Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Open Source Cross-Sectional Asset Pricing

Andrew Y. Chen and Tom Zimmermann

2021-037

Please cite this paper as:
Chen, Andrew Y., and Tom Zimmermann (2021). "Open Source Cross-Sectional Asset Pricing," Finance and Economics Discussion Series 2021-037. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2021.037.

Open Source Cross-Sectional Asset Pricing

Andrew Y. Chen
Federal Reserve Board
andrew.y.chen@frb.gov

Tom Zimmermann
University of Cologne
tom.zimmermann@uni-koeln.de

March 2021*

Abstract

We provide data and code that successfully reproduces nearly all cross-sectional stock return predictors. Our 319 characteristics draw from previous meta-studies, but we differ by comparing our t-stats to the original papers' results. For the 161 characteristics that were clearly significant in the original papers, 98% of our long-short portfolios find t-stats above 1.96. For the 44 characteristics that had mixed evidence, our reproductions find t-stats of 2 on average. A regression of reproduced t-stats on original long-short t-stats finds a slope of 0.90 and an R^2 of 83%. Mean returns are monotonic in predictive signals at the characteristic level. The remaining 114 characteristics were insignificant in the original papers or are modifications of the originals created by Hou, Xue, and Zhang (2020). These remaining characteristics are almost always significant if the original characteristic was also significant.

*First posted to SSRN: May 18, 2020. We thank Ivo Welch for encouraging us to work on this project, and thank Shane Corwin, Paul Schultz, Luis Palacios, Rabih Moussawi, and Denys Glushkov for making their code publically available. We also thank an anonymous referee, Guillaume Coqueret, Sebastian Hillenbrand, Liang Jiang, Yang Liu (from Tsinghua Finance), Yue Zhao, and anonymous contributors to our GitHub repo for helpful comments. We are grateful to Sebastian Weilbels and Arne Rodloff for excellent research assistance. The project received support from the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy - EXC2126/1-39083886. The views expressed herein are those of the authors and do not necessarily reflect the position of the Board of Governors of the Federal Reserve or the Federal Reserve System.

1. Introduction

Academic finance progresses through a mixture of open collaboration and closed competition. In this paper, we attempt to push the culture toward open collaboration by providing an “open source dataset” of hundreds of predictors of the cross-section of stock returns.

In our view, an open source dataset is essential because recent studies cast doubt on the credibility of the entire cross-sectional asset pricing literature. A series of influential critiques argue that *at least* 45% of the findings in this literature are false (Harvey, Liu, and Zhu 2016, Linnainmaa and Roberts 2018, Chordia, Goyal, and Saretto 2020).[1] Hou, Xue, and Zhang (2020) go further and claim that roughly 50% of the literature cannot be replicated—even when following the original methodologies.[2] This finding implies that much of the literature fails to survive reproduction, much less reanalysis.[3] These doubts echo the ongoing credibility crises in other fields ranging from medicine to management (Ioannidis 2005; Nosek, Spies, and Motyl 2012; Bettis 2012).

Our open source dataset takes a key step toward restoring the credibility of cross-sectional asset pricing. It shows that nearly 100% of the literature’s predictability results can be reproduced, including the predictability results for 100% of the characteristics studied in Hou, Xue, and Zhang (2020). Readers may be highly skeptical of our findings, given the critiques noted above. However, they do not need to take our word for it. Anyone with access to WRDS and Stata can perform this massive replication themselves using our code at https://github.com/OpenSourceAP/CrossSection.

Indeed, our code is written with the user in mind. The code is modular, so users can quickly examine a particular characteristic without worrying about most of the code. The code uses exception handling to gracefully move past er-

[1] Harvey, Liu, and Zhu’s (2016)“argue that most claimed researching findings in financial economics are likely false.” Chordia, Goyal, and Saretto (2020) “estimate the expected proportion of false rejections that researchers would produce if they failed to account for multiple hypothesis testing to be about 45%.” Linnainmaa and Roberts (2018) “show the majority of accounting-based return anomalies, including investment, are most likely an artifact of data snooping.”

