rest are essentially unaffected. Although the similarity of the means and medians in Table 2 suggests that this is not an issue, I address this possibility by examining post-filtering changes in CAPM alphas at the portfolio level.

Figure 4 plots the change in annualized anomaly portfolio alpha estimates after excluding stocks in the N th percentile of net inclusion (99th and 90th percentiles). I compute the change as the filtered alpha estimate less the unfiltered alpha estimate such that a positive value indicates an increase in alpha post-filtering, a value of zero indicates no change, and a negative value indicates a reduction.¹² Portfolios are sorted based on their unfiltered alpha p -values such that evidence of unfiltered abnormal returns is increasing along the x -axis. The results demonstrate that the reduction in alphas documented in Table 2 is not driven by extreme reductions in just a few portfolios; exclusion of the N th percentile net overlap stocks reduces alpha estimates for nearly all of the anomalies in my sample. It is also worth noting that both the reduction in alphas and the magnitude of unfiltered alphas are, on average, increasing in the magnitude of the unfiltered evidence of abnormal returns. Further, the few portfolios that experience an increase in alphas post-filtering are mostly portfolios that had insignificant full-sample alpha estimates pre-filtering.

The reduction in alphas is also not particular to a few “types” of anomalies. Table 4 reports the average alphas and alpha t -statistics across anomaly categories for unfiltered portfolios and for portfolios formed after excluding the 90th percentile net overlap stocks. The first two categorization schemes are taken directly from Chen and Zimmermann (2021). The “data” category categorizes anomalies based on the source of the data used to form their respective characteristics while the “economic” category offers a more fine-grained categorization. The third categorization scheme is based roughly on that used in Jensen et al. (2023) and is more fine-grained than the “data” scheme but less fine-grained than the “economic” scheme.¹³ The mappings between anomaly portfolios and categories, for each categorization scheme, are reported in the appendix.

All but one anomaly category, across any of the categorization schemes, experience a significant reduction in average alphas and average t -statistics. The one category that does not, “Informed Trading,” is part of the “economic” and “broad” categorizations schemes, includes only three highly related portfolios, and has an average unfiltered (filtered) t -statistic of -0.04 (0.05). Focusing first on the “data” categorization scheme, the most significant reduction in alphas is in the “Other” category, with an average unfiltered annualized alpha of 1.32% and an average filtered alpha of -0.12%. The two categories that account for the vast majority of

¹²Percent changes, though perhaps more natural, are not reported because a few anomaly portfolios with unfiltered alpha estimates of effectively zero have nonzero (but still very small) alpha estimates post-filtering such that the maximum percentage change is extremely large and the graph is visually uninterpretable.

¹³There is not a one-to-one mapping between the characteristics used in Jensen et al. (2023) and in Chen and Zimmermann (2021). I roughly replicate the categorization scheme in Jensen et al. (2023) by classifying characteristics based on the similarity of descriptions and/or variables used in their construction.

the 164 anomaly portfolios in my sample, “Accounting” and “Price” type anomalies, experience the two next largest reductions in average monthly alphas (66% and 49%, respectively). Of the five data type categories that have evidence of statistically significant alphas, on average, pre-filtering, only one category (“Trading”) continues to exhibit evidence of statistically significant alphas, on average, post-filtering.

Turning to the “broad” categorization scheme, the most significant reduction in average alphas is in the “Moments” category, with an average unfiltered annualized alpha of 1.32% and an average filtered alpha of 0.00%. “Accruals” and “Investment” type anomalies experience the second largest declines of 81% and 77%, respectively. Of the 13 categories, 10 have statistically significant alphas, on average, pre-filtering, while only three (“Institutions,” “Liquidity,” and “Momentum”) have statistically significant alphas, on average, post-filtering. Put differently, 77% of the “broad” anomaly categories (which account collectively for 80% of the 164 anomalies) have no evidence, on average, of abnormal returns post-filtering.

Overall, the evidence in this section suggests that the returns to anomaly portfolios are not “well-spread” across the stocks that comprise them. Rather, anomaly portfolio returns are concentrated in a relatively small subset of stocks – the stocks that are shared across the most anomaly portfolios.

3.3 Complementarity of Signals

The previous sections provide evidence that there is a relatively small set of stocks that is shared across nearly all anomalies in any given month and that these stocks are disproportionately responsible for anomaly performance. I now turn to examining whether these stocks yield generally complementary signals and, if so, whether this signal could be exploited by investors.

Figure 5 plots average returns as a function of anomaly portfolio overlap and suggests that characteristics are especially good predictors of returns when aggregated; average returns are increasing monotonically in long-leg inclusion and decreasing in short-leg inclusion. Though it is not surprising that characteristics, on average, are good return predictors, it perhaps is less obvious that they should be generally complimentary. That is, although characteristics within a given “family” may be closely related, there is no obvious reason, *ex ante*, to expect characteristics between families to be related. For example, although value-type anomalies likely generate similar signals, it is not obvious that value-type anomalies should yield contemporaneously similar signals as momentum-type anomalies. Moreover, many of the anomalies that I examine are strongly negatively correlated with one another (e.g., momentum vs reversal type anomalies). In this case it would be reasonable to expect returns to be *decreasing* in portfolio overlap, given that the two portfolios yield ostensibly conflicting signals – the exact opposite of what is suggested by Figure 5.

Panel A of Figure 5 plots average returns per decile of gross long (short) inclusion. Returns are monotonic

in gross long overlap and the spread between the 1st and 10th deciles is substantial (1.71% per month). The returns per decile of gross short overlap, however, are relatively flat across deciles 1-8, with the majority of the action contained in the 10th decile. The spread between deciles 1 and 10 of gross short inclusion is 0.65% per month while the spread between the 10th decile of gross long inclusion and the 10th decile of gross short inclusion is 1.72% per month. In other words, lack of gross long inclusion is a more stable predictor of low relative returns than is gross short leg inclusion. Panel B plots average returns per decile of net long (short) overlap. The returns per decile of net inclusion are strictly monotonic and result in a substantial spread between the 1st and 10th deciles (1.88% per month). The patterns in Panels A and B indicate that short leg inclusion, per se, is not an overly effective predictor of returns. Rather, short leg inclusion *conditional on* minimal long leg inclusion is a strong predictor of returns. Finally, Panel C plots average returns per decile of net long leg inclusion against the average return of each raw anomaly decile.¹⁴ On average, the spread in returns between the stocks in the 1st and 10th deciles of anomalies is relatively modest (0.45% per month) and is much smaller than the spread between the 1st and 10th deciles of either gross or net inclusion. In other words, inclusion in anomaly portfolios is a much stronger predictor of average returns than the average characteristic. Overall, the patterns presented in Figure 5 suggest that characteristics may be noisy proxies for a common signal, rather than distinct proxies for unrelated signals, which would help to explain not only why net (gross) inclusion is a better predictor of returns than average anomaly inclusion, but also why the spreads are both smoother and larger for net overlap stocks relative to gross overlap stocks.

Although all results are robust to the liquidity filters described in Section 2, a potential worry is that the complementarity suggested by Figure 5 is driven by a few small stocks. Figure 6 addresses this possibility by plotting firm size as a function of anomaly portfolio overlap. The stocks that are included in the most anomaly portfolios are not stocks in the extreme deciles of size. On average, the stocks that are included in the most anomaly portfolio short legs are in the sixth decile of firm size while the stocks that are included in the most portfolio long legs are in the fourth.¹⁵ In terms of raw market capitalization, rather than deciles, the stocks that are included in the most portfolio short-legs are, on average, average sized firms while those in the long-legs are slightly below-average sized. This is consistent with the findings in Israel and Moskowitz (2013) that firm size plays a larger role for the long legs of anomaly portfolios (i.e., long leg stocks tend to be smaller than short leg stocks) and with the results in panel B of Table 1 (i.e., stocks in the short leg of the overlap portfolio, on average, account for a greater proportion of anomaly market capitalization than stocks

¹⁴To compute the average return of each raw anomaly decile, I compute the average return per decile for each of my characteristics then take the average across deciles. For example, the average return for the first decile of raw anomaly inclusion is computed by taking the average of the average returns of firms in the first decile of size, in the first decile of B/M, the first decile of lag returns, etc.

¹⁵because I sign characteristics so that the 10th decile represents the long leg of the corresponding anomaly, my regular size ranking scheme ranks the smallest stocks as belonging to the 10th decile. The figure reports the inverse of the firm size ranking scheme used to construct anomalies; here, decile 10 corresponds to the largest stocks while decile 1 corresponds to the smallest.

in the long leg of the overlap portfolio).

To formally examine the complementarity of signals, I construct and evaluate the performance of the overlap portfolio – a portfolio that goes long the stocks that are above the N th percentile of net long leg inclusion and shorts stocks that are above the N th percentile of net short leg inclusion at the beginning of each month. Table 5 reports estimates of annualized alphas ranging from a high of 38.60% ($t = 14.39$) to a low of 5.76% ($t = 6.24$), all of which are significant at better than the 1% level. As a baseline estimate, the value-weighted 90th percentile overlap portfolio has an annualized CAPM alpha estimate of 12.58% ($t = 9.04$), an annualized FF3 alpha of 11.21% ($t = 8.75$), and an annualized CH4 alpha of 7.87% ($t = 6.33$). Note that the 90th percentile portfolio corresponds to the usual 10-1 decile portfolios used to form anomalies (and that the 80th percentile portfolio corresponds to a 5-1 quintile portfolio). I present percentile (rather than decile) portfolios for comparison with my main analysis.

Because anomaly portfolios are often heavily dependent on their short legs and many stocks face significant short-sale constraints, the returns achievable in practice are likely much lower than those reported in academic research (e.g., Stambaugh et al. (2012), Avramov et al. (2013), Chen and Velikov (2017)). Panel C of Table 5 suggests that this is not a large issue for the overlap portfolio; the returns to the long leg of the 90th percentile overlap portfolio are not only greater in magnitude than those of the short leg but also demonstrate stronger statistical significance. For example, the excess value-weighted 90th percentile overlap long leg has an average annualized return of 14.16%, an annualized Sharpe ratio of 0.79, and an annualized CAPM alpha of 7.74% ($t = 8.05$). In contrast, the excess value-weighted 90th percentile overlap short leg has an annualized return of 4.73%, and annualized Sharpe ratio of 0.20, and an annualized CAPM alpha of -4.85% ($t = -5.61$). Further, because the stocks in the short leg of the overlap portfolio are, on average, average-sized firms, they are potentially less likely to be subject to short-sale constraints relative to the stocks in most individual anomaly short legs.

Figure 7 plots the cumulative log returns to the 99th and 90th percentile overlap portfolios. The performance of the equal-weighted portfolio is remarkably stable over time, with none of the major crashes often associated with long-short anomaly portfolios. The performance of the value-weighted portfolio is less stable. Throughout most of the 2000's, the value-weighted overlap portfolio returns are relatively flat, and actually negative for the 99th percentile portfolio. This is perhaps not surprising given that anomaly performance, in general, has been quite poor throughout most of the 2000's and suggests that the evidence presented in Table 5 may not be representative of the evidence that would be garnered by using data from just the latter part of the sample (a possibility that I examine in the next section). Consistent with the results in Panel C of Table 5, the long legs of the overlap portfolios account for a significant portion of their total returns. In fact, the excess value-weighted 90th percentile long leg actually outperforms the value-weighted 90th percentile

long-short portfolio over the final decade of my sample.

3.4 Overlap Stocks Over Time

The evidence to this point strongly suggests that anomaly portfolio returns are driven by the stocks that are common across them and that these stocks yield generally complementary signals. A final potential worry, alluded to in the previous section, is that these results may not be consistent over time. It is well known that performance is sample-specific for many anomaly portfolios (e.g., McLean and Pontiff (2016), Hou et al. (2020)). As such, it may be that exclusion of the overlap stocks had a large negative impact on average anomaly returns for only a subset of my sample period. Similarly, it may be that the evidence of the abnormal returns of the overlap portfolio is driven by extremely strong performance in a subset of the time series. To examine whether this is the case, I reestimate my main results decade-by-decade.

Table 6 suggests that the effects documented so far are consistent over time. Panel A documents that the overlap stocks have been consistently responsible for the majority of anomaly abnormal returns. Panel B documents that the 90th percentile overlap portfolio has earned economically and statistically significant abnormal returns over the majority of the sample. As noted earlier, the poor performance of the value-weighted overlap portfolio over the last twelve years is perhaps not surprising given that anomaly performance in general has been poor throughout most of the 2000's. What is potentially surprising, however, is the extremely strong performance of the value-weighted portfolio over the first decade of the 2000's and the fact that exclusion of the overlap stocks still significantly reduces alphas in the period 2010-2022. Further, although I do not report the results in Table 6, the value-weighted 90th percentile portfolio does earn statistically significant FF3 and CH4 alphas in all decades. For example, over the period 2010-2022, the annualized FF3 alpha estimate is 8.09% ($t = 2.38$) and the CH4 alpha estimate is 5.86% ($t = 1.77$).

4 Implications for Factor Investing

The presence of a nonzero alpha in a regression of a portfolio on a factor model indicates that an investor could have achieved a higher Sharpe ratio by holding the ex post optimal combination of the right-hand side and left-hand side portfolios relative to the ex post optimal combination of the right-hand side portfolios (e.g., Gibbons et al. (1989), Cederburg et al. (2020)). For example, the positive CAPM alpha of the momentum portfolio indicates that an investor holding the market would have been better off, in mean-variance terms, by holding the market portfolio in combination with some positive weight on the momentum portfolio. It is precisely this benefit that makes anomalies attractive to investors. However, the theoretical improvement in Sharpe ratios implied by a nonzero alpha would not have necessarily been realizable by real-time investors.

That is, a nonzero alpha simply indicates that the left-hand side portfolio expands the efficient frontier. For an investor to have captured the benefit implied by this expansion, they would have had to have placed an *appropriate* weight on the left-hand side portfolio.¹⁶

Due to implementation costs, investors rarely hold anomaly portfolios that correspond exactly to those explored in academic research, and instead hold relatively more cost-effective portfolios that are constructed to mimic an underlying anomaly (e.g., Cremers et al. (2022)).¹⁷ Section 3.2 demonstrates that the exclusion of a relatively small number of stocks from the cross section drastically reduces anomaly alphas and suggests that investors seeking factor exposures may fail to capture the desirable features of anomalies if the portfolios they use to do so happen to exclude just a few key stocks in any given month. On the other hand, the results in Section 3.3 suggest that a significant portion of the desirable features of anomalies can be captured by the overlap portfolio. In this section I examine the effect that filtering has on the probability that a real-time investor could have improved the risk-return profile of their portfolio by investing in a given anomaly portfolio and whether or not the overlap portfolio would have offered better chances of improving an investor's Sharpe ratio relative to individual (unfiltered) anomaly portfolios.

