Assaying Anomalies*

Robert Novy-Marx† Mihail Velikov‡

February 4, 2025

Abstract

We propose a protocol for testing cross-sectional return predictors and describe turn-key tools for automatically implementing them. Our free-to-use web application generates text, tables, and figures analyzing candidate predictor performance, going far beyond direct inferences available from standard factor models. The protocol uncovers issues that commonly arise when testing equity strategies, paying particular attention to arbitrage limits that can make strategies look good on paper even when they cannot be profitably traded in practice. The protocol also identifies similar anomalies and places a proposed predictor in the context of the extensive "factor zoo." Using a case study, we document a new signal, taxes-to-debt, and demonstrate the protocol's ability to unmask novel, seemingly robust predictors as disguised versions of known factors. Using only two of the protocol's most stringent criteria eliminates over 98% of potential spurious and untradable signals from among more than 17,000 signals mined from firms' accounting data.

JEL classification: G11, G12, G14
Keywords: Anomalies, Performance Evaluation, Trading Costs, Factor Models

*Novy-Marx provides consulting services to Dimensional Fund Advisors, an investment firm headquartered in Austin, Texas with strong ties to the academic community. The thoughts and opinions expressed in this article are those of the authors alone, and no other person or institution has any control over its content. For helpful comments and suggestions, we thank Andrew Ang, Andrew Chen, Andrew Detzel, Ken French, Jonathan Lewellen, Juhani Linnainmaa, and Cam Harvey, along with conference and seminar participants at the CQA Annual Conference, Dartmouth College, Duke University, INQUIRE UK Autumn Residential Conference, Johns Hopkins University, Leibniz Institute for Financial Research, Midwest Finance Association, Penn State University, Portuguese Finance Association, Online Seminars in Finance, Rowan University, Tilburg University, University of Oklahoma, and University of Virginia.

†Simon Business School and NBER, University of Rochester, 500 Joseph C. Wilson Blvd., Box 270100, Rochester, NY 14627. Email: robert.novy-marx@simon.rochester.edu.

‡Smeal College of Business, Pennsylvania State University, 366 Business Building, University Park, PA 16802. Email: velikov@psu.edu.

assay (as•say) *Verb* • To analyze (something, such as an ore) for one or more specific components. • To judge the worth of.

Merriam-Webster Dictionary

1 Introduction

This paper describes a protocol, and easily-accessible, easily-implementable tools, for dissecting and understanding newly proposed cross-sectional equity return predictors. While simple tools cannot be completely exhaustive, they should identify the most important issues that arise in common tests of asset pricing strategies, while going far beyond the direct inferences available from the simple tests commonly employed using standard linear factor models. The tools described here automatically generate a complete paper, providing a transparent analysis along the lines of a thorough referee report, with little more effort than pushing a button.

After its introduction, the Fama and French (1993) three-factor model quickly became the dominant asset pricing model. It was swiftly adopted not only because it allowed researchers to quickly diagnose hidden tilts to the most common empirical phenomena of the day, but also because of the simplicity with which it could be operationalized.

The model's dominance has, however, created its own set of problems. Broad professional agreement on a standard model created a widely accepted, simple definition of an "anomaly," as any strategy that has abnormally high returns relative to the model. Anomalous returns became an almost necessary condition for publishing asset pricing papers. This created tremendous incentives for researchers to find these anomalies, often without any grounding in underlying economic theory, at a time when the profession was experiencing explosive growth, and technological innovation had lowered the bar on testing. The result is our current "factor zoo" (Cochrane, 2011), with hundreds of documented cross-section predictors of stock returns (Harvey et al., 2016).

The large, growing number of known predictors has itself also made it increasingly challenging to evaluate the contribution and robustness of newly proposed candidates for the factor zoo. The simple diagnostic tests allowed by the original Fama and French (1993) model and its extensions are now nowhere near as informative as they were upon its introduction. With the large number of known anomalies it is impossible for a three-factor model, or even a five- or six-factor model, to uncover all the potential tilts to known phenomena.

Perhaps more importantly, the profession has uncovered many techniques that appear to increase the magnitude of anomalous returns, but do so in ways that are economically uninteresting and can be difficult to detect. These are mostly driven by implementation issues, which limit the forces of arbitrage. The most obvious of these involve over-weighting small cap stocks. Anomalies almost always appear stronger, often far stronger, among small stocks, at least when one ignores transaction costs. While these strategies are more expensive to trade, and consequently don't represent a more attractive trading opportunity or attract significant arbitrage capital (e.g., Novy-Marx and Velikov, 2016), they generate large gross alphas, allowing researchers to report high statistical significance. Portfolio weighting schemes that overweight smaller stocks in non-obvious ways consequently gain popularity, because they contribute to impressive paper performance (see, e.g., Velikov and Novy-Marx, 2022). Similarly, more frequent trading can improve the paper performance of anomalies, but also entails significant, largely ignored costs. The tools presented here explicitly account for implementation costs and should consequently be of interest to practitioners as well as academics.

Increased computing power has enabled more sophisticated methods for summarizing and evaluating data. These machine learning techniques are especially important in the presence of the factor zoo, as they offer tools for imposing sparsity. They can help researchers select only an important subset of many potential factors under consideration to use in their analysis, or reduce dimensionality by coming up with particularly important combinations of factors. These techniques, while becoming more popular in finance, are still not a part of most researchers toolbox. While not our main focus, our protocol incorporates tests employing some of the machine learning techniques into our analysis.

The tools that run all these tests and automatically generate a report are available in two forms: a free web application, and a public github repository.[1] The web application allows users to test the robustness of a new predictive signal by uploading a .csv file with three columns: firm identifier, date, and signal. The application then generates a self-contained report testing the new anomaly, and emails the submitter latex files and .pdf documents for this report. This referee-style report includes extensive diagnostic and robustness results, as well as an estimation of a taxonomic rank that places the proposed anomaly in context in the factor zoo, described in our protocol. The advantage of this modality is its ease of use. It is accessible to everyone, and does not require any coding skills or the use of a particular platform.

The public github repository contains an extensive library of MATLAB code that implements the tests from scratch with just a couple of mouse clicks. While these tools are based on a specific platform, and using them requires a little more skill and subscriptions to the usual data vendors, this modality has several advantages.[2] These tools are more flexible. They can be modified and adapted by individual users

[1]A fully operational preliminary version of the web application, tutorials on the github repository, and additional samples of the automatically generated reports are available at http://assayinganomalies.com/. The github repository is available at https://github.com/velikov-mihail/AssayingAnomalies.

[2]Python implementation of the code will be available soon.

for their specific needs. This also allows for the tools' open-source evolution over time. Moreover, the github repository contains a far broader set of tools, offering functionality that goes far beyond the actual testing protocol proposed here. It includes tools for accessing and downloading data from common sources, organizing this data, and running common tests in the asset pricing literature. These are all well-documented, and designed so that their basic functionality requires minimal coding skill. This dramatically lowers the bar for researchers wanting to start serious empirical work, offering an easy on-ramp for those beginning their careers. Finally, the tools interface with related public github repositories, giving access to a growing library of replications of important (and not so important) papers in the literature, and code for running empirically driven finance classes.

In this paper, we provide a walk-through of the protocol which showcases its ability to find related signals and we demonstrate the protocol's ability to uncover the lack of robust predictability in a large set of 17,000+ COMPUSTAT signals commonly employed in the recent literature on data mining. We start by documenting a new signal, the ratio of Income Taxes (COMPUSTAT item TXC) to Total Debt (COMPUSTAT item DLTT), that initially appears as a robust predictor of returns, achieving impressive gross (net) Sharpe ratios of 0.36 (0.32) over the period from 1963 to 2023. At first glance, a researcher might plausibly hypothesize that this ratio represents a form of risk, with higher corporate tax burdens relative to debt leading to increased expected returns. However, the strength of our protocol lies in its ability to go beyond surface-level analysis. By systematically applying a series of rigorous tests, including the identification of closely related anomalies, our protocol reveals that the "taxes-to-debt" signal, despite its apparent robustness, is essentially a profitability-based signal in disguise. The protocol's comprehensive nature shows that this tax-to-debt signal is spanned by several previously published profitability-based signals, a crucial insight that might have been overlooked in less thorough analyses. This case study not only showcases the protocol's power in uncovering new potential predictors but also its critical role in preventing the proliferation of redundant or disguised anomalies in the literature. Furthermore, it demonstrates how our protocol ensures that essential tests, such as examining related anomalies—which might otherwise be conducted only at a referee's request—are systematically applied to all signals under evaluation. This approach significantly enhances the rigor and reliability of cross-sectional return predictability research.

To validate our protocol and provide a benchmark for evaluating newly proposed anomalies, we conduct a comprehensive data mining exercise using over 31,000 potential signals derived from COMPUSTAT variables. Our results, detailed in Section 3.2, demonstrate the protocol's effectiveness in filtering out spurious predictors. We find that while 41.60% of signals appear significant using simple equal-weighted portfolio sorts, only 9.92% remain significant under more conservative value-weighted sorts. After accounting for transaction costs and applying dual significance criteria, the percentage of robust predictors drops even further, with merely 1.85% of signals passing our most stringent tests. These findings underscore the importance of using conservative testing methods and highlight the discriminating power of our protocol. They also provide a valuable benchmark: any new predictor that survives our battery of tests can be considered more robust than the vast majority of signals generated through extensive data mining of accounting variables.

Our goal is to provide the option for any authors proposing new anomalies to freely implement our protocol with minimal effort. The potential impact of the project on the academic literature is extensive. The protocol provides a common, easily-accessible framework for the basic testing for new factors. This removes degrees of freedom that authors have when testing proposed new anomalies, thereby mitigating the overfitting concerns that have become increasingly pernicious for the profession. It makes it easy for reviewers to request, and authors to provide, a set of standard robustness checks for an online appendix. The project also has significant practical relevance due to its emphasis on accounting for implementation frictions, and can help bridge the gap between academic research on new factors and their application in industry.

2 Walk-through of the anomaly testing protocol

This section provides a brief overview of the actual tests performed and exhibits produced by the code when it automatically generates its report. Appendix A provides more details, presenting an actual example of an automatically generated report using input data for Income Taxes to Total Debt signal. The signal is constructed as the the ratio of Income Taxes (COMPUSTAT item TXC) to Total Debt (COMPUSTAT item DLTT). A walk-through of this report is provided below and the results are discussed in more detail in Section 3.1.

2.1 Section 1: Introduction

Section 1 briefly describes the report and how it is generated, referencing this paper's protocol. It states the specific version of the publicly available code that was used to produce the report.

2.2 Section 2: Signal diagnostics

Section 2 provides signal diagnostics. Figure 1 plots descriptive statistics for the proposed predictor (Panel A) and its coverage over time both as a fraction of total firms and total market capitalization (Panel B). The plot helps identify any obvious outliers and if there are any periods with poor data coverage.

2.3 Section 3: Does the signal predict returns?

Section 3 checks whether the signal reliably predicts cross-sectional differences in average returns. Table 1 reports time-series regression results employing the value-weighted returns to portfolios constructed from a quintile sort using NYSE breakpoints on the candidate predictor (TXCDLTT for this example). Univariate sorts like these are the main technique in the anomaly literature to test whether a signal predicts returns in the cross-section of equities. A version of this table is what most anomaly papers report, though they vary in the specific portfolio construction. Our choice of value-weighting and NYSE breakpoints is conservative, as anomalies are usually strongest among micro-cap stocks and thus generally look stronger when implemented using equal-weighted portfolio returns or name breakpoints (Fama and French, 2008). Our default choice of value-weighting and NYSE breaks provides results that are closer to what an actual investor might be able to achieve in practice.

