Interacting Anomalies

Karsten Müller and Simon N. M. Schmickler*

July 8, 2020

Abstract

An extensive literature studies interactions of stock market anomalies using double-sorted portfolios. But given hundreds of known candidate anomalies, examining selected interactions is subject to a data mining critique. In this paper, we conduct a comprehensive analysis of all possible double-sorted portfolios constructed from 102 underlying anomalies. We find hundreds of statistically significant anomaly interactions, even after accounting for multiple hypothesis testing. An out-of-sample trading strategy based on double-sorted portfolios performs on par with state-of-the-art machine learning strategies, suggesting that simple combinations of characteristics can capture a similar amount of variation in expected returns.

JEL Codes: G11; G12

*Karsten Müller is at Princeton University, Julis-Rabinowitz Center for Public Policy and Finance (e-mail: karstenm@princeton.edu). Simon Schmickler is at Princeton University, Department of Economics and Bendheim Center for Finance (e-mail: Simon.Schmickler@princeton.edu). For helpful comments and suggestions, we thank Markus Brunnermeier, Svetlana Bryzgalova, Matthias Büchner, Mehrshad Motahari, Wei Xiong, Motohiro Yogo, and Jason Zhang.

1 Introduction

An extensive literature documents hundreds of stock market anomalies, stock characteristics that predict returns beyond traditional risk factors. The sheer number of these characteristics gives rise to a data mining problem: testing potentially hundreds of hypotheses using conventional statistical significance thresholds likely produces many spurious signals (Harvey et al., 2015). In parallel, there is a sizable literature examining statistically significant *interactions* between selected individual anomalies, e.g. size and junk (Asness et al., 2018).[1] These interactions are important to understand because they can shed light on the economic forces behind the underlying anomalies. In addition, asset managers invest billions of dollars into strategies based on anomaly interactions.[2] However, with hundreds of candidate anomalies, there are tens of thousands of candidate anomaly combinations. This gives rise to an even greater data mining problem than that posed by individual anomalies.

We address this issue by taking an agnostic approach. The literature typically examines one specific anomaly interaction using double-sorted (DS) portfolios. Here, instead of analyzing individual DS portfolios in isolation, we conduct a comprehensive investigation of all possible DS strategies constructed from 102 underlying anomalies in a multiple hypothesis testing framework.

We find that, even after accounting for multiple hypothesis testing, hundreds of anomaly combinations generate excess returns beyond standard asset pricing factors and the factors corresponding to the underlying individual anomalies. We call these excess returns *interaction gains*. Using network graphs, we show that characteristics related to past returns (particularly short-term reversal) and limits to arbitrage (particularly size) are most likely to yield statistically significant interaction gains with other anomalies. Zooming in on the tails of the performance distribution, we find the top DS strategies generate equal-weighted (value-weighted) monthly average returns above 4% (3%) and Sharpe ratios above 2 (1.5). The best strategy combines short-term reversal and illiquidity, an interaction previously documented in Avramov et al. (2006). More broadly, the best-performing DS strategies have high turnover (e.g. short-term reversal) and high trading costs (e.g. illiquidity), and their returns are even higher in bad market states (high illiquidity, high volatility, low sentiment).

[1] Other examples include Ritter (1991), Lakonishok et al. (1994), Asness (1997), Daniel and Titman (1999), Hong et al. (2000), Lee and Swaminathan (2000), Bartov and Kim (2004), George and Hwang (2004), Jiang et al. (2005), Zhang (2006), Sadka (2006), Chan and Kot (2006), Avramov et al. (2006), Guo et al. (2006), Avramov et al. (2007), Fama and French (2008), Palmon et al. (2008), Hou et al. (2009), Fama and French (2012), Novy-Marx (2013), Asness et al. (2013), Anton and Polk (2014), Stambaugh et al. (2015), Zhu and Yung (2016), Lambert et al. (2016), Favilukis and Zhang (2019), Cho and Polk (2019), and Lou and Polk (2020).

[2] Examples of funds implementing such strategies include Dimensional's US Small Cap Value Portfolio and AQR's Small Cap Momentum Style Fund.

This suggests their stellar performance persists at least partially because they are difficult to arbitrage. Consistent with this interpretation, the best DS strategies have become less profitable since the early 2000s, when advances in financial technology and the introduction of decimalization reduced trading costs. When we focus on value-weighted returns and exclude micro caps, we also find a role for interactions of growth- or earnings-related anomalies with limits to arbitrage.

We also benchmark prominent DS strategies against the full distribution of possible portfolios. Six out of 24 selected strategies from the literature enter the top 5% of DS strategies. In other words, our approach uncovers a range of previously undocumented high-performing DS strategies. We also report which anomalies generate the highest interaction gains with major anomalies, such as size or value. This may shed light on the underlying economic forces behind these anomalies. Because space constraints prevent us from showing the full results of all trading strategies in this paper, we provide a visual tool to view them on a dedicated website at www.interactinganomalies.com. The website allows researchers to compare the performance of a range of trading strategies based on more than 10,000 anomaly combinations. It also provides the entire underlying input data.

Many DS strategies have similar average returns as state-of-the-art machine learning strategies that use similar samples and the same (and more) predictor variables (Bryzgalova et al., 2019; Chen et al., 2020; Freyberger et al., 2020; Gu et al., 2020b; Gu et al., 2020a; Kozak et al., 2020). This comparison, however, ignores that these machine learning strategies are constructed out-of-sample, whereas a trader could not have known ex-ante which sorting strategy to invest in. We thus consider DS-based strategies an investor could have implemented using only past data. Using the same input data and time period as in Gu et al. (2020b), we find that out-of-sample trading strategies that invest in the top backward-looking DS portfolios generate equal-weighted (value-weighted) monthly average returns of up to 4% (2.7%) or annualized Sharpe ratios of up to 2.7 (1.38). This is similar to the performance of state-of-the-art machine learning strategies: Gu et al. (2020b), for example, report equal-weighted (value-weighted) average returns of up to 3.3% (2.3%) and Sharpe ratios of up to 2.45 (1.35). The DS strategy remains profitable when we drop micro caps and only include anomalies after they were published in academic journals (Mclean and Pontiff, 2016). We also consider the role of trading costs and conclude they would likely have to be one order of magnitude higher than the estimates in Frazzini et al. (2018) to fully erode the profitability of these strategies.

The finding that DS-based strategies perform similarly to these machine learning strategies could suggest that these complex approaches largely gain their stellar performance by exploiting simple combinations of characteristics. In fact, we find a near-perfect overlap between the best-performing anomaly interactions and the characteristics the machine learning approaches in Gu et al. (2020b) and Avramov et al. (2020) pick as the most important variables for predicting returns. For the purpose of measuring risk premia, the similarity in performance also suggests that the mapping of characteristics into expected returns may be simpler than often assumed.

The similar performance of DS-based and machine learning strategies may, at first glance, seem puzzling. After all, machine learning models, such as random forests, nest DS strategies. We show, however, that modifying our DS approach to more closely resemble random forests leads to overfitting and decreased performance. In particular, our main results are based on *conditional* DS portfolios. Both conditional double-sorting and decision trees first split stocks according to a first anomaly (e.g. size). Conditional double-sorting then uses the same second anomaly within each portfolio for the second split (e.g. value). In contrast, decision trees can choose different anomalies (e.g. value and momentum) for the second split. When we modify the DS-based strategies to allow for this feature, we find drastically reduced performance. Picking the best long-only strategy for the long leg and the worst long-only strategy for the short leg–instead of choosing the best long-short strategy–reduces equal-weighted average returns from 4% to 1%. This suggests that requiring base-assets to perform in the long and the short leg increases the probability of selecting a true economic signal, which improves out-of-sample performance.

DS strategies also sidestep a potentially important source of look-ahead bias in the methods used to construct portfolios. A trader could not have trained many of the machine learning models available today in, say, 1990, because the computing power, software, and/or methods did not yet exist. In contrast, double-sorting stocks into portfolios is technically trivial and not subject to this critique. In addition, an investor would not have known which anomaly variables researchers would identify in the future. A basic version of our strategy, and the machine learning strategies cited above, are thus subject to look-ahead bias in variable selection. We address this problem by showing that our results are robust to excluding anomaly combinations before both of the underlying anomalies are published.

We argue that our work provides a useful new performance benchmark: if an investor could have generated abnormal returns of 4% or a Sharpe ratio of 2.7 by sorting stocks in a comparatively easy, transparent manner–with relatively limited concerns about look-ahead bias with respect to the methods used–black box machine learning strategies must perform substantially better to justify the loss in transparency. We interpret this finding as evidence that there is considerable room for progress in applying machine learning methods to asset pricing. Importantly, we do not claim that machine learning is not a useful tool for constructing portfolios; indeed, existing work shows that these methods hold great promise.

Our argument is that, when evaluating their efficacy, it is useful to keep in mind simple high-performance benchmarks that an investor could have implemented. Our results also suggest that one potentially fruitful way forward may be to impose more economic structure on machine learning models.

Related literature. This paper relates to a large body of work, cited above, that studies specific DS portfolios. We discuss these in more detail in Section 4.5. In contrast to this literature, we systematically analyze anomaly interactions in a multiple hypothesis testing framework and uncover hundreds of previously undocumented anomaly interactions. Our work is also related to the rapidly growing literature that constructs trading strategies using machine learning methods (Bryzgalova et al., 2019; Chen et al., 2020; Freyberger et al., 2020; Gu et al., 2020b; Gu et al., 2020a). These strategies generate impressive performance at the cost of largely being a "black box". Our results suggest a new performance benchmark for machine learning strategies.

Because decision trees nest conditional DS portfolios, this paper is also related to Bryzgalova et al. (2019), who use regression trees to build base assets from 10 anomalies for penalized mean-variance optimization. We directly analyze interaction strategies instead of using them as inputs to a mean-variance optimization step and also consider a much wider range of 102 anomalies in our analysis. Because DS portfolios are ubiquitous as test assets for asset pricing models and inputs for mean-variance optimization, the large set of interaction portfolios we make available should also be of independent interest for model testing or for constructing new strategies.

Our work is also related to Avramov et al. (2020), who show the profitability of machine learning strategies is substantially attenuated by trading costs. The results we present here suggest another, conceptually different limitation of current machine learning strategies: a much simpler way of interacting characteristics generates similar performance. Also related is Favilukis and Zhang (2019), who investigate the performance of interaction strategies that combine momentum with 36 other anomalies; our approach follows a similar spirit, but investigates over 10,000 anomaly combinations instead of 36. We further investigate the role of limits to arbitrage in explaining the performance of interactions with return-based anomalies and systematically compare the performance of DS-based versus machine learning strategies.

2 Empirical Strategy

2.1 Portfolio Construction

The central objects of this paper are conditional, 5×5 DS portfolios that we construct from 102 prominent anomaly variables.[3] 102 anomalies yield $102 \times (102 - 1) = 10,302$ possible, ordered combinations. Figure 1 illustrates why we form conditional DS portfolios. Panel A illustrates unconditional double-sorting. It shows a hypothetical scatter plot of the relative, cross-sectional ranks with respect to two anomaly variables. Forming unconditional DS portfolios means tracing out a 5×5, equally spaced grid and assigning stocks to their respective grid cell portfolio. We illustrate conditional DS portfolios in Panel B. The first portfolio sort is unconditional, as before, but the second sort splits stocks into quintiles within each quintile portfolio.

The hypothetical $firm \times time$ observations illustrate the problem with unconditional double sorts. When the ranks of two anomalies are correlated, the number of stocks assigned to a portfolio varies and–in extreme cases–can yield empty portfolios. For example, in many time periods, there are no large-cap stocks with low analyst coverage. This is not an issue in studies that focus on one particular interaction between two largely uncorrelated anomalies, which is why most papers listed in Table A1 construct unconditional DS portfolios. However, we consider all possible interactions between a large set of candidate anomalies, many of which are highly correlated. To avoid that the correlation of the underlying anomalies drives our performance comparison, we thus form conditional DS portfolios. By construction, all 5×5 conditional DS portfolios, contain an equal number of stocks. This makes the DS strategies more comparable and reduces the number of empty DS portfolios.[4] However, we consider unconditional double sorts in robustness exercises and find these yields highly similar results.

To keep the analysis tractable, we exclusively focus on the corner portfolios. HH, HL, LH, and LL denote corner portfolios depending on whether they are in the high or low quintile with respect to the first and second anomaly; i.e. a stock is assigned to the high-high (HH) portfolio if it has a high rank, greater than 0.8, in both sorts. We choose the sign of each anomaly to make them positive return predictive signals, so that HH portfolios take the advantageous side with regard to both underlying anomalies. We denote a long-short

[3] An exception are the seven anomalies that are indicator variables, such as IPO or sin stocks. Naturally, we cannot sort stocks into quintiles based on a binary variable. In these cases, we sort stocks into two rather than five portfolios.

[4] A small number of DS portfolios are still empty because the anomalies are incompatible. An example is IPO and long-term reversal: a stock cannot simultaneously be a recent IPO stock *and* have a non-missing 36-month long-term reversal return.

strategy by first naming the long and then the short portfolio. For example, HHLL is a long-short strategy that goes long HH and short LL.

We compute equal- and value-weighted returns for each portfolio and finance them with the risk-free rate. We run all exercises using all stocks and with a sample excluding microcaps, i.e. stocks in the bottom quintile of the cross-sectional market cap distribution, as in Gu et al. (2020b).[5] We also construct analogous single-sorted quintile portfolios to measure interaction gains of DS portfolios beyond their underlying anomalies, which we describe below.

Depending on the analysis, we report results for the long-only corner portfolios or long-short strategies constructed from the corner portfolios. We focus on two types of long-short DS strategies because they contain almost all long-short DS strategies in the literature. First, HHLL strategies take the advantageous side of both anomalies in the long and the short portfolio. An example is profitability and value as in Novy-Marx (2013). Second, HHHL, HHLH, HLLL, and LHLL strategies take the advantageous side of one anomaly but the same side of the other anomaly in the long and the short portfolio. These strategies typically take the same side of a limits-to-arbitrage characteristic which focuses the other anomaly on stocks that are harder to arbitrage. An example is momentum and volatility as in Zhang (2006).

2.2 Performance Evaluation

We compute performance metrics for the full panel and by decade. Further, we evaluate performance recursively to show performance over time and as an input to a DS-based out-of-sample trading strategy. We use standard performance metrics such as average returns, Sharpe ratios, CAPM- (Sharpe, 1964; Lintner, 1965), 3- (Fama and French, 1993), 5- (Fama and French, 2015), 6-factor alphas (Fama and French, 2018), and the respective information ratios.

Naturally, sorting on two characteristics gives the resulting portfolios exposure to both underlying anomalies. However, we are often interested in the alpha of a portfolio beyond these direct exposures. We thus define a new metric designed to evaluate the interaction between two anomalies, which we call *interaction gain*. In particular, we define the interaction gain of a trading strategy that combines anomalies i and j as the $\alpha_{i,j}$ with respect to a linear asset pricing model that includes standard risk factors and the factors corresponding to the underlying anomalies:

[5] When excluding microcaps (or imposing any other restrictions), we recompute portfolio breakpoints. Otherwise, the low size quintile portfolio would be empty by definition.

$$r_{i,j,t} = \alpha_{i,j} + \beta_{i,j} f_{i,t} + \gamma_{i,j} f_{j,t} + \delta'_{i,j} f_t + \epsilon_{i,j,t}. \tag{1}$$

$r_{i,j,t}$ is the return of the zero-investment DS strategy constructed from anomalies i and j in month t. f_t is a vector of standard risk factor returns. As a baseline, f_t contains the returns corresponding to a standard 6-factor model (Fama and French, 2018).[6] $f_{i,t}$ is the risk factor corresponding to anomaly i; the high minus low, long-short portfolio return constructed from quintile portfolios sorted on anomaly i.

2.3 Multiple Hypothesis Testing

We evaluate the performance of strategies based on more than 10,000 candidate combinations of anomalies. Therefore, hypotheses cannot be evaluated using conventional critical values for statistical significance. Harvey et al. (2015) discuss multiple appropriate methods. We choose the simplest and most conservative method, the Bonferroni correction. This testing strategy implies that we tolerate a high type-two error rate. As we show below, despite this conservative approach, we find hundreds of statistically significant interaction anomalies, more than we can discuss in detail here.

Concretely, instead of controlling the probability of a type I error for each hypothesis test individually, we control the probability of at least one type I error; i.e. the family-wise error rate. Using the Bonferroni correction, this means when testing whether any individual out of N strategies produces excess returns, we require a p-value less than α/N to reject the null of no excess returns at the α confidence level. For example, we consider $102 \times 101 \times 5 = 51{,}510$ long-short interaction strategies. Hence, when evaluating which of these strategies produce interaction gains, we require a p-value of $0.05/51,510$, i.e. a t-statistic of 4.9 to reject the null of no interaction gain at the 5% level.

3 Data

3.1 Stock Characteristics

We use the CRSP, Compustat, and IBES databases to construct a monthly stock-level panel dataset containing 102 prominent anomalies by applying the code provided by Green et al.

[6] One downside of this choice is that it implicitly treats DS strategies based on anomalies contained in the 6-factor model slightly differently. For example, it means we evaluate momentum-based strategies against a 5- instead of an 6-factor model, plus the underlying anomaly factors, because the additional momentum factor is redundant.

(2017). They restrict the sample to common stocks traded on the NYSE, AMEX or NASDAQ that have non-missing market and common equity values. To ensure that a trader could have used the data, they lag annual accounting data by 6 months and quarterly accounting data by 4 months.

We make minimal changes to their code. We extend the original sample from 1980-2014 to 1970-2017. We do not winsorize characteristics at the 1 and 99% level. This does not change quintile breakpoints, but it is necessary to construct value weights. The restrictions in Green et al. (2017) leave a few ETFs and closed-end funds in the sample. We exclude them by restricting the sample to share codes 10, 11, 12, and 18. This drops about 0.1% of observations. Common additional restrictions in the literature are to exclude firms with stock prices under $5, stocks of companies incorporated outside the US (share code 12), and REITs (share code 18). We do not exclude these firms in our main analysis to remain close to Green et al. (2017) and to allow a direct comparison to Gu et al. (2020b) in section 4.6, where we construct out-of-sample trading strategies. We consider these restrictions in robustness exercises in the appendix.

We flip the sign of negative return predictive signals to make the interpretation of the respective DS strategies uniform, so that HHLL is a DS strategy that takes the advantageous side in both anomalies. To detect the sign of an anomaly, we estimate univariate, cross-sectional, predictive regressions of returns on the cross-sectionally rank transformed characteristic in the sample from 1970 to the publication date of the anomaly. Table A11 explains all anomalies and reports the sign. This table is equivalent to the one in Green et al. (2017). They also provide a table that details the construction of each anomaly. We do not reproduce that table here because our data is identical. Where possible, we take anomaly publication dates from Mclean and Pontiff (2016), who study the impact of publication on anomaly performance. We take the remaining anomaly publication dates from Green et al. (2017).

We report results for two samples. As a baseline, we report the results for the full sample (1970-2017). When comparing the performance of an out-of-sample trading strategy to the machine learning approaches in Gu et al. (2020b), we adopt their testing sample (1987-2016). For the key results, we report equal- and value-weighted portfolios and present results that either include all stocks or exclude microcaps. For all exercises, we drop the performance of trading strategies that cover less than 2/3 of the time period we consider, because performance metrics computed from few return observations are prone to be driven by outliers and because we do not want performance comparisons to be driven by differences in sample periods.