For each of my 164 anomaly portfolios (and for each of my filtering schemes), I estimate the Sharpe ratio improvements that would have been obtained by an investor following a simple real-time mean-variance optimization strategy. Let $\hat{\mu} = \begin{bmatrix} \hat{\mu}_i & \hat{\mu}_{Mkt-RF} \end{bmatrix}'$ be the 2×1 vector of estimated mean returns of an anomaly portfolio i and the excess market portfolio, $\hat{\Sigma}$ be the estimated 2×2 variance-covariance matrix of returns, and γ be the investor's coefficient of risk aversion. The vector of ex post optimal weights for a mean-variance investor is given by

$$x^* = \begin{bmatrix} x_i^* \\ x_{Mkt-RF}^* \end{bmatrix} = \frac{1}{\gamma} \hat{\Sigma}^{-1} \hat{\mu} \quad (2)$$

which implicitly assumes that the investor has access to a risk-free asset (i.e., the weights need not sum to one). However, the resulting Sharpe ratio will be equivalent to that of a portfolio that scales the optimal weights x^* such that the investor splits 100% of their wealth between the anomaly portfolio and the market portfolio (i.e., the weights sum to one). The vector of scaled weights, which do not rely on the investor's risk

¹⁶For example, Cederburg et al. (2020) demonstrate that the implied portfolio benefits of many volatility managed portfolios would not have been easily exploited in real time due to difficulties in estimating optimal weights ex ante.

¹⁷Cremers et al. (2022) document that traded factor funds perform better the more closely they track the holdings of their respective academic factor portfolios. The analysis in this section helps to explain why this might be the case (i.e., traded factor funds that do not closely track the holdings of their target factors may be omitting overlap stocks).

aversion, is given by

$$x^* \propto w^* = \begin{bmatrix} w_i^* \\ w_{Mkt-RF}^* \end{bmatrix} = \frac{\hat{\Sigma}^{-1} \hat{\mu}}{|1_2' \hat{\Sigma}^{-1} \hat{\mu}|} \quad (3)$$

where 1_2 is a 2×1 vector of 1s.¹⁸ I assume that investor's follow equation 3 since it allows me to remain agnostic about the level of risk aversion and does not impact the interpretation of my analysis.

Because equation 3 yields ex post optimal weights, I assume that the investor estimates w^* period-by-period using historical data in an expanding window and uses this estimate to determine their weight in the anomaly portfolio for the following period. The return to the portfolio that blends anomaly i with the market in month t is thus

$$R_{blend,i,t} = w_{i,t-1}^* R_{i,t} + (1 - w_{i,t-1}^*) R_{Mkt-RF,t} \quad (4)$$

where $w_{i,t-1}^*$ is estimated using data from the periods $t = 0$ to $t - 1$.

Because investors could reasonably disagree on the appropriate initial estimation window, I examine results for initial windows set at 60 months, 120 months, and 200 months. I restrict weights to be between 0 and 1 by censoring the estimated weight in the anomaly portfolio and allocating the remainder to the market.¹⁹ That is, $w_i = \max(0, \min(1, w_i^*))$ and $w_{Mkt-RF} = 1 - w_i$. The restriction on short selling reflects the fact that an investor would not have reason, ex ante, to short any of the anomalies in my sample since they are signed so that they earn positive expected returns over their original sample periods. The restriction on leverage represents a realistic restriction for most investors (i.e., it is unlikely that the marginal investor would be either willing or able to take a position significantly greater than 1 in an anomaly portfolio).

The above procedure results in a time series of returns which I use to compute the Sharpe ratios of the blended portfolios. I then compute the differences between the Sharpe ratios of the portfolios that blend the anomaly with the market and the Sharpe ratio of a portfolio with a 100% allocation to the market portfolio. I compute the difference as,

$$\Delta S R_{blend,i} = \frac{\mu_{blend,i}}{\sigma_{blend,i}} - \frac{\mu_{Mkt-RF}}{\sigma_{Mkt-RF}} \quad (5)$$

¹⁸The coefficient of risk aversion in equation (4) affects the proportion of total capital allocated to the risky assets but does not affect the proportional weights in each of the risky assets (i.e., $x^* \propto w^*, \forall \gamma$). See, e.g., Cederburg et al. (2020).

¹⁹The optimization problem can be solved without censoring by solving a quadratic optimization problem via numerical methods. I opt for the censoring approach because of its simplicity and easy interpretation.

and assess the statistical significance of the difference using the test defined in Jobson and Korkie (1981).^{20,21}

The results presented in Table 7 reveal that real-time investors would have been relatively unlikely to have increased their Sharpe ratios by allocating to individual anomaly portfolios. For example, using a 60 month initial estimation window, only 18% (29 out of 164) of the unfiltered value-weighted anomaly portfolios would have increased the Sharpe ratio of a real-time investor. When anomaly portfolios are filtered at the 97.5th percentile, only 9% (15 out of 164) would have. Thus, even if investors knew ex ante which of the 86 value-weighted anomalies would have earned significant alphas over the full sample, they would have still only had a 33% chance of improving their Sharpe ratio by allocating to any of the individual 86 anomalies. If the portfolios that they used were similar to the 97.5th percentile filtered portfolios, they would have only had a 17% chance. Moreover, in untabulated results, among the 29 unfiltered value-weighted anomaly portfolios that could have improved an investor's Sharpe ratio, the average improvement (0.25) is less than half that of the value-weighted 90th percentile overlap portfolio (0.57). Additionally, the maximum improvement (0.43) is about 1.3 times lower than that of the value-weighted 90th percentile overlap portfolio (0.57).²² The results are qualitatively similar for equal-weighted portfolios and for each of the initial estimation windows.

Although I have restricted optimal weights to be between 0 and 1 it is still possible that the optimal weight in the 90th percentile overlap portfolio would have been much larger than the marginal investor would have realistically been comfortable with. Relatedly, it is possible that the documented Sharpe ratio improvement depends on a very tight range of weights in the 90th percentile overlap portfolio which an investor may have failed to estimate as optimal if, e.g., they had used a slightly different version of the overlap portfolio. To address these concerns I estimate the Sharpe ratio improvements that would have been realized by an investor that placed a static weight $w_i \in \{0.01, 0.02, \dots, 1.00\}$ in the 90th percentile overlap portfolio.

Figure 8 plots the Sharpe ratios of each of the portfolios that blend the 90th percentile net overlap portfolio with the excess market portfolio. The Sharpe ratio of the blended portfolio is greater than that of the market in all of the 100 weighting schemes for both the equal-weighted and value-weighted portfolios. For the equal-weighted blended portfolios, the differences are all significant at better than the 5% level. For the value-weighted portfolio, 88 out of 100 of the differences are significant at better than the 5% level, with the differences losing statistical significance (at the 5% level) past $w_i = 0.89$. Thus, a market investor could

²⁰Let $\hat{\mu}_i$, $\hat{\sigma}_i$, be the estimated mean and standard deviation of portfolio i , $\hat{\mu}_j$, $\hat{\sigma}_j$, be the estimated mean and standard deviation of portfolio j , and $\hat{\sigma}_{i,j}$ be the estimated covariance between portfolios i and j over a period T . The Jobson and Korkie (1981) test statistic corresponding to $H_0 : SR_i = SR_j$ is: $\hat{Z} = \frac{\hat{\mu}_i \hat{\sigma}_j - \hat{\mu}_j \hat{\sigma}_i}{\sqrt{\hat{\theta}}}$ where $\hat{\theta} = \frac{1}{T} \left(2\hat{\sigma}_i^2 \hat{\sigma}_j^2 - 2\hat{\sigma}_i \hat{\sigma}_j \hat{\sigma}_{i,j} + \frac{1}{2}\hat{\mu}_i^2 \hat{\sigma}_j^2 + \frac{1}{2}\hat{\mu}_j^2 \hat{\sigma}_i^2 - \frac{\hat{\mu}_i \hat{\mu}_j}{\hat{\sigma}_i \hat{\sigma}_j} \hat{\sigma}_{i,j}^2 \right)$.

$\hat{Z}$ incorporates the Memmel (2003) correction and is asymptotically standard normal.

²¹In a few cases, $w^* \approx 0$ for all periods resulting in $\hat{\theta} \approx -0$ or $\hat{\theta} = 0$ due to floating-point precision issues (e.g., $\hat{\theta} = -1 \times 10^{-19} \approx 0$). Because the Jobson and Korkie (1981) test statistic is undefined for $\hat{\theta} < 0$, I handle cases where $\hat{\theta} \approx -0$ by setting $\hat{\theta} = |\hat{\theta}|$. When $\hat{\theta} = 0$ I set $\hat{Z} = 0$. $\Delta SR = 0$ in all of these cases, so adjusting $\hat{\theta}$ does not affect the results (i.e., the only change is that $\hat{Z} = 0$ rather than NaN).

²²The maximum Sharpe ratio improvement for value-weighted anomaly portfolios using a 60 month initial estimation window corresponds to the characteristic dNoa of Hirshleifer et al. (2004).

have improved their Sharpe ratio by taking nearly any weight in the value-weighted 90th percentile overlap portfolio. Further, there are no weights for which they would have decreased their Sharpe ratio (either in nominal or statistical terms). For a more detailed view of the performance of the blended overlap portfolios, Table 8 reports average returns, standard deviations, Sharpe ratios, differences in Sharpe ratios, and the statistical significance of these differences for $w_i \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

The results in Table 8 highlight the degree to which a market investor could have improved the risk-return profile of their portfolio by investing in the 90th percentile overlap portfolio. Panel A suggests that an investor holding the full-sample optimal weight in the equal-weighted 90th percentile overlap portfolio could have achieved a Sharpe ratio nearly four times larger than that of the market (1.68 vs 0.43). Panel B suggests that a market investor could have more than doubled their Sharpe ratio by holding the full-sample optimal weight in the value-weighted 90th percentile portfolio (1.02 vs 0.43). Moreover, an investor could have significantly improved their Sharpe ratio even if they had chosen a weight very far from the optimal. For example, over the full sample the optimal weight in the value-weighted 90th percentile overlap portfolio for a market investor was approximately 60%, but any weight from 40% up to 80% would still have at least doubled their Sharpe ratio.

The results in this section suggest that investors could have easily failed to capture the portfolio benefits implied by the alphas of most anomaly portfolios. On the other hand, a market investor could have significantly improved the risk-return profile of their portfolio by taking nearly any weight in the 90th percentile overlap portfolio. Overall, the results in this section imply that the portfolio choice problem for factor investors can be substantially simplified; instead of estimating optimal weights across a large set of anomaly portfolios, investors may be better off simply allocating some portion of their wealth to the overlap portfolio.

5 Overlap Stocks and Investor Expectations

Prior research suggests that biased investor expectations and mispricing may partially explain anomaly returns. For example, Kozak et al. (2018) find that high-alpha stocks have larger earnings forecast errors, Engelberg et al. (2018) find that stocks associated with anomaly portfolios are more prone to analyst earnings forecast errors and earn outsized returns on news days, Engelberg et al. (2020) find that stocks associated with anomaly portfolios are more prone to analyst return forecast errors, Bordalo et al. (2024) find that earnings forecast errors subsume the explanatory power of a set of standard characteristics, and McLean and Pontiff (2016) and Jacobs and Müller (2020) both provide evidence that many anomaly returns decay post-publication in such a way that is consistent with mispricing. In this section, I examine whether biased expectations help explain the performance of overlap stocks by testing whether the degree of anomaly portfolio

inclusion is associated with analyst return forecast errors, earnings announcement day returns, or both.

I pull all 12-month analyst price targets for U.S. stocks from the the IBES Price Target database.²³ Following Engelberg et al. (2020), I compute firm-level median analyst price targets each month and define the stock-level return forecast as,

$$E_t(R_{i,t+12}) = \left(\frac{\tilde{E}_t(P_{i,t+12})}{P_{i,t}} \right) - 1 \quad (6)$$

where $\tilde{E}_t(P_{i,t+12})$ is the median 12-month price target for stock i in month t , $P_{i,t}$ is the actual price in month t , and $E_t(R_{i,t+12})$ is the 12-month forecasted return.²⁴ Again, following Engelberg et al. (2020), I drop return forecasts that are more than five standard deviations from the mean and then winsorize all forecasts at the 1% and 99% levels. After merging with my main dataset and accounting for my baseline data filters I am left with 333,135 firm-month return forecast observations. Table 9 presents average 12-month return forecasts and average (actual) annualized returns by decile of net anomaly inclusion.

Consistent with Engelberg et al. (2020), I find that analyst return forecasts are higher for stocks in the most anomaly short legs and lower for stocks in the most anomaly long legs. Stocks in the tenth decile of net anomaly short leg inclusion (1st decile of net anomaly long leg inclusion) have average realized annual returns of 12% but average forecasted annual returns of 47%. In contrast, stocks in the tenth decile of net anomaly long leg inclusion have average realized annual returns of 37% but average forecasted annual returns of 28%. The forecast errors are nearly monotonic in decile of anomaly portfolio inclusion, with analysts generally overestimating the returns of stocks more likely to be included in anomaly short legs and underestimating the returns of stocks more likely to be included in anomaly long legs. In terms of spreads, average annualized returns of a 10-1 portfolio are 24.19% but are forecasted to be -19.18%. This corresponds to a forecast error of -43.37%. Table 9 also shows that the number of analysts per stock forecast is relatively consistent across deciles of anomaly inclusion but that stocks in the most anomaly short legs are approximately three times more likely to be covered by analysts, which is consistent with analysts being more likely to cover larger stocks (on average, stocks in the 1st decile of net long inclusion (10th decile of net short inclusion) are in the 6th decile of size while stocks in the 10th decile of net long inclusion are in the 4th decile – see Figure 6).

To formally test the relation between anomaly inclusion and analyst forecast errors I estimate the following regression,

$$\begin{aligned} E_t(R_{i,t+12}) = & \alpha_i + \beta_1 Long_Leg_{i,t} + \beta_2 Short_Leg_{i,t} + \beta_3 R_t^m + \\ & + \beta_4 Long_Leg_{i,t} \times R_t^m + \beta_5 Short_Leg_{i,t} \times R_t^m + \epsilon_{i,t} \end{aligned} \quad (7)$$

²³IBES Price Target data begins in 1999-02 and ends in 2022-12.

²⁴I follow Engelberg et al. (2020) in using median analyst price targets but my results are nearly identical using average analyst price targets.

where $E_t(R_{i,t+12})$ is the median forecasted 12-month return as defined by equation 6, $Long_Leg_{i,t}$ is an indicator variable equal to one if stock i is in the 10th decile of net anomaly long leg inclusion for month t , $Short_Leg_{i,t}$ is an indicator variable equal to one if stock i is in the 10th decile of net anomaly short leg inclusion for month t , and R_t^m is the value weighted excess market return in month t .

The results in Table 10 indicate that analysts estimate annual returns to be 24% on average. For stocks in the tenth decile of net anomaly long leg inclusion, their estimates are approximately 4% above the average estimate ($0.240+0.038=0.278$), while they are approximately 24% above the average estimate for stocks in the tenth decile of net anomaly short leg inclusion ($0.240+0.238=0.448$). The regression results confirm the pattern presented in Table 9 – analysts expect the stocks in the most anomaly short legs to have much higher returns than the stocks in the most anomaly long legs.

It is important to note that analyst errors will contribute to market inefficiency and anomaly mispricing only if the investors that follow analyst forecasts exert significant price pressure. Although it is beyond the scope of this paper to estimate the impact of analyst forecasts on prices, prior literature suggests that the impact may not be trivial. For example, Brown et al. (2014) document that institutional investors herd into stocks with analyst upgrades and out of stocks with analyst downgrades and that these stocks “...experience a significant same-quarter price impact, followed by a sharp subsequent price reversal.” Relatedly, Kong et al. (2021) document that active institutional investors are net buyers (sellers) of stocks with positive (negative) analyst recommendations. If institutional investors follow analyst forecasts, the above results suggest that we would observe them overweighting the “incorrect” leg of anomaly portfolios. This is precisely what Edelen et al. (2016) observe – institutions, on average, overweight stocks that belong to anomaly short legs and underweight stocks that belong to anomaly long legs. At a minimum, if analyst forecasts do contribute to anomaly mispricing, we should expect to observe price corrections on earnings announcement days.