Table 2 reports results for various alternative construction methodologies. It varies the number of portfolios (five or ten), the type of portfolio breakpoints employed (NYSE, name, or capitalization), and the weighting of individual stocks within each portfolio (value- or equal-weighting). Panel B of Table 2 considers the impact of accounting for transaction costs. The trading cost calculation follows Detzel et al. (2022). The net-of-costs return on anomaly f in month t is given by:

$$f_t^{net} = f_t^{gross} - TC_{Long,t} - TC_{Short,t}$$

where

$$TC_{j,t} = \sum_{i \in I_{j,t}} |w_{i,t} - \tilde{w}_{i,t-1}| \cdot c_{i,t}$$

for $j \in \{Long, Short\}$ and $I_{j,t}$ indexes the stocks in portfolio j at time t; $c_{i,t}$ is the one-way trading cost of stock i at time t, measured as the high-frequency combination effective half-spreads from Chen and Velikov (2023); $w_{i,t}$ is the weight of stock i in its portfolio at time t after rebalancing and $\tilde{w}_{i,t-1} = \frac{w_{i,t-1}(1+r_{it})}{\sum_{k \in I_{j,t}} w_{k,t-1}(1+r_{kt})}$ is the weight of the stock in the portfolio before rebalancing.

Table 2, Panel B also reports the Novy-Marx and Velikov (2016) generalized alphas that account for trading costs. It reports these generalized alphas relative to five models: the CAPM, the Fama and French three- and five-factor models, and these three- and five-factor models augmented with the momentum factor UMD. The alphas are estimated as

$$w_{y,MVE_{\{X,y\}}}^{-1} MVE_{\{X,y\}} \;=\; \alpha^* + \beta^* \cdot MVE_{\{X\}} + \epsilon^*,$$

where $MVE_{\{X\}}$ denotes the ex-post mean-variance efficient portfolio of the assets X, where X are the factors in the model, and $w_{y,MVE_{\{X,y\}}}$ denotes the weight on asset y (the candidate factor) in $MVE_{\{X,y\}}$. Following Novy-Marx and Velikov (2016), α^* is defined to equal 0 when $w_{y,MVE_{\{X,y\}}} = 0$.

Table 3 explicitly accounts for the role of firm size in the strength of the candidate anomaly's performance. It does so by constructing strategies based on the candidate cross-sectional returns predictor within NYSE size quintiles. The table reports average portfolio returns, average number stocks, and average firm size, for twenty five portfolios constructed from a conditional double sort on size and and the proposed signal. It also reports the average returns and alphas for long/short trading strategies based on the signal within each size quintile.

2.4 Section 4: Signal performance relative to the factor zoo

Section 4 considers the strategy's performance in the context of the factor zoo. It does so by comparing the proposed factor's performance to that of up to 212 anomalies from the literature satisfying our criteria for inclusion in the testing protocol.[3]

Figure 2 plots histograms of gross and net Sharpe ratios for up to 212 known anomalies and places the candidate factor in these distributions (Panel A and B, respectively). To keep performances comparable, SR for anomalies in the factor zoo are calculated over the sample for which the candidate return predictor is provided.

Figure 3 plots the growth of a $1 invested in each of the 212 known anomalies, and compares those with the growth of a $1 invested in the test signal strategy (red lines), again on both a gross and net basis (top and bottom panels, respectively).[4]

Figure 4 shows how the candidate strategy performs relative to known anomalies in expanding the investment frontier spanned by common factor models. It plots the entire distribution, from lowest to highest, of the gross alphas (left panel) and the net generalized alphas (right hand panel) for each anomaly in the factor zoo relative to each of the five models used in Table 1 (CAPM and Fama-French three-, four-, five-, and six-factor models). It then places gross and net generalized alphas of the candidate strategy relative to each of these models in these distributions.

2.5 Section 5: Does the signal add relative to related anomalies?

Even if a candidate strategy has strong performance relative to most of the factors in the zoo, it may still not add significantly to the factor zoo. For example, a slight variation on one of the strongest know anomalies will itself have strong performance, but will not be a significant addition to the zoo already containing the strategy on which it is a variation. Section 5 accounts for this, by checking if the test signal adds information beyond that provided to the most closely related known anomalies.

[3]The anomalies come from the August, 2024 release of the Chen and Zimmermann (2022) open source asset pricing dataset.

[4]If an anomaly in the factor zoo starts later than the candidate strategy, then for that factor we assume that the dollar is invested in the candidate strategy up to the date the factor becomes available.

Figures 5 and 6 show how closely related the candidate strategy is to members of the factor zoo. Figure 5 plots a name histogram of the panel correlations of the test signal with the anomaly signals from the factor zoo. Figure 6 shows an agglomerative hierarchical cluster plot using Ward's minimum method and a maximum of 10 clusters.

Figure 7 shows how much the candidate signal adds relative to each individual member in the factor zoo. It plots histograms of t-statistics for predictability tests, which test the power of the test signal controlling for other individual known anomaly signals. Panel A reports t-statistics on the loading on the test signal, $t(\beta_S)$ from Fama-MacBeth regressions of the form:

$$r_{i,t} = \alpha + \beta_S S_{i,t} + \beta_X X_{i,t} + \epsilon_{i,t}$$

where X stands for one of the anomaly signals at a time, and S stands for the test signal.

Panel B plots t-statistics on α from spanning tests of the form:

$$r_{S,t} = \alpha + \beta r_{X,t} + \epsilon_t$$

where $r_{X,t}$ stands for the returns to one of the anomaly trading strategies at a time, and $r_{S,t}$ stands for the returns to the test signal trading strategy. The strategies employed in the spanning tests are constructed using quintile sorts, value-weighting, and NYSE breakpoints.

Panel C plots t-statistics on the average returns to strategies constructed by conditional double sorts. In each month, we sort stocks into quintiles based one of the anomaly signals at a time. Then, within each quintile, we sort stocks into quintiles based on the test signal. Stocks are finally grouped into five test-signal-portfolios by combining stocks within each anomaly sorting portfolio. The panel plots the t-statistics on the average returns of these conditional double-sorted trading strategies of the test signal conditioned on each of the anomalies.

Tables 4 and 5 control for the six most-closely related anomalies. To find the most closely related anomalies, we rank all anomalies based on:

$$\text{rank}(|\rho_{i,s}|) + rank(R^2_{r^i_t=\alpha+\beta r^i_s+\epsilon}),$$

where $\rho_{i,s}$ is the panel correlation of the underlying signal for anomaly i and the test signal s, and $R^2_{r^i_t=\alpha+\beta r^i_s+\epsilon}$ is R^2 from the spanning test of regressing the returns to the testing strategy exploiting anomaly i on the test signal s.

Table 4 reports Fama-MacBeth cross-sectional regressions of returns on the test signal controlling for the six most closely-related anomalies, both individually and jointly. Table 5 reports spanning tests results from time-series regressions of the returns to the test signal trading strategy onto the returns of trading strategies exploiting the six most closely-related anomalies and the six Fama-French factors.

2.6 Section 6: Does the signal add relative to the whole zoo?

Section 6 quantifies the extent to which the test signal increases the investment frontier beyond that spanned by the entire factor zoo.

Figure 8 plots the growth of a \$1 invested in trading strategies that combine multiple anomalies following Chen and Velikov (2023). We combine signals using a linear model of expected returns:

$$\mathbb{E}_t(r_{i,t+1}) = \beta_0 + \sum_{j=1}^{J} \beta_j x_{i,j,t},$$

where $r_{i,t+1}$ is the gross return of stock i in month $t+1$, J is the total number of predictors, β_j is the slope coefficient on predictor j, and $x_{i,j,t}$ is the standardized jth anomaly characteristic for stock i in month t.[5]

The figure shows results using six different methods for combining anomalies. The methods used are average rank (i.e., $\hat{\beta}_j = \frac{1}{J}$), weighted-average rank (i.e., $\hat{\beta}_j \propto \bar{r}^j$), Fama-MacBeth regression following Lewellen (2015), Partial Least Squares (PLS) filter following Light et al. (2017), Instrumented Principal Component Analysis (IPCA) following Kelly et al. (2019), and the Least Absolute Shrinkage and Selection Operator (LASSO) as implemented in Chen and Velikov (2023).

The figure compares the performance of combinations made using the broad cross-section of known anomalies, and the extent to which performance is improved by also including the proposed candidate.

3 Protocol validation

3.1 Case study: Income Taxes to Total Debt

To showcase the protocol's ability to uncover published predictors that explain the test signal, we document a new signal that robustly predicts returns, the ratio of Income Taxes (COMPUSTAT item TXC) to Total Debt (COMPUSTAT item DLTT). Appendix A shows that a value-weighted long/short trading strategy based on TXCDLTT achieves annualized gross (net) Sharpe ratios of 0.36 (0.32) and monthly average abnormal gross (net) returns of 23 (22) basis points relative to the Fama-French five-factor model plus momentum. The strategy's performance is robust to various portfolio construction methods and remains significant after accounting for transaction costs. TXCDLTT's predictive power persists across different size quintiles and outperforms many other documented anomalies.

Based on the initial results presented in this appendix, a researcher could formu-

[5]For these combination signals, we filter the 212 anomalies and require for each anomaly the average month to have at least 40% of the cross-sectional observations available for market capitalization on CRSP in the period for which the candidate signal is available.

late at least two contrasting hypotheses about the relationship between taxes-to-debt and stock returns:

Risk Hypothesis: The positive relationship between Income Taxes to Total Debt (TXCDLTT) and stock returns could be interpreted as evidence that higher taxes-to-debt ratios make firms riskier. This hypothesis would suggest that firms with higher TXCDLTT are exposed to greater financial or operational risks, perhaps due to less favorable debt structures or higher tax burdens relative to their debt capacity. As a result, investors demand higher returns to compensate for this increased risk.

Behavioral Hypothesis: Alternatively, the predictive power of TXCDLTT could be viewed through a behavioral finance lens. Under this hypothesis, higher taxes-to-debt might proxy for some positive firm characteristic or future performance indicator that is not yet fully reflected in stock prices. For instance, it could signal more efficient tax management or stronger overall financial health. The fact that trading on this signal leads to positive abnormal returns could suggest that the market is slow to incorporate this information, possibly due to investor inattention, limits to arbitrage, or other behavioral biases.

Both hypotheses could be supported by elements of the empirical evidence presented in the appendix, such as the strategy's consistent outperformance across various portfolio construction methods and its robustness to transaction costs. Without a comprehensive protocol to compare this signal to an exhaustive list in the literature, one could try to argue that this signal is a new, independent cross-sectional predictor.

The appendix's analysis of closely related anomalies, however, reveals that the TXCDLTT signal's predictive power is largely subsumed by existing profitability measures, suggesting that it may not represent a novel or independently valuable predictor of stock returns. This conclusion is evident from the results presented in Tables 4 and 5, which show the Fama-MacBeth regressions and spanning tests controlling for the six most closely related anomalies. Notably, these related anomalies include various measures of profitability such as "Operating profitability R&D adjusted," "Cash-based operating profitability," and "gross profits / total assets." When these profitability measures are included in the analysis, the TXCDLTT strategy's alpha becomes negative and statistically insignificant, dropping to -4 basis points per month with a t-statistic of -0.64 in the last column of Table 5. This suggests that TXCDLTT is essentially capturing information already contained in existing profitability metrics rather than providing a unique signal. Furthermore, the high R-squared value (67%) in the spanning test indicates that a large portion of TXCDLTT's variation can be explained by these related anomalies and standard risk factors. Given these findings, it would be difficult to justify publishing TXCDLTT as a new anomaly, as it appears to be largely redundant with existing profitability measures and does not contribute substantial new information to the cross-section of stock returns.

3.2 Data mining

To further demonstrate the protocol's power in uncovering signals that lack robustness, we replicate aspects of recent studies that employ data mining on COMPUSTAT variables (Yan and Zheng, 2017; Chordia et al., 2020; Chen et al., 2024). This exercise serves two purposes: (1) it demonstrates the protocol's ability to filter out spurious predictors, and (2) it provides a benchmark for evaluating the strength of newly proposed anomalies.