Data on the risk-free rate and standard factor returns are from Kenneth French's website.

We also report results by market state, similar to Avramov et al. (2020). In particular, we show results for above and below median market volatility, illiquidity and sentiment. Market volatility is the standard deviation of daily market factor returns during the month. Market illiquidity is linearly detrended value-weighted log Amihud illiquidity of NYSE stocks. We use the market sentiment data provided by Baker and Wurgler (2007).

3.2 Comparison to the Literature

While we choose one specification for all strategies, the literature contains a wide variety of implementations. This is why our results may at times differ quantitatively from papers using the DS portfolios we cite.

A few key differences are worth highlighting. First, many papers construct unconditional DS portfolios, while we focus on conditional DS portfolios to make anomaly combinations more comparable and to avoid empty portfolios (see Section 2.1). Second, many papers rebalance annually at the end of June. We rebalance monthly to accommodate short-lived anomalies such as short-term reversal. Third, many papers drop subsets of stocks, such as microcaps or stocks with low dollar prices. We show results for equal- and value-weighted portfolios for all stocks in the Green et al. (2017) data as a baseline and for alternative samples in the appendix. Fourth, anomaly definitions can vary. For example, while it is standard to drop the preceding month from the momentum trading signal, some papers include it (Zhu and Yung, 2016). Fifth, different analyses use different time periods. We use data from 1970 to 2017. We selected this start date because the 1970s are the first decade for which high-quality data are easily accessible for a wide cross-section of firms.

4 Empirical Results

4.1 The Performance Distribution of Interaction Strategies

Figure 2 plots the performance distribution of all long-short interaction strategies, a total of around 50,000 trading strategies. We show density plots for average returns and Sharpe ratios, as well as interaction gains (the $\alpha_{i,j}$ from equation 1) and their t-statistics. Returns are equal-weighted; we show the value-weighted versions in appendix Figure A1. For reference, we tag the performance of five established interaction strategies based on well-known anomalies: momentum and value (*MV,* Asness, 1997), momentum and turnover (*MT,* Lewellen and Shanken, 2000), short-term reversal and illiquidity (*RI,* Avramov et al., 2006), size and value (*SV,* Fama and French, 2012), and size and junk (*SJ,* Asness et al., 2018). For average returns and Sharpe ratios, we also plot the distribution of the underlying individual anomalies in gray.[7] For t-statistics, we also plot a standard normal density function in gray.

Average returns and Sharpe ratios have a positively skewed distribution around a median of 0.4. 90% of strategies deliver positive returns and the top 5%, i.e. 2,500 strategies, generate large average returns greater than 1.3% and Sharpe ratios greater than 1. The distribution also has fat tails. There are DS strategies that generate monthly average returns greater than 4% and annualized Sharpe ratios greater than 2. An example of such a strategy is short-term reversal and illiquidity (*RI,* Avramov et al., 2006). Other prominent strategies perform well, but not exceptionally so, clustering around the 95th percentile of the performance distribution. The distribution of single-sort strategies is very similar but lacks the fat tails, which is to be expected: 102 strategies generate less extreme values than 50,000 strategies.

To investigate which anomaly combinations generate alpha beyond standard asset pricing factors and the underlying anomaly factors, Panels C and D show the distribution of interaction gains from equation 1. The distribution of interaction gains and their t-statistics is centered around zero and approximately normal. Panel D shows the distribution of the t-statistics of the interaction gain, where we winsorize t-statistics at 10 for readability. Importantly, we adjust for multiple hypothesis testing: instead of controlling the probability of a type I error for each hypothesis test individually, we control the probability of at least one type I error. Applying a Bonferroni correction, we reject the null at the 5% confidence level if $t > 4.9$. With this conservative approach and under the null, the probability that more than 0 out of 50,000 DS strategies generate statistically significant interaction gains is less than 5%. However, we find that 797 (351) equal-weighted (value-weighted) strategies generate statistically significant interaction gains. Prominent examples are momentum and turnover (*MT,* Lewellen and Shanken, 2000) and short-term reversal and illiquidity (*RI,* Avramov et al., 2006). A large share of these strategies' profitability thus comes from the interaction, not the underlying anomalies.

Tables 1 and 2 zoom in on the tails of the Sharpe ratio and the interaction gain distributions. In particular, we show the top 10 strategies for equal-weighted and value-weighted portfolios, as well as value-weighted strategies excluding micro caps. To omit redundant information, we only show the best-performing strategy per anomaly combination.[8] The top 10 equal-weighted strategies are almost exclusively based on short-term reversal, which either interacts with variables capturing limits to arbitrage or other return-based anomalies. The single best-performing strategy, short-term reversal and illiquidity, generates average

[7] Note that there is no single-sort analogue for interaction gains.

[8] For each anomaly combination, there are multiple possible trading strategies. These are HHLL, HHHL, HHLH, HLLL and LHLL. In addition, there are two possible orders of each anomaly combinations because we form conditional DS portfolios.

monthly returns of 4.3%. 2.8% of this is due to the interaction gain, and thus not explained by the underlying anomalies and standard factors. Naturally, while the performance is stellar, it requires high turnover (short-term reversal) at high costs (illiquidity).

Despite this caveat, short-term reversal and illiquidity remains at the top when using value-weighted returns in Panel B, and average returns remain high at 3.5%. However, value-weighting also pushes additional return-based and growth-based DS strategies into the top 10. Finally, in Panel C, we drop micro caps and focus on value-weighted portfolios. This further decreases average returns and Sharpe ratios but yields a similar picture to that in Panel B, except that earnings surprises become important. Overall, the best-performing DS strategies exploit return, growth, or earnings surprise anomalies on a subset of stocks that is difficult to arbitrage.

Table 2 shows the same results sorted by interaction gain instead of Sharpe ratio. Again, interactions between return-based signals and limits-to-arbitrage anomalies stand out, particularly when using equal-weighted returns. In addition, the dividend-price ratio strongly interacts with sales to receivables, leverage, return on capital, investment, organizational capital, and momentum.[9]

4.2 Which Anomalies Generate Interaction Gains?

Table 2 zoomed in on the tail of the performance distribution. The evidence presented there, however, is limited to a handful of the best-performing strategies. To get a more comprehensive picture, we next display every single one of the hundreds of anomaly combinations that are statistically significant (again, taking into account multiple hypothesis testing).

We use network graphs to visualize the results in a compact way. We treat individual anomalies as nodes, and two anomalies are connected by an edge if their combination generates a statistically significant interaction gain. Figure 3 plots the resulting anomaly network. The size of an anomaly's name is proportional to its degree of network centrality. The color of the links turns red in proportion to the sum of the absolute values of the interaction gains, i.e. the alphas from equation 1.[10]

The top graph shows the anomaly interaction network for equal-weighted portfolios. As in Table 2, we find that return-based signals and limits to arbitrage generate statistically

[9]In appendix Tables A2 to A6, we show alternative versions of this exercise. We report the best long-only and the best short-only portfolios. For robustness, we then show the same analysis as in Table 1 for unconditional instead of conditional DS strategies. We also repeat the analysis excluding firms with stock prices under $5, stocks of companies incorporated outside the US, and REITs, which some papers in the literature exclude. We also report the best (single-sorting) anomaly strategies for comparison.

[10]Note that, for each anomaly combination, there are multiple possible trading strategies, and thus interaction gains. These are HHLL, HHHL, HHLH, HLLL and LHLL. In addition, there are two possible orders of each anomaly combinations because we form conditional DS portfolios.

significant interaction gains with many other anomalies. The return-based signals are primarily short-term reversal, but also momentum, momentum change and maximum return. The limits-to-arbitrage anomalies are primarily size, but also illiquidity, bid-ask spreads, turnover and (idiosyncratic) volatility. There are also many additional, smaller clusters. For example, earnings-based anomalies (on the left) interact with a wide variety of anomalies, and many growth anomalies (at the top) interact with many fundamental anomalies. The bottom graph repeats the exercise using value-weighted portfolios. The results are similar, but the number of statistically significant interactions decreases, the importance of short-term reversal decreases, and the importance of size and expected EPS increases.

4.3 Performance by Market State

Why do certain DS strategies generate such enormous returns? One key reason is that we examine the tail of the distribution. In addition, in the cross-section of strategies, the best-performing portfolios, such as short-term reversal and illiquidity, generate high trading costs. Table 3 uses time variation in market illiquidity, sentiment, and volatility to provide additional evidence consistent with this idea. Specifically, we split the sample into above and below median values of market illiquidity, sentiment, and volatility, and report the top-performing strategies by market state.

For each state variable, short-term reversal and illiquidity is the top strategy during bad times, but not during good times. More generally, the best DS portfolios perform better during bad than good times. This is similar to the results in Avramov et al. (2020), who find that machine learning strategies outperform predominantly in bad times. At first glance, this may appear as a positive: both DS strategies and machine learning strategies act as a hedge that pays out in bad times. However, a simpler explanation is that these strategies are more costly to trade in bad states of the world. As one example, short-term reversal and illiquidity are likely always costly to trade, but particularly so when markets are illiquid. Overall, comparing DS strategies in the cross-section and in the time series suggests that one reason why the best DS portfolios achieve large returns is because these returns are difficult to realize. We will return to this point in section 4.7.

4.4 Performance Over Time

After exploiting the time dimension using different market states, we now address how the performance of DS portfolios has changed over time. As a first exercise, Figure 4 plots the distribution of DS strategy performance over time. Panels (a) to (d) plot the 1st, 5th, 25th, 50th, 75th, 95th, and 99th percentiles of ten-year rolling average returns and Sharpe ratios for equal- and value-weighted strategies. This shows that the performance of DS portfolios has decreased overall and its distribution has become more compressed over time. The top 1% of DS strategies in particular experienced a performance decline in the early 2000s, with a drop in average returns from 2.5 to 1.5%. This also compresses the performance spread from 3.5% to 2.5%. These findings are similar across all panels.

Next, in Table 4, we zoom in on the top strategies for each decade. Strategies based on short-term reversal dominate from the 1970s to 1990s, as shown in Panels A to C. However, together with the overall performance decline in the new millennium, they disappear from the top 5 starting in the 2000s, replaced by earnings-based anomalies which mostly interact with limits-to-arbitrage characteristics. The statistical significance of anomaly interactions also decreased over time; the t-statistics often fail to pass the Bonferroni-thresholds starting in the 1990s.

Figure 5 shows this drop in statistical significance more comprehensively. We plot the time series of the fraction of strategies with statistically significant interaction gains in a 10-year rolling sample at the Bonferroni-corrected 1, 5 and 10% level. The left panel shows that in the 1990s, 0.5%, i.e. about 250 strategies, generate statistically significant interaction gains at the Bonferroni-corrected 1% level. Then, starting in the early 2000s, this fraction declines by an order of magnitude. The right panel shows that the pattern is qualitatively the same for value-weighted strategies.

4.5 Comparison With Known Anomaly Interactions

To get a sense of how our findings compare to the existing literature, we report the performance of selected anomaly combinations identified in well-known papers in Table 5. Starting with the list of selected anomaly interaction papers in appendix Table A1, we restrict our attention to anomaly combinations published in the top three finance journals, exclude analyst-based anomalies because they have the lowest coverage, and drop duplicates. This leaves us with 24 anomaly combinations, which we divide into combinations based on momentum, reversal, liquidity, and value.

Overall, previously-known DS portfolios perform well in our sample: all 24 strategies generate positive returns. However, while many strategies generate interaction gains that are statistically significant at conventional levels, many do not pass the higher, multiple testing-adjusted cutoff. Taking into account multiple hypothesis testing, we can only reject the null of no interaction gain alpha for momentum and turnover, the three short-term reversal based strategies, and asset growth and size.

Figure 6 compares the same set of prominent anomaly combinations to the full distribution of DS strategies over time. For each prominent anomaly combination, we show the time series of its relative rank in terms of its recursive Sharpe ratio. Panel (a) shows that the momentum-based strategies' ranks fall over time, largely driven by momentum crashes (Daniel and Moskowitz, 2016). The momentum and size combinations perform particularly poorly. At first glance, this seems to contradict the literature documenting that momentum works best in small caps. The reason for this apparent discrepancy is that momentum works best among the second, not the first, size quintile (Hong et al., 2000).

As before, Panel (b) shows that strategies based on short-term reversal rank at the very top throughout the entire sample. Panel (c) shows considerable heterogeneity for liquidity-based DS strategies. Two striking combinations are surprise earnings and illiquidity, and asset growth and size, which show marked improvements in performance over time. Finally, Panel (d) shows that value-based strategies deliver performance around the 90th percentile with little variation across the different strategies and over time. Overall, some DS strategies identified in the literature stand out among the around 50,000 possible DS strategies. However, many are not exceptional and at times considerably underperform compared to the universe of possible strategies.

Many papers examine DS portfolios to shed light on the economic mechanisms driving the underlying anomalies. For example, Hong et al. (2000), Zhang (2006), and Hou et al. (2009) combine momentum with analyst coverage or turnover to test behavioral theories that momentum is the result of slow information diffusion, investor underreaction, and limited attention. In this spirit, we use DS strategies to investigate the best "interactors" of major anomalies. For size, value, profitability, investment, and momentum, we separately report the other anomalies with which they generate the highest interaction gains in appendix table A7. For example, if momentum is generated by limited attention, momentum should generate large interaction gains with anomalies that capture limited attention. Indeed, size–one proxy for attention–is the third best anomaly to combine with momentum.

4.6 Out-Of-Sample Trading Strategy

In the previous sections, we document that many DS strategies generate large profits. For example, combining short-term reversal and illiquidity yields average returns of 4.3%. This would suggest a better performance than state-of-the-art machine learning strategies, which use the same and more predictor variables as inputs (Bryzgalova et al., 2019; Chen et al., 2020; Freyberger et al., 2020; Gu et al., 2020b; Gu et al., 2020a; Kozak et al., 2020). This comparison, however, ignores that these machine learning strategies are constructed out-of-sample, whereas an investor could not have known ex-ante which DS strategy to invest in. If a trader used DS portfolios, what performance outcomes could they attain? In this section, we address this question by constructing out-of-sample trading strategies.

At any point in time, an investor could have constructed all candidate DS strategies, but they would only have been able to evaluate them using past data. The trader would then have invested in strategies that did best in the past. We let the trader invest either into the one strategy that did best in the past, or into the top decile of strategies.[11] Note that the investor does not know which side of an unpublished anomaly is advantageous. Hence, for each anomaly combination, the trader chooses from all possible, i.e. $4 \times 3 = 12$, long-short corner portfolio combinations.

While the machine learning strategies mentioned above are constructed out-of-sample, they may still contain two subtle sources of look-ahead bias related to variable selection and methods. First, an investor would not know which anomaly variables academics would discover and publish in the future.[12] This critique applies both to the strategies discussed so far in this paper and the machine learning strategies cited above. To make our results comparable, we show results for a strategy that ignores this pre-publication bias. However, we also report post-publication results, where we restrict the investor's information set to published signals. Second, it would not have been feasible to implement many machine learning methods in, say, 1990. While a trader could have known about neural networks, techniques like Dropout (Srivastava et al., 2014) and Batch Normalization (Ioffe and Szegedy, 2015) had not yet been invented, the necessary computing resources were not developed, and there were no readily available software implementations.[13] Hence, machine learning strategies also contain look-ahead bias on a methodological level. This issue does not apply to DS portfolios, because sorting stocks is trivial. Out-of-sample DS strategies that only make use of anomaly returns after their publication are thus less likely to be affected by these biases.

We construct strategies to enable a direct comparison with Gu et al. (2020b). We use a subset of their predictor variables[14], the same time period from 1987 to 2016, the same universe of stocks, the same rebalancing frequency, and the same portfolio weighting. They find that neural nets perform best, with equal-weighted (value-weighted) average returns of

[11] We choose these two versions because we believe they are the most standard options. For robustness, we show results for all possible choices in appendix Figure A2.

[12] In fact, because only significant results are published, a trader could infer that the past relationship between the signal and returns will likely persist.

[13] As one example, the popular machine learning library Tensorflow was first released in November 2015 (Martin Abadi et al., 2015).

[14] Like this paper, Gu et al. (2020b) rely on data from Green et al. (2017). However, Gu et al. (2020b) drop eight anomalies; for this exercise, we thus also exclude them. Gu et al. (2020b) also add 2-digit SIC code indicators and macro variables, which are not part of our analysis. Hence, the information set used by our strategy is a strict subset.

up to 3.3% (2.3%) and Sharpe ratios of up to 2.45 (1.35).

Table 6 reports average returns and Sharpe ratios, as well as 3-, 5-, and 6-factor alphas for equal- and value-weighted implementations of our DS-based trading strategy. Panel A reports results for the extreme version of the strategy, namely investing into the one long-short strategy with the highest Sharpe ratio in the past. This strategy generates very large average returns of 4%. As shown above in Table 4 and Figure 6, the underlying anomaly combinations are based on short-term reversal, which consistently yields the highest Sharpe ratio since the beginning of the sample; Table 1 already showed that these strategies generate average returns of about 4%.

This finding implies that simply buying the backward-looking best-performing DS strategy generates higher average returns than the best machine learning strategy in Gu et al. (2020b). Value-weighting does not change this conclusion: we find value-weighted average returns of 2.7%, compared to 2.3%. Our results are also not explained by standard factor loadings. The simple DS strategy does, however, generate a lower Sharpe ratio than machine learning strategies, around 2 compared to 2.45. This is an obvious consequence of investing only in a single DS strategy, which is bound to generate volatile returns. Panel B considers a natural fix by investing into the top decile of strategies. This drastically reduces average returns to 1%, but lowers volatility even more, increasing the Sharpe ratio to 2.7 for equal-weighted portfolios.

Why would simple DS strategies perform on par with machine learning strategies that use the same inputs? After all, DS strategies are in fact nested by decision trees and random forests (e.g. Gu et al., 2020b; Bryzgalova et al., 2019). It would be natural to expect that a random forest should strictly dominate the DS strategies it nests. We investigate this question in panels C and D of Table 6. A major difference between the conditional DS portfolios we consider and a decision tree is that the latter can choose different predictors to split observations at nodes of the same depth. Conditional double-sorting and decision trees first split stocks according to a first anomaly (e.g. size). But while conditional double-sorting then uses the same second anomaly within each portfolio for the second split (e.g. value), decision trees can choose different anomalies (e.g. value and momentum). As such, trees add flexibility at the cost of economic structure: they can exploit asymmetries in the relationship between signals and returns at the risk of selecting spurious predictors.

In panel C and D, we modify the DS-based strategy to allow for this feature of decision trees. Instead of picking the best long-short strategy, we pick the best long-only portfolio for the long leg and the worst long-only portfolio for the short leg. This drastically reduces performance. Average returns drop from 4% in panel A to 1% in panel C and the Sharpe ratio drops from 2.7 in panel B to 1.3 in panel D. This suggests that increasing flexibility by modifying the DS-based strategy to more closely resemble strategies based on decision trees leads to overfitting. Of course, a spurious signal is much more likely to generate outstanding returns in one leg than in both legs. True economic signals, however, should be more likely to work in the long and the short portfolio. Because DS-based strategies, in contrast to trees, require base-assets to perform in the long *and* the short leg, this approach makes selecting spurious predictors less likely.