To examine whether anomaly inclusion is related to earnings announcement day returns, I examine average raw returns and CAPM alphas on earnings announcement days for each decile of net anomaly overlap. I pull data on earnings announcement days from the merged CRSP Compustat database, daily returns from CRSP, and short-window CAPM alphas from WRDS Beta Suite.²⁵ Table 11 presents the results.

Consistent with Engelberg et al. (2018) and Kozak et al. (2018), I find that the stocks in the most anomaly long legs have higher earnings announcement day returns than the stocks in the most anomaly short legs. In terms of raw returns, the stocks in the 10th decile of net long inclusion experience average announcement day returns of 0.72% compared to average announcement day returns of -0.10% for the stocks in the 10th decile of short inclusion (1st decile of long inclusion). The difference between the two legs is 0.82% and highly

²⁵The Merged CRSP Compustat earnings announcement data begins in 1962 and ends in 2023. I consider both 65 day and 120 day windows for beta estimates and incorporate the Scholes and Williams (1977) correction for nonsynchronous data.

significant ($t = -20.51$). The results are similar for abnormal returns; earnings announcement day CAPM alphas are decreasing monotonically in decile of net long inclusion with differences between long and short legs once again highly statistically significant.

To formally test the relationship between anomaly inclusion and earnings announcement day returns I estimate the following regression,

$$DepVar_{i,t} = \alpha_i + \beta_1 Long_Leg_{i,t} + \beta_2 Short_Leg_{i,t} + \epsilon_{i,t} \quad (8)$$

where $DepVar_{i,t}$ is either the raw return ($R_{i,t}$) or the CAPM alpha ($\alpha_{i,t}$) for stock i on announcement day t , $Long_Leg_{i,t}$ is an indicator variable equal to one if stock i is in the 10th decile of net long leg inclusion on announcement day t , and $Short_Leg_{i,t}$ is an indicator variable equal to one if stock i is in the 10th decile of net anomaly short leg inclusion on announcement day t .

The results in Table 12 confirm the pattern presented in Table 11 – stocks in the most anomaly long legs experience significantly higher earnings announcement day returns and alphas than the stocks in the most anomaly short legs – and are consistent with the notion that analyst forecasts impact prices. In fact, if analyst forecasts do impact prices, the results in Tables 11 and 12 may represent a lower bound on the effects of their forecasts: based on prior research, it is plausible that some investors actively exploit the mispricing caused by analyst forecasts and then largely resolve this mispricing prior to the earnings announcement dates (e.g., De Long et al. (1990b), Brunnermeier and Nagel (2004)). That is, if the mispricing is exacerbated by “rational bubble riders” and is largely resolved before earnings announcement dates (by, e.g., bubble riders exiting their positions), then the net effect of analyst-induced pricing errors will be much larger than those suggested by Tables 11 and 12. Such an explanation would align with the findings in Brown et al. (2014) and Kong et al. (2021) and suggests a potential avenue for future research.

The results in this section are not only consistent with prior literature but are obtained using methods and samples closely matching those in Engelberg et al. (2018) and Engelberg et al. (2020). The key contribution of the results in this section is their interaction with my main analysis. That is, I have provided evidence that the overlap stocks account for a disproportionately large portion of anomaly returns, which, in conjunction with evidence that the returns to these stocks can plausibly be explained by biased investor expectations, suggests that a large portion of the returns of nearly all cross-sectional anomalies can also be plausibly attributed to mispricing.²⁶

²⁶Some readers may find the results in this section to be at odds with the recommendation in Section 4: if the overlap stocks are driven largely by mispricing, one might argue that we should not expect the overlap portfolio to continue to perform well and thus that investors may be better off not allocating a portion of their wealth to it. Ignoring traditional limits to arbitrage, this would be true if the investor biases driving the returns to the overlap portfolio were not persistent. If they are, the superior returns of the overlap portfolio could be rationalized as compensation for exposure to a systematic risk *caused by* the existence of irrational “noise traders” and should continue to exist for as long as such investors remain systematically biased (e.g., De Long

6 Conclusion

I examine a possible source of commonality among anomaly portfolios: the stocks that constitute them. I find that a small set of stocks appear simultaneously in the long (short) legs of anomaly portfolios in any given month and that these stocks contribute asymmetrically to anomaly performance. For example, just 10% of stocks each month account for nearly 40% of CAPM alphas and nearly 80% of the cumulative returns to 164 anomaly portfolios from 1926 to present. Further, these stocks can be identified nonparametrically and used to form, in real time, portfolios that earn abnormal returns that are highly economically and statistically significant; estimates of annualized alphas range from a high of 38% to a low of 5% and are always significant at better than the 1% level. Additionally, I provide evidence consistent with prior research that the returns to the stocks that are shared across the most anomalies can be explained, at least partially, by biased investor expectations and mispricing. This paired with the evidence that the overlap stocks are disproportionately important for anomaly portfolio performance suggests that a large portion of all anomaly returns can be attributed to a single source of mispricing.

My results have implications for traditional asset pricing, behavioral finance, and factor investors. First, they support the notion that the return space is relatively low-dimensional, suggesting that efforts to describe the factor zoo may be more effective if focused on the stocks most important to anomaly portfolios – specifically, the overlap stocks – rather than examining anomalies individually or by type. Relatedly, my findings indicate that the collection of proposed cross-sectional return predictors are noisy proxies for a common signal, which helps explain why combination strategies often outperform individual ones. My findings also help explain why anomaly performance seems to be so sensitive to portfolio construction choices and suggest that real-time factor investors may easily fail to capture the desirable features of their target anomaly portfolios. On the other hand, my findings suggest that investors can capture many of the desirable features of anomalies by investing in a single composite portfolio: the overlap portfolio. Finally, my findings contribute to the debate on whether anomalies reflect latent risks or mispricing by providing evidence that investors have biased expectations over the stocks most responsible for anomaly performance.

et al. (1990a), Shleifer and Vishny (1997), Kozak et al. (2018)). Although estimating the persistence of the source of mispricing is beyond the scope of this paper, the biases documented in this section appear to be consistent over the sample for which IBES data are available. Regardless, if an investor believes that factor exposures are valuable, the findings in Section 4 remain applicable.

References

- Anarkulova, A., Cederburg, S., and Zhou, Y. (2025). The stock market’s two truths: Subjective beliefs and objective reality. *Available at SSRN*.
- Avramov, D., Chordia, T., Jostova, G., and Philipov, A. (2013). Anomalies and financial distress. *Journal of Financial Economics*, 108(1):139–159.
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1):3–18.
- Barroso, P. and Detzel, A. (2021). Do limits to arbitrage explain the benefits of volatility-managed portfolios? *Journal of Financial Economics*, 140(3):744–767.
- Boguth, O., Carlson, M., Fisher, A., and 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.
- Bordalo, P., Gennaioli, N., La Porta, R., and Shleifer, A. (2024). Finance without exotic risk. Technical report, National Bureau of Economic Research.
- Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411–3447.
- Brown, N. C., Wei, K. D., and Wermers, R. (2014). Analyst recommendations, mutual fund herding, and overreaction in stock prices. *Management Science*, 60(1):1–20.
- Brunnermeier, M. and Nagel, S. (2004). Hedge funds and the technology bubble. *The journal of Finance*, 59(5):2013–2040.
- Cederburg, S., O’Doherty, M. S., Wang, F., and Yan, X. S. (2020). On the performance of volatility-managed portfolios. *Journal of financial Economics*, 138(1):95–117.
- Chen, A. Y. and Velikov, M. (2017). Accounting for the anomaly zoo: A trading cost perspective. *Available at SSRN*, 3073681.
- Chen, A. Y. and Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review*, *Forthcoming*.
- Clarke, C. (2022). The level, slope, and curve factor model for stocks. *Journal of Financial Economics*, 143(1):159–187.

- Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108.
- Cremers, M., Liu, Y., and Riley, T. B. (2022). Factor investing funds: Replicability of academic factors and after-cost performance. *Yuekun and Riley, Timothy Brandon, Factor Investing Funds: Replicability of Academic Factors and After-Cost Performance (November 1, 2022)*.
- De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990a). Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738.
- De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990b). Positive feedback investment strategies and destabilizing rational speculation. *the Journal of Finance*, 45(2):379–395.
- Edelen, R. M., Ince, O. S., and Kadlec, G. B. (2016). Institutional investors and stock return anomalies. *Journal of Financial Economics*, 119(3):472–488.
- Engelberg, J., McLean, R. D., and Pontiff, J. (2018). Anomalies and news. *The Journal of Finance*, 73(5):1971–2001.
- Engelberg, J., McLean, R. D., and Pontiff, J. (2020). Analysts and anomalies. *Journal of Accounting and Economics*, 69(1):101249.
- Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152.
- Hirshleifer, D., Hou, K., Teoh, S. H., and Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of accounting and economics*, 38:297–331.
- Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of financial studies*, 33(5):2019–2133.
- Israel, R. and Moskowitz, T. J. (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics*, 108(2):275–301.
- Ivkovich, Z., Kim, Y. H., and Muravyev, D. (2021). Causal effect of information costs on asset pricing anomalies. *Available at SSRN 3921785*.
- Jacobs, H. and Müller, S. (2020). Anomalies across the globe: Once public, no longer existent? *Journal of Financial Economics*, 135(1):213–230.
- Jensen, T. I., Kelly, B., and Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78(5):2465–2518.

- Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance*, pages 889–908.
- Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.
- Kong, D., Lin, C., Liu, S., and Tan, W. (2021). Whose money is smart? individual and institutional investors’ trades based on analyst recommendations. *Journal of Empirical Finance*, 62:234–251.
- Kozak, S., Nagel, S., and Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, 73(3):1183–1223.
- Lettau, M. and Pelger, M. (2020). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies*, 33(5):2274–2325.
- Lewellen, J. (2015). The cross-section of expected stock returns. *critical finance review* 4 (1): 1–44.
- McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.
- Memmel, C. (2003). Performance hypothesis testing with the sharpe ratio. *Available at SSRN 412588*.
- Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of financial economics*, 108(1):1–28.
- Piotroski, J. D. and So, E. C. (2012). Identifying expectation errors in value/glamour strategies: A fundamental analysis approach. *The Review of Financial Studies*, 25(9):2841–2875.
- Scholes, M. and Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of financial economics*, 5(3):309–327.
- Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of finance*, 52(1):35–55.
- Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of financial economics*, 104(2):288–302.
- Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.
- Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *The review of financial studies*, 30(4):1270–1315.

Table 1:**Overlap Stocks: Percent of MCAP**

Average percent of anomaly portfolio market capitalization accounted for by the N th percentile gross overlap stocks. For each anomaly portfolio, I calculate the percent of market capitalization accounted for by the N th percentile gross overlap stocks that are included in the portfolio for the month. I then take the average of the time-series for each anomaly. The statistics below report the minimum, 25th percentile, mean, 75th percentile, maximum, and standard deviation of the time-series means. ‘N Stocks’ reports the average total number of gross overlap portfolio stocks that are included in anomaly portfolios each month.

Panel A: Anomaly Portfolios							
	Percent of MCAP						
Percentile	Min	25th	Mean	75th	Max	StDev	N Stocks
99	0.14	1.14	3.21	5.02	16.74	2.78	16
97.5	0.28	2.88	6.85	10.07	23.82	5.25	34
95	0.84	5.92	12.20	18.13	37.15	8.36	60
90	3.07	12.05	21.57	29.87	56.07	12.41	105
80	9.65	26.71	38.07	48.06	77.50	15.45	176
Panel B: Anomaly Portfolio Legs							
	Percent of MCAP						
Percentile	Min	25th	Mean	75th	Max	StDev	N Stocks
Long Leg							
99	0.04	0.37	1.17	1.47	10.37	1.28	7
97.5	0.06	1.04	2.91	4.08	12.17	2.51	16
95	0.11	2.58	5.99	8.28	21.15	4.47	28
90	0.37	6.98	12.93	18.25	37.13	7.82	49
80	1.18	18.92	28.10	37.44	59.82	11.98	83
Short Leg							
99	0.42	2.58	5.54	7.77	16.26	3.64	8
97.5	0.68	5.51	11.21	15.74	30.65	6.81	18
95	1.86	10.63	18.89	25.69	45.95	10.33	32
90	5.53	20.22	30.99	41.61	66.13	14.38	56
80	13.73	36.39	49.16	62.70	86.06	16.70	93

Table 2:
Effects of Excluding Overlap Stocks

Means and medians of average returns, Sharpe ratios, CAPM alphas, CAPM alpha t -statistics, the average number of stocks in a portfolio, and the number of portfolios with alphas that are significant at the 5% level or better. Returns, alphas, and Sharpe ratios are annualized. ‘Unfiltered’ reports results for 164 long-short anomaly portfolios formed using all available stocks while ‘Filtered’ reports results for portfolios formed after excluding stocks above the N th percentile of net long/short overlap each month. Regressions use the full sample of anomaly portfolio returns (1926-2023 for the longest series and twenty years of monthly data for the shortest series). All results are robust to using data beginning in 1964 and decade-by-decade.

Panel A: Equal-Weighted											
Percentile	Returns		Sharpe Ratio		Alpha		t -statistic		N Stocks		N Sig
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	
Unfiltered:	4.83	4.86	0.41	0.38	5.70	5.84	3.78	3.91	414	439	110
Filtered:											
99	4.01	4.01	0.33	0.31	4.78	5.01	3.10	3.26	396	422	106
97.5	3.42	3.38	0.28	0.28	4.09	3.87	2.58	2.56	373	393	103
95	2.86	2.81	0.22	0.23	3.43	3.12	2.08	2.06	341	345	95
90	2.05	2.00	0.15	0.17	2.46	1.98	1.38	1.42	285	282	69
80	1.02	0.98	0.06	0.09	1.25	1.11	0.58	0.66	194	186	54
Change(%):											
99	-16.89	-17.37	-18.12	-17.23	-16.12	-14.28	-17.93	-16.62	-4.30	-3.82	-3.63
97.5	-29.19	-30.48	-31.79	-25.99	-28.22	-33.72	-31.88	-34.48	-9.70	-10.35	-6.36
95	-40.81	-42.06	-44.68	-38.48	-39.86	-46.55	-44.87	-47.41	-17.61	-21.39	-13.64
90	-57.49	-58.85	-63.20	-54.42	-56.75	-66.06	-63.59	-63.67	-31.14	-35.73	-37.27
80	-78.86	-79.76	-84.57	-76.27	-78.03	-80.93	-84.63	-83.25	-53.08	-57.63	-50.91
Panel B: Value-Weighted											
Percentile	Returns		Sharpe Ratio		Alpha		t -statistic		N Stocks		N Sig
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	
Unfiltered:	2.94	2.52	0.17	0.15	3.82	3.71	1.85	1.92	414	439	86
Filtered:											
99	2.43	2.12	0.14	0.13	3.23	3.14	1.54	1.49	396	422	74
97.5	2.11	1.81	0.12	0.11	2.84	2.82	1.34	1.36	373	393	65
95	1.75	1.52	0.10	0.09	2.40	2.20	1.10	1.07	341	345	55
90	1.25	0.92	0.06	0.05	1.79	1.50	0.77	0.76	285	282	31
80	0.45	0.38	0.02	0.02	0.90	0.52	0.32	0.24	194	186	26
Change(%):											
99	-17.32	-15.84	-19.01	-9.31	-15.39	-15.27	-16.75	-22.21	-4.30	-3.82	-13.95
97.5	-27.99	-28.19	-29.94	-24.32	-25.69	-24.01	-27.59	-28.88	-9.70	-10.35	-24.42
95	-40.53	-39.57	-43.80	-38.34	-37.13	-40.81	-40.29	-44.00	-17.61	-21.39	-36.05
90	-57.27	-63.48	-62.64	-65.60	-53.22	-59.46	-58.24	-60.23	-31.14	-35.73	-63.95
80	-84.67	-84.76	-90.31	-84.31	-76.38	-86.11	-82.69	-87.44	-53.08	-57.63	-69.77

Table 3:**Overlap Stocks and Anomaly Cumulative Returns**

Cumulative returns from 1926-2023 of a portfolio that earns a monthly return equal to the mean (median) monthly return of all 164 anomaly portfolios. ‘Stocks Excluded’ reports the average raw number as well as the percent of stocks excluded from the investable universe in a given month. On average, the investable universe includes 2,498 stocks per month.