Table I presents the results of our data mining exercise. Panel A outlines our filtering process, which reduces an initial set of 31,460 signals to 17,074 filtered signals. We apply several criteria to ensure data quality and sufficient historical coverage, including excluding redundant signals, requiring a minimum of 30 stocks per signal-month, data availability through December 2023, and at least 360 months of historical data.

Panels B and D report the number and percentage of signals that pass single significance tests under various portfolio construction methods for gross and net returns, respectively. We consider different combinations of test statistics (excess returns or alphas), portfolio sorts (quintiles or deciles), breakpoint methods (NYSE or name), and weighting schemes (equal- or value-weighted).

For gross returns (Panel B), we find that 41.60% of signals are significant when using equal-weighted quintile sorts with NYSE breakpoints and excess returns as the test statistic. However, this percentage drops substantially to 9.92% when using value-weighted quintile sorts with NYSE breakpoints. This stark difference highlights the importance of using conservative portfolio construction methods, as equal-weighting and name breakpoints can lead to an inflated number of apparently significant predictors.

When we account for transaction costs and examine net returns (Panel D), the percentage of significant predictors decreases further. For instance, using value-weighted quintile sorts with NYSE breakpoints, only 4.99% of signals remain significant after accounting for trading costs.

Panels C and E present results for signals that pass dual significance criteria, providing a more stringent test of robustness. These panels show the percentage of signals that are simultaneously significant under pairs of different portfolio construction methods. For gross returns (Panel C), we find that only 7.32% of signals are significant under both equal- and value-weighted quintile sorts using NYSE breakpoints. When we consider net returns (Panel E), this percentage drops to just 2.66

The most conservative estimate in our analysis comes from signals that are simultaneously significant under value-weighted quintile sorts using NYSE breakpoints for both excess returns and Fama-French 6-factor alphas, after accounting for transaction costs. Only 1.85% of signals pass this stringent dual criterion for net returns.

These results demonstrate the effectiveness of our protocol in filtering out spurious predictors. The vast majority of potential signals do not survive our battery of tests, especially when using conservative portfolio construction methods and accounting for transaction costs. This underscores the importance of applying rigorous and comprehensive testing procedures, such as those outlined in our protocol, when evaluating new candidate predictors for cross-sectional stock returns.

Furthermore, these findings provide a useful benchmark for assessing the strength of newly proposed anomalies. Any new predictor that survives our protocol's tests can be considered more robust than the vast majority of signals generated through data mining of accounting variables.

4 Discussion and caveats

First, as mentioned previously, the protocol is not meant to be completely exhaustive. The tools provide a thorough, transparent analysis, going far beyond the tests commonly employed using standard linear factor models. They cannot, however, account for everything, and that is not their purpose. They are meant to effortlessly identify the most important issues that arise in common tests of asset pricing strategies. They are also specifically designed to identify strategies constructed to exploit market frictions that limit arbitrage in ways that strengthen paper performance without improving the opportunities available to actual investors.

Discrete signals also present a challenge, especially when the set of possible values is small. Many of our tests rely on non-parametric methods involving assigning stocks to portfolios on the basis of some signal. When many firms share the same signal, then some firms with the same signal must be assigned to different portfolios or the resulting portfolios will be unbalanced. In our context, where the tests are designed to run independently without requiring human judgement specific to the signal being tested, thin portfolio present a real risk. When multiple firms can naturally be assigned to two different portfolios, we consequently let nature chose which firms are assigned to each in a manner that ensures a similar degree of portfolio diversification. That is, we have some random assignment among firms with identical signals. While there is nothing inherently wrong with this procedure, it is somewhat arbitrary, and complicates the interpretation of results involving discrete signals.

Finally, even when the protocol uncovers serious inconsistencies across differently constructed strategies formed on the basis of a candidate predictor, the underlying signal may still be interesting. Several of the tests are designed to identify difficulties that may arise exploiting the strategy in practice due to market frictions. Results that reveal significant differences in performance across construction methods point to significant implementation issues related to limits to arbitrage. While this does suggest the strategy may be of limited interest to practitioners as an investment opportunity, the existence and nature of the limits to arbitrage that impact strategies based on the signal may themselves be highly interesting.

5 Conclusion

This paper describes a protocol for testing potential cross-sectional equity return predictors. This protocol goes far beyond the simple tests commonly employed using standard linear factor models, and identifies the most important issues that arise testing asset pricing strategies. It also describes turn-key tools for implementing this protocol, which produce a thorough, transparent analysis, along the lines of a referee report, with little more effort than pushing a button. These are part of broader package of freely available tools offering functionality that goes far beyond the testing protocol proposed here. These are well-documented and require minimal coding skill, dramatically lowering the bar for researchers wanting to start serious empirical work and offering an easy on-ramp for those beginning their careers.

References

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

Chen, A. Y., Lopez-Lira, A., and Zimmermann, T. (2024). Does peer-review research help predict stock returns? *Working paper*.

Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, 27(2):207–264.

Chordia, T., Goyal, A., and Saretto, A. (2020). Anomalies and false rejections. *Review of Financial Studies*, 33(5):2134–2179.

Cochrane, J. (2011). Discount rate. *Journal of Finance*, 66(4):1047–1108.

Detzel, A., Novy-Marx, R., and Velikov, M. (2022). Model comparison with transaction costs. *Journal of Finance, Forthcoming*.

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

Fama, E. F. and French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, 63(4):1653–1678.

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

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

Lewellen, J. (2015). The cross-section of expected returns. *Critical Finance Review*, 4(1):1–44.

Light, N., Maslov, D., and Rytchkov, O. (2017). Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of Financial Studies*, 30:1339–1381.

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

Velikov, M. and Novy-Marx, R. (2022). Betting against betting against beta. *Journal of Financial Economics*, 143(1):80–106.

Yan, X. S. and Zheng, L. (2017). Fundamental analysis and the cross-section of stock returns: A data mining approach. *Review of Financial Studies*, 30(4):1382–1423.

Table I: Data mining summary
This table presents results from our data mining exercise on COMPUSTAT variables. Panel A outlines the filtering process, reducing 31,460 initial signals to 17,074 filtered signals. Panel B (D) displays the number and percentage of signals passing single gross (net) significance tests under various portfolio trading strategy construction methods, for both gross and net returns. Panel C (E) shows the cumulative effect of applying dual gross (net) significance criteria. Criteria include t-statistics > 1.96 for different strategies which vary the statistic of interest (excess return, r^e, or alpha, α), portfolio sort (Quintile or Decile), breakpoints (NYSE or Name), and weighting scheme (Equal or Value).

Panel A: Filtering

Filter	# of Signals	% of Filtered
Initial set	31460	
Exclude redundant signals	29315	
Require 30 stocks	25852	
Require data until 12/2023	19834	
Require 360 months	17074	100.0%

Panel B: Single gross significance criteria

	r^e Quint NYSE Equal	r^e Dec NYSE Value	r^e Quint Name Value	r^e Quint NYSE Value	α_{FF6} Quint NYSE Value
# of significant	7102	1880	1705	1693	4220
% of significant	41.60%	11.01%	9.99%	9.92%	24.72%

Panel C: Dual gross significance criteria

				r^e Quint NYSE Equal	r^e Dec NYSE Value	r^e Quint Name Value	r^e Quint NYSE Value	α_{FF6} Quint NYSE Value
r^e	Quintile	NYSE	Equal		7.32%	6.93%	6.70%	9.66%
r^e	Decile	NYSE	Value	1249		6.18%	5.68%	3.00%
r^e	Quintile	Name	Value	1183	1056		7.00%	3.02%
r^e	Quintile	NYSE	Value	1144	969	1195		3.55%
α_{FF6}	Quintile	NYSE	Value	1649	512	515	606	

Table I (Cont'd): Data mining summary

Panel D: Single net significance criteria								
				r^e	r^e	r^e	r^e	α_{FF6}
				Quint	Dec	Quint	Quint	Quint
				NYSE	NYSE	Name	NYSE	NYSE
				Equal	Value	Value	Value	Value
# of significant				4,589	1,045	817	852	2,335
% of significant				26.88%	6.12%	4.79%	4.99%	13.68%

Panel E: Dual net significance criteria								
				r^e	r^e	r^e	r^e	α_{FF6}
				Quint	Dec	Quint	Quint	Quint
				NYSE	NYSE	Name	NYSE	NYSE
				Equal	Value	Value	Value	Value
r^e	Quintile	NYSE	Equal		2.66%	2.10%	1.89%	2.34%
r^e	Decile	NYSE	Value	454		3.15%	2.99%	1.50%
r^e	Quintile	Name	Value	359	537		3.21%	1.36%
r^e	Quintile	NYSE	Value	322	510	549		1.85%
α_{FF6}	Quintile	NYSE	Value	399	256	233	316	

Appendix A Output example

This Appendix provides the actual output from the tools that implement our protocol. These take a flat .csv data file with three columns—firm identifier, date, and signal—and generate a .tex file for a referee report or internet appendix that tests the proposed signal. For the actual implementation provided here the signal we use is Income Taxes to Total Debt signal (TXCDLTT). The signal is constructed as the the ratio of Income Taxes (COMPUSTAT item TXC) to Total Debt (COMPUSTAT item DLTT). The following report is the direct output of the tools that results from inputing the .csv file containing firm-date observations on TXCDLTT.

Online Appendix for Assaying Anomalies: Income Taxes to Total Debt and the Cross Section of Stock Returns

Robert Novy-Marx   Mihail Velikov

February 4, 2025

Abstract

This report studies the asset pricing implications of Income Taxes to Total Debt (TXCDLTT), and its robustness in predicting returns in the cross-section of equities using the protocol proposed by Novy-Marx and Velikov (2023). A value-weighted long/short trading strategy based on TXCDLTT achieves an annualized gross (net) Sharpe ratio of 0.36 (0.33), and monthly average abnormal gross (net) return relative to the Fama and French (2015) five-factor model plus a momentum factor of 24 (22) bps/month with a t-statistic of 2.82 (2.72), respectively. Its gross monthly alpha relative to these six factors plus the six most closely related strategies from the factor zoo (Operating profitability R&D adjusted, net income / book equity, Cash-based operating profitability, gross profits / total assets, Taxable income to income, operating profits / book equity) is -4 bps/month with a t-statistic of -0.61.

1 Introduction

The following automatically generated report tests the asset pricing implications of Income Taxes to Total Debt (TXCDLTT), and its robustness in predicting returns in the cross-section of equities. It is produced using the methodology of Novy-Marx and Velikov (2023), from input data consisting of firm-month observations for the proposed predictor.[1]

2 Signal diagnostics

Figure 1 plots descriptive statistics for the TXCDLTT signal. Panel A plots the time-series of the mean, median, and interquartile range for TXCDLTT. On average, the cross-sectional mean (median) TXCDLTT is 3.66 (0.06) over the 1963 to 2023 sample, where the starting date is determined by the availability of the input TXCDLTT data. The signal's interquartile range spans -0.00 to 0.49. Panel B of Figure 1 plots the time-series of the coverage of the TXCDLTT signal for the CRSP universe. On average, the TXCDLTT signal is available for 4.34% of CRSP names, which on average make up 5.83% of total market capitalization.