Figure 7 illustrates our findings graphically. We plot the cumulative returns of the equal-weighted out-of-sample trading strategies in red. The black and gray lines report cumulative log returns on the long and the short leg, each financed by the risk-free rate. Panel (a) shows the returns of investing in the single best DS strategy. The returns are exceptional but decrease markedly in the early 2000s, a finding similar to Gu et al. (2018). Panel (b) shows the clearly inferior performance once we change to long-only base assets. The cumulative returns of the long portfolio decrease by a factor of 2.5 and the short portfolio now has negative returns. Panels (c) and (d) make the same comparison for the top decile strategy. Panel (c) illustrates the low volatility of the strategy. Panel (d) shows that changing to long-only base-assets does not decrease returns, but increases volatility.

Up to now, we have ignored the two types of look-ahead bias identified above: a trader would not have been able to implement sophisticated machine learning methods before their invention and could not have known which anomalies academics would publish. While the DS strategies presented above do not contain look-ahead bias with respect to methods, they are subject to the critique regarding variable selection. Table 7 reports results for strategies an investor could have implemented by excluding unpublished anomalies: the trading strategy now only uses DS strategies once both underlying anomalies are published. This exercise yields similar average returns, but somewhat lower Sharpe ratios. The top decile strategy's Sharpe ratio drops from 2.7 to 2.25, but the results seem overall robust. The reason for this result is likely that, as shown in Figure 4, strategies based on short-term reversal perform best throughout, including its interaction with momentum. These anomalies have been known since at least Jegadeesh (1990), close to the beginning of the performance evaluation sample.

4.7 Discussion

The results we have documented so far suggest that simple DS-based strategies can generate similar performance to that of state-of-the-art machine learning strategies. This raises two key questions: Are these strategies still profitable after trading costs? And what do our results imply about the promise of complex methods for creating profitable trading strategies?

The evidence presented above shows that the best-performing DS strategies are likely costly to trade. These strategies have high turnover (e.g. short-term reversal) and high trading costs (e.g. illiquidity), and they generate these returns in bad market states (high illiquidity, high volatility, low sentiment) when these limits to arbitrage are even harder to overcome. As a result, limits to arbitrage may at least partially explain why these trading opportunities persist. Consistent with this interpretation, their performance somewhat decreased in the early 2000s (see Figures 4 and 7), which coincides with increased market liquidity, driven by advances in financial technology and quantitative trading, and the introduction of decimalization.

Nevertheless, it seems unlikely that trading costs alone can explain the high returns we observe. Implementable strategies based on published anomalies generate 6-factor alphas of 4% (equal-weighted) and 3% (value-weighted) per month. This finding also holds when we exclude micro cap stocks, as in Table A8 in the appendix, where we still find highly statistically significant, value-weighted 6-factor alphas of 1.5%. To consider the role of trading costs, we conduct a back-of-the-envelope calculation. Frazzini et al. (2018) find that the trading costs for optimally executed, low-frequency trading strategies are about 0.1% for large cap and 0.2% for small cap stocks. Even with a 100% turnover in small cap stocks, this only drops the lowest alpha we find from 1.5% to 1.3%. This implies that trading costs would need to be one order of magnitude larger than documented by Frazzini et al. (2018) to fully account for the performance we find by using value-weighted returns and excluding micro caps. A remaining possibility is that short-selling costs would account for the remaining excess return of 1.3%. Indeed, after dropping micro caps, alphas are largely driven by the short leg. However, even the long-only, post-publication, value-weighted strategy excluding micro caps delivers a 6-factor alpha of 0.7%. To account for the entirety of our findings, sophisticated investors would thus have to face considerably higher trading costs than documented in Frazzini et al. (2018).

Second, depending on the specification, our out-of-sample, equal-weighted strategy generates a Sharpe ratio of 2.7 or average returns of 4%. This is very close to the equal-weighted performance of various machine learning strategies.[15] Bryzgalova et al. (2019), Chen et al. (2020), Freyberger et al. (2020), Gu et al. (2020b), and Gu et al. (2020a) find maximum Sharpe ratios of 2.4, 2.6, 2.75, 2.45 and 2.6, and average returns of 0.9%, 3.3%, 3.8% and 3.3% using penalized mean-variance optimization with random tree base assets, general adversarial neural networks, adaptive group LASSO, feed forward neural nets, and autoencoder neural nets.[16] Mitigating trading costs by value-weighting does not change this conclusion,

[15] These papers construct portfolios either from return predictions or they construct mean-variance efficient portfolios. So the correct comparison metric is average returns for the former and Sharpe ratios for the latter.

[16] Bryzgalova et al. (2019) use value-weighted base assets. They do not report average returns; 0.9% is the 3-factor alpha. Gu et al. (2020a) report neither, so the respective value is missing in the list.

which yields average returns of 2.7% and a Sharpe ratio of 1.38 for our DS-based strategy. Gu et al. (2020b) find very similar value-weighted average returns of 2.3% and a Sharpe ratio of 1.35.

The fact that DS-based strategies perform similarly to these machine learning strategies could suggest that these complex approaches largely gain their stellar performance by exploiting relatively simple combinations of known anomalies. As one indication, Gu et al. (2020b) report that the two most important characteristics driving their results are size and short-term reversal, followed by idiosyncratic volatility, 6-month momentum, volatility, turnover, momentum, bid-ask spread, and illiquidity. Our strategy loads on the same anomalies, in the same order. This interpretation is also consistent with the finding in Avramov et al. (2020) that machine learning portfolios load on the individual characteristics of well-established return anomalies. It also suggests that the mapping of characteristics into expected returns may be simpler than often assumed. Although neural nets, for example, can exploit higher order interactions and nonlinearities, they do not clearly dominate a simple DS strategy that only uses first- and second-order terms.

Moreover, machine learning techniques come with the additional cost that they are mostly a "black box": in contrast to simple, transparent DS strategies, machine learning methods are usually difficult to interpret. We thus propose a new benchmark: for machine learning methods to claim success, they should be able to considerably outperform simple approaches such as the DS-based strategy presented here. Importantly, we do not claim that machine learning is not a useful tool for asset pricing. The existing literature shows that these methods indeed show great promise for portfolio construction, as well as other tasks such as optimal execution and risk management (Bartram et al., 2019). Our argument is that, when evaluating their performance, it is useful to keep simple benchmarks in mind that an investor could have implemented. Our takeaway is that the potential benefits of machine learning have not yet materialized into clearly superior performance of equity portfolios.

5 Conclusion

Interactions of stock market anomalies have been extensively studied, but the large number of candidate anomalies creates a potentially severe data mining problem. We address this issue by comprehensively investigating all possible double-sorted portfolios constructed from 102 underlying anomalies. Even when we account for multiple hypothesis testing, we find that hundreds of anomaly combinations generate statistically significant interaction gains, with signals related to past returns and limits to arbitrage taking center stage. We present the full results for hundreds of thousands of strategies based on more than 10,000 anomaly combinations on www.interactinganomalies.com.

In addition, we examine DS strategies from the perspective of a trader. We show that investing in the best backward-looking DS strategies recursively out-of-sample is about as profitable as state-of-the-art machine learning strategies. One interpretation of this finding is that the performance gains generated by current machine learning methods may largely stem from simple interactions of characteristics, and that DS strategies capture a similar amount of variation in expected returns. Our analysis also establishes a new performance benchmark: if an investor could have generated a Sharpe ratio of 2.7 without sophisticated methods, then "black box" machine learning techniques must perform significantly better to justify their lack of transparency. At least in their current form, machine learning methods may be less uniquely useful for creating profitable equity portfolios than previous work suggests. Future work should investigate whether imposing economic restrictions–as conditional double-sorts implicitly do–helps improve the performance of machine learning approaches.

References

Anton, Miguel and Christopher Polk (2014). "Connected Stocks". In: *The Journal of Finance* 69.3, pp. 1099–1127.

Asness, Clifford, Tobias Moskowitz, and Lasse Heje Pedersen (2013). "Value and Momentum Everywhere". In: *The Journal of Finance* 68.3, pp. 929–985.

Asness, Clifford, Andrea Frazzini, Ronen Israel, Tobias J. Moskowitz, and Lasse H. Pedersen (2018). "Size Matters, If You Control Your Junk". In: *Journal of Financial Economics* 129.3, pp. 479–509.

Asness, Clifford S. (1997). "The Interaction of Value and Momentum Strategies". In: *Financial Analysts Journal* 53.2, pp. 29–36.

Avramov, Doron, Tarun Chordia, and Amit Goyal (2006). "Liquidity and Autocorrelations in Individual Stock Returns". In: *The Journal of Finance* 61.5, pp. 2365–2394.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov (2007). "Momentum and Credit Rating". In: *The Journal of Finance* 62.5, pp. 2503–2520.

Avramov, Doron, Si Cheng, and Lior Metzker (2020). "Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability". In: *Unpublished Working Paper*.

Baker, Malcolm and Jeffrey Wurgler (2007). "Investor Sentiment in the Stock Market". In: *Journal of Economic Perspectives* 21.2, pp. 129–152.

Bartov, Eli and Myungsun Kim (2004). "Risk, Mispricing, and Value Investing". In: *Review of Quantitative Finance and Accounting* 23.4, pp. 353–376.

Bartram, Söhnke M., Jürgen Branke, and Mehrshad Motahari (2019). "Artificial Intelligence in Asset Management". In: *Unpublished Working Paper*.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu (2019). “Forest Through the Trees: Building Cross-Sections of Stock Returns”. In: *Unpublished Working Paper*.

Chan, Kalok and Hung Wan Kot (2006). “Can Contrarian Strategies Improve Momentum Profits?” In: *Journal of Investment Management* 4.1, pp. 70–89.

Chen, Luyang, Markus Pelger, and Jason Zhu (2020). “Deep Learning in Asset Pricing”. In: *Unpublished Working Paper*.

Cho, Thummim and Christopher Polk (2019). “Asset Pricing with Price Levels”. In: *Unpublished Working Paper*.

Daniel, Kent and Tobias J. Moskowitz (2016). “Momentum crashes”. In: *Journal of Financial Economics* 122.2, pp. 221–247.

Daniel, Kent and Sheridan Titman (1999). “Market Efficiency in an Irrational World”. In: *Financial Analysts Journal* 55.6, pp. 28–40.

Fama, Eugene F. and Kenneth R. French (1993). “Common Risk-Factors in the Returns of Stocks and Bonds”. In: *Journal of Financial Economics* 33.1, pp. 3–56.

Fama, Eugene F. and Kenneth R. French (2008). “Dissecting Anomalies”. In: *The Journal of Finance* 63.4, pp. 1653–1678.

Fama, Eugene F. and Kenneth R. French (2012). “Size, Value, and Momentum in International Stock Returns”. In: *Journal of Financial Economics* 105.3, pp. 457–472.

Fama, Eugene F. and Kenneth R. French (2015). “A Five-factor Asset Pricing Model”. In: *Journal of Financial Economics* 116.1, pp. 1–22.

Fama, Eugene F. and Kenneth R. French (2018). “Choosing factors”. In: *Journal of Financial Economics* 128.2, pp. 234–252.

Favilukis, Jack Y. and Terry Zhang (2019). “One Anomaly to Explain Them All”. In: *Unpublished Working Paper*.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz (2018). “Trading Costs”. In: *Unpublished Working Paper*.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (2020). “Dissecting Characteristics Nonparametrically”. In: *The Review of Financial Studies* 33.5, pp. 2326–2377.

Fruchterman, Thomas M. J. and Edward M. Reingold (1991). “Graph Drawing by Force-directed Placement”. In: *Software: Practice and Experience* 21.11, pp. 1129–1164.

George, Thomas and Chuan-yang Hwang (2004). “The 52-Week High and Momentum Investing”. In: *The Journal of Finance* 59.5, pp. 2145–2176.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang (2017). “The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns”. In: *The Review of Financial Studies* 30.12, pp. 4389–4436.

Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu (June 2018). *Empirical Asset Pricing via Machine Learning*. en. SSRN Scholarly Paper ID 3159577. Rochester, NY: Social Science Research Network.

Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu (2020a). “Autoencoder Asset Pricing Models”. In: *Journal of Econometrics* Forthcoming.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2020b). "Empirical Asset Pricing via Machine Learning". In: *The Review of Financial Studies* 33.5, pp. 2223–2273.

Guo, Re-Jin, Baruch Lev, and Charles Shi (2006). "Explaining the Short- and Long-Term IPO Anomalies in the US by R&D". In: *Journal of Business Finance & Accounting* 33.3-4, pp. 550–579.

Harvey, Campbell R., Yan Liu, and Heqing Zhu (2015). ". . . and the Cross-Section of Expected Returns". In: *The Review of Financial Studies* 29.1, pp. 5–68.

Hong, Harrison, Terence Lim, and Jeremy C. Stein (2000). "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies". In: *The Journal of Finance* 55.1, pp. 265–295.

Hou, Kewei, Wei Xiong, and Lin Peng (2009). "A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum". In: *Unpublished Working Paper*.

Ioffe, Sergey and Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *arXiv:1502.03167 [cs]*.

Jegadeesh, Narasimhan (1990). "Evidence of Predictable Behavior of Security Returns". In: *The Journal of Finance* 45.3, pp. 881–898.

Jiang, Guohua, Charles M. C. Lee, and Yi Zhang (2005). "Information Uncertainty and Expected Returns". In: *Review of Accounting Studies* 10.2, pp. 185–221.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2020). "Shrinking the cross-section". In: *Journal of Financial Economics* 135.2, pp. 271–292.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny (1994). "Contrarian Investment, Extrapolation, and Risk". In: *The Journal of Finance* 49.5, pp. 1541–1578.

Lambert, Marie, Boris Fays, and Georges Hübner (2016). "Size and Value Matter, But Not the Way You Thought". In: *Unpublished Working Paper*.

Lee, Charles M. C. and Bhaskaran Swaminathan (2000). "Price Momentum and Trading Volume". In: *The Journal of Finance* 55.5, pp. 2017–2069.

Lewellen, Jonathan and Jay Shanken (2000). "Estimation Risk, Market Efficiency, and the Predictability of Returns". In: *Unpublished Working Paper*. NBER working paper series working paper 7699.

Lintner, John (1965). "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets". In: *The Review of Economics and Statistics* 47.1, pp. 13–37.

Lou, Dong and Christopher Polk (2020). "Comomentum: Inferring Arbitrage Activity from Return Correlations". In: *Unpublished Working Paper*.

Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*.

Mclean, R. David and Jeffrey Pontiff (2016). "Does Academic Research Destroy Stock Return Predictability?" In: *The Journal of Finance* 71.1, pp. 5–32.

Novy-Marx, Robert (2013). "The Other Side of Value: The Gross Profitability Premium". In: *Journal of Financial Economics* 108.1, pp. 1–28.

Palmon, Dan, Ephraim F. Sudit, and Ari Yezegel (2008). "The Accruals Anomaly and Company Size". In: *Financial Analysts Journal* 64.5, pp. 47–60.

Ritter, Jay R. (1991). "The Long-Run Performance of initial Public Offerings". In: *The Journal of Finance* 46.1, pp. 3–27.

Sadka, Ronnie (2006). "Momentum and Post-earnings-announcement Drift Anomalies: The Role of Liquidity Risk". In: *Journal of Financial Economics* 80.2, pp. 309–349.

Sharpe, William F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*". In: *The Journal of Finance* 19.3, pp. 425–442.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: p. 30.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan (2015). "Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle". In: *The Journal of Finance* 70.5, pp. 1903–1948.

Zhang, X. Frank (2006). "Information Uncertainty and Stock Returns". In: *The Journal of Finance* 61.1, pp. 105–137.

Zhu, Zhaobo and Kenneth Yung (2016). "The Interaction of Short-Term Reversal and Momentum Strategies". In: *The Journal of Portfolio Management* 42.4, pp. 96–107.

Table 1: Top 10 Anomaly Interactions by Sharpe Ratio

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Equal-weighted										
ST rev	Illiquidity	HHLH	4.3	2.4	2.8	**4.5**	13.8	2.9	**2.8**	16.1
ST rev	Volume	HHLH	3.8	2.3	2.5	**4.0**	12.7	2.4	**2.3**	14.1
6m mom	ST rev	LHLL	4.8	2.1	2.5	**5.1**	11.6	2.7	**2.9**	14.4
Mom	ST rev	LHLL	4.6	2.1	2.4	**4.9**	11.1	2.6	**2.8**	14.6
ST rev	Size	HHLH	3.9	2.0	2.3	**4.2**	10.3	2.1	**2.2**	12.3
0 tr. days	ST rev	HHHL	2.6	1.9	2.2	**2.7**	12.1	1.7	**1.5**	10.0
Vola	ST rev	LHLL	4.5	1.9	2.2	**4.8**	9.6	2.2	**2.5**	11.0
Idio. vol	ST rev	HHHL	4.2	1.8	2.1	**4.5**	9.2	2.0	**2.1**	11.2
LT debt gr.	SUE	HHLL	1.8	1.8	1.7	**1.5**	10.7	0.3	0.2	1.0
Bid-ask	ST rev	LHLL	4.2	1.8	2.2	**4.6**	9.4	2.0	**2.2**	10.6
Panel B: Value-weighted										
Illiquidity	ST rev	HHHL	3.5	1.9	2.2	**3.8**	10.1	2.5	**3.9**	11.2
ST rev	Size	HHLH	3.2	1.8	2.1	**3.5**	9.3	2.2	**3.1**	12.2
Volume	ST rev	HHHL	3.0	1.7	2.0	**3.2**	9.9	2.1	**3.1**	10.5
Illiquidity	Max ret	HHHL	2.3	1.5	1.6	**2.2**	9.5	1.5	**1.9**	8.8
CAPX/assets	Size	HHLH	1.6	1.4	1.5	**1.6**	8.6	1.4	**1.4**	8.9
Asset gr.	Size	HHLH	1.9	1.4	1.4	**1.8**	9.3	1.4	**1.7**	8.6
Sales gr.	Size	HHLH	1.5	1.4	1.3	**1.3**	8.6	1.2	**1.2**	8.0
LT debt gr.	Size	HHLH	1.3	1.3	1.2	**1.1**	7.2	1.1	**1.0**	7.2
Volume	Δ 6m mom	HHHL	1.5	1.3	1.4	**1.6**	8.1	1.3	**1.4**	7.7
Op. asset gr.	Size	HHLH	1.5	1.3	1.2	**1.4**	7.9	1.2	**1.3**	7.5
Panel C: Value-weighted, Excluding Micro Caps										
Illiquidity	Max ret	HHHL	2.1	1.5	2.0	**2.0**	13.1	2.0	**1.8**	12.0
Illiquidity	SUE	HHHL	1.2	1.4	1.2	**1.0**	8.0	1.1	**0.9**	7.0
Illiquidity	ST rev	HHHL	1.8	1.3	1.5	**1.9**	8.7	1.7	**1.9**	9.8
Δ tax exp.	Size	HHLH	1.0	1.3	1.2	**0.9**	7.4	1.2	**0.9**	7.5
Volume	ST rev	HHHL	1.6	1.3	1.5	**1.7**	8.4	1.6	**1.6**	9.4
Size	SUE	HHHL	1.2	1.3	1.2	**1.0**	7.2	1.1	**1.0**	6.8
Size	LT debt gr.	HHHL	0.9	1.3	1.1	**0.8**	7.2	1.1	**0.8**	7.3
SUE	Volume	HHLH	0.9	1.3	1.1	**0.8**	7.0	1.0	**0.7**	6.4
Illiquidity	Mom	HHHL	1.9	1.3	1.3	**1.3**	7.7	1.2	**1.2**	7.2
Size	CAPX/assets	HHHL	1.0	1.2	1.1	**0.9**	7.2	1.1	**0.9**	7.2