Percentile	Cumulative Mean		Cumulative Median		Stocks Excluded	
	EW	VW	EW	VW	N	%
Unfiltered:						
	80.38	17.84	69.41	12.53	0	0
Filtered:						
99	39.91	11.09	28.57	7.49	50	2
97.5	23.52	7.59	16.82	6.13	125	5
95	14.45	5.21	9.98	4.35	250	10
90	6.95	2.87	4.32	2.21	500	20
80	2.00	0.72	1.72	0.20	999	40
Change(%):						
99	-50.35	-37.80	-58.84	-40.24	50	2
97.5	-70.73	-57.42	-75.76	-51.11	125	5
95	-82.02	-70.80	-85.63	-65.29	250	10
90	-91.35	-83.88	-93.78	-82.34	500	20
80	-97.52	-95.97	-97.52	-98.38	999	40

Table 4:**Overlap Stocks Across Anomaly Categories**

Average alphas (annualized) and t -statistics of anomaly portfolio categories formed using all available stocks and again after excluding stocks in the 90th percentile of net long/short overlap. Categories and category types for panels A and B are taken directly from Chen and Zimmermann (2021). Panel C reports anomaly categories based roughly on those reported in Jensen et al. (2023). ‘N Ports’ reports the number of anomaly portfolios out of 164 that are included in the category.

Panel A: Data Categories							
Category	CAPM Alpha			t -statistic			N Ports
	Unfiltered	Filtered	Change(%)	Unfiltered	Filtered	Change(%)	
13F	4.44	2.64	-40.54	1.82	1.10	-39.89	1
Accounting	6.24	2.04	-66.36	4.67	1.30	-72.21	86
Analyst	5.52	2.88	-47.22	2.59	1.48	-42.72	8
Event	5.76	3.24	-43.62	1.47	0.75	-48.56	1
Options	2.88	1.44	-48.31	1.39	0.76	-45.53	9
Other	1.32	-0.12	-105.41	0.80	0.05	-93.16	6
Price	4.08	2.16	-49.08	2.42	1.10	-54.7	41
Trading	12.00	8.04	-32.78	6.49	4.03	-37.92	12
Panel B: Economic Categories							
Category	CAPM Alpha			t -statistic			N Ports
	Unfiltered	Filtered	Change(%)	Unfiltered	Filtered	Change(%)	
R&D	9.12	5.28	-42.00	5.54	3.00	-45.78	2
Accruals	5.52	1.08	-81.19	5.07	0.77	-84.83	5
Asset Composition	4.92	2.64	-46.12	3.06	1.48	-51.57	4
Cash Flow Risk	-9.00	-6.84	-23.89	-4.31	-3.14	-27.11	1
Composite Accounting	2.28	1.08	-53.14	1.67	0.84	-49.88	2
Earnings Event	11.40	9.72	-15.46	10.6	8.52	-19.68	1
Earnings Forecast	6.60	3.24	-50.31	3.08	1.67	-45.69	4
Earnings Growth	6.12	4.56	-25.32	4.74	3.27	-30.95	3
External Financing	9.12	3.12	-65.44	7.01	1.93	-72.45	9
Info Proxy	-8.16	-7.56	-8.02	-2.76	-2.51	-9.14	1
Informed Trading	-0.24	-0.36	22.21	-0.04	0.05	-217.11	3
Investment	10.80	3.00	-71.84	8.89	2.17	-75.54	7
Investment Alt	6.12	1.08	-82.75	5.30	0.48	-90.96	10
Investment Growth	8.28	1.56	-81.00	8.74	1.50	-82.85	3
Lead Lag	5.64	4.08	-27.8	3.70	2.29	-38.08	7
Leverage	2.64	1.20	-55.59	1.03	-0.02	-102.06	4
Liquidity	10.20	6.96	-31.39	5.27	3.25	-38.37	9
Long Term Reversal	6.96	2.76	-60.43	3.49	1.30	-62.81	6
Momentum	8.88	5.28	-41.06	4.35	2.53	-41.81	9
Option Risk	3.12	1.32	-58.72	1.78	0.76	-57.42	4
Panel C: Broad Categories							
Category	CAPM Alpha			t -statistic			N Ports
	Unfiltered	Filtered	Change(%)	Unfiltered	Filtered	Change(%)	
Accounting - Other	2.28	1.08	-53.14%	1.67	0.84	-49.88	2
Accruals	5.52	1.08	-81.19	5.07	0.77	-84.83	5
Informed Trading	-0.24	-0.36	22.21	-0.04	0.05	-217.11	3
Institutions	4.80	2.52	-47.50	3.14	1.97	-37.26	7
Investment	8.04	1.80	-77.38	7.07	1.23	-82.68	20
Leverage	6.60	2.52	-61.12	4.67	1.37	-70.78	17
Liquidity	10.68	7.32	-31.77	5.57	3.50	-37.07	16
Moments	1.32	0.00	-100.98	0.92	0.03	-96.38	16
Momentum	8.88	5.28	-41.06	4.35	2.53	-41.81	9
Other	3.00	1.68	-44.90	1.95	0.87	-55.44	27
Profitability	3.72	1.32	-64.24	2.91	0.96	-67.04	19
Reversal	6.96	2.76	-60.43	3.49	1.30	-62.81	6
Value	7.56	2.88	-61.68	4.53	1.76	-61.27	17

Table 5:**Performance of Overlap Portfolios**

Annualized mean returns, Sharpe ratios, alphas, and number of stocks (monthly), for long-short N th percentile net overlap portfolios. Alphas are reported for the market model, Fama-French three-factor model, and the Carhart four-factor model (FF3 plus momentum). t -statistics are reported in parentheses. ‘N’ reports the average total number of stocks included in the portfolio each month (long plus short). Panel A reports results for equal-weighted long-short portfolios, panel B reports results for value-weighted long-short portfolios, and panel C reports results for 90th percentile excess long and short legs separately (both equal-weighted and value-weighted). All regressions use the full time series of data (1926-2023) but results are robust to using data beginning in 1964 and decade-by-decade. ***, **, * indicate significance at better than the 1%, 5%, and 10% levels, respectively.

Panel A: Equal-Weighted						
Percentile	Return	Sharpe	CAPM	FF3	CH4	N
99	32.99	1.13	38.60*** (14.39)	36.94*** (14.13)	31.37*** (12.12)	50
97.5	26.54	1.18	31.63*** (16.18)	30.16*** (16.14)	24.90*** (13.84)	124
95	21.66	1.12	26.05*** (15.51)	24.68*** (15.48)	19.83*** (13.11)	248
90	17.84	1.11	21.65*** (15.76)	20.41*** (15.82)	16.47*** (13.45)	495
80	13.81	1.06	16.85*** (15.03)	15.83*** (15.03)	12.23*** (12.51)	990
Panel B: Value-Weighted						
Percentile	Return	Sharpe	CAPM	FF3	CH4	N
99	18.62	0.57	24.25*** (7.89)	22.73*** (7.55)	16.44*** (5.50)	50
97.5	14.96	0.61	19.25*** (8.28)	17.35*** (7.84)	12.96*** (5.89)	124
95	12.74	0.65	16.34*** (8.96)	14.64*** (8.60)	10.47*** (6.29)	248
90	9.74	0.65	12.58*** (9.04)	11.21*** (8.75)	7.87*** (6.33)	495
80	7.32	0.65	9.42*** (9.05)	8.45*** (8.81)	5.76*** (6.24)	990
Panel C: Overlap Portfolio Legs (90th Percentile)						
Percentile	Return	Sharpe	CAPM	FF3	CH4	N
EW Long	20.59	0.89	12.50*** (9.17)	10.66*** (13.20)	10.45*** (12.60)	248
EW Short	2.75	0.09	-9.15*** (-6.69)	-9.75*** (-9.94)	-6.02*** (-6.80)	248
VW Long	14.46	0.79	7.74*** (8.05)	7.28*** (8.39)	5.53*** (6.41)	248
VW Short	4.73	0.20	-4.85*** (-5.61)	-3.93*** (-5.15)	-2.34*** (-3.09)	248

Table 6:**Overlap Stocks Decade-by-Decade**

Panel A reports CAPM alphas of unfiltered anomaly portfolios and of anomaly portfolios formed after excluding stocks above the 90th percentile of net long/short inclusion for the month. I require that an anomaly have at least 100 monthly observations in the decade to be included in my analysis for the decade. ‘N Ports’ reports the number of anomaly portfolios used in each decade. Panel B reports the CAPM alphas and t -statistics of the equal- and value-weighted 90th percentile net overlap portfolios. ***, **, * indicate significance at better than the 1%, 5%, and 10% levels, respectively.

Panel A: Average Annualized CAPM Alphas (%) Pre- and Post-Filtering (90th Percentile)									
	1926-1939	1940-1949	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2022
Unfiltered:									
EW	5.93	4.24	4.25	5.10	5.47	8.47	7.61	6.25	4.07
VW	3.86	3.24	4.19	3.82	4.59	5.85	4.01	5.18	2.04
Filtered:									
EW	4.29	2.27	2.26	2.05	2.43	3.79	3.35	2.89	1.61
VW	1.70	1.78	2.78	1.68	2.11	3.19	1.70	2.71	1.49
Change(%):									
EW	-27.71	-46.54	-46.75	-59.88	-55.65	-55.19	-55.97	-53.78	-60.45
VW	-55.99	-44.99	-33.62	-56.20	-53.99	-45.40	-57.76	-47.56	-26.81
N Ports	39	44	60	90	120	139	149	154	155
Panel B: CAPM Alphas (%) of 90th Percentile Common Stock Portfolios									
	1926-1939	1940-1949	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2022
EW	16.61*** (2.63)	13.73*** (4.15)	13.00*** (7.26)	18.19*** (7.58)	18.94*** (5.57)	31.00*** (11.03)	31.64*** (11.25)	28.57*** (5.42)	20.73*** (5.06)
VW	11.66** (2.20)	7.94*** (3.15)	8.41*** (3.49)	10.74*** (3.46)	14.50*** (3.18)	15.67*** (4.25)	12.26*** (3.08)	17.54*** (3.50)	6.79 (1.57)

Table 7:**Sharpe Ratio Improvement: Real-Time Mean-Variance Optimization**

Sharpe ratio improvements [deteriorations] for a mean-variance investor estimating optimal weights via equation (5). Panels A (equal-weighted) and B (value-weighted) report the number of unfiltered (filtered) blended anomaly portfolios that result in statistically significant improvements [deteriorations] in Sharpe ratios relative to the market portfolio. Results are reported for initial estimation windows of 60, 120, and 200 months. Panel C reports annualized returns, standard deviations, Sharpe ratios, Sharpe ratio improvements, and the statistical significance of Sharpe ratio improvements for the equal-weighted and value-weighted 90th percentile blended overlap portfolios. The statistical significance of the difference in Sharpe ratios is computed using the test defined by Jobson and Korkie (1981) and incorporates the correction proposed by Memmel (2003). Statistical significance is set at the 5% level.

Panel A: Equal-Weighted						
Initial Window	Unfiltered	99	97.5	95	90	80
60	78 [1]	65 [1]	55 [1]	42 [1]	21 [1]	10 [0]
120	77 [0]	67 [0]	60 [0]	45 [0]	21 [0]	8 [1]
200	73 [0]	66 [0]	55 [1]	42 [1]	29 [2]	13 [1]
Panel B: Value-Weighted						
Initial Window	Unfiltered	99	97.5	95	90	80
60	29 [0]	24 [1]	15 [1]	16 [0]	15 [0]	11 [0]
120	37 [0]	33 [0]	26 [0]	22 [0]	24 [2]	14 [1]
200	26 [0]	23 [0]	24 [0]	24 [0]	25 [1]	17 [1]
Panel C: Blended Overlap Portfolio						
	Equal-Weighted			Value-Weighted		
	60	120	200	60	120	200
R	14.38	13.92	14.55	9.26	8.87	9.36
σ	8.40	7.22	6.95	8.77	8.34	8.25
SR_{blend}	1.71	1.93	2.09	1.06	1.06	1.13
SR_{Mkt-RF}	0.43	0.43	0.43	0.43	0.43	0.43
ΔSR	1.23	1.44	1.55	0.57	0.57	0.59
	(9.25)	(10.32)	(11.53)	(4.81)	(4.88)	(5.21)

Table 8:**Sharpe Ratio Improvement: Overlap Portfolio**

Average annualized returns, standard deviations and Sharpe ratios of a portfolio that takes a weight of $w_j \in \{0.0, 0.1, 0.2, \dots, 1.0\}$ in the 90th percentile overlap portfolio and a weight of $1 - w_j$ in the market portfolio. $\Delta SR = SR_{blend} - SR_{Mkt-RF}$ reports the difference in Sharpe ratios between the blended portfolio and the market portfolio. The Jobson and Korkie (1981) test statistic of the difference between the Sharpe ratios is reported in parentheses and incorporates the correction proposed by Memmel (2003). Note that $w_j = 0.0$ corresponds to a portfolio with a 100% allocation to the excess market portfolio while $w_j = 1.0$ corresponds to a portfolio with a 100% allocation to the overlap portfolio. ***, **, and * indicate significance at better than the 1%, 5%, and 10% levels, respectively.

Panel A: Equal-Weighted											
	Mkt-RF	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Overlap
R	7.97	8.96	9.95	10.93	11.92	12.91	13.89	14.88	15.86	16.85	17.84
σ	18.58	15.89	13.36	11.09	9.28	8.25	8.28	9.37	11.21	13.50	16.04
SR	0.43	0.56	0.74	0.99	1.28	1.56	1.68	1.59	1.42	1.25	1.11
ΔSR	0.00	0.13***	0.32***	0.56***	0.85***	1.14***	1.25***	1.16***	0.99***	0.82***	0.68***
	(0.00)	(14.79)	(14.48)	(13.97)	(13.08)	(11.63)	(9.63)	(7.57)	(5.89)	(4.66)	(3.77)
Panel B: Value-Weighted											
	Mkt-RF	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Overlap
R	7.97	8.15	8.33	8.50	8.68	8.86	9.03	9.21	9.38	9.56	9.74
σ	18.58	16.12	13.81	11.74	10.07	9.03	8.83	9.52	10.95	12.86	15.07
SR	0.43	0.51	0.60	0.72	0.86	0.98	1.02	0.97	0.86	0.74	0.65
ΔSR	0.00	0.08***	0.17***	0.30***	0.43***	0.55***	0.59***	0.54***	0.43***	0.31*	0.22
	(0.00)	(8.70)	(8.43)	(8.01)	(7.37)	(6.42)	(5.18)	(3.88)	(2.76)	(1.89)	(1.25)

Table 9:**Analyst Return Forecasts and Anomaly Overlap**

Average analyst return forecasts (month t), average annualized realized returns (month $t + 1$), and forecast errors ('return forecast' - 'annual return') by decile of net anomaly inclusion. Returns and forecasts are multiplied by 100. ' t ' reports the t -statistic from a paired t-test of the difference in means between actual and forecasted returns, 'Analyst/Forecast' reports the average number of analysts per stock-level forecast, and 'N' reports the total number of median analyst forecasts per decile. '10-1' reports the difference in means and the t -statistic from a paired t-test of the differences in means between deciles 10 and 1. Note that the net long measure is the inverse of the net short measure such that the first decile of net long inclusion corresponds to the tenth decile of net short inclusion. Data runs from 1999-03 to 2022-12.