[2] Hou, Xue, and Zhang (2020) (HXZ) state “[r]epeating our tests on the shorter samples in the original studies, we find that 65.4% of anomalies cannot clear the single test hurdle of $|t| \geq 1.96$ with NYSE breakpoints and value-weighted returns. The failure rate drops to 43.1% if we allow microcaps to run amok with NYSE-Amex-NADSQ breakpoints and equal-weighted returns.

[3] Following Welch’s (2019) terminology, we use the term “reproduction” to refer to an attempt to produce the same result using the same sample with the same code. Reanalysis refers to a broader concept that includes extensions and re-examinations.

rors and semantic versioning to make updates easy to understand.[4] Last, we are committed to updating these data on an annual basis. We hope this demonstration of open collaboration will inspire others to open up their analyses, and further support the credibility of academic finance.[5]

Our code produces 319 firm-level characteristics, and only *three* fail to reproduce the original paper's evidence of statistical significance for long-short portfolio returns. Our t-stats even match the originals quantitatively: A regression of reproduced t-stats on hand-collected t-stats finds a slope of 0.90 and an R^2 of 83%. We also find that mean returns are nicely monotonic in the predictors. The probability that predictor decile k has a higher mean return than decile $k-2$ is roughly 96%.

This near-100% reproduction success rate was made possible by a careful examination of the original papers. For each of our 319 signals, we hand-collect the key table, empirical test, the sign of predictability, t-stat, and other details from the original papers. This hand-collected data is also part of our open source dataset, and available for public inspection.[6]

The hand-collected data shows that characteristics in previous meta-studies vary wildly in their original predictability evidence. They include everything from dividend seasonality, which was shown to produce a long-short t-stat of 16.2 (Hartzmark and Solomon 2013), to R&D to sales, which utterly failed to predict returns (Chan, Lakonishok, and Sougiannis 2001). Many others were never explicitly tested for return predictability (e.g. Francis et al.'s (2004) accrual quality). As a result, judging the reproducibility of these characteristics cannot come down to a single t-stat > 1.96 rule.[7]

Instead, we apply the t-stat > 1.96 rule only when it is appropriate. This rule is appropriate for the 161 characteristics we categorize as "clear predictors." For these characteristics, the original papers showed clear evidence of significant predictability for our long-short portfolios. We find that 158 of 161 clear predic-

[4]The current version of the code is v1.0.0, indicating a major revision to the previous v0.1.2. The modular structure and exception handling was part of this revision.

[5]Jensen, Kelly, and Pedersen (2021) is another open source project that provides code for many cross-sectional predictors. Harvey and Liu (2019) provides crowd-sourced hand-collected data on cross-sectional asset pricing t-stats.

[6]Hand collected data is here:
https://github.com/OpenSourceAP/CrossSection/raw/master/SignalDocumentation.xlsx

[7]Throughout the paper, we sign predictor portfolios to have positive mean returns based on the original studies. Thus, the rule does not involve the absolute value, and t-stat < -1.96 typically indicates a reproduction failure.

tors meet t-stat > 1.96, and one of the three which did not still achieved a t-stat of 1.93.

A more subtle evaluation is required for our 44 characteristics that had mixed evidence of predictability in the original papers. These "likely predictors" include 52-week high, with its original long-short t-stat of 2.00 (George and Hwang 2004), and sales growth over inventory growth, with its original t-stat of 2.4 in a multivariate regression (Abarbanell and Bushee 1998). Evaluating the reproducibility of these t-stats is a subtle exercise, as an immaterial change can cause the 52-week high to dip below $t = 1.96$, and we do not employ Abarbanell and Bushee's (1998) control variables in our portfolio sorts. Complicating issues further is the fact that a few of our likely predictors deviate significantly from the original papers in terms of methodology. In a replication project as large as ours, a few large deviations are unavoidable, and simply dropping these reproductions (or attempted reproductions) is inconsistent with our philosophy of openness.

To examine the reproducibility of likely predictors, we first examine the subset for which the original papers provide long-short t-stats. All but four of these predictors lead to similar t-stats in our long-short portfolios. The exceptions are almost all accounted for by significant deviations between our methodology and the original papers, necessitated by the scale and openness of our project.[8] The likely predictors that did not have long-short t-stats largely had borderline regression results. Our reproductions of these predictors find similarly borderline significance in long-short portfolios. Once again, we can trace deviations in results to deviations in methodology. For example, some likely predictors that perform poorly in our single-sort long-short portfolios were studied in multivariate regressions with many significant controls in the original papers.