This table plots key performance measures for the top 10 long-short DS trading strategies. We show monthly average returns, 6-factor alphas, and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. The underlying number of tests is $102 \times 101 \times 5 = 51{,}510$. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table 2: Top 10 Anomaly Interactions by Interaction Gain

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Equal-weighted										
Illiquidity	ST rev	HHHL	4.8	2.3	2.6	**5.0**	11.8	2.7	**3.0**	15.8
6m mom	ST rev	LHLL	4.8	2.1	2.5	**5.1**	11.6	2.7	**2.9**	14.4
Mom	ST rev	LHLL	4.6	2.1	2.4	**4.9**	11.1	2.6	**2.8**	14.6
Volume	ST rev	HHHL	4.3	2.3	2.6	**4.5**	12.1	2.5	**2.6**	14.1
Vola	ST rev	LHLL	4.5	1.9	2.2	**4.8**	9.6	2.2	**2.5**	11.0
Div/P	Sales/rec.	HHHL	3.0	0.6	0.6	2.6	3.0	0.5	2.4	2.8
Illiquidity	Size	HHHL	3.1	1.5	1.9	**3.6**	9.4	2.0	**2.3**	12.3
Size	ST rev	HHHL	4.2	1.9	2.3	**4.6**	9.8	2.1	**2.2**	12.1
Bid-ask	ST rev	LHLL	4.2	1.8	2.2	**4.6**	9.4	2.0	**2.2**	10.6
Div/P	Leverage	HHHL	1.9	0.5	0.5	1.9	2.4	0.5	2.2	2.9
Panel B: Value-weighted										
Illiquidity	ST rev	HHHL	3.5	1.9	2.2	**3.8**	10.1	2.5	**3.9**	11.2
Size	ST rev	HHHL	3.3	1.6	2.0	**3.6**	8.5	2.1	**3.2**	10.9
Volume	ST rev	HHHL	3.0	1.7	2.0	**3.2**	9.9	2.1	**3.1**	10.5
Div/P	ROC	HHHL	2.6	0.5	0.5	2.6	3.0	0.5	2.7	3.0
Div/P	Sales/rec.	HHHL	3.0	0.6	0.5	2.6	3.0	0.5	2.5	2.9
Mom	Vola	LHLL	2.5	1.1	1.2	**2.5**	6.9	1.2	**2.1**	6.5
Bid-ask	Max ret	LHLL	1.9	0.9	1.1	**2.2**	6.7	1.1	**2.1**	6.7
Bid-ask	Vola	LHLL	2.2	0.9	1.0	**2.4**	6.3	0.9	**2.1**	5.5
ST rev	6m mom	HLLL	2.0	0.8	1.0	**2.3**	5.7	1.1	**2.0**	6.0
Mom	Max ret	LHLL	2.1	1.0	1.2	**2.1**	7.4	1.2	**2.0**	7.4
Panel C: Value-weighted, Excluding Micro Caps										
Div/P	Sales/rec.	HHHL	2.6	0.6	0.6	2.3	3.0	0.6	2.3	2.9
Div/P	Δ Inv.	HHHL	2.0	0.6	0.7	2.5	4.1	0.7	2.3	4.1
Div/P	CAPX gr.	HHHL	1.6	0.5	0.5	1.9	2.9	0.6	2.0	3.0
Mom	Vola	LHLL	2.3	1.1	1.3	**2.2**	8.1	1.4	**2.0**	8.0
Div/P	Org. capital	HHHL	2.0	0.5	0.6	2.2	3.3	0.5	2.0	3.0
Gr. profit.	Sin stocks	HHLH	2.4	0.5	0.5	2.0	2.6	0.5	1.9	2.4
Illiquidity	ST rev	HHHL	1.8	1.3	1.5	**1.9**	8.7	1.7	**1.9**	9.8
Div/P	6m mom	HHHL	1.8	0.4	0.4	1.6	1.6	0.5	1.8	2.0
Mom	Max ret	LHLL	1.9	1.1	1.3	**1.9**	7.4	1.4	**1.8**	7.4
Illiquidity	Max ret	HHHL	2.1	1.5	2.0	**2.0**	13.1	2.0	**1.8**	12.0

This table plots key performance measures for the top 10 long-short trading strategies. We show monthly average returns, 6-factor alphas, and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. The underlying number of tests is 102 × 101 × 5 = 51,510. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table 3: Top 5 Anomaly Interactions by Market State

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A.1: Market Illiquidity - High										
ST rev	Illiquidity	HHLH	5.4	3.2	3.2	**5.0**	15.2	2.9	**2.9**	10.6
ST rev	Volume	HHLH	4.9	3.0	3.1	**4.5**	13.7	2.5	**2.5**	9.3
6m mom	ST rev	LHLL	5.8	2.8	2.9	**5.4**	12.9	2.3	**2.6**	7.4
ST rev	Size	HHLH	4.9	2.7	2.8	**4.7**	12.6	2.1	**2.3**	8.3
Mom	ST rev	LHLL	5.5	2.6	2.8	**5.1**	12.4	2.2	**2.5**	7.6
Panel A.2: Market Illiquidity - Low										
Δ $\mathbb{E}$ EPS	Vol. vola	HHLH	1.4	2.3	2.1	**1.2**	8.5	1.7	**0.8**	5.6
SUE	Δ Inv.	HHLL	1.8	2.3	2.1	**1.6**	9.2	0.5	0.3	2.1
Δ $\mathbb{E}$ EPS	Illiquidity	HHLH	1.4	2.2	2.0	**1.3**	8.7	1.6	**0.9**	5.4
Asset gr.	% Accruals	HHLL	1.8	2.2	2.0	**1.3**	8.9	1.2	**0.6**	5.1
# Analysts	SUE	HHHL	1.7	2.2	1.8	**1.3**	5.3	0.6	0.3	1.8
Panel B.1: Market Sentiment - High										
CAPX/assets	% Accruals	HHLL	1.9	2.3	2.2	**1.5**	9.8	1.2	0.6	4.2
Mom	ST rev	HHLL	3.7	2.3	2.2	**3.0**	9.2	2.0	**1.1**	8.0
ST rev	Illiquidity	HHLH	4.3	2.3	2.8	**4.7**	9.3	3.2	**3.1**	12.9
6m mom	ST rev	HHLL	3.5	2.3	2.0	**2.8**	8.7	2.0	**1.3**	6.5
% Accruals	LT debt gr.	HHLL	1.9	2.2	2.3	**1.6**	10.2	1.1	0.6	3.5
Panel B.2: Market Sentiment - Low										
ST rev	Illiquidity	HHLH	4.3	2.6	2.8	**4.3**	12.2	2.5	**2.5**	9.9
ST rev	Volume	HHLH	4.0	2.5	2.7	**4.1**	11.7	2.3	**2.3**	8.9
Mom	ST rev	LHLL	4.4	2.2	2.5	**4.5**	10.7	2.2	**2.4**	8.4
6m mom	ST rev	LHLL	4.3	2.2	2.5	**4.3**	10.5	2.1	**2.2**	7.7
ST rev	Size	HHLH	3.9	2.2	2.4	**4.0**	10.1	1.9	**1.9**	7.5
Panel C.1: Market Volatility - High										
ST rev	Illiquidity	HHLH	4.8	2.4	3.0	**5.4**	11.1	3.3	**3.4**	13.8
Volume	ST rev	HHHL	4.9	2.2	2.7	**5.5**	9.7	2.7	**3.1**	12.3
6m mom	ST rev	LHLL	5.3	2.0	2.8	**6.2**	9.8	3.1	**3.4**	13.2
Sales gr.	SUE	HHLL	2.1	2.0	1.7	**1.6**	7.8	0.2	0.1	1.0
Mom	ST rev	LHLL	5.1	1.9	2.6	**5.9**	9.4	2.9	**3.3**	12.1
Panel C.2: Market Volatility - Low										
Mom	ST rev	HHLL	3.4	2.9	2.5	**2.6**	9.7	1.6	**0.8**	4.9
ST rev	6m mom	HHLL	3.0	2.7	2.4	**2.4**	9.3	1.8	**0.9**	5.7
ST rev	Illiquidity	HHLH	3.8	2.7	2.6	**3.6**	11.4	2.2	**2.0**	7.9
ST rev	Volume	HHLH	3.3	2.6	2.4	**3.1**	10.2	1.9	**1.6**	6.9
ST rev	Size	HHLH	3.6	2.4	2.4	**3.4**	10.2	1.8	**1.7**	6.5

This table plots key performance measures for the top 5 long-short trading equal-weighted strategies by Sharpe ratio. t-statistics are computed using robust standard errors. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. With $102 \times 101 \times 5 = 51{,}510$ tests, the Bonferroni 5% critical t-value is 4.9. Market illiquidity is linearly detrended value-weighted log Amihud illiquidity of NYSE stocks. Market sentiment is as in Baker and Wurgler (2007). Market volatility is the standard deviation of daily market factor returns during the month (Avramov et al., 2020). To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table 4: Top 5 Anomaly Interactions by Decade

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: 1970s										
Illiquidity	ST rev	HHHL	6.8	3.7	3.9	**4.7**	9.6	3.1	**2.8**	6.5
Max ret	ST rev	HHLL	3.9	3.7	3.7	**3.5**	9.5	1.3	0.7	2.6
ST rev	Volume	HHLH	5.5	3.5	2.9	**3.3**	6.3	1.8	1.6	3.5
Size	ST rev	HHHL	5.9	3.5	3.3	**4.0**	8.8	2.0	1.7	4.2
CF/P	ST rev	HHLL	4.2	3.4	2.5	**2.4**	7.0	-0.6	-0.3	-1.3
Panel B: 1980s										
6m mom	ST rev	LHLL	5.5	4.2	3.6	**7.7**	9.7	3.8	**3.9**	7.7
Idio. vol	ST rev	HHHL	5.1	4.0	3.0	**6.4**	8.0	2.5	**2.5**	5.8
Mom	ST rev	LHLL	4.8	3.7	3.6	**7.4**	9.6	3.8	**3.8**	8.5
ST rev	Max ret	HHLL	3.5	3.6	1.9	**2.4**	5.8	1.6	1.0	3.6
ST rev	Vola	HLLL	4.6	3.6	3.2	**6.7**	8.3	2.7	**3.0**	6.0
Panel C: 1990s										
ST rev	Illiquidity	HHLH	6.8	4.2	1.6	3.0	4.6	2.2	**2.1**	5.7
Volume	ST rev	HHHL	6.9	4.0	1.5	3.1	4.1	1.8	2.1	4.8
0 tr. days	ST rev	HHHL	4.2	3.4	1.6	2.1	4.3	1.7	1.6	4.7
ST rev	Size	HHLH	6.4	3.4	1.3	3.1	3.8	1.7	1.9	4.7
ST rev	6m mom	HLLL	6.9	3.3	1.2	3.4	3.4	1.7	**2.0**	4.9
Panel D: 2000s										
Δ $\mathbb{E}$ EPS	Vol. vola	HHLH	1.6	2.1	0.8	0.6	2.2	0.5	0.3	1.4
B/M	SUE	HHHL	1.9	2.0	1.2	1.0	2.7	0.4	0.3	0.9
Δ $\mathbb{E}$ EPS	Illiquidity	HHLH	1.6	2.0	1.2	0.9	3.3	0.9	0.6	2.4
Δ $\mathbb{E}$ EPS	Volume	HHLH	1.7	2.0	1.1	0.8	2.7	0.8	0.5	1.9
# Analysts	SUE	HHHL	1.7	2.0	1.4	1.2	3.3	0.4	0.2	1.0
Panel E: 2010s										
Vol. vola	SUE	HHHL	1.8	2.1	1.9	1.6	4.5	1.3	0.9	2.8
0 tr. days	SUE	HHHL	1.4	2.0	1.9	**1.2**	5.1	1.1	0.6	2.3
SUE	Turn. vola	HLLL	1.3	1.9	1.8	1.0	4.9	1.1	0.5	2.4
Illiquidity	Volume	HHHL	2.2	1.8	2.8	**2.9**	6.7	1.7	1.5	3.3
CF/P	Cur. ratio	HHLL	2.0	1.8	2.1	**1.7**	5.8	2.0	1.1	4.7

This table plots key performance measures for the top 5 long-short trading strategies by Sharpe ratio. Returns are equal-weighted. We show monthly average returns, 6-factor alphas, and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. The underlying number of tests is $102 \times 101 \times 5 = 51{,}510$. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table 5: Performance of Known Anomaly Interactions

			Stats		FF5 + Mom			Interaction Gain		
			$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Momentum-based										
Mom	B/M	HHLL	1.9	0.9	0.6	0.7	2.8	0.4	0.3	2.4
Mom	Turnover	HHLL	2.1	0.9	1.0	**1.2**	5.1	1.4	**0.9**	8.3
Mom	CF vola	HLLL	1.1	0.5	0.0	0.1	0.2	0.1	0.1	0.5
Mom	Age	HLLL	0.9	0.4	-0.1	-0.1	-0.4	-0.1	-0.0	-0.4
Mom	Vola	HLLL	0.9	0.4	0.1	0.1	0.4	0.2	0.2	1.0
Illiquidity	Mom	HHHL	0.5	0.2	-0.2	-0.3	-0.9	-0.2	-0.2	-1.2
Size	Mom	HHHL	0.0	0.0	-0.5	-0.9	-2.2	-0.7	-0.8	-4.4
Panel B: Reversal-based										
ST rev	Illiquidity	HHLH	4.3	2.4	2.8	**4.5**	13.8	2.9	**2.8**	16.1
Mom	ST rev	HHLL	3.3	2.0	2.0	**2.6**	11.9	1.8	**1.0**	9.4
Turnover	ST rev	HHHL	2.3	1.8	2.1	**2.4**	12.4	1.6	**1.3**	8.3
LT rev	Turnover	HHLL	1.2	0.6	0.5	0.8	3.3	0.3	0.2	1.5
Panel C: Liquidity-based										
Asset gr.	Size	HHLH	2.4	1.6	1.6	**2.3**	9.9	0.9	**0.8**	5.2
Illiquidity	SUE	HHHL	1.3	1.2	1.1	**1.1**	6.4	0.9	0.8	4.3
B/M	Size	HHLH	1.5	0.9	0.9	1.1	4.2	0.7	0.6	4.2
CAPX/assets	Size	HHLL	1.8	0.8	1.1	**2.0**	5.6	0.3	0.2	2.0
Profit.	Size	HHLL	0.8	0.4	0.6	1.0	3.1	-0.1	-0.0	-0.4
Age	IPO	HLLL	1.0	0.4	0.1	0.2	0.3	0.0	0.1	0.2
Illiquidity	Turnover	HLLL	0.3	0.2	0.2	0.4	1.0	-0.1	-0.1	-0.8
Panel D: Value-based										
B/M	Sales gr.	HHLL	1.7	1.1	0.9	**1.0**	5.7	0.0	0.0	0.3
Profit.	B/M	HHLL	1.2	0.9	0.7	0.7	4.5	0.4	0.2	2.4
B/M	CF/P	HHLL	1.5	0.8	0.7	0.9	3.8	0.5	0.4	2.8
CF/P	Sales gr.	HHLL	1.2	0.8	0.5	0.5	2.5	0.0	0.0	0.0
E/P	B/M	HHLL	1.4	0.8	0.6	0.8	3.5	0.5	0.3	2.8
Sales gr.	E/P	HHLL	1.0	0.6	0.2	0.3	1.3	-0.3	-0.2	-2.0

This table plots key performance measures for selected DS strategies from the literature. We report results for the ordering of the two anomalies that generates the higher Sharpe ratio. Returns are equal-weighted. We show monthly average returns, 6-factor alphas, and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. There are $102 \times 101 \times 5 = 51{,}510$ candidate anomaly interaction strategies. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table 6: Out-Of-Sample Trading Strategy

	Equal-weighted					Value-weighted				
	$\mathbb{E}r_t$	SR	3F	5F	6F	$\mathbb{E}r_t$	SR	3F	5F	6F
Panel A: Top 1, Long-short Base Assets										
Long-short	4.00	1.98	3.87***	4.12***	4.33***	2.70	1.38	2.52***	2.79***	3.02***
			(0.37)	(0.45)	(0.45)			(0.35)	(0.47)	(0.49)
Long	3.28	1.15	2.51***	2.94***	3.24***	2.09	0.80	1.30***	1.65***	1.93***
			(0.42)	(0.50)	(0.49)			(0.37)	(0.47)	(0.49)
Short	0.73	0.38	1.36***	1.18***	1.10***	0.62	0.38	1.22***	1.14***	1.09***
			(0.24)	(0.24)	(0.25)			(0.19)	(0.20)	(0.21)
Panel B: Top Decile, Long-short Base Assets										
Long-short	1.09	2.70	1.09***	1.02***	0.98***	0.67	0.84	0.82***	0.51***	0.34***
			(0.07)	(0.07)	(0.07)			(0.12)	(0.11)	(0.07)
Long	1.28	0.65	0.46***	0.65***	0.79***	0.79	0.55	0.07*	0.09**	0.07
			(0.17)	(0.20)	(0.20)			(0.04)	(0.04)	(0.04)
Short	-0.19	-0.10	0.63***	0.36**	0.19	-0.11	-0.06	0.75***	0.42***	0.26***
			(0.16)	(0.18)	(0.16)			(0.13)	(0.12)	(0.09)
Panel C: Top - Bottom 1, Long-only Base Assets										
Long-short	1.09	1.25	1.26***	0.97***	0.82***	0.72	0.57	0.96***	0.49***	0.31***
			(0.12)	(0.12)	(0.09)			(0.15)	(0.13)	(0.09)
Long	1.25	0.83	0.54***	0.60***	0.66***	0.79	0.64	0.13***	0.06*	0.05
			(0.11)	(0.12)	(0.13)			(0.04)	(0.04)	(0.04)
Short	-0.15	-0.07	0.71***	0.37*	0.16	-0.07	-0.03	0.83***	0.43***	0.27***
			(0.19)	(0.22)	(0.19)			(0.14)	(0.13)	(0.10)
Panel D: Top - Bottom Decile, Long-only Base Assets										
Long-short	1.13	1.28	1.29***	1.00***	0.85***	0.75	0.58	1.00***	0.52***	0.33***
			(0.12)	(0.12)	(0.09)			(0.16)	(0.14)	(0.09)
Long	1.26	0.84	0.56***	0.61***	0.67***	0.79	0.64	0.13***	0.07*	0.05
			(0.11)	(0.12)	(0.13)			(0.04)	(0.04)	(0.04)
Short	-0.13	-0.06	0.73***	0.39*	0.18	-0.04	-0.02	0.87***	0.45***	0.28***
			(0.19)	(0.22)	(0.19)			(0.15)	(0.14)	(0.10)

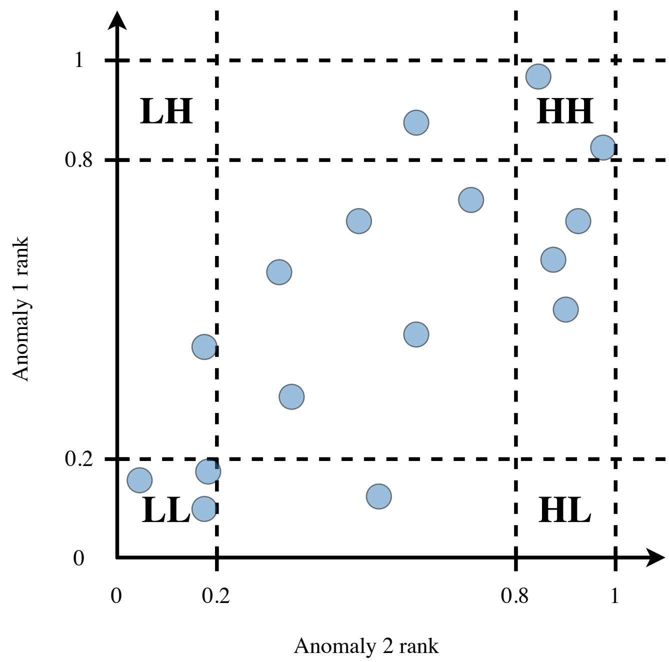
This table plots key performance measures for different zero-investment, long-short, out-of-sample trading strategies. The strategy recursively invests into the DS strategies that generated the highest Sharpe ratios in the past. It either chooses among long-short DS strategies (panel A and B) or among long-only DS corner portfolios (panel C and D). For the former, the strategies are already financed, so it only invests into the top strategies. For the latter, it goes long the top and short the bottom corner portfolios. We show results for investing only into the single best-performing strategy (panel A and C) and for strategies in the top decile (panel B and D). We show monthly mean returns and three, five, and six-factor alphas in percent with standard errors in parenthesis below; for equal and for value-weighted portfolios. The sample is monthly from 1988 to 2017 as in Gu et al. (2020b). Robust standard errors are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level.