Decile	Annual Return	Return Forecast	Forecast Error	t	Analyst/Forecast	N
1 (short)	12.24	46.95	34.71	23.37	3.76	40,671
2	19.16	34.46	15.30	12.44	3.68	42,106
3	19.81	26.69	6.88	6.41	3.64	44,639
4	18.92	22.29	3.37	3.21	3.47	39,645
5	19.54	20.88	1.34	1.31	3.34	40,201
6	20.65	19.10	-1.55	-1.25	3.10	33,110
7	21.18	19.37	-1.81	-1.55	3.18	30,959
8	23.24	19.42	-3.82	-3.05	3.06	25,518
9	25.37	22.46	-2.91	-2.09	2.95	22,253
10 (long)	36.43	27.77	-8.66	-4.17	2.66	13,993
10-1	24.19 (9.60)	-19.18 (-15.17)	-43.37 (-15.56)			

Table 10:**Analyst Return Forecasts and Anomaly Overlap - Regression Results**

Results from a pooled OLS regression of forecasted returns on indicators for long leg/short leg common stock portfolio inclusion. Columns 2 and 3 control for the level of market returns and the interaction between the level of market returns and long/short leg inclusions. ***, **, * indicate significance at better than the 1%, 5%, and 10% levels, respectively. Data runs from 1999-03 to 2022-12.

	(1)	(2)	(3)
Intercept	0.241*** (126.243)	0.237*** (125.52)	0.240*** (125.25)
<i>Long_Leg</i>	0.041*** (4.77)	0.041*** (4.77)	0.039*** (4.46)
<i>Short_Leg</i>	0.233*** (44.28)	0.233*** (44.27)	0.234*** (44.27)
R^m		-0.074** (-2.01)	-0.051 (-1.26)
$Long \times Rm$			0.351** (1.94)
$Short \times Rm$			-0.313** (-2.79)
R^2	0.006	0.006	0.006
N	333,135	333,135	333,135

Table 11:**Earnings Announcement Day Returns and Anomaly Overlap**

Average earnings announcement day returns and CAPM alphas by decile of net anomaly inclusion. Alphas and returns are multiplied by 100. Alphas are computed using betas from a 65 (120) day estimation window, incorporating the Scholes and Williams (1977) correction for nonsynchronous data. ‘10-1’ reports the difference between deciles 10 and 1 along with the t -statistic from a paired t -test of the difference in means. ‘N’ reports the number of earnings announcement return observations per decile. Note that the net long measure is the inverse of the net short measure such that the first decile of net long inclusion corresponds to the tenth decile of net short inclusion. Data runs from 1971-07 to 2022-12.

Decile	Earnings Return	Earnings Alpha (65)	Earnings Alpha (120)	N
1 (short)	-0.101	-0.125	-0.139	72205
2	0.048	-0.002	0.013	70805
3	0.107	0.067	0.072	69783
4	0.203	0.118	0.158	66342
5	0.158	0.129	0.132	65627
6	0.220	0.189	0.196	61366
7	0.252	0.225	0.229	55647
8	0.357	0.280	0.326	59055
9	0.393	0.318	0.368	60702
10 (long)	0.717	0.666	0.691	60134
10-1	0.818	0.791	0.830	
	(-20.55)	(-16.84)	(-21.20)	

Table 12:**Earnings Announcement Day Returns and Anomaly Overlap - Regression Results**

Results from a pooled OLS regression of announcement day returns and CAPM alphas on indicators for long leg/short leg 90th percentile overlap portfolio inclusion. Alphas are computed using betas from a 65 (120) day estimation window, incorporating the Scholes and Williams (1977) correction for nonsynchronous data. Dependent variables are multiplied by 100. ***, **, * indicate significance at better than the 1%, 5%, and 10% levels, respectively.

	$R_{i,t}$	$\alpha_{i,t}(65)$	$\alpha_{i,t}(120)$
Intercept	0.211*** (24.71)	0.159*** (15.06)	0.179*** (21.50)
<i>Long_Leg</i>	0.511*** (19.32)	0.508*** (15.56)	0.512*** (19.93)
<i>Short_Leg</i>	-0.311*** (-12.88)	-0.283*** (-9.42)	-0.318*** (-13.43)
R^2	0.001	0.001	0.001
N	641,666	648,376	641,666

Stock	B/M Decile	Unf. Weight	Long Percentile	Short Percentile	Filt. Weight
A	1	-0.25	25	75	-0.50
B	1	-0.25	10	90	0
C	1	-0.25	2	98	0
D	1	-0.25	37	63	-0.50
...	...	...	...	...	...
E	10	0.25	90	10	0
F	10	0.25	92	8	0
G	10	0.25	50	50	0.50
H	10	0.25	72	28	0.50

Figure 1: Portfolio Construction Example

Illustrative example of how equal-weighted unfiltered and filtered anomaly portfolios are constructed. ‘Unf. Weight’ reports the weight in each stock required to form an equal-weighted value portfolio, ‘Long (Short) Percentile’ reports the net long (short) percentile of anomaly overlap, and ‘Filt. Weight’ reports the weight in each stock required to form the 90th percentile filtered value portfolio. The procedure is identical for value-weighted portfolios except that stocks are weighted relative to the total market capitalization of the set of stocks from which they can be selected.

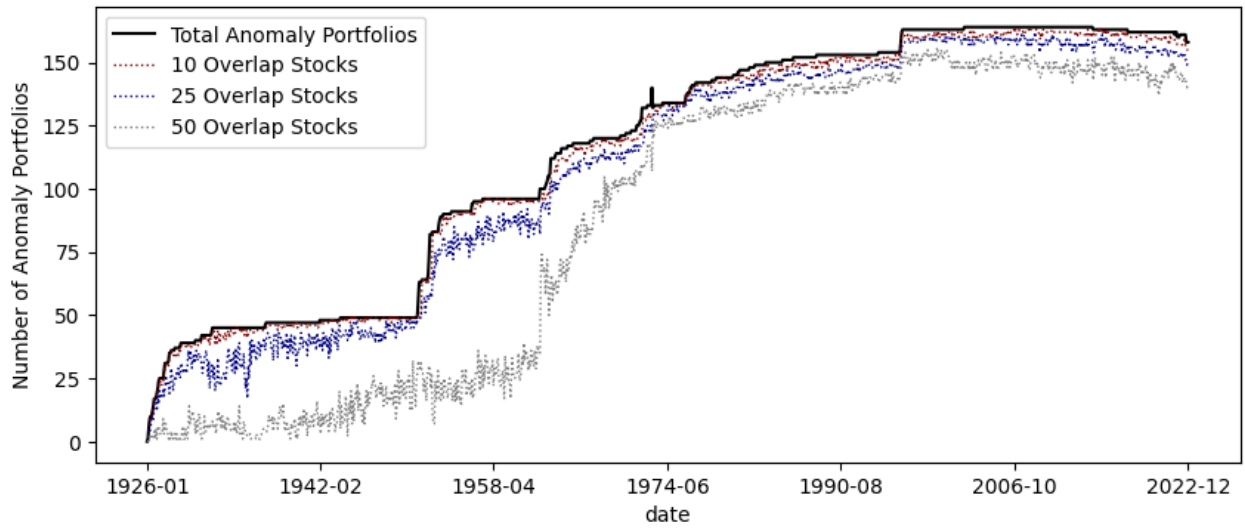


Figure 2: Number of Anomalies That Include N Overlap Stocks.

Total number of anomaly portfolios with valid returns each month and the number of anomaly portfolios that include at least N stocks that are above the 90th percentile of net portfolio overlap for the month. Note that not all anomalies have sufficient data to construct strictly continuous return series. For example, some anomalies have valid initial observations in 1973 but not in the following few months.

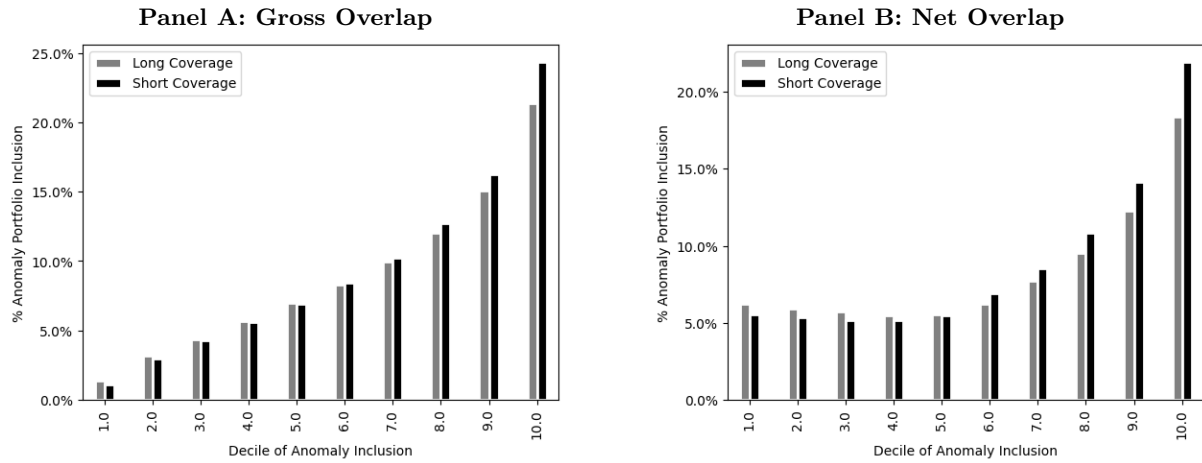


Figure 3: Decile of Anomaly Inclusion and Portfolio Overlap

The average percent of anomaly long (short) legs that a given stock in the N th decile of long (short) leg inclusion is included in each month. Panel A reports results for gross inclusion while Panel B reports results for net inclusion.

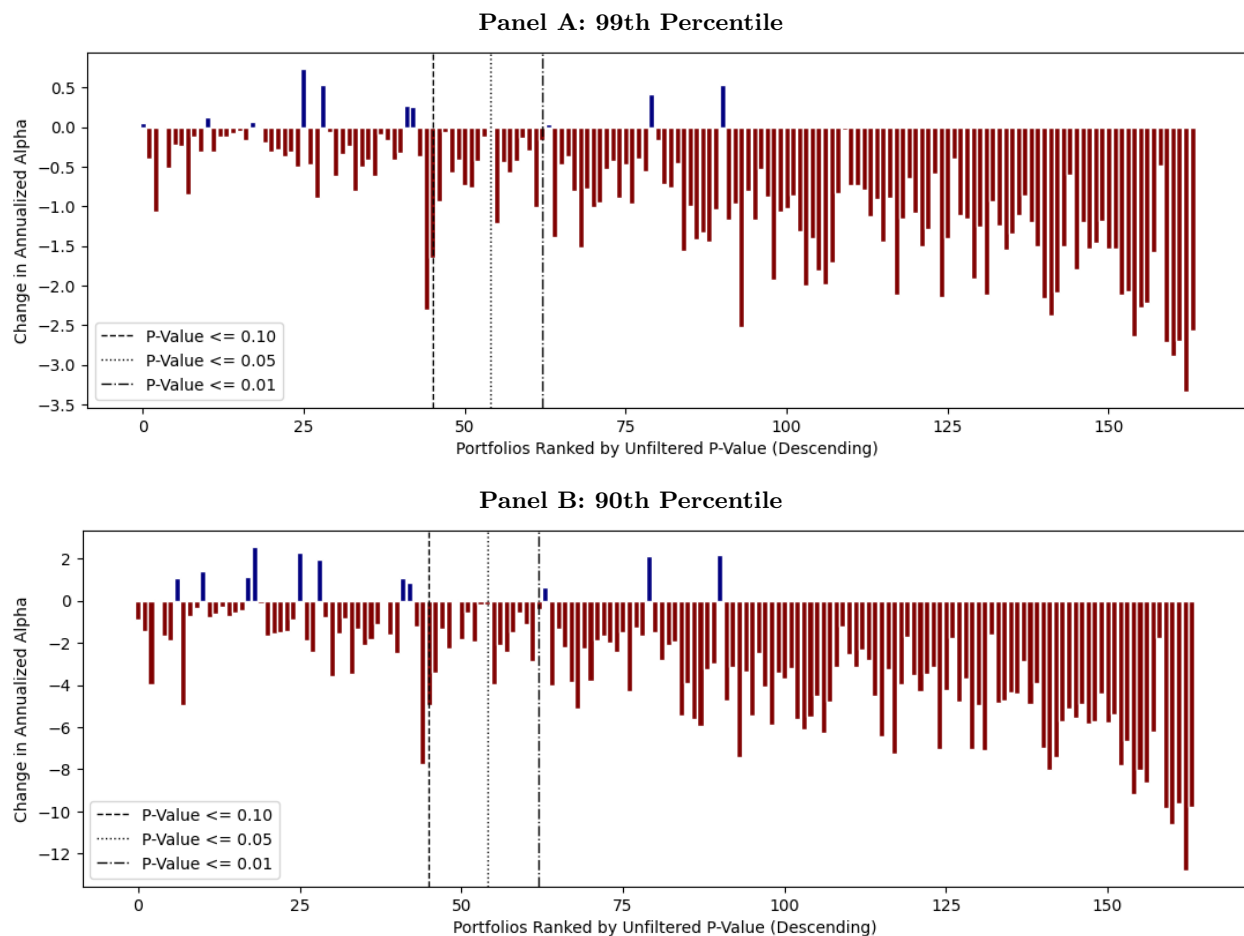


Figure 4: Anomaly-Level Changes in CAPM Alphas Post-Filtering

Changes in annualized alphas of value-weighted portfolios after excluding stocks in the N th percentile of net portfolio overlap. Changes are computed as filtered alphas less unfiltered alphas such that a positive value indicates an increase in alpha post-filtering, a value of zero indicates no change, and a negative value indicates a reduction. Portfolios are sorted based on the statistical significance of their unfiltered CAPM alpha estimates such that evidence of unfiltered abnormal returns is increasing along the x-axis. The vertical lines indicate cutoffs past which unfiltered portfolios have alphas that are significant at better than the 10%, 5%, and 1% levels, respectively. Note that changes are plotted in percent format (e.g., -2.0 corresponds to -2%). Results are qualitatively and quantitatively similar for equal-weighted portfolios.