Overall, only three of our 205 clear and likely predictors failed to achieve the predictability evidence found in the original papers. These three predictors are R&D ability from Cohen, Diether, and Malloy (2013) and the two shareholder activism measures from Cremers and Nair (2005). These failures should not be taken as a criticism of these papers, however, as it is quite likely that there are remaining deviations or coding errors among our hundreds of reproduced charac-

[8]For example, our simple regression version of slope-based price delay produces a raw return t-stat of 2.01, far smaller than Hou and Moskowitz's (2005) characteristic adjusted t-stat of 7.7 from their two-stage shrinkage estimate. However, Hou and Moskowitz (2005) also show a similar difference in t-stats for their baseline R^2-based price delay when the move from a simple regression estimate with no adjustment (t-stat = 3.4) to a two-stage estimate with characteristic adjustments (t-stat = 8.0).

teristics. Indeed, a previous version of this paper failed to reproduce Harvey and Siddique's (2000) coskewness, but we have since fixed an error which restored the power of this predictor.

The rest of the dataset consists of 14 "not-predictors" that produced clearly *in*significant predictability in the original papers, and 100 "indirect signals" which had only suggestive predictability evidence. Most of the indirect signals are variations on other characteristics created by Hou, Xue, and Zhang (2020). We refrain from precisely judging the reproducibility of these characteristics, as assessing the success of indirect signals requires multiple stages of judgment. Nevertheless, visual inspection of the reproduced t-stats for not-predictors and indirect signals suggests that the cross-sectional literature is not only shockingly replicable, but robust. Almost all of the indirect signals that are modifications of clear predictors also produced significant predictability in our reproductions.

We also provide several supporting demonstrations of the quality of our 205 clear and likely predictors. Pairwise correlations demonstrate that the dataset consists of many distinct predictors. The dataset also displays intuitive properties with respect to rebalancing frequencies and liquidity screens. Notably, using either value-weighting or screening out stocks below the 20th percentile of NYSE market equity leads to in-sample mean returns that are about 30% (20 bps per month) lower, consistent with Chen and Velikov's (2019) in-sample results.

Relation to the Literature Our results contrast with Hou, Xue, and Zhang (2020) (HXZ), who find that "most anomalies fail to replicate." HXZ emphasize a 65% "failure" rate from value-weighting and NYSE breakpoints, but they also find replicability is quite poor when they use methods that are close to the modal equal-weighted implementation of the original papers (Green, Hand, and Zhang 2013; McLean and Pontiff 2016). Specifically, HXZ state that "[t]he failure rate drops to 43.1% if we allow microcaps to run amok with NYSE-Amex-NASDAQ breakpoints and equal-weighted returns."

We find that HXZ's findings are driven by their permissive definition of an "anomaly." HXZ analyze 452 "anomalies," but these derive from only 240 characteristics, as 212 of these anomalies are just different rebalancing frequencies of the 240 basic strategies. And of the 240 characteristics, only 118 showed clear evidence of significance for long-short returns in the original papers. In fact, our reproductions find that 117 out of these 118 clear predictors achieve t-stats

> 1.96, and the remaining predictor has a t-stat of 1.93. In other words, much of HXZ's "replication failures" are simply due to misclassification: these "anomalies" never had long-short portfolio significance to replicate in the first place.

Our reproduction rates also contrast with Chang and Li (2018), who find replication rates of 30-50% for 67 papers from general interest and macroeconomics journals. Our studies differ in several ways, but a key difference is that Chang and Li define a successful replication as "when the authors or journals provide data and code that allow [them] to qualitatively reproduce key results of the paper." Indeed, lack of author-provided data and code is the most common reason Chang and Li cite for their unsuccessful replications. In contrast, we did not use author-provided code at all, and simply wrote code based on the text and exhibits in the original papers.[9] These results suggest that the widespread replication failures documented in previous studies (e.g. Dewald, Thursby, and Anderson 1986, McCullough, McGeary, and Harrison 2006) could be remedied by applying more effort.