Table 7: Out-Of-Sample Trading Strategy — Excluding Pre-Publication

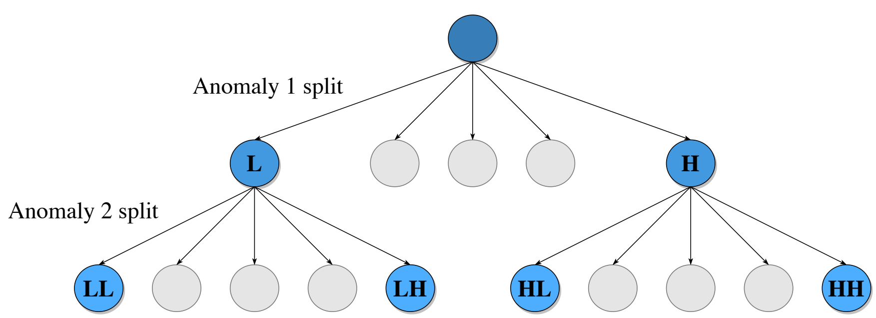
	Equal-weighted					Value-weighted				
	$\mathbb{E}r_t$	SR	3F	5F	6F	$\mathbb{E}r_t$	SR	3F	5F	6F
Panel A: Top 1, Long-short Base Assets										
Long-short	4.20	1.73	4.13***	4.30***	4.44***	2.76	1.38	2.66***	2.98***	3.21***
			(0.45)	(0.58)	(0.64)			(0.37)	(0.48)	(0.49)
Long	3.14	0.98	2.38***	2.93***	3.27***	2.31	0.82	1.59***	2.10***	2.39***
			(0.49)	(0.66)	(0.68)			(0.41)	(0.51)	(0.52)
Short	1.06	0.49	1.74***	1.36***	1.16***	0.45	0.24	1.07***	0.87***	0.82***
			(0.26)	(0.26)	(0.25)			(0.22)	(0.22)	(0.22)
Panel B: Top Decile, Long-short Base Assets										
Long-short	1.28	2.25	1.27***	1.22***	1.19***	0.78	0.87	0.93***	0.66***	0.43***
			(0.10)	(0.11)	(0.11)			(0.15)	(0.16)	(0.10)
Long	1.46	0.73	0.64***	0.86***	1.01***	0.84	0.57	0.11*	0.18***	0.15**
			(0.19)	(0.22)	(0.21)			(0.06)	(0.06)	(0.06)
Short	-0.18	-0.09	0.64***	0.36*	0.19	-0.06	-0.03	0.82***	0.48***	0.28**
			(0.16)	(0.19)	(0.17)			(0.15)	(0.16)	(0.11)
Panel C: Top - Bottom 1, Long-only Base Assets										
Long-short	1.22	1.22	1.41***	1.18***	1.04***	0.85	0.63	1.11***	0.66***	0.45***
			(0.12)	(0.13)	(0.11)			(0.17)	(0.16)	(0.11)
Long	1.40	0.92	0.72***	0.83***	0.91***	0.86	0.72	0.20***	0.17***	0.16***
			(0.13)	(0.15)	(0.15)			(0.05)	(0.05)	(0.05)
Short	-0.19	-0.08	0.69***	0.34	0.13	-0.01	-0.00	0.91***	0.49***	0.29**
			(0.20)	(0.23)	(0.20)			(0.17)	(0.16)	(0.12)
Panel D: Top - Bottom Decile, Long-only Base Assets										
Long-short	1.27	1.26	1.47***	1.23***	1.10***	0.90	0.65	1.17***	0.71***	0.48***
			(0.12)	(0.13)	(0.11)			(0.17)	(0.17)	(0.12)
Long	1.43	0.94	0.75***	0.87***	0.94***	0.87	0.72	0.21***	0.18***	0.16***
			(0.13)	(0.15)	(0.15)			(0.05)	(0.05)	(0.05)
Short	-0.17	-0.07	0.71***	0.37	0.16	0.03	0.01	0.96***	0.53***	0.31**
			(0.20)	(0.22)	(0.20)			(0.17)	(0.17)	(0.12)

This table plots key performance measures for different zero-investment, long-short, out-of-sample trading strategies. At any point in time, the strategy only uses published anomalies. The strategy recursively invests into the DS strategies that generated the highest Sharpe ratios in the past. It either chooses among long-short DS strategies (panel A and B) or among long-only DS corner portfolios (panel C and D). For the former, the strategies are already financed, so it only invests into the top strategies. For the latter, it goes long the top and short the bottom corner portfolios. We show results for investing only into the single best-performing strategy (panel A and C) and for strategies in the top decile (panel B and D). We show monthly mean returns and three, five, and six-factor alphas in percent with standard errors in parenthesis below; for equal and for value-weighted portfolios. The sample is monthly from 1988 to 2017 as in Gu et al. (2020b). Robust standard errors are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level.

Figure 1: Unconditional vs. Conditional Double-Sorted Portfolios



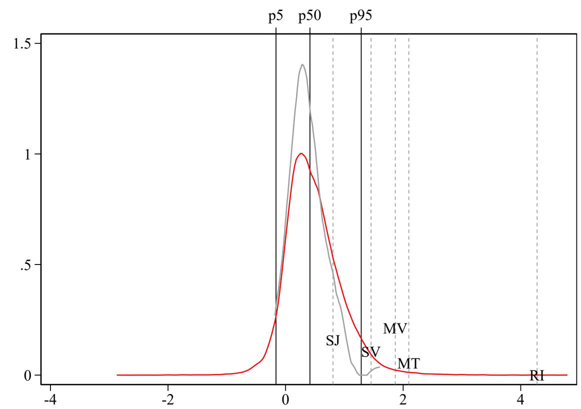
(a) Unconditional



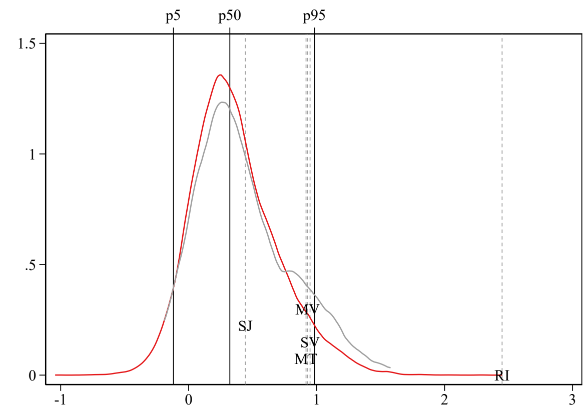
(b) Conditional

These figures illustrate the difference between conditional and unconditional DS portfolios. The top graph shows a hypothetical unconditional DS portfolio by plotting a scatter plot of the relative, cross-sectional ranks with respect to two anomaly variables. A stock is assigned to the high, high (HH) portfolio if it has a high rank (greater than 0.8) with respect to both anomalies. Stocks are similarly sorted into the LH, HL, and LL corner portfolios. The bottom graph illustrates conditional DS portfolios. The first sort is unconditional, as before. The second sort splits stocks into quintiles within each quintile portfolio.

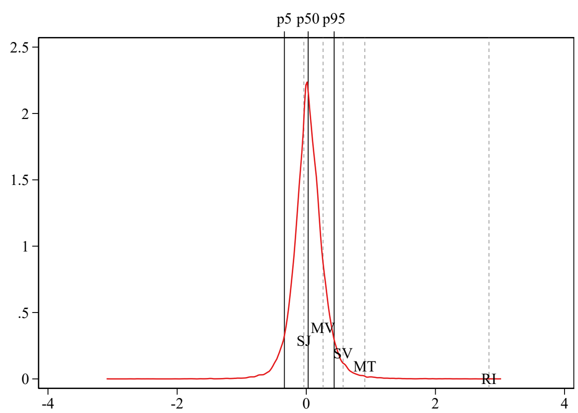
Figure 2: The Performance Distribution of Interaction Strategies



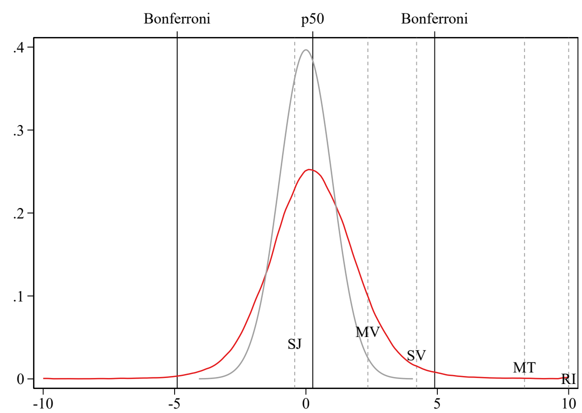
(a) Average Return



(b) Sharpe Ratio



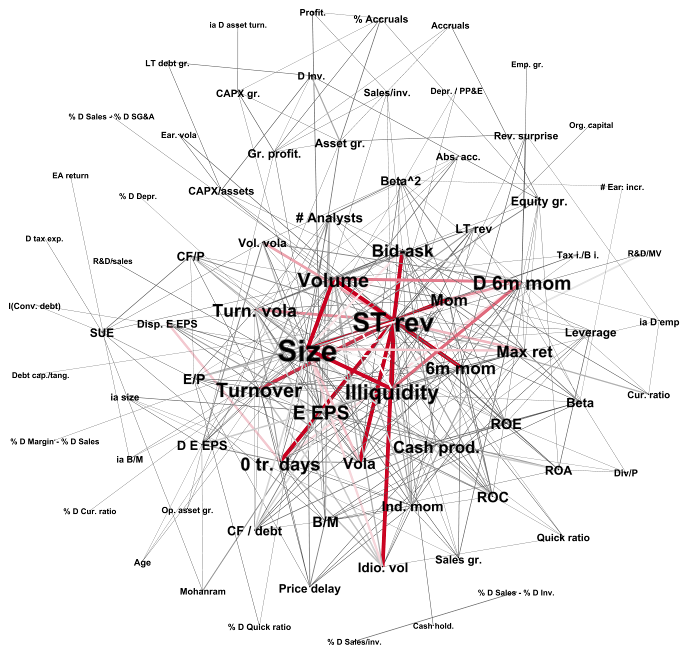
(c) Interaction Gain



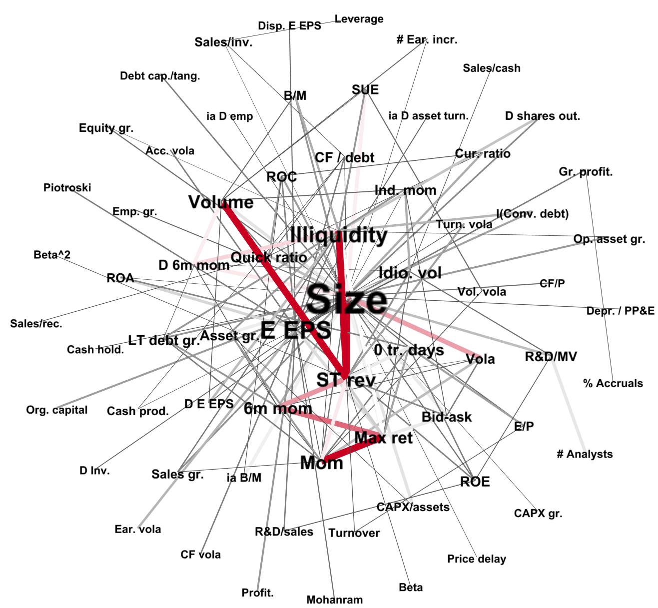
(d) t-statistic of IG

This figure shows Epanechnikov kernel density plots for different performance measures over all long-short trading strategies. Returns are equal-weighted. For average returns and Sharpe ratios, we show kernel density plots for single-sorted strategies in gray in the background. For t-statistics, we show a standard normal distribution in gray. We winsorize t-statistics at 10. We indicate 5th, 50th, and 95th percentiles or Bonferroni-corrected 5% critical t-statistics. The underlying number of tests is $102 \times 101 \times 5 = 51{,}510$. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. We also indicate the positions of selected strategies: size and value (SV), size and junk (SJ), momentum and value (MV), momentum and turnover (MT), and short-term reversal and illiquidity (RI). The sample is 1970 to 2017.

Figure 3: Which Anomalies Generate Interaction Gains?



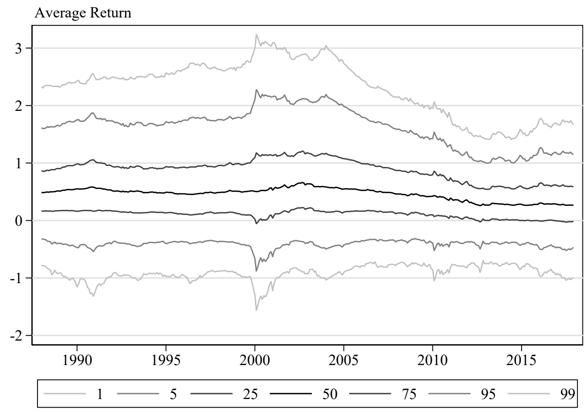
(a) Equal-weighted



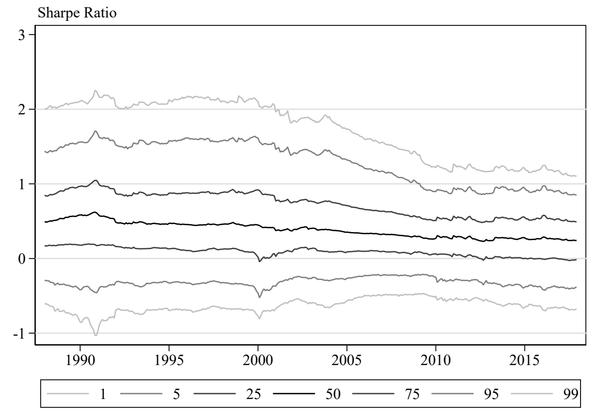
(b) Value-weighted

This figure illustrates which anomalies interact most with other anomalies using a network graph. Anomalies are nodes and two anomalies are connected by an edge if they have an interaction strategy that generates a statistically significant interaction gain (after taking into account multiple testing with the Bonferroni correction). The underlying number of tests is $102 \times 101 \times 5 = 51{,}510$. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. The size of the anomaly name is proportional to its degree of network centrality. The color of the link turns red in proportion to the absolute value of the interaction gain. Network positions are chosen using the Fruchterman and Reingold (1991) algorithm. We expand the network graph until all labels are visible. The sample is 1970 to 2017.

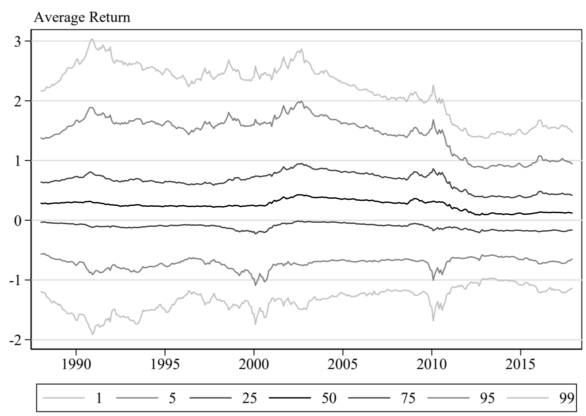
Figure 4: Performance Distribution Over Time



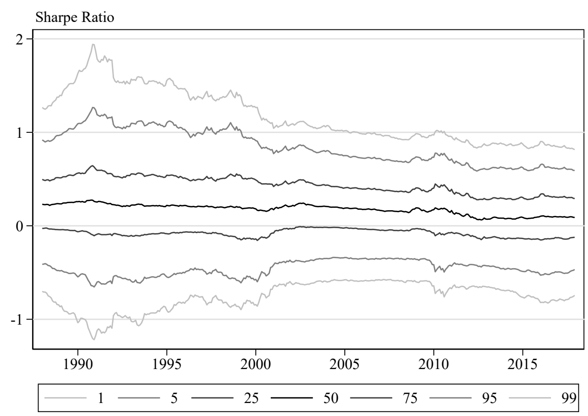
(a) Average Return, Equal-weighted



(b) Sharpe Ratio, Equal-weighted



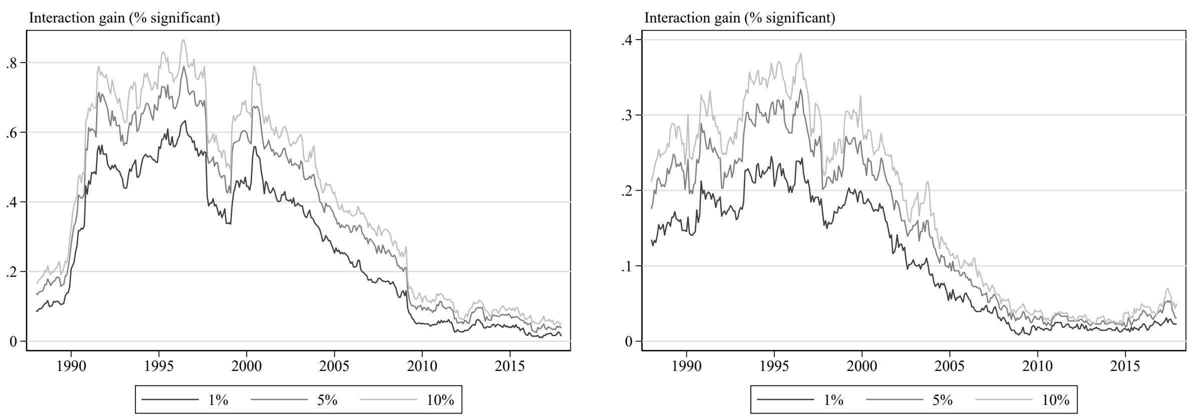
(c) Average Return, Value-weighted



(d) Sharpe Ratio, Value-weighted

This figure shows the distribution of 10-year rolling performance measures over time for all long-short trading strategies. We show 1st, 5th, 25th, 50th, 75th, 95th, and 99th percentiles of monthly average returns in percent and annualized Sharpe ratios, for equal- and value-weighted portfolios. The sample is 1988 to 2017, as in Gu et al. (2020b).