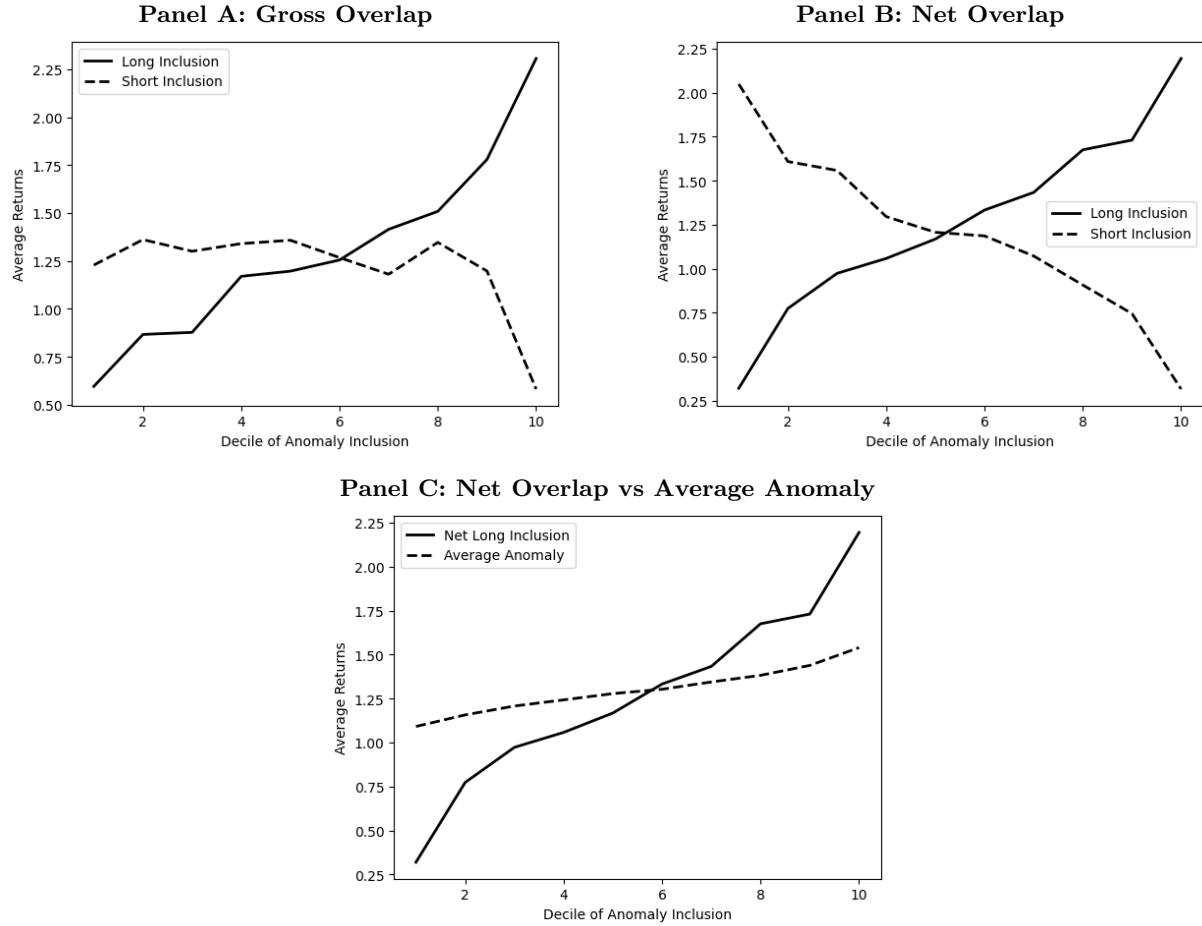


Figure 5: Average Returns as a Function of Anomaly Overlap

Average monthly returns as a function of anomaly portfolio inclusion. Stocks in the 1st decile of long (short) anomaly overlap are those that are present in very few anomaly long (short) legs for the month while those that are in the 10th are present in many. Panel A presents results for gross overlap, Panel B presents results for net overlap, and Panel C presents results for net long overlap versus the average return per decile of raw anomaly inclusion. Average returns per decile of raw anomaly inclusion report the average of the average returns per decile for each anomaly. For example, the average return for the first decile of raw anomaly inclusion is computed by taking the average of the average returns of firms in the first decile of size, in the first decile of B/M, the first decile of lag returns, etc. Note that the long and short net figures are the inverse of one another (such that, e.g., the 10th decile of net long inclusion is equal to the 1st decile of net short inclusion, and vice versa), with some small differences due to the handling of ties at decile cutoffs.

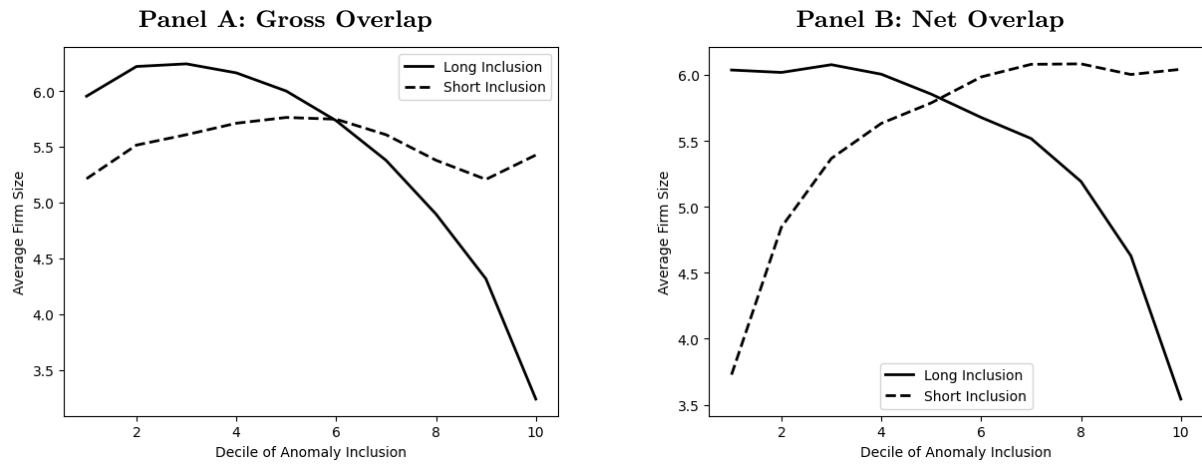


Figure 6: Average Firm Size as a Function of Anomaly Overlap

Average firm size as a function of anomaly portfolio overlap. Stocks in the 1st decile of long (short) anomaly overlap are those that are present in very few anomaly long (short) legs for the month while those that are in the 10th are present in many. Firm size is reported as the average firm size decile within a given decile of portfolio overlap. The figure reports the inverse of the firm size ranking scheme used to construct anomalies; here, decile 10 corresponds to the largest stocks while decile 1 corresponds to the smallest. Note that the long and short net figures are the inverse of one another, with some small differences due to the handling of ties at decile cutoffs.

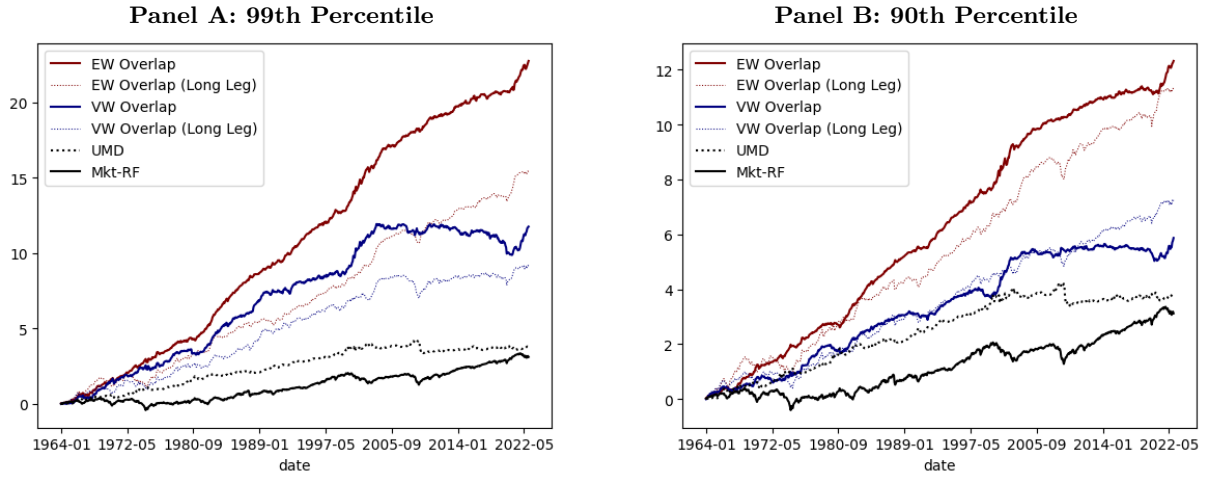


Figure 7: Cumulative Log Returns to Overlap Portfolios

Cumulative log returns to the long-short 99th (90th) percentile net overlap portfolios, the 99th (90th) percentile excess long legs, the Fama-French UMD factor, and the excess market portfolio. Returns are plotted beginning in 1964 for visual clarity. The return series are qualitatively similar (and quantitatively stronger) for the full sample.

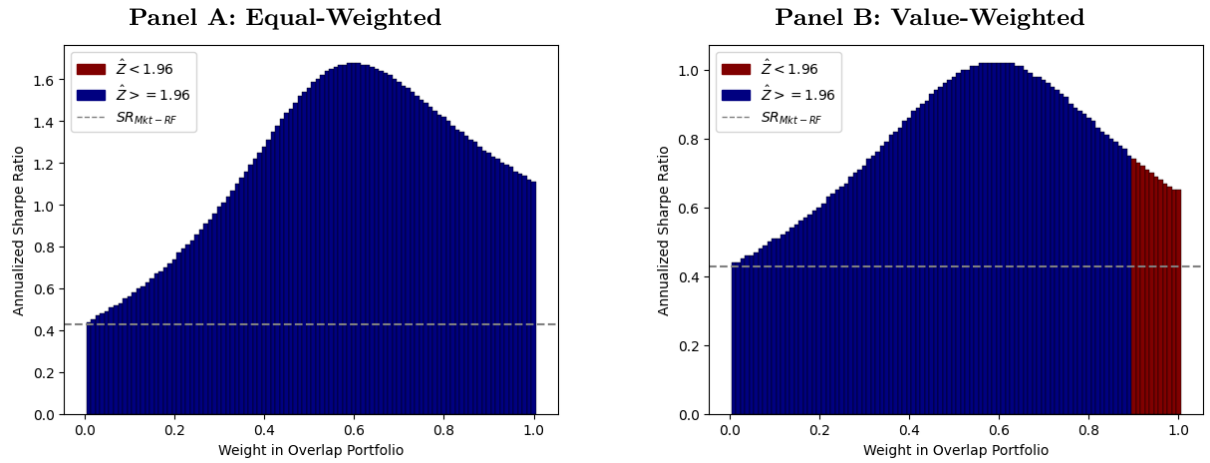


Figure 8: Sharpe Ratios of 90th Percentile Overlap Blended Portfolios

Annualized Sharpe ratios of portfolios that take a weight of $w_j \in \{0.01, 0.02, \dots, 1.00\}$ in the 90th percentile net overlap portfolio and a weight of $1 - w_j$ in the excess market portfolio. Columns are colored red for Sharpe ratios that are not significantly different (at the 5% level) from that of the market portfolio and blue for those that are significantly different. The horizontal dashed line corresponds to the Sharpe ratio of the excess market portfolio.

7 Appendix

Table A13:

Acronyms, original papers, and sample periods for the 164 characteristics used to form the anomaly portfolios in my main sample. ‘First Observation’ reports the first month in which the anomaly portfolio associated with the characteristic has a valid return. ‘Last Observation’ reports the last. ‘N Obs’ reports the total number of months in which the anomaly portfolio has a valid return. ‘Acronym’, ‘Authors’, and ‘Description’ are pulled directly from Chen and Zimmermann (2021).

Acronym	Authors	Description	First Observation	Last Observation	N Obs
AbnormalAccruals	Xie (2001)	Abnormal Accruals	1972-02-01	2022-12-01	611
Accruals	Sloan (1996)	Accruals	1952-05-01	2022-12-01	848
AdExp	Chan, Lakonishok and Sougiannis (2001)	Advertising Expense	1967-04-01	2022-12-01	669
AgeIPO	Ritter (1991)	IPO and age	1981-01-01	2022-12-01	479
AM	Fama and French (1992)	Total assets to market	1951-04-01	2022-12-01	861
AnalystValue	Frankel and Lee (1998)	Analyst Value	1973-01-01	2022-12-01	558
AnnouncementReturn	Chan, Jegadeesh and Lakonishok (1996)	Earnings announcement return	1971-11-01	2022-12-01	614
AOP	Frankel and Lee (1998)	Analyst Optimism	1973-01-01	2022-12-01	558
AssetGrowth	Cooper, Gulen and Schill (2008)	Asset growth	1952-04-01	2022-12-01	849
Beta	Fama and MacBeth (1973)	CAPM beta	1928-04-01	2022-12-01	1137
BetaFP	Frazzini and Pedersen (2014)	Frazzini-Pedersen Beta	1929-03-01	2022-12-01	1126
BetaLiquidityPS	Pastor and Stambaugh (2003)	Pastor-Stambaugh liquidity beta	1965-09-01	2022-12-01	688
BetaTailRisk	Kelly and Jiang (2014)	Tail risk beta	1932-02-01	2022-12-01	1091
betaVIX	Ang et al. (2006)	Systematic volatility	1986-03-01	2021-11-01	429
BidAskSpread	Amihud and Mendelsohn (1986)	Bid-ask spread	1926-03-01	2022-12-01	1162
BM	Stattman (1980)	Book to market, original (Stattman 1980)	1951-04-01	2022-12-01	861
BMdec	Fama and French (1992)	Book to market using December ME	1952-08-01	2022-12-01	845
BookLeverage	Fama and French (1992)	Book leverage (annual)	1951-05-01	2022-12-01	860

Continued on next page

Acronym	Authors	Description	First Observation	Last Observation	N Obs
BPEBM	Penman, Richardson and Tuna (2007)	Leverage component of BM	1962-08-01	2022-12-01	725
BrandInvest	Belo, Lin and Vitorino (2014)	Brand capital investment	1972-08-01	2022-12-01	581
Cash	Palazzo (2012)	Cash to assets	1971-12-01	2022-12-01	590
CashProd	Chandrashekar and Rao (2009)	Cash Productivity	1951-04-01	2022-12-01	861
CBOperProf	Ball et al. (2016)	Cash-based operating profitability	1962-08-01	2022-12-01	725
CF	Lakonishok, Shleifer, Vishny (1994)	Cash flow to market	1951-04-01	2022-12-01	861
cfp	Desai, Rajgopal, Venkatachalam (2004)	Operating Cash flows to price	1964-08-01	2022-12-01	701
ChAssetTurnover	Soliman (2008)	Change in Asset Turnover	1953-04-01	2022-12-01	837
ChEQ	Lockwood and Prombutr (2010)	Growth in book equity	1963-08-01	2022-12-01	713
ChInv	Thomas and Zhang (2002)	Inventory Growth	1952-04-01	2000-11-01	518
ChInvIA	Abarbanell and Bushee (1998)	Change in capital inv (ind adj)	1952-06-01	2022-12-01	847
ChNNCOA	Soliman (2008)	Change in Net Noncurrent Op Assets	1952-05-01	2022-12-01	848
ChNWC	Soliman (2008)	Change in Net Working Capital	1952-05-01	2022-12-01	848
ChTax	Thomas and Zhang (2011)	Change in Taxes	1963-04-01	2022-12-01	660
CompEquIss	Daniel and Titman (2006)	Composite equity issuance	1931-02-01	2022-12-01	1103
CompositeDebtIssuance	Lyandres, Sun and Zhang (2008)	Composite debt issuance	1956-06-01	2022-12-01	799
CoskewACX	Ang, Chen and Xing (2006)	Coskewness using daily returns	1962-09-01	2022-12-01	724
Coskewness	Harvey and Siddique (2000)	Coskewness	1927-08-01	2022-12-01	1145
CPVolSpread	Bali and Hovakimian (2009)	Call minus Put Vol	1996-03-01	2022-12-01	322
CustomerMomentum	Cohen and Frazzini (2008)	Customer momentum	1982-10-01	2022-12-01	455
dCPVolSpread	An, Ang, Bali, Cakici (2014)	Change in put vol minus change in call vol	1996-04-01	2022-12-01	321
DelBreadth	Chen, Hong and Stein (2002)	Breadth of ownership	1982-04-01	2022-12-01	470