3 Does TXCDLTT predict returns?

Table 1 reports the performance of portfolios constructed using a value-weighted, quintile sort on TXCDLTT using NYSE breaks. The first two lines of Panel A report monthly average excess returns for each of the five portfolios and for the long/short portfolio that buys the high TXCDLTT portfolio and sells the low TXCDLTT portfolio. The rest of Panel A reports the portfolios' monthly abnormal returns relative to the five most common factor models: the CAPM, the Fama and French (1993)

[1] It used version v0.4.1 of the publicly available code repository at https://github.com/velikov-mihail/AssayingAnomalies. See more details at http://AssayingAnomalies.com.

three-factor model (FF3) and its variation that adds momentum (FF4), the Fama and French (2015) five-factor model (FF5), and its variation that adds momentum factor used in Fama and French (2018) (FF6). The table shows that the long/short TXCDLTT strategy earns an average return of 0.30% per month with a t-statistic of 2.81. The annualized Sharpe ratio of the strategy is 0.36. The alphas range from 0.24% to 0.52% per month and have t-statistics exceeding 2.82 everywhere. The lowest alpha is with respect to the FF6 factor model.

Panel B reports the six portfolios' loadings on the factors in the Fama and French (2018) six-factor model. The long/short strategy's most significant loading is 0.68, with a t-statistic of 17.46 on the RMW factor. Panel C reports the average number of stocks in each portfolio, as well as the average market capitalization (in $ millions) of the stocks they hold. In an average month, the five portfolios have at least 354 stocks and an average market capitalization of at least $679 million.

Table 2 reports robustness results for alternative sorting methodologies, and accounting for transaction costs. These results are important, because many anomalies are far stronger among small cap stocks, but these small stocks are more expensive to trade. Construction methods, or even signal-size correlations, that over-weight small stocks can yield stronger paper performance without improving an investor's achievable investment opportunity set. Panel A reports gross returns and alphas for the long/short strategies made using various different protfolio constructions. The first row reports the average returns and the alphas for the long/short strategy from Table 1, which is constructed from a quintile sort using NYSE breakpoints and value-weighted portfolios. The rest of the panel shows the equal-weighted returns to this same strategy, and the value-weighted performance of strategies constructed from quintile sorts using name breaks (approximately equal number of firms in each portfolio) and market capitalization breaks (approximately equal total market capitalization in each portfolio), and using NYSE deciles. The average return is lowest for the quintile sort using cap breakpoints and value-weighted portfolios, and equals 24 bps/month with a t-statistics of 2.30. Out of the twenty-five alphas reported in Panel A, the t-statistics for twenty-three exceed two, and for seventeen exceed three.

Panel B reports for these same strategies the average monthly net returns and the generalized net alphas of Novy-Marx and Velikov (2016). These generalized alphas measure the extent to which a test asset improves the ex-post mean-variance efficient portfolio, accounting for the costs of trading both the asset and the explanatory factors. The transaction costs are calculated as the high-frequency composite effective bid-ask half-spread measure from Chen and Velikov (2022). The net average returns reported in the first column range between 13-30bps/month. The lowest return, (13 bps/month), is achieved from the quintile sort using NYSE breakpoints and equal-weighted portfolios, and has an associated t-statistic of 1.09. Out of the twenty-five construction-methodology-factor-model pairs reported in Panel B, the TXCDLTT trading strategy improves the achievable mean-variance efficient frontier spanned by the factor models in twenty-four cases, and significantly expands the achievable frontier in twenty-one cases.

Table 3 provides direct tests for the role size plays in the TXCDLTT strategy performance. Panel A reports the average returns for the twenty-five portfolios constructed from a conditional double sort on size and TXCDLTT, as well as average returns and alphas for long/short trading TXCDLTT strategies within each size quintile. Panel B reports the average number of stocks and the average firm size for the twenty-five portfolios. Among the largest stocks (those with market capitalization greater than the 80^{th} NYSE percentile), the TXCDLTT strategy achieves an average return of 36 bps/month with a t-statistic of 3.26. Among these large cap stocks, the alphas for the TXCDLTT strategy relative to the five most common factor models range from 27 to 53 bps/month with t-statistics between 2.83 and 5.19.

4 How does TXCDLTT perform relative to the zoo?

Figure 2 puts the performance of TXCDLTT in context, showing the long/short strategy performance relative to other strategies in the "factor zoo." It shows Sharpe ratio histograms, both for gross and net returns (Panel A and B, respectively), for 212 documented anomalies in the zoo.[2] The vertical red line shows where the Sharpe ratio for the TXCDLTT strategy falls in the distribution. The TXCDLTT strategy's gross (net) Sharpe ratio of 0.36 (0.33) is greater than 79% (93%) of anomaly Sharpe ratios, respectively.

Figure 3 plots the growth of a $1 invested in these same 212 anomaly trading strategies (gray lines), and compares those with the growth of a $1 invested in the TXCDLTT strategy (red line).[3] Ignoring trading costs, a $1 invested in the TXCDLTT strategy would have yielded $5.33 which ranks the TXCDLTT strategy in the top 6% across the 212 anomalies. Accounting for trading costs, a $1 invested in the TXCDLTT strategy would have yielded $4.12 which ranks the TXCDLTT strategy in the top 4% across the 212 anomalies.

Figure 4 plots percentile ranks for the 212 anomaly trading strategies in terms of gross and Novy-Marx and Velikov (2016) net generalized alphas with respect to the CAPM, and the Fama-French three-, four-, five-, and six-factor models from Table 1, and indicates the ranking of the TXCDLTT relative to those. Panel A shows that the TXCDLTT strategy gross alphas fall between the 69 and 88 percentiles across the five factor models. Panel B shows that, accounting for trading costs, a large fraction of anomalies have not improved the investment opportunity set of an investor with

[2]The anomalies come from March, 2022 release of the Chen and Zimmermann (2022) open source asset pricing dataset.

[3]The figure assumes an initial investment of $1 in T-bills and $1 long/short in the two sides of the strategy. Returns are compounded each month, assuming, as in Detzel et al. (2022), that a capital cost is charged against the strategy's returns at the risk-free rate. This excess return corresponds more closely to the strategy's economic profitability.

access to the factor models over the 196306 to 202306 sample. For example, 47% (53%) of the 212 anomalies would not have improved the investment opportunity set for an investor having access to the Fama-French three-factor (six-factor) model. The TXCDLTT strategy has a positive net generalized alpha for five out of the five factor models. In these cases TXCDLTT ranks between the 87 and 96 percentiles in terms of how much it could have expanded the achievable investment frontier.

5 Does TXCDLTT add relative to related anomalies?

With so many anomalies, it is possible that any proposed, new cross-sectional predictor is just capturing some combination of known predictors. It is consequently natural to investigate to what extent the proposed predictor adds additional predictive power beyond the most closely related anomalies. Closely related anomalies are more likely to be formed on the basis of signals with higher absolute correlations. Figure 5 plots a name histogram of the correlations of TXCDLTT with 208 filtered anomaly signals.[4] Figure 6 also shows an agglomerative hierarchical cluster plot using Ward's minimum method and a maximum of 10 clusters.

A closely related anomaly is also more likely to price TXCDLTT or at least to weaken the power TXCDLTT has predicting the cross-section of returns. Figure 7 plots histograms of t-statistics for predictability tests of TXCDLTT conditioning on each of the 208 filtered anomaly signals one at a time. Panel A reports t-statistics on $\beta_{TXCDLTT}$ from Fama-MacBeth regressions of the form $r_{i,t} = \alpha + \beta_{TXCDLTT} TXCDLTT_{i,t} + \beta_X X_{i,t} + \epsilon_{i,t}$, where X stands for one of the 208 fil-

[4] When performing tests at the underlying signal level (e.g., the correlations plotted in Figure 5), we filter the 212 anomalies to avoid small sample issues. For each anomaly, we calculate the common stock observations in an average month for which both the anomaly and the test signal are available. In the filtered anomaly set, we drop anomalies with fewer than 100 common stock observations in an average month.

tered anomaly signals at a time. Panel B plots t-statistics on α from spanning tests of the form: $r_{TXCDLTT,t} = \alpha + \beta r_{X,t} + \epsilon_t$, where $r_{X,t}$ stands for the returns to one of the 208 filtered anomaly trading strategies at a time. The strategies employed in the spanning tests are constructed using quintile sorts, value-weighting, and NYSE breakpoints. Panel C plots t-statistics on the average returns to strategies constructed by conditional double sorts. In each month, we sort stocks into quintiles based one of the 208 filtered anomaly signals. Then, within each quintile, we sort stocks into quintiles based on TXCDLTT. Stocks are finally grouped into five TXCDLTT portfolios by combining stocks within each anomaly sorting portfolio. The panel plots the t-statistics on the average returns of these conditional double-sorted TXCDLTT trading strategies conditioned on each of the 208 filtered anomalies.

Table 4 reports Fama-MacBeth cross-sectional regressions of returns on TXCDLTT and the six anomalies most closely-related to it. The six most-closely related anomalies are picked as those with the highest combined rank where the ranks are based on the absolute value of the Spearman correlations in Panel B of Figure 5 and the R^2 from the spanning tests in Figure 7, Panel B. Controlling for each of these signals at a time, the t-statistics on the TXCDLTT signal in these Fama-MacBeth regressions exceed -2.61, with the minimum t-statistic occurring when controlling for Operating profitability R&D adjusted. Controlling for all six closely related anomalies, the t-statistic on TXCDLTT is -3.17.

Similarly, Table 5 reports results from spanning tests that regress returns to the TXCDLTT strategy onto the returns of the six most closely-related anomalies and the six Fama-French factors. Controlling for the six most-closely related anomalies individually, the TXCDLTT strategy earns alphas that range from 1-21bps/month. The minimum t-statistic on these alphas controlling for one anomaly at a time is 0.12, which is achieved when controlling for Operating profitability R&D adjusted. Controlling for all six closely-related anomalies and the six Fama-French factors simultaneously, the TXCDLTT trading strategy achieves an alpha of -4bps/month with a t-statistic of -0.61.

6 Does TXCDLTT add relative to the whole zoo?

Finally, we can ask how much adding TXCDLTT to the entire factor zoo could improve investment performance. Figure 8 plots the growth of $1 invested in trading strategies that combine multiple anomalies following Chen and Velikov (2022). The combinations use either the 154 anomalies from the zoo that satisfy our inclusion criteria (blue lines) or these 154 anomalies augmented with the TXCDLTT signal.[5] We consider six different methods for combining signals.

Panel A shows results using "Average rank" as the combination method. This method sorts stocks on the basis of forecast excess returns, where these are calculated on the basis of their average cross-sectional percentile rank across return predictors, and the predictors are all signed so that higher ranks are associated with higher average returns. For this method, $1 investment in the 154-anomaly combination strategy grows to $3258.25, while $1 investment in the combination strategy that includes TXCDLTT grows to $3672.45.

Panel B shows results using "Weighted-Average rank" as the combination method. This method sorts stocks on the basis of forecast excess returns, where these are calculated as weighted-average cross-sectional percentile rank across return predictors, and the predictors are all signed so that higher ranks are associated with higher average returns and the weights are determined by the average returns over the past ten years to the long/short strategies based on the individual signals. For this method, $1 investment in the 154-anomaly combination strategy grows to $193.64, while $1 investment in the combination strategy that includes TXCDLTT grows to $252.22.

[5] We filter the 207 Chen and Zimmermann (2022) anomalies and require for each anomaly the average month to have at least 40% of the cross-sectional observations available for market capitalization on CRSP in the period for which TXCDLTT is available.

Panel C shows results using "Fama-MacBeth" as the combination method. This method sorts stocks on the basis of forecast excess returns, where these are calculated from Fama and MacBeth (1973) regressions following Haugen and Baker (1996) and Lewellen (2015) using only data in the investor's information set at the time of portfolio formation. The estimation uses rolling ten years of data, so the actual strategies begin ten years later for this combination method. For this method, \$1 investment in the 154-anomaly combination strategy grows to \$160154.97, while \$1 investment in the combination strategy that includes TXCDLTT grows to \$195460.94.