Our reproduction rates may also look high compared to McLean and Pontiff (2016), who find that 12% of their 97 predictor portfolios produce t-stats < 1.5. These results are reconciled, however, by the fact that 24 of MP's predictors have what we describe as borderline evidence of statistical significance in the original papers. Among the 85 of our characteristics that are also studied by MP, 14% of our reproductions lead to t-stats < 1.96, quite close to MP's numbers. Most of these low t-stats come from "likely predictors," however, and we do not judge them as reproduction failures. We should note that MP's inclusion of likely predictors is entirely valid for the goals of their study.

Indeed, our open source data is highly consistent with MP's findings. Like MP, we find that returns decay post-publication but remain positive, and that this decay is stronger for predictors that are stronger in-sample. Our results are quantitatively similar to those in MP, even for the subset of our predictors that are not studied by MP. These results echo previous papers that replicate MP in larger sets of predictors (Chen and Zimmermann 2020; Jacobs and Müller 2020).

Our paper adds to the evidence that the cross-sectional predictability literature is actually quite credible. These studies include out-of-sample tests (McLean and Pontiff 2016; Jacobs and Müller 2020), as well as multiple-testing adjustments (Chen and Zimmermann 2020; Chen 2020; Jensen, Kelly, and Ped-

[9]For one characteristic, we download characteristic data from the authors' website.

ersen 2021). Relative to these papers, ours adds novel evidence that the cross-sectional literature is extremely replicable by comparing reproduced t-stats to hand-collected t-stats.

An important caveat is that we do not address the distinct but related question of whether the literature offers implementable trading profits. In a closely related paper, Chen and Velikov (2019) suggest that the answer is no. Building on Novy-Marx and Velikov (2016), Chen and Velikov find that effective bid-ask spreads wipe out most of the post-publication returns for a large set of anomalies.

The paper proceeds as follows. Section 2 describes our methodology, including our characteristics selection and predictor category definitions. Section 3 contains the main results: literature-level reproduction performance. Section 4 takes a closer look, examining reproductions at the characteristic level. Section 5 provides additional evidence supporting the quality of our dataset. Section 6 concludes.

2. Methods: Characteristic Selection, Variable Construction, and Predictability Categories

This section describes our methodology. We describe how we select characteristics, how we construct characteristics and portfolios, and how we classify characteristics based on predictability evidence. Table 1 provides a broad overview of the dataset. We explain the terminology and numbers in this table in what follows.

[Table 1: "Overview of Open Source Asset Pricing Data" around here]

2.1. Characteristic Selection

We select characteristics to balance three objectives: (1) comprehensive coverage of previous meta-studies, (2) comprehensive coverage of firm-level cross-sectional predictors, and (3) completion of high quality code and data with a reasonable amount of economist-hours.

To balance these goals, our list begins with the 240 characteristics used in

Hou, Xue, and Zhang (2020) (HXZ).[10] We focus on HXZ because their study also examines the replicability of cross-sectional predictors. We then add an additional 49 characteristics for near-complete coverage of McLean and Pontiff (2016) (MP) and Green, Hand, and Zhang (2017) (GHZ). Finally, we add 30 firm-level stock return predictors from Harvey, Liu, and Zhu (2016) (HLZ).

This collection leads to a total of 319 characteristics drawn from 153 papers. Covering HXZ requires that we sometimes include many characteristics based on a single study.

We do not add all characteristics from MP and GHZ due to our third goal of completing high quality code in a reasonable amount of time. We omit the mergers and SEO characteristics from MP because they use the difficult-to-integrate SDC dataset. From the GHZ dataset, we omit seven characteristics from unpublished working papers (Asness, Porter, and Stevens 2000; Gettleman and Marks 2006; Lerman, Livnat, and Mendenhall 2008; Bandyopadhyay, Huang, and Wirjanto 2010), and one predictor from a retracted paper (Chen and Zhang 2010).