Continued on next page

Acronym	Authors	Description	First Observation	Last Observation	N Obs
DelCOA	Richardson et al. (2005)	Change in current operating assets	1952-04-01	2022-12-01	849
DelCOL	Richardson et al. (2005)	Change in current operating liabilities	1952-05-01	2022-12-01	848
DelDRC	Prakash and Sinha (2013)	Deferred Revenue	1973-01-01	2022-12-01	252
DelEqu	Richardson et al. (2005)	Change in equity to assets	1963-08-01	2022-12-01	713
DelFINL	Richardson et al. (2005)	Change in financial liabilities	1961-08-01	2022-12-01	561
DelNetFin	Richardson et al. (2005)	Change in net financial assets	1953-01-01	2022-12-01	743
dNoa	Hirshleifer, Hou, Teoh, Zhang (2004)	change in net operating assets	1963-08-01	2022-12-01	713
DolVol	Brennan, Chordia, Subra (1998)	Past trading volume	1926-05-01	2022-12-01	1160
dVolCall	An, Ang, Bali, Cakici (2014)	Change in call vol	1996-04-01	2022-12-01	321
dVolPut	An, Ang, Bali, Cakici (2014)	Change in put vol	1996-04-01	2022-12-01	321
EarningsConsistency	Alwathainani (2009)	Earnings consistency	1953-08-01	2022-12-01	833
EarningsForecastDisparity	Da and Warachka (2011)	Long-vs-short EPS forecasts	1982-02-01	2022-12-01	491
EarningsStreak	Loh and Warachka (2012)	Earnings surprise streak	1985-03-01	2022-12-01	454
EarningsSurprise	Foster, Olsen and Shevlin (1984)	Earnings Surprise	1964-01-01	2022-12-01	708
EarnSupBig	Hou (2007)	Earnings surprise of big firms	1964-05-01	2020-11-01	501
EBM	Penman, Richardson and Tuna (2007)	Enterprise component of BM	1962-08-01	2022-12-01	725
EntMult	Loughran and Wellman (2011)	Enterprise Multiple	1951-05-01	2022-12-01	860
EP	Basu (1977)	Earnings-to-Price Ratio	1951-04-01	2022-12-01	861
EquityDuration	Dechow, Sloan and Soliman (2004)	Equity Duration	1963-08-01	2022-12-01	713
FEPS	Cen, Wei, and Zhang (2006)	Analyst earnings per share	1976-03-01	2022-12-01	562
fgr5yrLag	La Porta (1996)	Long-term EPS forecast	1973-01-01	2022-12-01	486
FirmAge	Barry and Brown (1984)	Firm age based on CRSP	1985-09-01	2022-12-01	422

Continued on next page

Acronym	Authors	Description	First Observation	Last Observation	N Obs
FirmAgeMom	Zhang (2006)	Firm Age - Momentum	1927-01-01	2022-12-01	1126
ForecastDispersion	Diether, Malloy and Scherbina (2002)	EPS Forecast Dispersion	1976-03-01	2022-12-01	562
FR	Franzoni and Marin (2006)	Pension Funding Status	1981-03-01	2022-12-01	382
Frontier	Nguyen and Swanson (2009)	Efficient frontier index	1963-08-01	2022-12-01	713
GP	Novy-Marx (2013)	gross profits / total assets	1951-05-01	2022-12-01	860
GrAdExp	Lou (2014)	Growth in advertising expenses	1970-08-01	2022-12-01	629
grcapx	Anderson and Garcia-Feijoo (2006)	Change in capex (two years)	1953-05-01	2022-12-01	836
grcapx3y	Anderson and Garcia-Feijoo (2006)	Change in capex (three years)	1954-05-01	2022-12-01	824
GrLTNOA	Fairfield, Whisenant and Yohn (2003)	Growth in long term operating assets	1952-04-01	2022-12-01	849
GrSaleToGrInv	Abarbanell and Bushee (1998)	Sales growth over inventory growth	1952-04-01	2022-12-01	849
GrSaleToGrOverhead	Abarbanell and Bushee (1998)	Sales growth over overhead growth	1952-05-01	2022-12-01	848
Herf	Hou and Robinson (2006)	Industry concentration (sales)	1959-03-01	2022-12-01	703
HerfAsset	Hou and Robinson (2006)	Industry concentration (assets)	1952-03-01	2022-12-01	840
HerfBE	Hou and Robinson (2006)	Industry concentration (equity)	1952-03-01	2022-12-01	719
High52	George and Hwang (2004)	52 week high	1926-04-01	2022-12-01	1161
hire	Bazdresch, Belo and Lin (2014)	Employment growth	1975-05-01	2022-06-01	280
IdioVol3F	Ang et al. (2006)	Idiosyncratic risk (3 factor)	1926-09-01	2022-12-01	1156
IdioVolAHT	Ali, Hwang, and Trombley (2003)	Idiosyncratic risk (AHT)	1926-12-01	2022-12-01	1153
Illiquidity	Amihud (2002)	Amihud's illiquidity	1927-02-01	2022-12-01	1151
IndMom	Grinblatt and Moskowitz (1999)	Industry Momentum	1926-07-01	2015-11-01	955
IndRetBig	Hou (2007)	Industry return of big firms	1926-06-01	2022-11-01	939
IntanBM	Daniel and Titman (2006)	Intangible return using BM	1967-08-01	2022-12-01	665

Continued on next page

Acronym	Authors	Description	First Observation	Last Observation	N Obs
IntanCFP	Daniel and Titman (2006)	Intangible return using CFtoP	1956-04-01	2022-12-01	801
IntanEP	Daniel and Titman (2006)	Intangible return using EP	1956-04-01	2022-12-01	801
IntanSP	Daniel and Titman (2006)	Intangible return using Sale2P	1956-04-01	2022-12-01	801
IntMom	Novy-Marx (2012)	Intermediate Momentum	1927-02-01	2022-12-01	1151
Investment	Titman, Wei and Xie (2004)	Investment to revenue	1953-04-01	2022-12-01	837
InvestPPEInv	Lyandres, Sun and Zhang (2008)	change in ppe and inv/assets	1952-04-01	2022-12-01	849
InvGrowth	Belo and Lin (2012)	Inventory Growth	1952-04-01	2022-12-01	849
Leverage	Bhandari (1988)	Market leverage	1951-05-01	2022-12-01	860
LRreversal	De Bondt and Thaler (1985)	Long-run reversal	1929-02-01	2022-12-01	1127
MaxRet	Bali, Cakici, and Whitelaw (2011)	Maximum return over month	1926-03-01	2022-12-01	1161
MeanRankRevGrowth	Lakonishok, Shleifer, Vishny (1994)	Revenue Growth Rank	1957-04-01	2022-12-01	789
Mom12m	Jegadeesh and Titman (1993)	Momentum (12 month)	1927-01-01	2022-12-01	1152
Mom12mOffSeason	Heston and Sadka (2008)	Momentum without the seasonal part	1926-08-01	2022-12-01	1157
Mom6m	Jegadeesh and Titman (1993)	Momentum (6 month)	1926-07-01	2022-12-01	1158
Mom6mJunk	Avramov et al (2007)	Junk Stock Momentum	1979-01-01	2017-04-01	460
MomOffSeason	Heston and Sadka (2008)	Off season long-term reversal	1928-01-01	2022-12-01	1140
MomOffSeason06YrPlus	Heston and Sadka (2008)	Off season reversal years 6 to 10	1932-01-01	2022-12-01	1092
MomOffSeason11YrPlus	Heston and Sadka (2008)	Off season reversal years 11 to 15	1937-01-01	2022-12-01	1032
MomOffSeason16YrPlus	Heston and Sadka (2008)	Off season reversal years 16 to 20	1944-01-01	2022-12-01	948
MomSeason	Heston and Sadka (2008)	Return seasonality years 2 to 5	1928-02-01	2022-12-01	1137
MomSeason06YrPlus	Heston and Sadka (2008)	Return seasonality years 6 to 10	1932-02-01	2022-12-01	1089
MomSeason11YrPlus	Heston and Sadka (2008)	Return seasonality years 11 to 15	1937-02-01	2022-12-01	1029
					Continued on next page

Acronym	Authors	Description	First Observation	Last Observation	N Obs
MomSeason16YrPlus	Heston and Sadka (2008)	Return seasonality years 16 to 20	1942-02-01	2022-12-01	969
MomSeasonShort	Heston and Sadka (2008)	Return seasonality last year	1927-02-01	2022-12-01	1147
MRreversal	De Bondt and Thaler (1985)	Medium-run reversal	1927-08-01	2022-12-01	1145
NetDebtFinance	Bradshaw, Richardson, Sloan (2006)	Net debt financing	1972-02-01	2022-04-01	364
NetDebtPrice	Penman, Richardson and Tuna (2007)	Net debt to price	1963-05-01	2022-12-01	716
NetEquityFinance	Bradshaw, Richardson, Sloan (2006)	Net equity financing	1972-02-01	2022-12-01	559
NetPayoutYield	Boudoukh et al. (2007)	Net Payout Yield	1953-03-01	2022-12-01	838
NOA	Hirshleifer et al. (2004)	Net Operating Assets	1963-03-01	2022-12-01	718
OperProf	Fama and French (2006)	operating profits / book equity	1963-05-01	2022-12-01	716
OperProfRD	Ball et al. (2016)	Operating profitability R&D adjusted	1963-02-01	2022-12-01	719
OPLeverage	Novy-Marx (2011)	Operating leverage	1951-05-01	2022-12-01	860
OptionVolume1	Johnson and So (2012)	Option to stock volume	1996-04-01	2022-12-01	321
OptionVolume2	Johnson and So (2012)	Option volume to average	1996-05-01	2022-12-01	320
OrderBacklog	Rajgopal, Shevlin, Venkatachalam (2003)	Order backlog	1971-02-01	2022-12-01	623
OrderBacklogChg	Baik and Ahn (2007)	Change in order backlog	1972-02-01	2022-12-01	611
OrgCap	Eisfeldt and Papanikolaou (2013)	Organizational capital	1951-08-01	2022-12-01	857
PayoutYield	Boudoukh et al. (2007)	Payout Yield	1953-03-01	2022-12-01	838
PctAcc	Hafzalla, Lundholm, Van Winkle (2011)	Percent Operating Accruals	1964-08-01	2022-12-01	701
PctTotAcc	Hafzalla, Lundholm, Van Winkle (2011)	Percent Total Accruals	1973-01-01	2022-12-01	417
PredictedFE	Frankel and Lee (1998)	Predicted Analyst forecast error	1973-01-01	2022-12-01	474
PriceDelayRsqr	Hou and Moskowitz (2005)	Price delay r square	1927-09-01	2022-09-01	1141
PriceDelaySlope	Hou and Moskowitz (2005)	Price delay coeff	1927-09-01	2022-09-01	1141

Continued on next page

Acronym	Authors	Description	First Observation	Last Observation	N Obs
PriceDelayTstat	Hou and Moskowitz (2005)	Price delay SE adjusted	1927-09-01	2022-09-01	1141
ProbInformedTrading	Easley, Hvidkjaer and O'Hara (2002)	Probability of Informed Trading	1994-02-01	2014-01-01	240
RDAbility	Cohen, Diether and Malloy (2013)	R&D ability	1971-08-01	2022-12-01	617
RDS	Landsman et al. (2011)	Real dirty surplus	1973-01-01	2022-12-01	552
realestate	Tuzel (2010)	Real estate holdings	1970-02-01	2022-12-01	635
RealizedVol	Ang et al. (2006)	Realized (Total) Volatility	1926-09-01	2022-12-01	1156
ResidualMomentum	Blitz, Huij and Martens (2011)	Momentum based on FF3 residuals	1930-07-01	2022-12-01	1110
retConglomerate	Cohen and Lou (2012)	Conglomerate return	1976-03-01	2022-12-01	545
ReturnSkew	Bali, Engle and Murray (2015)	Return skewness	1926-03-01	2022-12-01	1162
ReturnSkew3F	Bali, Engle and Murray (2015)	Idiosyncratic skewness (3F model)	1926-09-01	2022-12-01	1156
REV6	Chan, Jegadeesh and Lakonishok (1996)	Earnings forecast revisions	1976-10-01	2022-12-01	555
RevenueSurprise	Jegadeesh and Livnat (2006)	Revenue Surprise	1964-02-01	2022-12-01	707
RVolSpread	Bali and Hovakimian (2009)	Realized minus Implied Vol	1996-03-01	2022-12-01	322
roaq	Balakrishnan, Bartov and Faurel (2010)	Return on assets (qtrly)	1971-08-01	2022-12-01	617
RoE	Haugen and Baker (1996)	net income / book equity	1962-08-01	2022-12-01	725
sfe	Elgers, Lo and Pfeiffer (2001)	Earnings Forecast to price	1976-05-01	2022-12-01	560
ShareIss1Y	Pontiff and Woodgate (2008)	Share issuance (1 year)	1963-03-01	2022-12-01	413
ShareIss5Y	Daniel and Titman (2006)	Share issuance (5 year)	1959-02-01	2022-12-01	672
ShortInterest	Dechow et al. (2001)	Short Interest	1973-03-01	2022-12-01	598
Size	Banz (1981)	Size	1926-02-01	2022-12-01	1163
skew1	Xing, Zhang and Zhao (2010)	Volatility smirk near the money	1996-03-01	2022-12-01	322
SmileSlope	Yan (2011)	Put volatility minus call volatility	1996-03-01	2022-12-01	322

Continued on next page

Acronym	Authors	Description	First Observation	Last Observation	N Obs
SP	Barbee, Mukherji and Raines (1996)	Sales-to-price	1951-04-01	2022-12-01	861
std_turn	Chordia, Subra, Anshuman (2001)	Share turnover volatility	1928-02-01	2022-12-01	1139
tang	Hahn and Lee (2009)	Tangibility	1951-05-01	2022-12-01	860
Tax	Lev and Nissim (2004)	Taxable income to income	1951-04-01	2013-07-01	514
TotalAccruals	Richardson et al. (2005)	Total accruals	1952-05-01	2022-12-01	848
TrendFactor	Han, Zhou, Zhu (2016)	Trend Factor	1926-04-01	2022-12-01	1161
VarCF	Haugen and Baker (1996)	Cash-flow to price variance	1953-03-01	2022-12-01	838
VolMkt	Haugen and Baker (1996)	Volume to market equity	1926-12-01	2022-12-01	1153
VolSD	Chordia, Subra, Anshuman (2001)	Volume Variance	1928-02-01	2022-12-01	1139
VolumeTrend	Haugen and Baker (1996)	Volume Trend	1928-08-01	2022-12-01	1133
XFIN	Bradshaw, Richardson, Sloan (2006)	Net external financing	1972-02-01	2022-12-01	611
zerotrade	Liu (2006)	Days with zero trades	1926-08-01	2022-12-01	1157
zerotradeAlt1	Liu (2006)	Days with zero trades	1926-04-01	2022-12-01	1161
zerotradeAlt12	Liu (2006)	Days with zero trades	1927-03-01	2022-12-01	1150

Table A14:

Acronyms, original papers, and category assignments for the 164 anomaly portfolios in my main sample. ‘Acronym’, ‘Authors’, ‘Data’, and ‘Economic’ are pulled directly from Chen and Zimmermann (2021). ‘Broad’ approximates the categorization scheme used in Jensen et al. (2023).