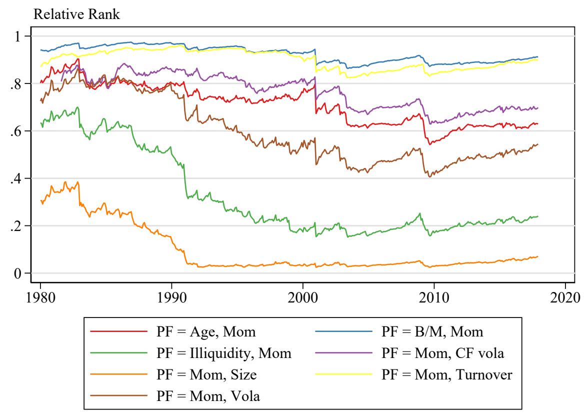
Figure 5: Fraction of Statistically Significant Anomaly Interactions Over Time



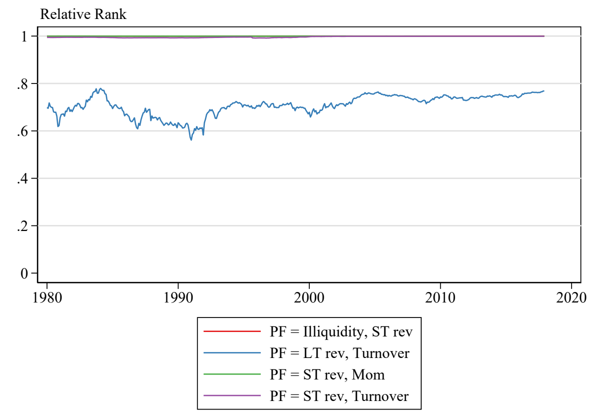
(a) Equal-weighted (b) Value-weighted

This figures plots the fraction of long-short trading strategies with statistically significant interaction gains in a 10-year rolling sample at the 1, 5 and 10% level after adjusting for multiple testing. The underlying number of tests is $102 \times 101 \times 5 = 51{,}510$. Hence, the Bonferroni-corrected critical t-values are 5.2, 4.9 and 4.76, respectively. We show results for equal- and value-weighted portfolios. The sample is 1988 to 2017, as in Gu et al. (2020b).

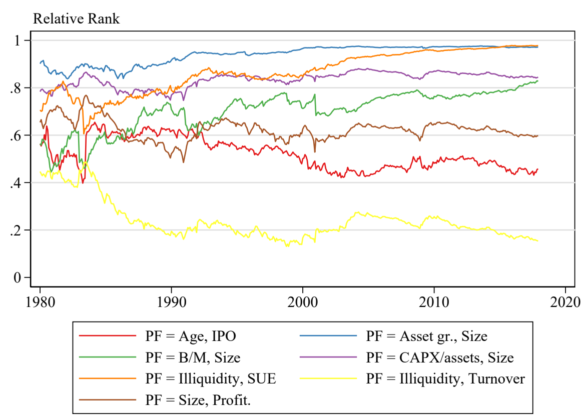
Figure 6: Sharpe Ratio Ranks of Known Anomaly Interactions



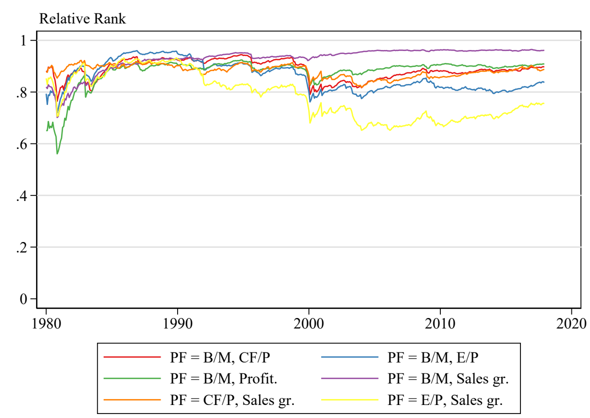
(a) Momentum-based DS PFs



(b) Reversal-based DS PFs



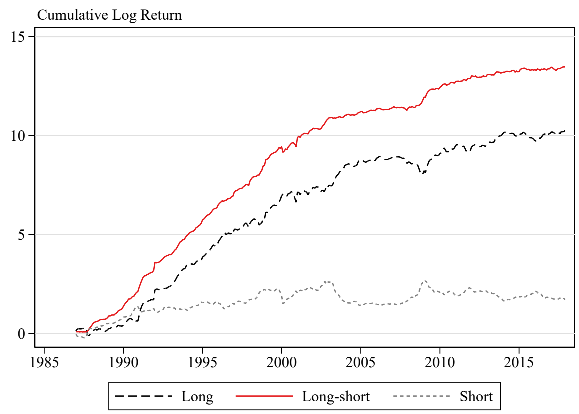
(c) Liquidity-based DS PFs



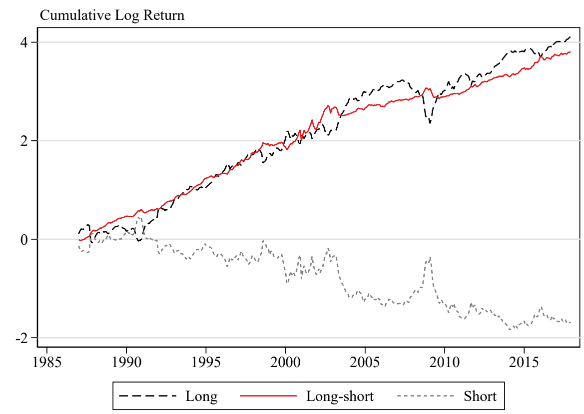
(d) Value-based DS PFs

This figure plots the relative rank of selected DS strategies from the literature among all long-short trading strategies. The relative rank is defined based on recursive Sharpe ratios. Returns are equal-weighted. The sample is 1988 to 2017, as in Gu et al. (2020b).

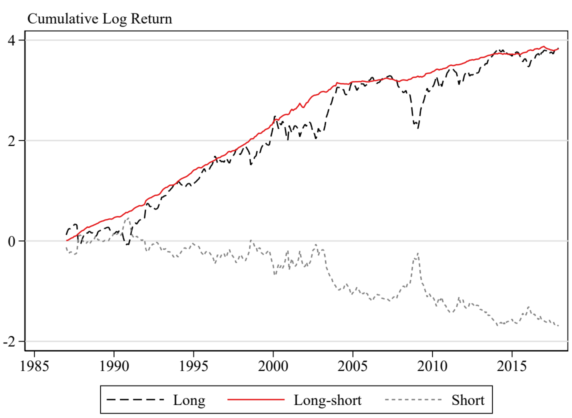
Figure 7: Cumulative Returns of Out-Of-Sample Trading Strategy



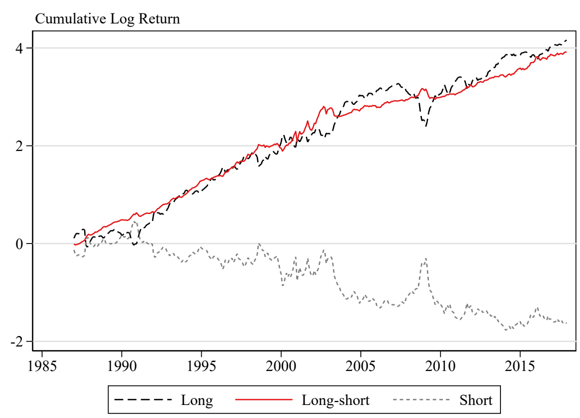
(a) Top 1, Long-short Base Assets



(b) Top 1, Long-only Base Assets



(c) Top Decile, Long-short Base Assets



(d) Top Decile, Long-only Base Assets

This figure plots the cumulative log returns of zero-investment, long-short, out-of-sample trading strategies in red. The black and the gray lines report cumulative log returns on the long and the short leg, each financed by the risk-free rate. The strategy recursively invests into the DS strategies that generated the highest Sharpe ratios in the past. It either choses among long-short DS strategies (a and c) or among long-only DS corner portfolios (b and d). For the former, the strategies are already financed, so it only invests into the top strategies. For the latter, it goes long the top and short the bottom corner portfolios. We show results for investing only into the single best-performing strategy (a and b) and for strategies in the top decile (c and d). The sample is 1988 to 2017, as in Gu et al., 2018.

A Tables

Table A1: Major Anomaly Interactions From the Literature

Reference	Journal	Anomaly 1	Anomaly 2	Strategy
Ritter, 1991	JF	IPO	Age	LHLL
Lakonishok et al., 1994	JF	Sales gr.	CF/P	HHLL
		Sales gr.	B/M	HHLL
		Sales gr.	E/P	HHLL
		E/P	B/M	HHLL
		B/M	CF/P	HHLL
Asness, 1997	FAJ	Mom	B/M	HHLL
		Mom	Div/P	HHLL
Daniel and Titman, 1999	FAJ	Mom	B/M	HLLL
Hong et al., 2000	JF	Mom	Size	HHLH
		Mom	# Analysts	HHLH
Lee and Swaminathan, 2000	JF	Mom	Turnover	HHLL
		LT rev	Turnover	HHLL
Bartov and Kim, 2004	RQFA	B/M	Accruals	HHLL
		B/M	P<10$	HHLH
George and Hwang, 2004	JF	Mom	YearHigh Dev.	HHLL
Jiang et al., 2005	RAS	Mom	Age	HLLL
		Mom	Vola	HLLL
		Mom	Turnover	HLLL
		SUE	Age	HLLL
		SUE	Vola	HLLL
		SUE	Turnover	HLLL
Avramov et al., 2006	JF	ST rev	Illiquidity	HHLH
		ST rev	Turnover	HHLH
		Illiquidity	Turnover	HLLL
Chan and Kot, 2006	JIM	Mom	LT rev	HHLL
Guo et al., 2006	JBFA	IPO	R&D increase	LHLL
Sadka, 2006	JFE	Mom	Illiquidity	HHLH
		SUE	Illiquidity	HHLH
Zhang, 2006	JF	Mom	Age	HLLL
		Mom	Size	HHLH

Table A1: Major Anomaly Interactions From the Literature (continued)

Reference	Journal	Anomaly 1	Anomaly 2	Strategy
		Mom	# Analysts	HHLH
		Mom	Analyst Disp.	HLLL
		Mom	Vola	HLLL
		Mom	CF vola	HLLL
		Analyst Revision	Age	HLLL
		Analyst Revision	Size	HHLH
		Analyst Revision	# Analysts	HHLH
		Analyst Revision	Analyst Disp.	HLLL
		Analyst Revision	Vola	HLLL
		Analyst Revision	CF vola	HLLL
Avramov et al., 2007	JF	Mom	Rating	HHLH
Fama and French, 2008	JF	Asset gr.	Size	HHLH
		Profit.	Size	HHLH
Palmon et al., 2008	FAJ	Accruals	Size	HHLL
Hou et al., 2009	WP	Mom	Turnover	HLLL
		SUE	Turnover	HHLH
Fama and French, 2012	JFE	B/M	Size	HHLH
		Mom	Size	HHLH
Asness et al., 2013	JF	Mom	B/M	HHLL
Novy-Marx, 2013	JFE	Profit.	B/M	HHLL
		Profit.	Size	HHLL
Anton and Polk, 2014	JF	Connected ret	ST rev	HHLL
Stambaugh et al., 2015	JF	Mispricing	Idio. vol	HLLL
Lambert et al., 2016	WP	Size	B/M	LLHL
Zhu and Yung, 2016	JPM	Mom	ST rev	HHLL
		6m mom	ST rev	HHLL
Asness et al., 2018	JFE	Size	Profit.	HHLL
		Size	CAPX/assets	HHLL
Cho and Polk, 2019	WP	Profit.	B/M	HHLL
Favilukis and Zhang, 2019	WP	Mom	36 anomalies	-
Lou and Polk, 2020	WP	Mom	Comomentum	HHLL

Table A2: Top 10 Anomaly Interactions — Best Long Portfolios

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Sorted by Sharpe Ratio										
Volume	ST rev	HH	3.7	1.3	1.7	**3.4**	8.5	1.8	**1.5**	10.0
Illiquidity	ST rev	HH	3.6	1.2	1.6	**3.4**	7.4	1.9	**1.7**	10.7
E/P	Beta	HH	1.2	1.2	1.1	**0.6**	6.4	1.1	**0.6**	6.4
E/P	Beta^2	HH	1.2	1.1	1.1	**0.6**	6.3	1.1	**0.6**	6.3
ST rev	Size	HH	3.5	1.1	1.6	**3.4**	7.6	1.8	**1.4**	10.8
0 tr. days	ST rev	HH	2.5	1.1	1.4	**1.9**	8.0	0.8	0.8	4.3
Beta^2	# Ear. incr.	HH	1.3	1.1	1.0	**0.7**	5.7	0.9	**0.6**	5.3
Volume	Size	HH	3.2	1.1	1.4	**2.9**	7.4	1.8	**1.6**	12.0
Beta	# Ear. incr.	HH	1.3	1.1	1.0	**0.7**	5.6	0.9	**0.6**	5.2
CF/P	Beta	HH	1.3	1.1	1.0	**0.7**	6.5	1.0	**0.7**	6.7
Panel B: Sorted by Average Returns										
Size	ST rev	HH	3.7	1.1	1.6	**3.7**	7.3	1.6	**1.4**	9.2
Volume	ST rev	HH	3.7	1.3	1.7	**3.4**	8.5	1.8	**1.5**	10.0
Size	Sin stocks	HH	3.7	0.5	0.4	3.1	3.0	0.1	0.7	1.1
Illiquidity	ST rev	HH	3.6	1.2	1.6	**3.4**	7.4	1.9	**1.7**	10.7
R&D/MV	ST rev	HH	3.2	1.0	1.6	**3.0**	7.9	0.3	0.3	1.6
Illiquidity	Size	HH	3.2	1.1	1.4	**3.0**	7.1	2.0	**1.7**	12.2
Volume	Size	HH	3.2	1.1	1.4	**2.9**	7.4	1.8	**1.6**	12.0
Vola	ST rev	LH	3.2	0.9	1.4	**3.1**	6.3	2.0	**1.7**	11.6
Bid-ask	ST rev	LH	3.1	0.9	1.5	**3.3**	6.4	1.6	**1.4**	9.0
6m mom	ST rev	LH	3.1	0.9	1.6	**3.2**	7.0	1.4	**1.5**	7.5
Panel C: Sorted by Interaction Gain										
Vola	ST rev	LH	3.2	0.9	1.4	**3.1**	6.3	2.0	**1.7**	11.6
Illiquidity	Size	HH	3.2	1.1	1.4	**3.0**	7.1	2.0	**1.7**	12.2
Bid-ask	Size	LH	2.9	0.9	1.4	**3.1**	6.6	1.9	**1.7**	11.9
Illiquidity	ST rev	HH	3.6	1.2	1.6	**3.4**	7.4	1.9	**1.7**	10.7
Vola	Size	LH	2.9	0.9	1.2	**2.8**	6.1	2.0	**1.7**	11.4
Idio. vol	Size	HH	3.1	0.9	1.3	**2.9**	6.6	1.8	**1.7**	11.3
Volume	Size	HH	3.2	1.1	1.4	**2.9**	7.4	1.8	**1.6**	12.0
Mom	ST rev	LH	2.8	0.8	1.4	**3.0**	6.4	1.7	**1.6**	9.1
Div/P	Sales/rec.	HH	2.0	0.5	0.4	1.4	2.4	0.5	1.6	3.1
6m mom	ST rev	LH	3.1	0.9	1.6	**3.2**	7.0	1.4	**1.5**	7.5

This table plots key performance measures for the top 10 long-only, DS corner portfolios financed with the risk-free rate. Returns are equal-weighted. We show monthly average returns, 6-factor alphas, and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. The underlying number of tests is $102 \times 101 \times 4 = 41{,}208$. Hence, the Bonferroni-corrected 5% critical t-value is 4.85. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table A3: Top 10 Anomaly Interactions — Best Short Portfolios

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Sorted by Sharpe Ratio										
Mom	ST rev	LL	-1.7	-0.8	-1.8	**-1.9**	-11.3	-1.8	**-1.2**	-10.1
6m mom	ST rev	LL	-1.7	-0.8	-1.9	**-1.9**	-11.9	-1.8	**-1.4**	-9.6
ST rev	Vola	LL	-1.4	-0.6	-1.4	**-1.8**	-8.9	-1.2	**-0.9**	-6.0
Illiquidity	ST rev	HL	-1.1	-0.6	-1.4	**-1.7**	-8.6	-2.1	**-1.3**	-11.7
Ind. mom	ST rev	LL	-1.0	-0.5	-1.6	**-1.3**	-9.6	-1.2	**-0.8**	-6.4
ST rev	Max ret	LL	-1.2	-0.5	-1.3	**-1.5**	-8.3	-0.9	-0.6	-4.5
Idio. vol	ST rev	HL	-1.1	-0.4	-1.2	**-1.5**	-7.4	-1.2	**-0.7**	-6.7
Bid-ask	ST rev	LL	-1.1	-0.4	-1.0	**-1.3**	-6.5	-1.3	**-0.8**	-6.8
Div/P	Abs. acc.	HH	-1.6	-0.4	-0.8	-2.9	-4.3	-0.9	-2.4	-4.2
ROA	ST rev	LL	-1.0	-0.4	-1.0	**-1.2**	-5.9	-0.2	-0.2	-1.3
Panel B: Sorted by Average Returns										
Mom	ST rev	LL	-1.7	-0.8	-1.8	**-1.9**	-11.3	-1.8	**-1.2**	-10.1
6m mom	ST rev	LL	-1.7	-0.8	-1.9	**-1.9**	-11.9	-1.8	**-1.4**	-9.6
Div/P	Abs. acc.	HH	-1.6	-0.4	-0.8	-2.9	-4.3	-0.9	-2.4	-4.2
ST rev	Vola	LL	-1.4	-0.6	-1.4	**-1.8**	-8.9	-1.2	**-0.9**	-6.0
ST rev	Max ret	LL	-1.2	-0.5	-1.3	**-1.5**	-8.3	-0.9	-0.6	-4.5
Idio. vol	ST rev	HL	-1.1	-0.4	-1.2	**-1.5**	-7.4	-1.2	**-0.7**	-6.7
Illiquidity	ST rev	HL	-1.1	-0.6	-1.4	**-1.7**	-8.6	-2.1	**-1.3**	-11.7
Div/P	Accruals	HH	-1.1	-0.3	-0.7	-2.3	-3.7	-0.9	-2.5	-4.4
Bid-ask	ST rev	LL	-1.1	-0.4	-1.0	**-1.3**	-6.5	-1.3	**-0.8**	-6.8
Div/P	Age	HH	-1.0	-0.3	-0.7	-2.3	-3.7	-0.7	-1.8	-3.4
Panel C: Sorted by Interaction Gain										
Div/P	Accruals	HH	-1.1	-0.3	-0.7	-2.3	-3.7	-0.9	-2.5	-4.4
Div/P	Abs. acc.	HH	-1.6	-0.4	-0.8	-2.9	-4.3	-0.9	-2.4	-4.2
Div/P	Age	HH	-1.0	-0.3	-0.7	-2.3	-3.7	-0.7	-1.8	-3.4
ia Δ emp	Div/P	HH	-0.8	-0.2	-0.3	-1.1	-2.0	-0.5	-1.5	-3.3
6m mom	ST rev	LL	-1.7	-0.8	-1.9	**-1.9**	-11.9	-1.8	**-1.4**	-9.6
Div/P	ia % Δ CAPX	HH	-0.8	-0.2	-0.4	-1.5	-2.4	-0.5	-1.3	-2.4
Illiquidity	ST rev	HL	-1.1	-0.6	-1.4	**-1.7**	-8.6	-2.1	**-1.3**	-11.7
CF vola	Div/P	LH	-0.4	-0.1	-0.4	-1.4	-2.0	-0.4	-1.3	-2.3
IPO	ia Δ emp	LH	-0.6	-0.2	-0.4	-0.9	-2.4	-0.6	-1.2	-3.6
Mom	ST rev	LL	-1.7	-0.8	-1.8	**-1.9**	-11.3	-1.8	**-1.2**	-10.1