Panel D shows results using "Partial Least Squares" as the combination method. This method sorts stocks on the basis of forecast excess returns, where these are calculated from partial least squares (PLS) filtering procedure following Light et al. (2017) using only data in the investor's information set at the time of portfolio formation. The estimation uses rolling ten years of data, so the actual strategies begin ten years later for this combination method. For this method, \$1 investment in the 154-anomaly combination strategy grows to \$173.86, while \$1 investment in the combination strategy that includes TXCDLTT grows to \$198.14.

Panel E shows results using "IPCA" as the combination method. This method sorts stocks on the basis of forecast excess returns, where these are calculated from the instrumented principal component analysis (IPCA) procedure of Kelly et al. (2019) using only data in the investor's information set at the time of portfolio formation. The estimation uses rolling ten years of data, so the actual strategies begin ten years later for this combination method. For this method, \$1 investment in the 154-anomaly combination strategy grows to \$18077.21, while \$1 investment in the combination strategy that includes TXCDLTT grows to \$14756.88.

Panel F shows results using "LASSO" as the combination method. This method sorts stocks on the basis of forecast excess returns, where these are estimated by least absolute shrinkage and selection operator (LASSO) using only data in the investor's information set at the time of portfolio formation. Following Chen and Velikov (2022), LASSO penalty (λ) is selected by minimizing the mean squared error (MSE) estimated by 5-fold cross validation. The estimation uses rolling ten years of data, so the actual strategies begin ten years later for this combination method. For this method, \$1 investment in the 154-anomaly combination strategy grows to \$50687.59, while \$1 investment in the combination strategy that includes TXCDLTT grows to \$47179.84.

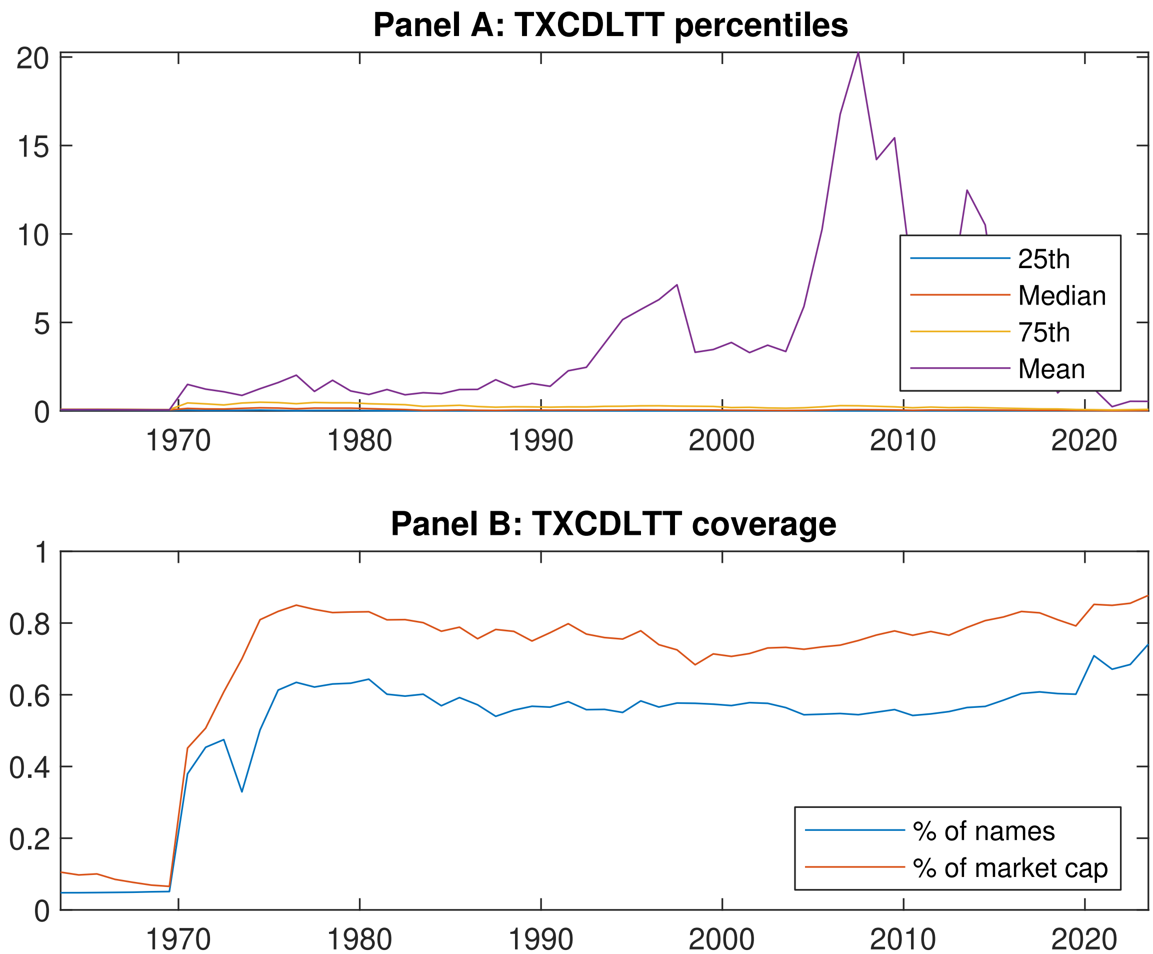

Figure 1: Times series of TXCDLTT percentiles and coverage.
This figure plots descriptive statistics for TXCDLTT. Panel A shows cross-sectional percentiles of TXCDLTT over the sample. Panel B plots the monthly coverage of TXCDLTT relative to the universe of CRSP stocks with available market capitalizations.

Table 1: Basic sort: VW, quintile, NYSE-breaks

This table reports average excess returns and alphas for portfolios sorted on TXCDLTT. At the end of each month, we sort stocks into five portfolios based on their signal using NYSE breakpoints. Panel A reports average value-weighted quintile portfolio (L,2,3,4,H) returns in excess of the risk-free rate, the long-short extreme quintile portfolio (H-L) return, and alphas with respect to the CAPM, Fama and French (1993) three-factor model, Fama and French (1993) three-factor model augmented with the Carhart (1997) momentum factor, Fama and French (2015) five-factor model, and the Fama and French (2015) five-factor model augmented with the Carhart (1997) momentum factor following Fama and French (2018). Panel B reports the factor loadings for the quintile portfolios and long-short extreme quintile portfolio in the Fama and French (2015) five-factor model. Panel C reports the average number of stocks and market capitalization of each portfolio. T-statistics are in brackets. The sample period is 196306 to 202306.

Panel A: Excess returns and alphas on TXCDLTT-sorted portfolios						
	(L)	(2)	(3)	(4)	(H)	(H-L)
r^e	0.28	0.44	0.61	0.58	0.58	0.30
	[1.43]	[2.52]	[3.60]	[3.53]	[3.52]	[2.81]
α_{CAPM}	-0.33	-0.10	0.07	0.05	0.05	0.38
	[-3.92]	[-1.39]	[1.16]	[0.95]	[0.93]	[3.65]
α_{FF3}	-0.40	-0.18	0.04	0.05	0.12	0.52
	[-4.89]	[-2.58]	[0.63]	[0.94]	[2.16]	[5.40]
α_{FF4}	-0.32	-0.11	0.06	0.05	0.15	0.47
	[-3.91]	[-1.63]	[1.03]	[0.98]	[2.68]	[4.83]
α_{FF5}	-0.22	-0.16	-0.05	-0.07	0.04	0.26
	[-3.11]	[-2.35]	[-0.76]	[-1.36]	[0.72]	[3.12]
α_{FF6}	-0.16	-0.11	-0.01	-0.06	0.07	0.24
	[-2.32]	[-1.55]	[-0.25]	[-1.08]	[1.32]	[2.82]
Panel B: Fama and French (2018) 6-factor model loadings for TXCDLTT-sorted portfolios						
β_{MKT}	1.03	1.00	1.02	0.99	0.96	-0.08
	[62.06]	[59.86]	[71.52]	[76.05]	[72.12]	[-3.90]
β_{SMB}	0.05	-0.09	-0.12	-0.05	-0.11	-0.16
	[1.90]	[-3.71]	[-5.62]	[-2.40]	[-5.76]	[-5.42]
β_{HML}	0.06	0.13	-0.00	-0.09	-0.23	-0.29
	[1.95]	[4.13]	[-0.00]	[-3.68]	[-9.14]	[-7.71]
β_{RMW}	-0.55	-0.11	0.08	0.21	0.13	0.68
	[-16.80]	[-3.52]	[2.84]	[8.22]	[5.13]	[17.46]
β_{CMA}	0.12	0.17	0.28	0.26	0.18	0.06
	[2.58]	[3.66]	[7.06]	[7.09]	[4.95]	[1.13]
β_{UMD}	-0.08	-0.08	-0.04	-0.02	-0.05	0.03
	[-4.92]	[-5.04]	[-3.18]	[-1.67]	[-3.77]	[1.61]
Panel C: Average number of firms (n) and market capitalization (me)						
n	775	394	354	375	516	
me ($\$10^6$)	679	1005	1390	1935	2668	

Table 2: Robustness to sorting methodology & trading costs

This table evaluates the robustness of the choices made in the TXCDLTT strategy construction methodology. In each panel, the first row shows results from a quintile, value-weighted sort using NYSE break points as employed in Table 1. Each of the subsequent rows deviates in one of the three choices at a time, and the choices are specified in the first three columns. For each strategy construction methodology, the table reports average excess returns and alphas with respect to the CAPM, Fama and French (1993) three-factor model, Fama and French (1993) three-factor model augmented with the Carhart (1997) momentum factor, Fama and French (2015) five-factor model, and the Fama and French (2015) five-factor model augmented with the Carhart (1997) momentum factor following Fama and French (2018). Panel A reports average returns and alphas with no adjustment for trading costs. Panel B reports net average returns and Novy-Marx and Velikov (2016) generalized alphas as prescribed by Detzel et al. (2022). T-statistics are in brackets. The sample period is 196306 to 202306.

Panel A: Gross Returns and Alphas								
Portfolios	Breaks	Weights	r^e	α_{CAPM}	α_{FF3}	α_{FF4}	α_{FF5}	α_{FF6}
Quintile	NYSE	VW	0.30	0.38	0.52	0.47	0.26	0.24
			[2.81]	[3.65]	[5.40]	[4.83]	[3.12]	[2.82]
Quintile	NYSE	EW	0.29	0.37	0.44	0.34	0.18	0.11
			[2.42]	[3.18]	[4.06]	[3.14]	[1.83]	[1.16]
Quintile	Name	VW	0.34	0.42	0.54	0.50	0.30	0.28
			[2.90]	[3.61]	[5.16]	[4.69]	[3.22]	[2.98]
Quintile	Cap	VW	0.24	0.27	0.44	0.41	0.29	0.28
			[2.30]	[2.61]	[4.91]	[4.47]	[3.43]	[3.17]
Decile	NYSE	VW	0.29	0.38	0.53	0.48	0.30	0.28
			[2.15]	[2.83]	[4.29]	[3.86]	[2.61]	[2.38]
Panel B: Net Returns and Novy-Marx and Velikov (2016) generalized alphas								
Portfolios	Breaks	Weights	r^e_{net}	α^*_{CAPM}	α^*_{FF3}	α^*_{FF4}	α^*_{FF5}	α^*_{FF6}
Quintile	NYSE	VW	0.27	0.35	0.46	0.43	0.24	0.22
			[2.53]	[3.31]	[4.79]	[4.50]	[2.88]	[2.72]
Quintile	NYSE	EW	0.13	0.20	0.24	0.19	0.01	
			[1.09]	[1.62]	[2.15]	[1.70]	[0.06]	
Quintile	Name	VW	0.30	0.38	0.48	0.45	0.28	0.26
			[2.56]	[3.24]	[4.56]	[4.34]	[2.97]	[2.83]
Quintile	Cap	VW	0.21	0.24	0.39	0.37	0.26	0.25
			[2.08]	[2.37]	[4.31]	[4.09]	[3.03]	[2.89]
Decile	NYSE	VW	0.25	0.33	0.46	0.43	0.27	0.25
			[1.83]	[2.46]	[3.73]	[3.51]	[2.36]	[2.21]

Table 3: Conditional sort on size and TXCDLTT

This table presents results for conditional double sorts on size and TXCDLTT. In each month, stocks are first sorted into quintiles based on size using NYSE breakpoints. Then, within each size quintile, stocks are further sorted based on TXCDLTT. Finally, they are grouped into twenty-five portfolios based on the intersection of the two sorts. Panel A presents the average returns to the 25 portfolios, as well as strategies that go long stocks with high TXCDLTT and short stocks with low TXCDLTT .Panel B documents the average number of firms and the average firm size for each portfolio. The sample period is 196306 to 202306.