In adding characteristics from HLZ, we require that the characteristic was clearly shown to predict firm-level stock returns in the original paper. The firm-level requirement eliminates many papers that target only portfolio returns, which was a prevalent feature of asset pricing papers before the 1990s (e.g. Chan, Chen, and Hsieh 1985; Chen, Roll, and Ross 1986) and of macroeconomics-motivated papers more generally (e.g. Vassalou 2003; Balvers and Huang 2007). This requirement is motivated by our second goal of a covering firm-level predictors.

However, our first goal of meta-study coverage motivates us to not require clear firm-level predictability on characteristics from HXZ, MP, or GHZ. As a result, our characteristics vary greatly in their predictive power based purely on the results in the original studies.

2.2. Characteristic Construction

We try to follow the original papers as closely as possible. We aim for "reproductions" in the sense of Welch (2019), and avoid reassessing the validity of the original papers. In our view, extensions and reexaminations require focused studies (e.g. Pontiff and Singla 2019, Novy-Marx and Velikov Forthcom-

[10] HXZ use these 240 characteristics to generate 452 "anomalies" by varying rebalancing frequencies.

ing), and are inappropriate for a large-scale meta-study like ours. Indeed, our large-scale study necessitates deviations from the original papers, as we explain shortly.[11] As a result, our reproductions may be better described as "attempted reproductions," but for ease of reading we typically drop the "attempted" modifier throughout the text.

We standardize several of our reproduction procedures to maintain the transparency of our literature-level results, as well as the usability of our data. All characteristics are computed at a monthly frequency. For variables that are updated at a lower frequency, the monthly value is the most recently observed value. This allows for portfolios that update at a lower frequency, while maintaining flexibility for alternative implementations.[12] However, our monthly characteristics imply that we deviate from the handful of studies that use weekly rebalancing.

For almost all characteristics, we use the standard six-month lag for annual accounting data availability and a one-quarter lag for quarterly accounting data availability. In a couple cases, we use the earnings reporting date (RDQ) to indicate availablility of quarterly data in order to more closely match the original papers. For IBES, we assume earnings estimates are available by the statistical period end date. Other data is assumed to be available following the original papers. As a consequence of this standardization, we deviate from a few accounting studies, which use a shorter data lag.[13]

Many characteristics were only shown to be predictive in particular subsets of the data. We try to put off this subsetting until the portfolio generation step. Thus, the characteristics code and data omit price and exchange filters, which are instead imposed in portfolio generation.

Other filters, however, are quite diverse and difficult to implement at the portfolio stage. Several papers exclude stocks based on SIC codes or missing accounting data. Still others find predictability only in subsets of stocks based on specific characteristics (Piotroski 2000; Elgers, Lo, and Pfeiffer 2001). To accommodate these filters in a manageable fashion, we set to missing stock-months that don't satisfy these filters in the characteristics code.

[11] Even for a small-scale study, a perfect reproduction is likely impossible. WRDS' reproduction of HML obtains a correlation of 98.9% with Ken French's data, but the monthly returns still have deviations of up to 1% in particular months (Vora and Palacios 2010)

[12] For a handful of studies, we enforce the timing of the updating in the characteristics code. This was done help us find the code that closely matches the original results in a handful of difficult cases.

[13] For users of the code: the accounting data lag is imposed in the data download step, and is not visible in the files that generate characteristics.

In a few cases, we deviate from the original papers to trade off costs and benefits. For most of these cases, we deviate by not acquiring the data used in the original papers. For example, Barber et al. (2001) and Jegadeesh et al. (2004) use Zack's analyst recommendations data, which goes back much further than the easily accessible IBES recommendations on WRDS. Similarly, we do not obtain the NYSE archive data required for Barry and Brown (1984) or the Fitch's Quotation data used by Amihud and Mendelson (1986).

But other deviations are more idiosyncratic. We include a "number of consecutive earnings increases" predictor that is somewhat distant from the earnings streak predictor in Loh and Warachka (2012) in order to cover a predictor in Green, Hand, and Zhang (2017). We do, however, also include a characteristic that is closely-based on Loh and Warachka's (2012) earnings streaks. All of our price delay predictors (Hou and Moskowitz 2005) use rolling regressions of daily individual stock returns, while the original paper typically uses a 2-stage procedure that first estimates a noisy measure of price delay on individual stocks and then reduces the noise by running a second set of regressions on portfolios formed from the first stage. We use the simpler estimates of price delay largely to due to our economist-hours budget.