Acronym	Authors	Category		
		Broad	Data	Economic
FR	Franzoni and Marin (2006)	Accounting - other	Accounting	Composite accounting
RDS	Landsman et al. (2011)	Accounting - other	Accounting	Composite accounting
AbnormalAccruals	Xie (2001)	Accruals	Accounting	Accruals
Accruals	Sloan (1996)	Accruals	Accounting	Accruals
OrderBacklogChg	Baik and Ahn (2007)	Accruals	Accounting	Accruals
PctAcc	Hafzalla, Lundholm, Van Winkle (2011)	Accruals	Accounting	Accruals
PctTotAcc	Hafzalla, Lundholm, Van Winkle (2011)	Accruals	Accounting	Accruals
dCPVolSpread	An, Ang, Bali, Cakici (2014)	Informed trading	Options	Informed trading
dVolCall	An, Ang, Bali, Cakici (2014)	Informed trading	Options	Informed trading
dVolPut	An, Ang, Bali, Cakici (2014)	Informed trading	Options	Informed trading
DelBreadth	Chen, Hong and Stein (2002)	Institutions	13f	Ownership
PredictedFE	Frankel and Lee (1998)	Institutions	Accounting	Earnings forecast
EarningsForecastDisparity	Da and Warachka (2011)	Institutions	Analyst	Earnings forecast
fgr5yrLag	La Porta (1996)	Institutions	Analyst	Earnings forecast
REV6	Chan, Jegadeesh and Lakonishok (1996)	Institutions	Analyst	Earnings forecast
FirmAge	Barry and Brown (1984)	Institutions	Other	Info proxy
AnnouncementReturn	Chan, Jegadeesh and Lakonishok (1996)	Institutions	Price	Earnings event
AssetGrowth	Cooper, Gulen and Schill (2008)	Investment	Accounting	Investment
ChEQ	Lockwood and Prombutr (2010)	Investment	Accounting	Investment
DelEqu	Richardson et al. (2005)	Investment	Accounting	Investment
dNoa	Hirshleifer, Hou, Teoh, Zhang (2004)	Investment	Accounting	Investment
GrLTNOA	Fairfield, Whisenant and Yohn (2003)	Investment	Accounting	Investment
Investment	Titman, Wei and Xie (2004)	Investment	Accounting	Investment
InvestPPEInv	Lyandres, Sun and Zhang (2008)	Investment	Accounting	Investment
BrandInvest	Belo, Lin and Vitorino (2014)	Investment	Accounting	Investment alt
ChInv	Thomas and Zhang (2002)	Investment	Accounting	Investment alt
ChNNCOA	Soliman (2008)	Investment	Accounting	Investment alt
ChNWC	Soliman (2008)	Investment	Accounting	Investment alt
DelCOA	Richardson et al. (2005)	Investment	Accounting	Investment alt
DelDRC	Prakash and Sinha (2013)	Investment	Accounting	Investment alt
DelNetFin	Richardson et al. (2005)	Investment	Accounting	Investment alt
GrAdExp	Lou (2014)	Investment	Accounting	Investment alt

Continued on next page

Acronym	Authors	Broad	Data	Economic
TotalAccruals	Richardson et al. (2005)	Investment	Accounting	Investment alt
ChInvIA	Abarbanell and Bushee (1998)	Investment	Accounting	Investment growth
grcapx	Anderson and Garcia-Feijoo (2006)	Investment	Accounting	Investment growth
grcapx3y	Anderson and Garcia-Feijoo (2006)	Investment	Accounting	Investment growth
hire	Bazdresch, Belo and Lin (2014)	Investment	Other	Investment alt
Cash	Palazzo (2012)	Leverage	Accounting	Asset composition
NOA	Hirshleifer et al. (2004)	Leverage	Accounting	Asset composition
realestate	Tuzel (2010)	Leverage	Accounting	Asset composition
tang	Hahn and Lee (2009)	Leverage	Accounting	Asset composition
CompEquIss	Daniel and Titman (2006)	Leverage	Accounting	External financing
CompositeDebtIssuance	Lyandres, Sun and Zhang (2008)	Leverage	Accounting	External financing
DelCOL	Richardson et al. (2005)	Leverage	Accounting	External financing
DelFINL	Richardson et al. (2005)	Leverage	Accounting	External financing
NetDebtFinance	Bradshaw, Richardson, Sloan (2006)	Leverage	Accounting	External financing
NetEquityFinance	Bradshaw, Richardson, Sloan (2006)	Leverage	Accounting	External financing
ShareIss1Y	Pontiff and Woodgate (2008)	Leverage	Accounting	External financing
ShareIss5Y	Daniel and Titman (2006)	Leverage	Accounting	External financing
XFIN	Bradshaw, Richardson, Sloan (2006)	Leverage	Accounting	External financing
BookLeverage	Fama and French (1992)	Leverage	Accounting	Leverage
BPEBM	Penman, Richardson and Tuna (2007)	Leverage	Accounting	Leverage
Leverage	Bhandari (1988)	Leverage	Accounting	Leverage
NetDebtPrice	Penman, Richardson and Tuna (2007)	Leverage	Accounting	Leverage
OptionVolume1	Johnson and So (2012)	Liquidity	Options	Volume
OptionVolume2	Johnson and So (2012)	Liquidity	Options	Volume
BetaLiquidityPS	Pastor and Stambaugh (2003)	Liquidity	Price	Liquidity
Size	Banz (1981)	Liquidity	Price	Size
BidAskSpread	Amihud and Mendelsohn (1986)	Liquidity	Trading	Liquidity
Illiquidity	Amihud (2002)	Liquidity	Trading	Liquidity
ProbInformedTrading	Easley, Hvidkjaer and O'Hara (2002)	Liquidity	Trading	Liquidity
std_turn	Chordia, Subra, Anshuman (2001)	Liquidity	Trading	Liquidity
VolSD	Chordia, Subra, Anshuman (2001)	Liquidity	Trading	Liquidity
zerotrade	Liu (2006)	Liquidity	Trading	Liquidity
zerotradeAlt1	Liu (2006)	Liquidity	Trading	Liquidity
zerotradeAlt12	Liu (2006)	Liquidity	Trading	Liquidity
ShortInterest	Dechow et al. (2001)	Liquidity	Trading	Short sale constraints
DolVol	Brennan, Chordia, Subra (1998)	Liquidity	Trading	Volume
VolMkt	Haugen and Baker (1996)	Liquidity	Trading	Volume
VolumeTrend	Haugen and Baker (1996)	Liquidity	Trading	Volume

Continued on next page

Acronym	Authors	Broad	Data	Economic
ForecastDispersion	Diether, Malloy and Scherbina (2002)	Moments	Analyst	Volatility
CPVolSpread	Bali and Hovakimian (2009)	Moments	Options	Optionrisk
RIVolSpread	Bali and Hovakimian (2009)	Moments	Options	Optionrisk
skew1	Xing, Zhang and Zhao (2010)	Moments	Options	Optionrisk
SmileSlope	Yan (2011)	Moments	Options	Optionrisk
Beta	Fama and MacBeth (1973)	Moments	Price	Risk
BetaTailRisk	Kelly and Jiang (2014)	Moments	Price	Risk
CoskewACX	Ang, Chen and Xing (2006)	Moments	Price	Risk
Coskewness	Harvey and Siddique (2000)	Moments	Price	Risk
ReturnSkew	Bali, Engle and Murray (2015)	Moments	Price	Risk
ReturnSkew3F	Bali, Engle and Murray (2015)	Moments	Price	Risk
betaVIX	Ang et al. (2006)	Moments	Price	Volatility
IdioVol3F	Ang et al. (2006)	Moments	Price	Volatility
IdioVolAHT	Ali, Hwang, and Trombley (2003)	Moments	Price	Volatility
MaxRet	Bali, Cakici, and Whitelaw (2011)	Moments	Price	Volatility
RealizedVol	Ang et al. (2006)	Moments	Price	Volatility
FirmAgeMom	Zhang (2006)	Momentum	Price	Momentum
High52	George and Hwang (2004)	Momentum	Price	Momentum
IndMom	Grinblatt and Moskowitz (1999)	Momentum	Price	Momentum
IntMom	Novy-Marx (2012)	Momentum	Price	Momentum
Mom12m	Jegadeesh and Titman (1993)	Momentum	Price	Momentum
Mom6m	Jegadeesh and Titman (1993)	Momentum	Price	Momentum
Mom6mJunk	Avramov et al (2007)	Momentum	Price	Momentum
ResidualMomentum	Blitz, Huij and Martens (2011)	Momentum	Price	Momentum
TrendFactor	Han, Zhou, Zhu (2016)	Momentum	Price	Momentum
EarnSupBig	Hou (2007)	Other	Accounting	Lead lag
ChTax	Thomas and Zhang (2011)	Other	Accounting	Other
OPLeverage	Novy-Marx (2011)	Other	Accounting	Other
RDAbility	Cohen, Diether and Malloy (2013)	Other	Accounting	Other
Tax	Lev and Nissim (2004)	Other	Accounting	Other
AOP	Frankel and Lee (1998)	Other	Analyst	Other
AgeIPO	Ritter (1991)	Other	Event	Other
CustomerMomentum	Cohen and Frazzini (2008)	Other	Other	Lead lag
Herf	Hou and Robinson (2006)	Other	Other	Other
HerfAsset	Hou and Robinson (2006)	Other	Other	Other
HerfBE	Hou and Robinson (2006)	Other	Other	Other
IndRetBig	Hou (2007)	Other	Price	Lead lag
PriceDelayRsq	Hou and Moskowitz (2005)	Other	Price	Lead lag

Continued on next page

Acronym	Authors	Broad	Data	Economic
PriceDelaySlope	Hou and Moskowitz (2005)	Other	Price	Lead lag
PriceDelayTstat	Hou and Moskowitz (2005)	Other	Price	Lead lag
retConglomerate	Cohen and Lou (2012)	Other	Price	Lead lag
BetaFP	Frazzini and Pedersen (2014)	Other	Price	Other
Mom12mOffSeason	Heston and Sadka (2008)	Other	Price	Other
MomOffSeason	Heston and Sadka (2008)	Other	Price	Other
MomOffSeason06YrPlus	Heston and Sadka (2008)	Other	Price	Other
MomOffSeason11YrPlus	Heston and Sadka (2008)	Other	Price	Other
MomOffSeason16YrPlus	Heston and Sadka (2008)	Other	Price	Other
MomSeason	Heston and Sadka (2008)	Other	Price	Other
MomSeason06YrPlus	Heston and Sadka (2008)	Other	Price	Other
MomSeason11YrPlus	Heston and Sadka (2008)	Other	Price	Other
MomSeason16YrPlus	Heston and Sadka (2008)	Other	Price	Other
MomSeasonShort	Heston and Sadka (2008)	Other	Price	Other
VarCF	Haugen and Baker (1996)	Profitability	Accounting	Cash flow risk
EarningsConsistency	Alwathainani (2009)	Profitability	Accounting	Earnings growth
EarningsStreak	Loh and Warachka (2012)	Profitability	Accounting	Earnings growth
EarningsSurprise	Foster, Olsen and Shevlin (1984)	Profitability	Accounting	Earnings growth
CBOperProf	Ball et al. (2016)	Profitability	Accounting	Profitability
GP	Novy-Marx (2013)	Profitability	Accounting	Profitability
InvGrowth	Belo and Lin (2012)	Profitability	Accounting	Profitability
OperProf	Fama and French (2006)	Profitability	Accounting	Profitability
OperProfRD	Ball et al. (2016)	Profitability	Accounting	Profitability
roaq	Balakrishnan, Bartov and Faurel (2010)	Profitability	Accounting	Profitability
RoE	Haugen and Baker (1996)	Profitability	Accounting	Profitability
CashProd	Chandrashekar and Rao (2009)	Profitability	Accounting	Profitability alt
ChAssetTurnover	Soliman (2008)	Profitability	Accounting	Sales growth
GrSaleToGrInv	Abarbanell and Bushee (1998)	Profitability	Accounting	Sales growth
GrSaleToGrOverhead	Abarbanell and Bushee (1998)	Profitability	Accounting	Sales growth
MeanRankRevGrowth	Lakonishok, Shleifer, Vishny (1994)	Profitability	Accounting	Sales growth
OrderBacklog	Rajgopal, Shevlin, Venkatachalam (2003)	Profitability	Accounting	Sales growth
RevenueSurprise	Jegadeesh and Livnat (2006)	Profitability	Accounting	Sales growth
FEPS	Cen, Wei, and Zhang (2006)	Profitability	Analyst	Profitability
IntanBM	Daniel and Titman (2006)	Reversal	Accounting	Long term reversal
IntanCFP	Daniel and Titman (2006)	Reversal	Accounting	Long term reversal
IntanEP	Daniel and Titman (2006)	Reversal	Accounting	Long term reversal
IntanSP	Daniel and Titman (2006)	Reversal	Accounting	Long term reversal
LRreversal	De Bondt and Thaler (1985)	Reversal	Price	Long term reversal

Continued on next page

Acronym	Authors	Broad	Data	Economic
MRreversal	De Bondt and Thaler (1985)	Reversal	Price	Long term reversal
AdExp	Chan, Lakonishok and Sougiannis (2001)	Value	Accounting	R&d
OrgCap	Eisfeldt and Papanikolaou (2013)	Value	Accounting	R&d
AM	Fama and French (1992)	Value	Accounting	Valuation
BM	Stattman (1980)	Value	Accounting	Valuation
BMdec	Fama and French (1992)	Value	Accounting	Valuation
CF	Lakonishok, Shleifer, Vishny (1994)	Value	Accounting	Valuation
cfp	Desai, Rajgopal, Venkatachalam (2004)	Value	Accounting	Valuation
EBM	Penman, Richardson and Tuna (2007)	Value	Accounting	Valuation
EntMult	Loughran and Wellman (2011)	Value	Accounting	Valuation
EP	Basu (1977)	Value	Accounting	Valuation
EquityDuration	Dechow, Sloan and Soliman (2004)	Value	Accounting	Valuation
Frontier	Nguyen and Swanson (2009)	Value	Accounting	Valuation
NetPayoutYield	Boudoukh et al. (2007)	Value	Accounting	Valuation
PayoutYield	Boudoukh et al. (2007)	Value	Accounting	Valuation
SP	Barbee, Mukherji and Raines (1996)	Value	Accounting	Valuation
AnalystValue	Frankel and Lee (1998)	Value	Analyst	Valuation
sfe	Elgers, Lo and Pfeiffer (2001)	Value	Analyst	Valuation