Panel A: portfolio average returns and time-series regression results

	TXCDLTT Quintiles					TXCDLTT Strategies					
Size quintiles	(L)	(2)	(3)	(4)	(H)	r^e	α_{CAPM}	α_{FF3}	α_{FF4}	α_{FF5}	α_{FF6}
(1)	0.48	0.18	0.60	0.80	0.82	0.34	0.44	0.47	0.48	0.19	0.22
	[1.69]	[0.58]	[2.37]	[3.48]	[3.43]	[2.43]	[3.21]	[3.50]	[3.47]	[1.50]	[1.67]
(2)	0.60	0.67	0.65	0.79	0.72	0.11	0.22	0.24	0.26	0.03	0.06
	[2.19]	[2.82]	[2.93]	[3.46]	[3.23]	[0.82]	[1.68]	[1.80]	[1.97]	[0.22]	[0.48]
(3)	0.48	0.68	0.73	0.79	0.67	0.19	0.26	0.31	0.29	0.07	0.06
	[2.02]	[3.52]	[3.60]	[3.84]	[3.19]	[1.54]	[2.20]	[2.59]	[2.33]	[0.59]	[0.54]
(4)	0.46	0.58	0.68	0.72	0.66	0.20	0.20	0.28	0.24	0.07	0.05
	[2.20]	[3.10]	[3.58]	[3.73]	[3.23]	[1.79]	[1.79]	[2.54]	[2.14]	[0.66]	[0.48]
(5)	0.24	0.51	0.58	0.49	0.60	0.36	0.39	0.53	0.47	0.31	0.27
	[1.33]	[2.91]	[3.48]	[2.99]	[3.54]	[3.26]	[3.49]	[5.19]	[4.49]	[3.26]	[2.83]

Panel B: Portfolio average number of firms and market capitalization

	TXCDLTT Quintiles					TXCDLTT Quintiles				
	Average n					Average market capitalization ($\$10^6$)				
Size quintiles	(L)	(2)	(3)	(4)	(H)	(L)	(2)	(3)	(4)	(H)
(1)	251	258	260	267	259	17	16	20	24	24
(2)	77	78	77	80	76	39	40	40	42	40
(3)	56	56	56	57	57	70	72	72	73	72
(4)	49	49	49	49	50	159	163	161	168	165
(5)	47	47	47	47	47	809	1030	1151	1405	1813

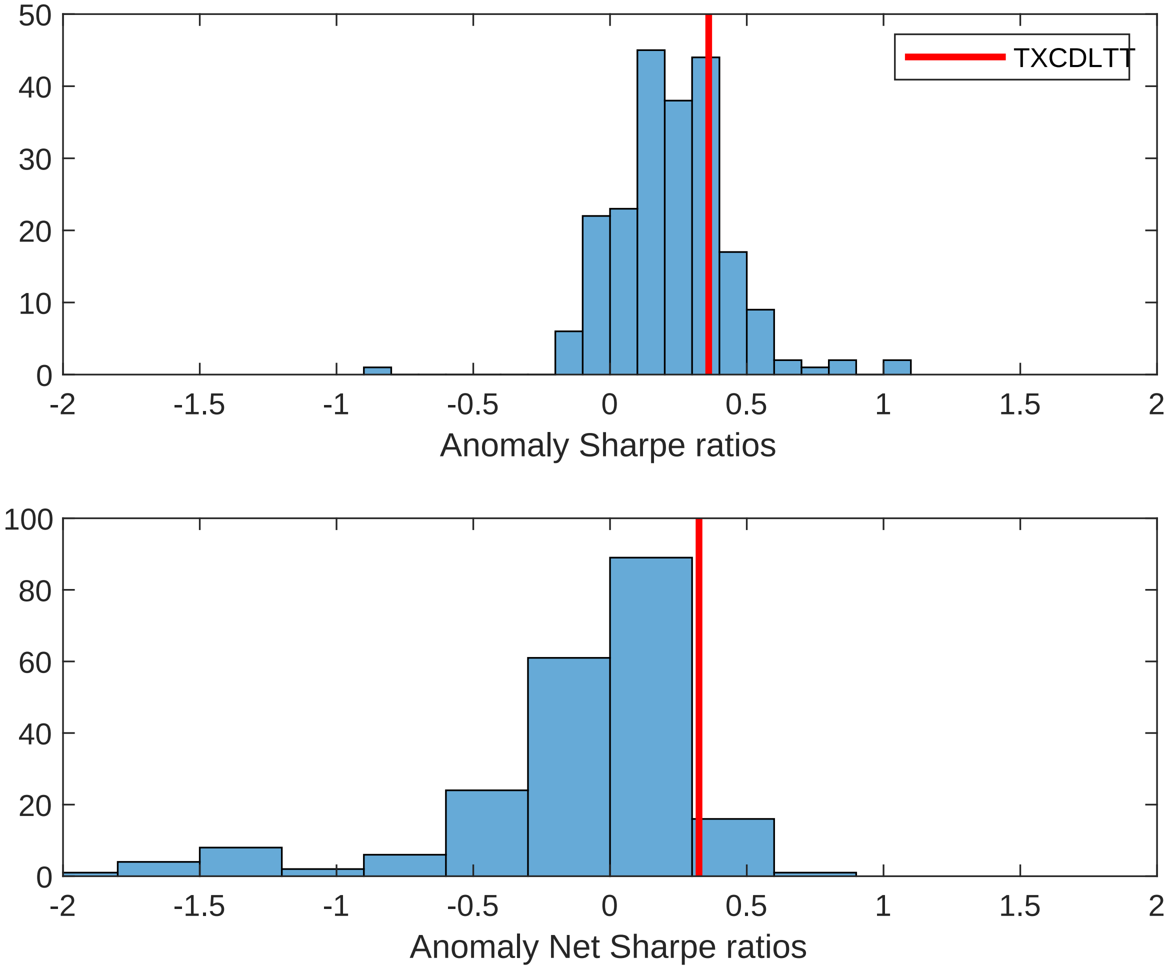

Figure 2: Distribution of Sharpe ratios.
This figure plots a histogram of Sharpe ratios for 212 anomalies, and compares the Sharpe ratio of the TXCDLTT with them (red vertical line). Panel A plots results for gross Sharpe ratios. Panel B plots results for net Sharpe ratios.

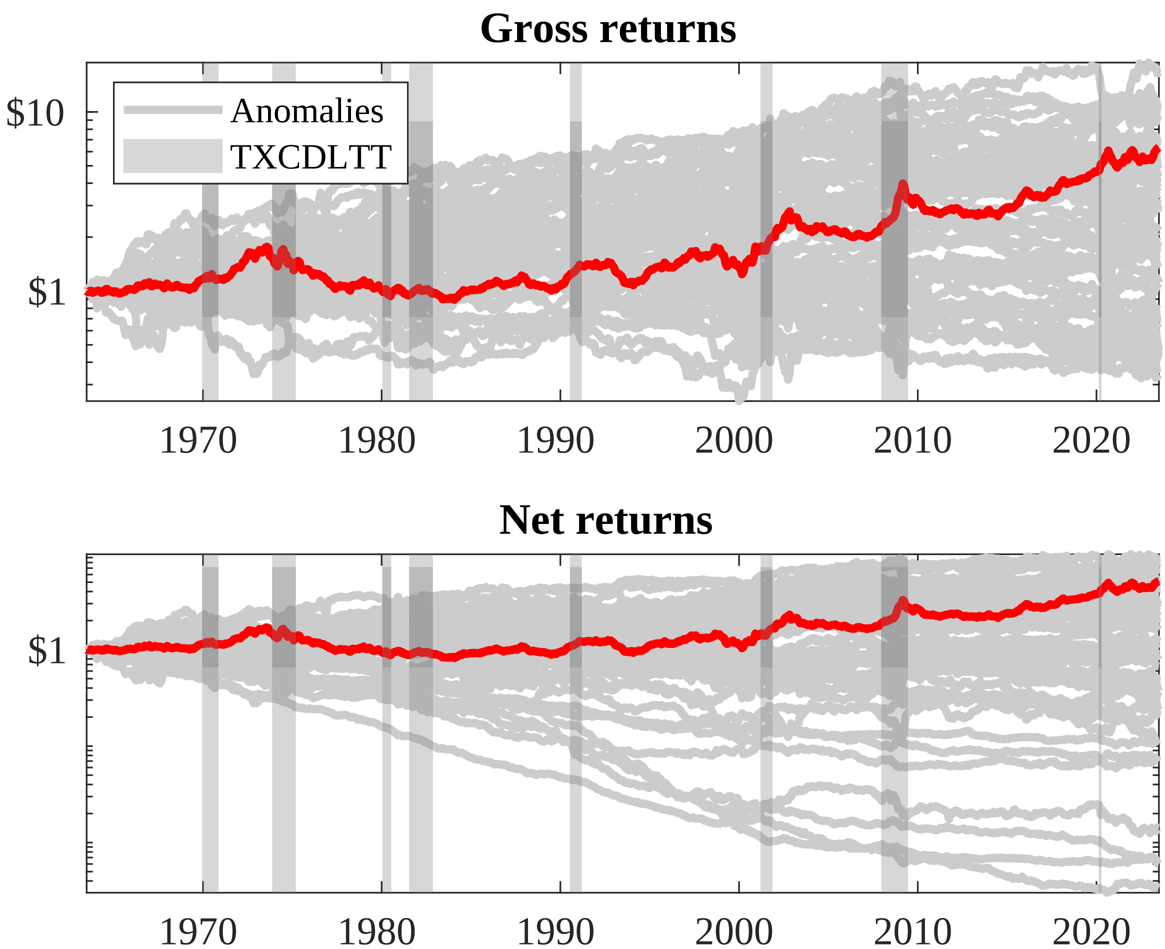

Figure 3: Dollar invested.
This figure plots the growth of a $1 invested in 212 anomaly trading strategies (gray lines), and compares those with the TXCDLTT trading strategy (red line). The strategies are constructed using value-weighted quintile sorts using NYSE break-points. Panel A plots results for gross strategy returns. Panel B plots results for net strtaegy returns.