This table plots key performance measures for the bottom 10 long-only, DS corner portfolios financed with the risk-free rate. Returns are equal-weighted. We show monthly average returns, 6-factor alphas, and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. The underlying number of tests is 102 × 101 × 4 = 41,208. Hence, the Bonferroni-corrected 5% critical t-value is 4.85. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table A4: Top 10 Anomaly Combinations – Unconditional DS Strategies

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Equal-weighted										
Illiquidity	ST rev	HHHL	3.9	2.4	2.7	**4.0**	12.4	2.7	**2.4**	15.3
Volume	ST rev	HHHL	3.7	2.2	2.5	**3.9**	12.0	2.3	**2.2**	13.6
ST rev	6m mom	HLLL	3.9	2.1	2.5	**4.1**	11.7	2.7	**2.3**	13.0
ST rev	Mom	HLLL	3.6	2.0	2.4	**3.8**	11.1	2.6	**2.1**	13.9
Asset gr.	SUE	HHLL	2.0	2.0	1.8	**1.6**	10.6	0.7	0.5	1.6
ST rev	Size	HHLH	3.2	1.9	2.3	**3.4**	10.1	2.0	**1.7**	11.3
ST rev	0 tr. days	HHLH	2.7	1.9	2.1	**2.8**	10.9	1.6	**1.6**	9.8
Δ Inv.	SUE	HHLL	1.5	1.9	1.7	**1.3**	11.2	0.7	0.4	1.8
CAPX/assets	SUE	HHLL	1.8	1.9	1.7	**1.4**	10.4	0.6	0.3	1.3
LT debt gr.	SUE	HHLL	1.7	1.8	1.6	**1.3**	10.4	-0.2	-0.1	-1.0
Panel B: Value-weighted										
Illiquidity	ST rev	HHHL	3.0	1.8	2.0	**3.1**	11.0	2.3	**3.2**	11.9
ST rev	Size	HHLH	2.6	1.7	2.0	**2.8**	8.7	2.1	**2.5**	10.8
Volume	ST rev	HHHL	2.5	1.6	1.9	**2.7**	10.3	2.0	**2.6**	10.7
Size	SUE	HHHL	1.1	1.5	1.3	**0.9**	8.5	1.4	**0.9**	8.7
Illiquidity	LT debt gr.	HHHL	1.0	1.3	1.2	**0.9**	7.4	1.2	**0.9**	7.3
Δ shares out.	Illiquidity	HHLH	1.2	1.2	1.0	**1.0**	6.9	1.0	**0.9**	6.3
CAPX/assets	Size	HHLH	1.1	1.2	1.1	**0.9**	6.9	1.1	**0.9**	6.7
LT debt gr.	Size	HHLH	0.9	1.2	1.0	**0.7**	5.9	0.9	**0.7**	6.0
Illiquidity	Ind. mom	HHHL	1.6	1.1	0.8	**1.0**	5.7	1.0	**1.2**	6.0
Δ 6m mom	Volume	HHLH	1.3	1.1	1.3	**1.5**	7.0	1.2	**1.3**	6.7
Panel C: Value-weighted, Excluding Micro Caps										
Illiquidity	SUE	HHHL	1.0	1.5	1.3	**0.8**	8.5	1.2	**0.8**	7.4
Size	SUE	HHHL	1.0	1.4	1.2	**0.8**	7.8	1.2	**0.7**	7.1
LT debt gr.	Size	HHLH	0.8	1.3	1.1	**0.7**	6.9	1.1	**0.7**	7.1
Δ tax exp.	Size	HHLH	1.1	1.3	1.2	**1.0**	7.1	1.2	**1.0**	7.1
Illiquidity	ST rev	HHHL	1.7	1.2	1.4	**1.8**	7.9	1.6	**1.8**	8.7
CAPX/assets	Size	HHLH	1.0	1.2	1.1	**0.8**	7.0	1.1	**0.8**	7.0
Volume	SUE	HHHL	0.8	1.2	1.1	**0.7**	6.4	1.0	**0.6**	5.8
Asset gr.	Size	HHLH	1.1	1.2	1.0	**0.8**	5.8	1.0	**0.8**	5.9
Illiquidity	LT debt gr.	HHHL	0.8	1.2	1.0	**0.7**	6.5	1.0	**0.6**	5.9
Volume	ST rev	HHHL	1.5	1.1	1.3	**1.6**	7.4	1.4	**1.5**	7.9

This table plots key performance measures for the top 10 long-short trading strategies by Sharpe ratio. This table is the same as table 1, except using unconditional instead of conditional double sorted portfolios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. With (1/2) $\times$ 102 $\times$ 101 $\times$ 5 = 25,755 tests, the Bonferroni-corrected 5% critical t-value is 4.76. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table A5: Top 10 Anomaly Combinations – With Common Sample Restrictions

Anomalies			Stats		FF5 + Mom			Interaction Gain		
1	2	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Equal-weighted										
6m mom	Max ret	LHLL	2.1	1.9	2.4	**1.9**	14.2	2.2	**1.4**	11.8
Illiquidity	ST rev	HHHL	1.7	1.9	2.0	**1.7**	12.7	1.9	**1.3**	11.7
0 tr. days	SUE	HHHL	1.0	1.9	1.8	**0.9**	11.2	1.2	**0.6**	7.1
Volume	ST rev	HHHL	1.6	1.8	2.0	**1.6**	12.5	1.8	**1.2**	11.4
Mom	Max ret	LHLL	1.9	1.8	2.2	**1.7**	13.1	1.8	**1.1**	8.7
Mom	ST rev	HHLL	2.6	1.8	1.8	**2.0**	10.6	1.5	**0.8**	8.5
Illiquidity	Mom	HHHL	1.6	1.8	1.9	**1.3**	11.6	1.6	**1.0**	9.0
Turnover	ST rev	HHHL	1.3	1.7	2.0	**1.4**	12.6	1.8	**1.1**	10.3
Volume	Mom	HHHL	1.5	1.7	1.8	**1.2**	10.5	1.5	**1.0**	8.2
6m mom	ST rev	HHLL	2.3	1.7	1.6	**1.8**	9.3	1.7	**0.8**	10.0
Panel B: Value-weighted										
Illiquidity	ST rev	HHHL	1.7	1.6	1.7	**1.6**	10.9	1.8	**1.6**	11.3
Volume	ST rev	HHHL	1.6	1.6	1.7	**1.6**	10.7	1.7	**1.5**	10.9
Illiquidity	Max ret	HHHL	1.5	1.5	2.2	**1.7**	13.5	2.1	**1.6**	13.7
Size	SUE	HHHL	1.2	1.5	1.5	**1.1**	9.5	1.4	**1.1**	9.2
Size	Mom	HHHL	1.6	1.4	1.6	**1.2**	9.8	1.6	**1.2**	9.8
SUE	Illiquidity	HHLH	1.1	1.4	1.2	**0.9**	7.9	1.2	**0.9**	7.5
Illiquidity	Mom	HHHL	1.4	1.3	1.3	**1.1**	8.4	1.3	**1.1**	8.4
Size	ROE	HHHL	1.2	1.3	1.2	**1.0**	7.5	1.2	**1.0**	7.3
Illiquidity	ROE	HHHL	1.2	1.3	1.2	**1.0**	7.6	1.2	**1.1**	7.7
Size	ST rev	HHHL	1.4	1.3	1.4	**1.4**	8.6	1.4	**1.3**	8.7

This table plots key performance measures for the top 10 long-short trading strategies by Sharpe ratio. This table is the same as table 1, except that it drops shares codes beyond 10 and 11 and firms with a dollar price less than \$5. We show monthly average returns, 6-factor alphas, and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. The underlying number of tests is $102 \times 101 \times 5 = 51{,}510$. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table A6: Top 10 Single-Sorted Portfolios

	Stats		CAPM Model			FF5 + Mom		
Anomaly	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Equal-weighted								
LT debt gr.	0.8	1.6	1.8	**0.8**	12.3	1.6	**0.6**	9.5
CAPX/assets	0.9	1.3	1.5	**1.0**	10.5	1.4	**0.7**	8.0
Δ Inv.	0.6	1.3	1.4	**0.7**	9.9	1.4	**0.6**	8.7
Asset gr.	1.0	1.2	1.2	**1.1**	8.8	1.3	**0.9**	8.1
Δ shares out.	0.9	1.1	1.5	**1.0**	10.1	1.0	**0.5**	5.4
Op. asset gr.	0.8	1.1	1.2	**0.8**	8.6	1.1	**0.6**	7.1
ST rev	1.6	1.1	1.0	**1.4**	7.0	1.4	**1.8**	5.9
% Δ Sales/inv.	0.4	1.0	1.0	**0.3**	7.1	0.9	**0.3**	5.8
% Δ Sales - % Δ Inv.	0.3	1.0	1.0	**0.3**	6.8	0.9	**0.3**	5.8
Sales gr.	0.6	1.0	1.1	**0.7**	8.1	0.9	**0.5**	6.1
Panel B: Value-weighted								
Δ shares out.	0.5	0.7	1.0	**0.7**	6.5	0.4	0.2	2.6
Δ Inv.	0.5	0.6	0.8	**0.5**	5.3	0.4	0.2	2.8
Mom	1.0	0.5	0.6	**1.2**	4.5	0.2	0.2	1.4
CAPX/assets	0.4	0.5	0.6	**0.5**	4.0	0.2	0.1	1.1
Accruals	0.4	0.5	0.5	0.4	3.4	0.6	**0.5**	4.0
# Ear. incr.	0.3	0.5	0.4	0.3	2.9	0.5	0.2	3.0
SUE	0.5	0.5	0.5	**0.5**	3.5	0.4	0.4	2.5
Asset gr.	0.4	0.5	0.6	**0.5**	4.1	-0.1	-0.0	-0.6
Sales/P	0.5	0.4	0.4	0.5	3.0	-0.3	-0.2	-2.1
Δ 6m mom	0.6	0.4	0.3	0.4	2.3	0.5	0.6	3.1
Panel C: Value-weighted, Excluding Micro Caps								
Δ shares out.	0.5	0.7	1.0	**0.6**	6.4	0.4	0.2	2.6
SUE	0.5	0.6	0.6	**0.5**	4.3	0.5	0.4	3.4
Δ Inv.	0.4	0.6	0.8	**0.5**	5.3	0.5	0.2	2.9
# Ear. incr.	0.3	0.5	0.5	0.3	3.3	0.5	0.3	3.4
Mom	0.9	0.5	0.6	**1.0**	4.0	0.1	0.1	0.9
Accruals	0.4	0.5	0.5	0.4	3.4	0.7	**0.5**	4.2
Asset gr.	0.4	0.5	0.7	**0.6**	4.6	-0.0	-0.0	-0.1
CAPX/assets	0.4	0.5	0.6	**0.5**	3.9	0.2	0.1	0.9
% Δ Sales/inv.	0.3	0.5	0.5	0.3	3.4	0.2	0.1	1.5
Op. asset gr.	0.4	0.5	0.5	0.4	3.2	0.0	0.0	0.0

This table plots key performance measures for the top 10 zero-investment, long-short, single-sort trading strategies. We show monthly average returns, CAPM- and 6-factor alphas with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the Bonferroni corrected 5% level. With 102 tests, the Bonferroni 5% critical t-value is 3.49. The sample is monthly from 1970 to 2017. The factor models include factors that correspond to some of the evaluated anomaly portfolios. Still, alphas are not exactly zero, because they are different from the factors provided on Kenneth French's website.

Table A7: Highest Interaction Gains With Major Anomalies

Anomalies		Stats		FF5 + Mom			Interaction Gain		
1	PF	$\mathbb{E}r_t$	SR	IR	α	t	IR	α	t
Panel A: Size									
ST rev	HHHL	4.2	1.9	2.3	**4.6**	9.8	2.1	**2.2**	12.1
Volume	HHHL	1.5	0.9	0.9	**1.3**	4.6	1.4	**1.5**	8.8
Illiquidity	HHHL	1.1	0.7	0.6	0.9	3.4	1.1	**1.3**	7.0
Δ 6m mom	HHHL	1.9	1.4	1.6	**2.1**	9.2	1.3	**1.2**	8.6
SUE	HHHL	1.3	0.9	0.8	**1.2**	6.1	0.6	0.8	2.0
Panel B: Book-to-market									
R&D/sales	HHLH	1.5	1.0	1.0	**1.4**	6.7	0.8	**1.0**	4.1
Cash prod.	HLLL	1.6	1.4	1.5	**1.5**	9.6	1.2	**0.9**	6.7
Sin stocks	HHLH	1.3	0.5	0.4	1.1	2.9	0.3	0.9	2.1
SUE	HHHL	1.2	1.3	1.2	**1.1**	7.4	1.1	**0.9**	5.4
ROE	HHHL	0.8	0.6	0.5	0.6	3.0	1.0	**0.8**	5.0
Panel C: Profitability									
Div/P	HHLH	1.7	0.5	0.3	1.0	1.8	0.3	1.0	1.8
Sales/inv.	HHHL	0.6	0.7	1.1	**0.8**	7.3	1.1	**0.7**	6.2
R&D/sales	HHLH	0.4	0.3	0.3	0.5	1.6	0.4	0.6	2.5
% Accruals	HHLH	0.4	0.5	0.6	**0.5**	3.9	0.8	**0.6**	4.6
CF vola	HHLH	0.7	0.4	0.3	0.5	1.9	0.3	0.5	2.0
Panel D: Investment									
ROE	HHLL	1.9	1.3	0.9	**1.1**	5.6	0.6	0.7	2.7
ROA	HHLL	2.1	1.4	1.3	**1.4**	6.9	1.0	**0.7**	5.5
Size	HHLH	1.8	1.5	1.5	**1.7**	8.6	0.7	**0.6**	4.5
SUE	HHLL	1.9	1.7	1.5	**1.6**	9.6	0.7	0.6	3.0
Asset gr.	LHLL	1.2	1.2	1.1	**0.9**	6.5	0.8	**0.5**	4.4
Panel E: Momentum									
ST rev	LHLL	4.6	2.1	2.4	**4.9**	11.1	2.6	**2.8**	14.6
SUE	LHLL	1.0	0.7	0.9	**1.2**	6.4	1.1	1.5	3.7
Size	LHLL	2.5	1.0	1.1	**2.6**	5.9	1.2	**1.3**	7.4
Δ 6m mom	LHLL	1.9	1.3	1.6	**2.1**	8.5	1.3	**1.2**	8.4
Max ret	LHLL	1.7	1.0	1.0	**1.5**	5.8	1.2	**1.2**	6.0

This table plots key performance measures for the top 5 long-short trading strategies by Sharpe ratio. Returns are equal-weighted. We show monthly average returns, 6-factor alphas and interaction gains with t-statistics computed using robust standard errors and annualized Sharpe and information ratios. Bold alphas are statistically significance at the 5% level after applying the Bonferroni correction. As the first anomaly is fixed, the underlying number of tests is $101 \times 5 = 505$. Hence, the Bonferroni-corrected 5% critical t-value is 3.9. To omit redundant information, we only show the best strategy per anomaly combination. The sample is monthly from 1970 to 2017.

Table A8: Out-Of-Sample Trading Strategy — Excluding Microcaps & Pre-Publication

	Equal-weighted					Value-weighted				
	$\mathbb{E}r_t$	SR	3F	5F	6F	$\mathbb{E}r_t$	SR	3F	5F	6F
Panel A: Top 1, Long-short Base Assets										
Long-short	2.09	1.17	2.23***	1.99***	1.61***	1.24	0.79	1.18***	1.41***	1.56***
			(0.31)	(0.31)	(0.31)			(0.30)	(0.36)	(0.36)
Long	1.12	0.51	0.26	0.37**	0.30	0.97	0.42	0.13	0.50	0.71**
			(0.18)	(0.18)	(0.20)			(0.27)	(0.33)	(0.32)
Short	0.96	0.41	1.97***	1.62***	1.32***	0.27	0.14	1.05***	0.91***	0.85***
			(0.22)	(0.23)	(0.19)			(0.19)	(0.19)	(0.19)
Panel B: Top Decile, Long-short Base Assets										
Long-short	1.02	1.37	1.09***	0.87***	0.69***	0.73	0.81	0.87***	0.61***	0.39***
			(0.13)	(0.15)	(0.12)			(0.14)	(0.16)	(0.10)
Long	0.96	0.56	0.14*	0.20**	0.23***	0.77	0.52	0.03	0.08	0.04
			(0.08)	(0.08)	(0.08)			(0.05)	(0.06)	(0.05)
Short	0.06	0.03	0.96***	0.67***	0.47***	-0.04	-0.02	0.84***	0.53***	0.35***
			(0.16)	(0.19)	(0.16)			(0.14)	(0.14)	(0.11)
Panel C: Top - Bottom 1, Long-only Base Assets										
Long-short	1.04	0.79	1.29***	0.91***	0.70***	0.74	0.55	0.99***	0.58***	0.37***
			(0.17)	(0.18)	(0.15)			(0.16)	(0.16)	(0.11)
Long	0.94	0.73	0.26***	0.22***	0.23***	0.78	0.64	0.11**	0.06	0.03
			(0.05)	(0.06)	(0.06)			(0.05)	(0.05)	(0.05)
Short	0.10	0.04	1.04***	0.69***	0.47***	-0.03	-0.02	0.89***	0.51***	0.34***
			(0.18)	(0.20)	(0.17)			(0.15)	(0.15)	(0.11)
Panel D: Top - Bottom Decile, Long-only Base Assets										
Long-short	1.06	0.80	1.33***	0.94***	0.72***	0.77	0.56	1.03***	0.61***	0.39***
			(0.17)	(0.19)	(0.15)			(0.17)	(0.17)	(0.11)
Long	0.94	0.74	0.27***	0.24***	0.24***	0.79	0.65	0.12**	0.07	0.04
			(0.05)	(0.06)	(0.06)			(0.05)	(0.05)	(0.05)
Short	0.12	0.05	1.06***	0.70***	0.48***	-0.01	-0.01	0.92***	0.53***	0.35***
			(0.18)	(0.21)	(0.17)			(0.16)	(0.15)	(0.11)

This table plots key performance measures for different zero-investment, long-short, out-of-sample trading strategies. This table is the same as table 7, except that we exclude micro cap stocks. At any point in time, the strategy only uses published anomalies. The strategy recursively invests into the DS strategies that generated the highest Sharpe ratios in the past. It either choses among long-short DS strategies (panel A and B) or among long-only DS corner portfolios (panel C and D). For the former, the strategies are already financed, so it only invests into the top strategies. For the latter, it goes long the top and short the bottom corner portfolios. We show results for investing only into the single best-performing strategy (panel A and C) and for strategies in the top decile (panel B and D). We show monthly mean returns and three, five, and six-factor alphas in percent with standard errors in parenthesis below; for equal and for value-weighted portfolios. The sample is monthly from 1988 to 2017 as in Gu et al., 2018. Robust standard errors are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level.