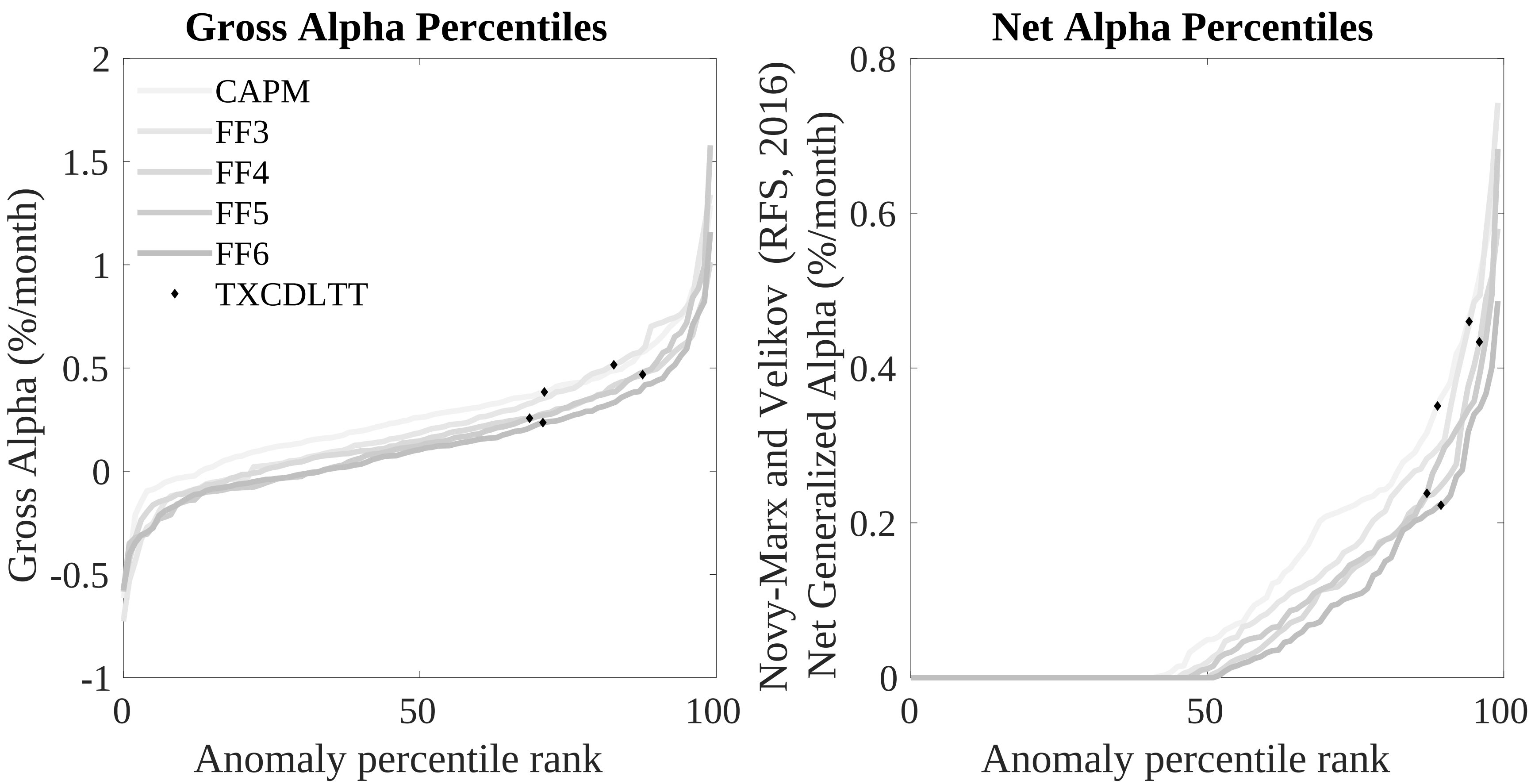

Figure 4: Gross and generalized net alpha percentiles of anomalies relative to factor models

This figure plots the percentile ranks for 212 anomaly trading strategies in terms of alphas (solid lines), and compares those with the TXCDLTT trading strategy alphas (diamonds). The strategies are constructed using value-weighted quintile sorts using NYSE breakpoints. The alphas include those with respect to the CAPM, Fama and French (1993) three-factor model, Fama and French (1993) three-factor model augmented with the Carhart (1997) momentum factor, Fama and French (2015) five-factor model, and the Fama and French (2015) five-factor model augmented with the Carhart (1997) momentum factor following Fama and French (2018). The left panel plots alphas with no adjustment for trading costs. The right panel plots Novy-Marx and Velikov (2016) net generalized alphas.

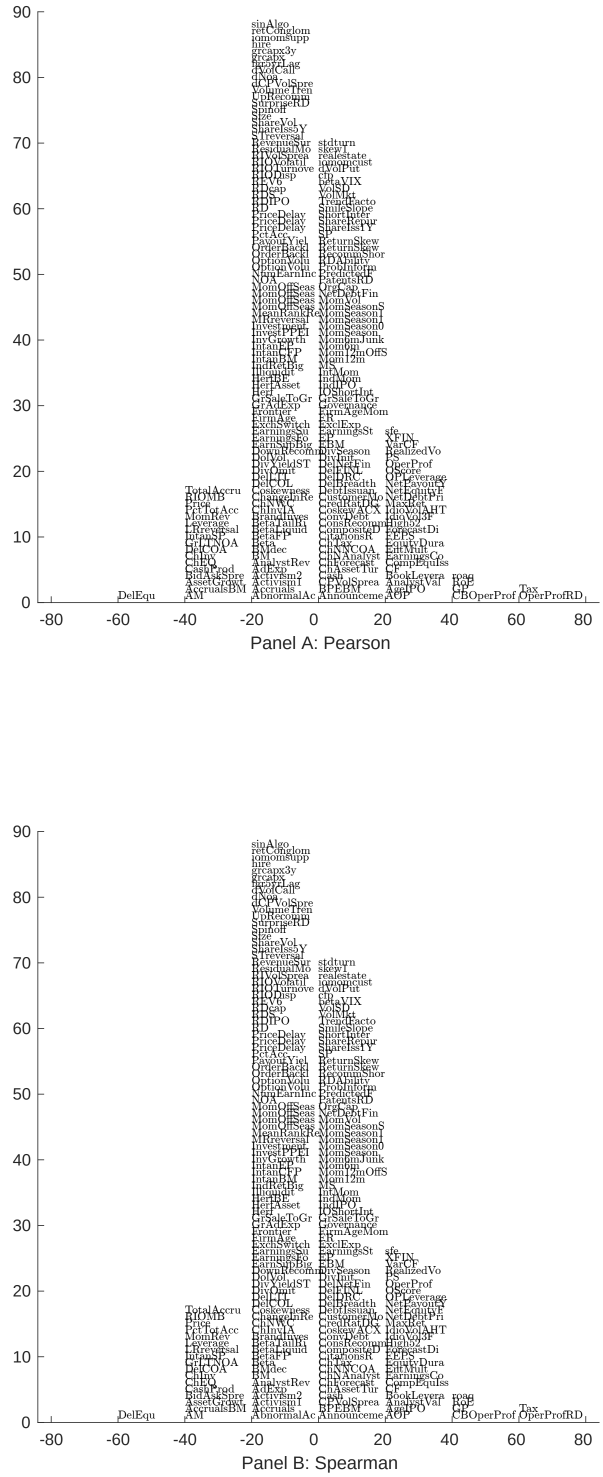

Figure 5: Distribution of correlations.
This figure plots a name histogram of correlations of 208 filtered anomaly signals with TXCDLTT. The correlations are pooled. Panel A plots Pearson correlations, while Panel B plots Spearman rank correlations.

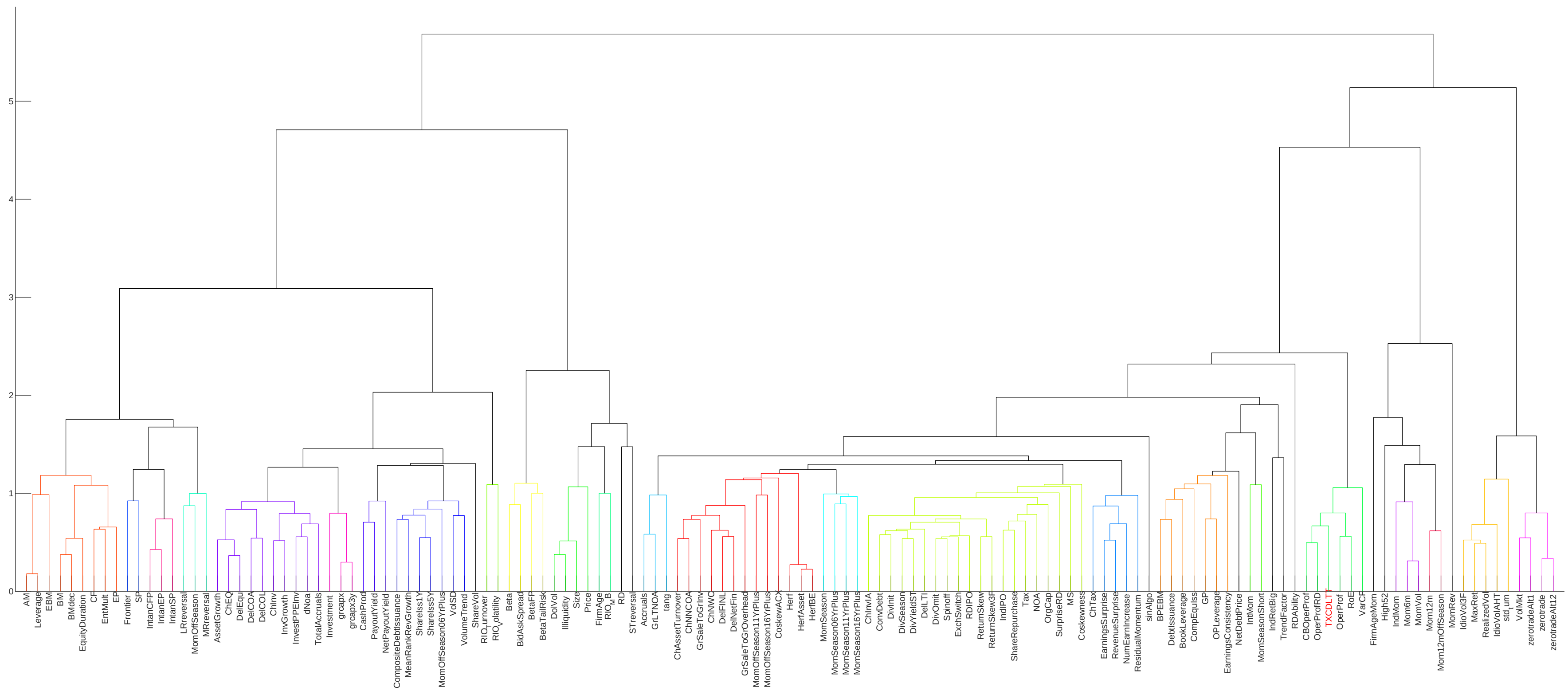

Figure 6: Agglomerative hierarchical cluster plot

This figure plots an agglomerative hierarchical cluster plot using Ward's minimum method and a maximum of 10 clusters.