Table A9: Out-Of-Sample Trading Strategy
— Common Sample Restrictions & Excluding Pre-Publication

	Equal-weighted					Value-weighted				
	$\mathbb{E}r_t$	SR	3F	5F	6F	$\mathbb{E}r_t$	SR	3F	5F	6F
Panel A: Top 1, Long-short Base Assets										
Long-short	1.99	1.29	2.08***	2.09***	1.83***	1.04	0.97	0.98***	1.10***	1.01***
			(0.27)	(0.28)	(0.34)			(0.20)	(0.21)	(0.21)
Long	1.23	0.59	0.45**	0.63***	0.58**	1.04	0.61	0.30**	0.32**	0.29*
			(0.18)	(0.20)	(0.24)			(0.15)	(0.16)	(0.16)
Short	0.75	0.39	1.63***	1.46***	1.25***	0.00	0.00	0.67***	0.78***	0.73***
			(0.16)	(0.17)	(0.15)			(0.17)	(0.19)	(0.20)
Panel B: Top Decile, Long-short Base Assets										
Long-short	0.77	1.36	0.84***	0.74***	0.61***	0.54	0.74	0.64***	0.51***	0.34***
			(0.10)	(0.11)	(0.10)			(0.12)	(0.13)	(0.09)
Long	0.96	0.61	0.21***	0.23***	0.21***	0.80	0.56	0.09*	0.14**	0.09*
			(0.05)	(0.05)	(0.05)			(0.05)	(0.06)	(0.05)
Short	-0.19	-0.11	0.64***	0.51***	0.41***	-0.26	-0.15	0.54***	0.37***	0.25***
			(0.09)	(0.10)	(0.09)			(0.10)	(0.10)	(0.08)
Panel C: Top - Bottom 1, Long-only Base Assets										
Long-short	0.81	0.70	1.06***	0.79***	0.66***	0.56	0.45	0.80***	0.47***	0.32***
			(0.12)	(0.12)	(0.10)			(0.14)	(0.13)	(0.10)
Long	0.94	0.82	0.33***	0.26***	0.24***	0.80	0.70	0.17***	0.11***	0.07**
			(0.05)	(0.05)	(0.05)			(0.04)	(0.04)	(0.04)
Short	-0.14	-0.07	0.74***	0.53***	0.42***	-0.24	-0.11	0.63***	0.35***	0.24***
			(0.10)	(0.11)	(0.09)			(0.12)	(0.10)	(0.09)
Panel D: Top - Bottom Decile, Long-only Base Assets										
Long-short	0.83	0.71	1.09***	0.82***	0.69***	0.58	0.46	0.82***	0.49***	0.33***
			(0.12)	(0.12)	(0.10)			(0.14)	(0.13)	(0.10)
Long	0.95	0.83	0.34***	0.27***	0.26***	0.80	0.70	0.17***	0.12**	0.08*
			(0.05)	(0.05)	(0.05)			(0.04)	(0.05)	(0.04)
Short	-0.12	-0.06	0.75***	0.55***	0.43***	-0.21	-0.10	0.66***	0.37***	0.26***
			(0.10)	(0.11)	(0.09)			(0.12)	(0.11)	(0.09)

This table plots key performance measures for different zero-investment, long-short, out-of-sample trading strategies. This table is the same as table 7, except that we exclude shares codes beyond 10 and 11 and firms with a dollar price less than $5. At any point in time, the strategy only uses published anomalies. The strategy recursively invests into the DS strategies that generated the highest Sharpe ratios in the past. It either choses among long-short DS strategies (panel A and B) or among long-only DS corner portfolios (panel C and D). For the former, the strategies are already financed, so it only invests into the top strategies. For the latter, it goes long the top and short the bottom corner portfolios. We show results for investing only in the single best-performing strategy (panel A and C) and the strategies in the top decile (panel B and D). We show monthly mean returns and three, five, and six-factor alphas in percent with standard errors in parenthesis below; for equal and for value-weighted portfolios. The sample is monthly from 1988 to 2017 as in Gu et al., 2018. Robust standard errors are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level.

Table A10: Out-Of-Sample Trading Strategy
— Unconditional DS Strategies & Excluding Pre-Publication

	Equal-weighted					Value-weighted				
	$\mathbb{E}r_t$	SR	3F	5F	6F	$\mathbb{E}r_t$	SR	3F	5F	6F
Panel A: Top 1, Long-short Base Assets										
Long-short	3.01	1.50	2.91***	3.07***	3.18***	1.80	1.08	1.68***	1.99***	2.18***
			(0.37)	(0.47)	(0.51)			(0.31)	(0.40)	(0.41)
Long	2.72	0.82	1.91***	2.49***	2.87***	1.99	0.72	1.23***	1.74***	2.03***
			(0.51)	(0.69)	(0.71)			(0.41)	(0.51)	(0.52)
Short	0.29	0.12	0.99***	0.58	0.31	-0.19	-0.09	0.46	0.25	0.15
			(0.32)	(0.37)	(0.36)			(0.28)	(0.29)	(0.29)
Panel B: Top Decile, Long-short Base Assets										
Long-short	1.18	2.31	1.17***	1.10***	1.07***	0.68	0.81	0.81***	0.56***	0.34***
			(0.09)	(0.10)	(0.10)			(0.14)	(0.16)	(0.10)
Long	1.40	0.71	0.58***	0.80***	0.94***	0.81	0.54	0.08	0.16**	0.12**
			(0.18)	(0.21)	(0.20)			(0.06)	(0.06)	(0.06)
Short	-0.22	-0.11	0.59***	0.30	0.13	-0.13	-0.07	0.73***	0.40***	0.21**
			(0.17)	(0.19)	(0.17)			(0.15)	(0.15)	(0.11)
Panel C: Top - Bottom 1, Long-only Base Assets										
Long-short	1.10	1.15	1.29***	1.04***	0.90***	0.76	0.57	1.02***	0.58***	0.36***
			(0.12)	(0.12)	(0.10)			(0.17)	(0.16)	(0.12)
Long	1.29	0.88	0.61***	0.70***	0.77***	0.80	0.67	0.13***	0.09*	0.07
			(0.12)	(0.13)	(0.13)			(0.05)	(0.05)	(0.05)
Short	-0.19	-0.09	0.68***	0.34	0.13	-0.04	-0.02	0.88***	0.49***	0.30**
			(0.19)	(0.21)	(0.19)			(0.16)	(0.16)	(0.12)
Panel D: Top - Bottom Decile, Long-only Base Assets										
Long-short	1.14	1.17	1.33***	1.08***	0.94***	0.79	0.58	1.05***	0.61***	0.38***
			(0.12)	(0.13)	(0.11)			(0.17)	(0.17)	(0.12)
Long	1.30	0.89	0.62***	0.71***	0.77***	0.80	0.67	0.14***	0.09*	0.07
			(0.12)	(0.13)	(0.13)			(0.05)	(0.05)	(0.05)
Short	-0.16	-0.07	0.71***	0.37*	0.16	-0.02	-0.01	0.92***	0.52***	0.32**
			(0.19)	(0.22)	(0.19)			(0.17)	(0.17)	(0.13)

This table plots key performance measures for different zero-investment, long-short, out-of-sample trading strategies. This table is the same as table 7, except that we unconditional instead of conditional double sorted portfolios. At any point in time, the strategy only uses published anomalies. The strategy recursively invests into the DS strategies that generated the highest Sharpe ratios in the past. It either choses among long-short DS strategies (panel A and B) or among long-only DS corner portfolios (panel C and D). For the former, the strategies are already financed, so it only invests into the top strategies. For the latter, it goes long the top and short the bottom corner portfolios. We show results for investing only in the single best-performing strategy (panel A and C) and for strategies in the top decile (panel B and D). We show monthly mean returns and three, five, and six-factor alphas in percent with standard errors in parenthesis below; for equal and for value-weighted portfolios. The sample is monthly from 1988 to 2017 as in Gu et al. (2020b). Robust standard errors are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level.

Table A11: Variable Definitions from Green et al., 2017

Variable Name	Firm characteristic	Sign
Abs. acc.	Absolute accruals	+
Accruals	Working capital accruals	-
Abn. EA volume	Abnormal earnings announcement volume	+
Age	# years since first Compustat coverage	+
Asset gr.	Asset growth	-
Bid-ask	Bid-ask spread	-
Beta	Beta	-
Beta^2	Beta squared	-
B/M	Book-to-market	+
ia B/M	Industry-adjusted book to market	+
Cash hold.	Cash holdings	+
CF / debt	Cash flow to debt	+
Cash prod.	Cash productivity	-
CF/P	Cash flow to price ratio	+
ia CF/P	Industry-adjusted cash flow to price ratio	+
ia Δ asset turn.	Industry-adjusted change in asset turnover	+
Δ shares out.	Change in shares outstanding	-
ia Δ emp	Industry-adjusted change in employees	-
Δ $\mathbb{E}$ EPS	Change in forecasted EPS	+
Δ Inv.	Change in inventory	-
Δ 6m mom	Change in 6-month momentum	-
Δ # Analysts	Change in number of analysts	+
ia Δ mar	Industry-adjusted change in profit margin	+
Δ tax exp.	Change in tax expense	+
Investment	Corporate investment	+
I(Conv. debt)	Convertible debt indicator	-
Cur. ratio	Current ratio	+
Depr. / PP&E	Depreciation / PP&E	+
Disp. $\mathbb{E}$ EPS	Dispersion in forecasted EPS	-
Div. init.	Dividend initiation	-
Div. om.	Dividend omission	+
Volume	Dollar trading volume	-

Table A11: Variable Definitions from Green et al., 2017 (continued)

Variable Name	Firm characteristic	Sign
Div/P	Dividend to price	-
EA return	Earnings announcement return	+
Equity gr.	Growth in common shareholder equity	-
E/P	Earnings to price	+
$\mathbb{E}$ EPS gr.	Forecasted growth in 5-year EPS	+
Gr. profit.	Gross profitability	+
CAPX gr.	Growth in capital expenditures	-
Op. asset gr.	Growth in long term net operating assets	-
Ind. sales conc.	Industry sales concentration	-
Emp. gr.	Employee growth rate	-
Idio. vol	Idiosyncratic return volatility	+
Illiquidity	Illiquidity	+
Ind. mom	Industry momentum	+
CAPX/assets	Capital expenditures and inventory	-
IPO	New equity issue	-
Leverage	Leverage	+
LT debt gr.	Growth in long-term debt	-
Max ret	Maximum daily return	-
Mom	12-month momentum	+
ST rev	1-month momentum	-
LT rev	36-month momentum	-
6m mom	6-month momentum	+
Mohanram	Financial statement score	+
Size	Size	-
ia size	Industry-adjusted size	-
# Analysts	Number of analysts covering stock	-
# Ear. incr.	Number of earnings increases	+
Profit.	Operating profitability	+
Org. capital	Organizational capital	+
ia % Δ CAPX	Industry-adjusted % change in capital expenditures	-
% Δ Cur. ratio	% change in current ratio	+
% Δ Depr.	% change in depreciation	+

Table A11: Variable Definitions from Green et al., 2017 (continued)

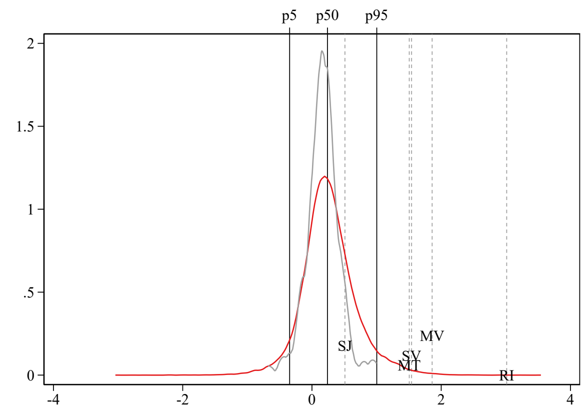
Variable Name	Firm characteristic	Sign
% Δ Margin - % Δ Sales	% change in gross margin - % change in sales	+
% Δ Quick ratio	% change in quick ratio	+
% Δ Sales - % Δ Inv.	% change in sales - % change in inventory	+
% Δ Sales - % Δ A/R	% change in sales - % change in A/R	+
% Δ Sales - % Δ SG&A	% change in sales - % change in SG&A	+
% Δ Sales/inv.	% change sales-to-inventory	+
% Accruals	Percent accruals	-
Price delay	Price delay	+
Piotroski	Financial statements score	+
Quick ratio	Quick ratio	+
R&D increase	R&D increase	+
R&D/MV	R&D to market capitalization	+
R&D/sales	R&D to sales	+
RE holdings	Real estate holdings	-
Vola	Return volatility	-
ROA	Return on assets	+
Ear. vola	Earnings volatility	+
ROE	Return on equity	+
ROC	Return on invested capital	+
Rev. surprise	Revenue surprise	+
Sales/cash	Sales to cash	+
Sales/inv.	Sales to inventory	+
Sales/rec.	Sales to receivables	+
Secured debt	Secured debt	+
I(Sec. debt)	Secured debt indicator	-
$\mathbb{E}$ EPS	Scaled earnings forecast	-
Sales gr.	Sales growth	-
Sin stocks	Sin stocks	+
Sales/P	Sales to price	+
Vol. vola	Volatility of liquidity (dollar trading volume)	+
Turn. vola	Volatility of liquidity (share turnover)	+
Acc. vola	Accrual volatility	-

Table A11: Variable Definitions from Green et al., 2017 (continued)

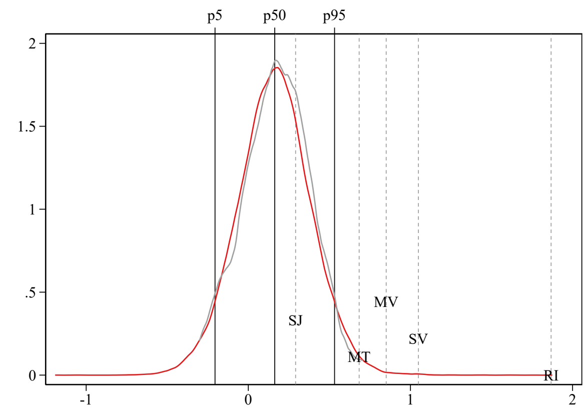
Variable Name	Firm characteristic	Sign
CF vola	Cash flow volatility	-
SUE	Unexpected quarterly earnings	+
Debt cap./tang.	Debt capacity/firm tangibility	+
Tax i./B i.	Tax income to book income	+
Turnover	Share turnover	-
0 tr. days	Zero trading days	+

B Figures

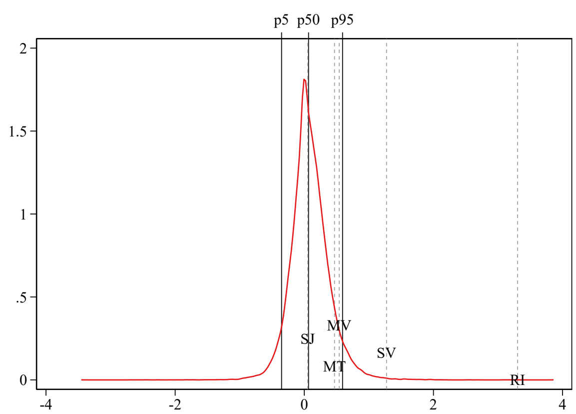
Figure A1: The Performance Distribution of Interaction Strategies – Value-weighted



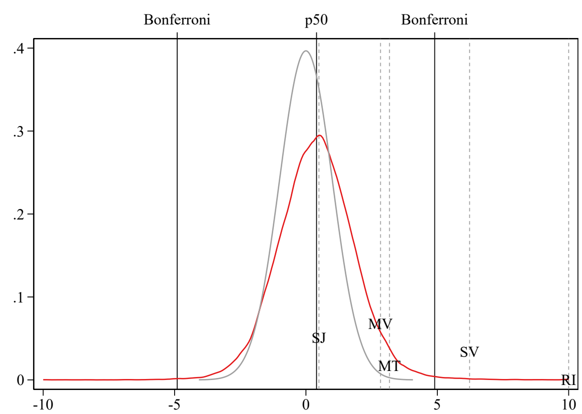
(a) Average Return



(b) Sharpe Ratio



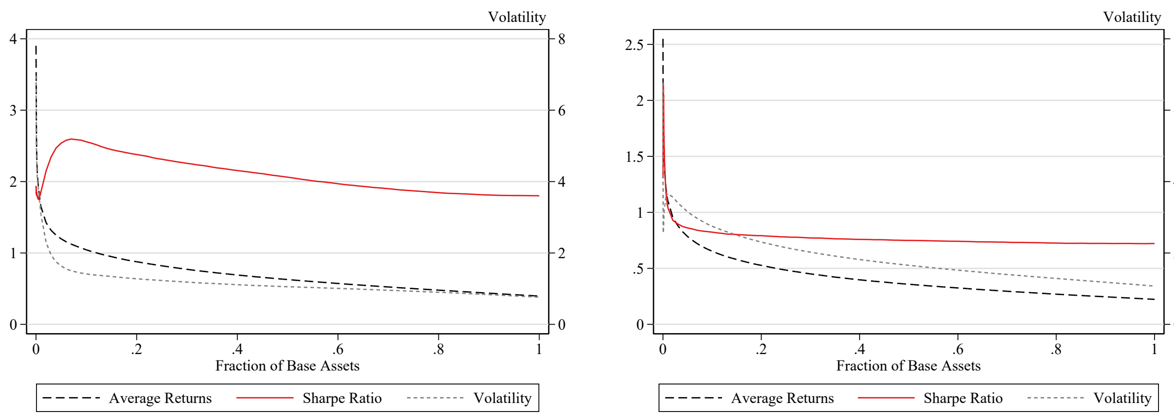
(c) Interaction Gain



(d) t-statistic of IG

This figure shows Epanechnikov kernel density plots for different performance measures over all long-short trading strategies. Returns are value-weighted. For average returns and Sharpe ratios, we show kernel density plots for single-sorted strategies in gray in the background. For t-statistics, we show a standard normal distribution in gray. We cut off t-statistics at 10. We indicate 5th, 50th and 95th percentiles or Bonferroni-corrected 5% critical t-statistics. The underlying number of tests is 102 $\times$ 101 $\times$ 5 = 51,510. Hence, the Bonferroni-corrected 5% critical t-value is 4.9. We also indicate the positions of selected strategies: size and value (SV), size and junk (SJ), momentum and value (MV), momentum and turnover (MT), and short-term reversal and illiquidity (RI). The sample is 1970 to 2017.

Figure A2: Out-Of-Sample Trading Strategies by Fraction of Base Assets Chosen



(a) Equal-weighted

(b) Value-weighted

This figure shows performance measures for different versions of the out-of-sample trading strategy. The strategy recursively invests into the long-short DS strategies that generated the highest Sharpe ratios in the past. The fraction of strategies the composite strategy invests in is on the x-axis. The step size is 0.0001 from 0 to 0.01 and 0.01 from 0.01 to 1. It starts at 1/# of strategies. A fraction of 1 means the composite strategy invests into all 102 $\times$ 101 $\times$ 6 = 61,812 candidate strategies. 6 is the number of unordered combinations of the 4 corner portfolios and the strategy takes the backward-looking profitable side in each one. Monthly average returns and annualized Sharpe ratios are on the left y-axis; monthly volatility is on the right y-axis. This figure corresponds to table 6. Table 6 reports more detailed results for the special cases where the fraction of strategies is 0.1 or 1/# of strategies. The sample is 1988 to 2017, as in Gu et al. (2020b).