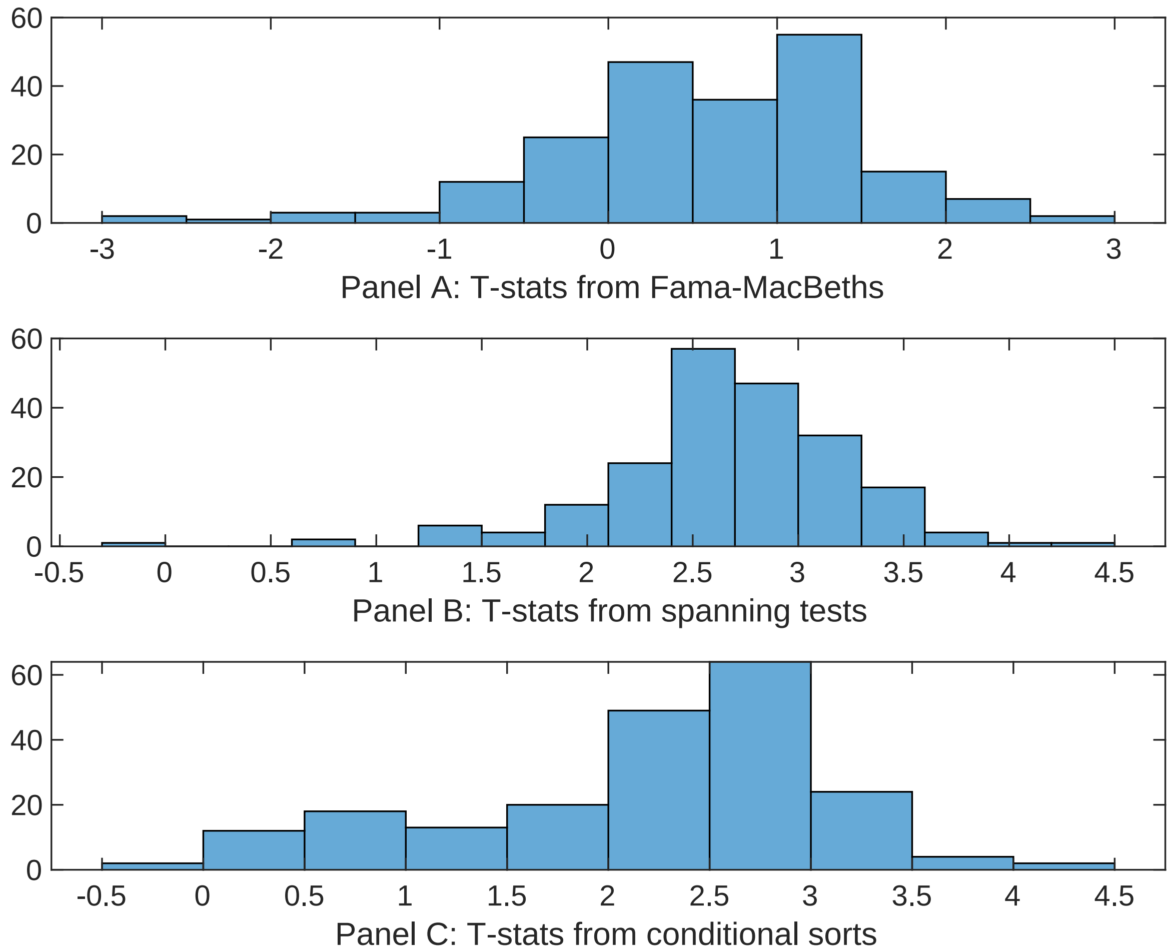

Figure 7: Distribution of t-stats on conditioning strategies
This figure plots histograms of t-statistics for predictability tests of TXCDLTT conditioning on each of the 208 filtered anomaly signals one at a time. Panel A reports t-statistics on $\beta_{TXCDLTT}$ from Fama-MacBeth regressions of the form $r_{i,t} = \alpha + \beta_{TXCDLTT} TXCDLTT_{i,t} + \beta_X X_{i,t} + \epsilon_{i,t}$, where X stands for one of the 208 filtered anomaly signals at a time. Panel B plots t-statistics on α from spanning tests of the form: $r_{TXCDLTT,t} = \alpha + \beta r_{X,t} + \epsilon_t$, where $r_{X,t}$ stands for the returns to one of the 208 filtered anomaly trading strategies at a time. The strategies employed in the spanning tests are constructed using quintile sorts, value-weighting, and NYSE breakpoints. Panel C plots t-statistics on the average returns to strategies constructed by conditional double sorts. In each month, we sort stocks into quintiles based one of the 208 filtered anomaly signals at a time. Then, within each quintile, we sort stocks into quintiles based on TXCDLTT. Stocks are finally grouped into five TXCDLTT portfolios by combining stocks within each anomaly sorting portfolio. The panel plots the t-statistics on the average returns of these conditional double-sorted TXCDLTT trading strategies conditioned on each of the 208 filtered anomalies.

Table 4: Fama-MacBeths controlling for most closely related anomalies

This table presents Fama-MacBeth results of returns on TXCDLTT. and the six most closely related anomalies. The regressions take the following form: $r_{i,t} = \alpha + \beta_{TXCDLTT} TXCDLTT_{i,t} + \sum_{k=1}^{s} ix\beta_{X_k} X_{i,t}^{k} + \epsilon_{i,t}$. The six most closely related anomalies, X, are Operating profitability R&D adjusted, net income / book equity, Cash-based operating profitability, gross profits / total assets, Taxable income to income, operating profits / book equity. These anomalies were picked as those with the highest combined rank where the ranks are based on the absolute value of the Spearman correlations in Panel B of Figure 5 and the R^2 from the spanning tests in Figure 7, Panel B. The sample period is 196306 to 202306.

Intercept	0.11	0.12	0.97	0.86	0.11	0.10	0.67
	[3.69]	[4.97]	[3.83]	[3.53]	[4.24]	[4.15]	[2.45]
TXCDLTT	-0.18	0.19	0.45	0.11	0.26	-0.41	-0.18
	[-2.61]	[1.48]	[0.15]	[0.37]	[0.29]	[-0.55]	[-3.17]
Anomaly 1	0.16						0.46
	[3.37]						[0.82]
Anomaly 2		-0.37					-0.20
		[-1.04]					[-1.14]
Anomaly 3			0.19				0.16
			[2.71]				[4.29]
Anomaly 4				0.73			0.32
				[1.18]			[1.62]
Anomaly 5					0.10		0.43
					[3.06]		[1.76]
Anomaly 6						0.37	0.14
						[3.20]	[0.84]
# months	631	684	673	684	636	631	631
$\bar{R}^2(\%)$	1	1	1	0	0	0	0

Table 5: Spanning tests controlling for most closely related anomalies

This table presents spanning tests results of regressing returns to the TXCDLTT trading strategy on trading strategies exploiting the six most closely related anomalies. The regressions take the following form: $r_t^{TXCDLTT} = \alpha + \sum_{k=1}^{6} \beta_{X_k} r_t^{X_k} + \sum_{j=1}^{6} \beta_{f_j} r_t^{f_j} + \epsilon_t$, where X_k indicates each of the six most-closely related anomalies and f_j indicates the six factors from the Fama and French (2015) five-factor model augmented with the Carhart (1997) momentum factor. The six most closely related anomalies, X, are Operating profitability R&D adjusted, net income / book equity, Cash-based operating profitability, gross profits / total assets, Taxable income to income, operating profits / book equity. These anomalies were picked as those with the highest combined rank where the ranks are based on the absolute value of the Spearman correlations in Panel B of Figure 5 and the R^2 from the spanning tests in Figure 7, Panel B. The sample period is 196306 to 202306.

Intercept	0.04	0.21	0.01	0.08	0.18	0.18	-0.04
	[0.50]	[2.60]	[0.12]	[1.05]	[2.27]	[2.22]	[-0.61]
Anomaly 1	44.31						13.06
	[13.68]						[2.94]
Anomaly 2		31.27					9.86
		[7.03]					[2.38]
Anomaly 3			50.50				21.10
			[14.60]				[4.68]
Anomaly 4				45.36			26.00
				[15.25]			[8.28]
Anomaly 5					36.44		21.32
					[10.11]		[6.45]
Anomaly 6						29.39	-2.70
						[7.74]	[-0.68]
mkt	-0.16	-3.71	-3.19	-8.81	-7.25	-4.21	-2.42
	[-0.08]	[-1.82]	[-1.78]	[-5.07]	[-3.87]	[-2.11]	[-1.40]
smb	2.00	-5.91	1.74	-18.65	-14.94	-8.01	-2.15
	[0.70]	[-1.90]	[0.62]	[-7.41]	[-5.53]	[-2.72]	[-0.76]
hml	-15.60	-26.90	-15.87	-10.97	-29.76	-27.79	-7.63
	[-4.34]	[-7.20]	[-4.50]	[-3.08]	[-8.31]	[-7.46]	[-2.30]
rmw	38.91	39.87	48.02	42.46	51.81	41.02	19.43
	[9.42]	[7.10]	[12.86]	[11.13]	[12.84]	[7.84]	[3.91]
cma	6.97	14.04	-1.85	13.63	6.10	6.88	9.16
	[1.37]	[2.53]	[-0.37]	[2.77]	[1.16]	[1.26]	[1.96]
umd	-1.17	2.02	-1.30	1.82	1.82	0.02	-1.38
	[-0.66]	[1.06]	[-0.73]	[1.06]	[0.98]	[0.01]	[-0.86]
# months	716	720	716	720	720	716	716
$\bar{R}^2$(%)	57	50	59	60	53	50	67

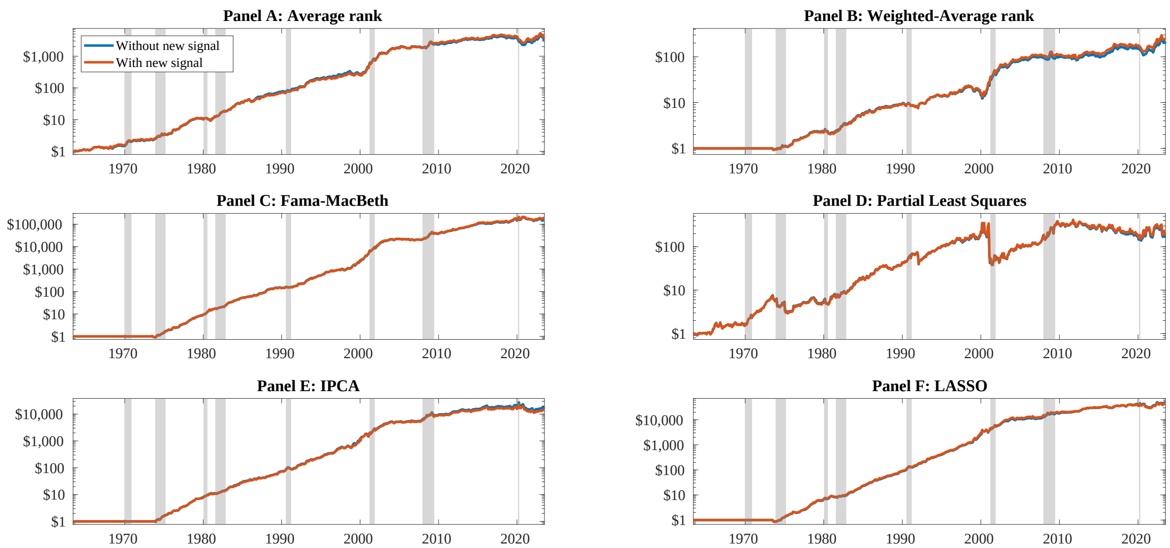

Figure 8: Combination strategy performance

This figure plots the growth of a $1 invested in trading strategies that combine multiple anomalies following Chen and Velikov (2022). In all panels, the blue solid lines indicate combination trading strategies that utilize 154 anomalies. The red solid lines indicate combination trading strategies that utilize the 154 anomalies as well as TXCDLTT. Panel A shows results using "Average rank" as the combination method. Panel B shows results using "Weighted-Average rank" as the combination method. Panel C shows results using "Fama-MacBeth" as the combination method. Panel D shows results using "Partial Least Squares" as the combination method. Panel E shows results using "IPCA" as the combination method. Panel F shows results using "LASSO" as the combination method. See Section 6 for details on the combination methods.

References

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52:57–82.

Chen, A. and Velikov, M. (2022). Zeroing in on the expected returns of anomalies. *Journal of Financial and Quantitative Analysis*, Forthcoming.

Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, 27(2):207–264.

Detzel, A., Novy-Marx, R., and Velikov, M. (2022). Model comparison with transaction costs. *Journal of Finance, Forthcoming*.

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

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

Fama, E. F. and French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, 128(2):234–252.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, 81(3):607–636.

Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3):401–439.

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

Lewellen, J. (2015). The cross-section of expected returns. *Critical Finance Review*, 4(1):1–44.

Light, N., Maslov, D., and Rytchkov, O. (2017). Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of Financial Studies*, 30:1339–1381.

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

Novy-Marx, R. and Velikov, M. (2023). Assaying anomalies. *Working paper*.