These deviations imply that, in a handful of cases, our reproduced portfolios are very far from the originals. We include these attempted reproductions to maintain openness, which we believe is important for the credibility of the literature.

2.3. Portfolio Construction

The core of our portfolio data consists of predictive long-short portfolios formed following the original papers. Like other predictability studies, we implicitly assume data at the end of month t can be used to make trades in closing auctions on the last day of month t.[14] These are the portfolios we use in our evaluations of reproduction success.

Using the original paper's results, we select the stock-weighting, rebalancing frequency, and quantile sort (if applicable). Each portfolio implementation is listed in our hand-collected data. Once again, we deviate from the original pa-

[14] As pointed out in Chen and Velikov (2019), the hypothetical traders in our portfolio tests would add demand to the closing auctions for the long legs, thereby increasing the buying prices and reducing trading profits.

pers for a handful of portfolios in the spirit of standardizing our code and results. The most notable deviation is that we use a simple equal-weighted decile sort portfolio for reproduction of Frazzini and Pedersen's (2014) betting-against-beta instead of the original construction where each stock is weighted depending on its beta ranking.

We also offer portfolios implementations with alternative rebalancing frequencies and liquidity screens, as well as decile sorts and value-weighted only portfolios. We caution the user, however, in that these alternative implementations are not as closely examined as the portfolios that follow the original papers.

In our baseline data, we do not sign our characteristics, but do sign our portfolios. That is, a higher idiosyncratic volatility characteristic implies a lower mean return, but the corresponding long-short portfolio has a positive mean return. The hand-collected data has this sign information, which users can apply to transform characteristics if this fits their applications.[15] For simplicity, we do not sign portfolios if the original papers do not provide direct evidence on the correct sign.

2.4. Predictability Categories

To determine predictability categories, we compare results from the original papers to our characteristic reproduction code.

The original results are hand-collected. We hand-collect the key table demonstrating predictive power, the empirical test used, the sign of predictability, mean return, t-stat, individual stock weighting, portfolio sort quantile, rebalancing frequency, among other notes. Figure 1 shows an excerpt of our spreadsheet for illustration.[16]

Comparing the hand-collected data to our code, we assign our characteristics to four categories:

- "Clear Predictor": Our characteristic is expected to achieve statistically significant mean raw returns in long-short portfolios (e.g. t-stat > 2.5 in a long-short portfolio, monotonic portfolio sort with 80 bps spread, t-stat > 4 in a regression, t-stat > 3 in 6-month event study).

[15] We also provide a dataset of signed predictors for direct download.

[16] The full spreadsheet is found at
https://github.com/OpenSourceAP/CrossSection/raw/master/SignalDocumentation.xlsx.

- "Likely Predictor": Our characteristic is expected to achieve borderline evidence for the significance of mean raw returns in long-short portfolios (e.g. t-stat = 2.0 in long-short with factor adjustments, t-stat between 2 and 3 in a regression, large t-stat in 3-day event study).
- "Not-Predictor": Expected to be statistically insignificant in long-short portfolios. (e.g. t-stat = 1.5 in long-short, t-stat = 1 in a regression).
- "Indirect Signal": Only suggestive evidence of predictive power (e.g. correlated with earnings/price, modified version of a different characteristic, in-sample evidence only).

These categorizations are necessary in order to measure reproduction success. For clear predictors, the measurement is straightforward, as a t-stat > 1.96 easily identifies a success. Note that we sign our long-short portfolios based on the original results, so this rule does not involve the absolute value. For not-predictors, the sign flips, and a t-stat < 1.96 is indicative of a success.

Measuring reproduction success is more subtle for the other categories, however. Some likely predictors had t-stats very close to 1.96 in a raw long-short portfolio in the original paper. Given that data updates will surely move the t-stat up or down, one should then place 50/50 odds that the reproduced t-stat will be above or below 1.96. Similarly, if the original paper found a t-stat of 2.6 in a univariate regression, it's hard to say if the reproduced long-short portfolio should also produce a t-stat > 1.96. Deviations between our characteristics code and the original recipes also lead to likely predictors, and necessitate subtle judgment calls.

Indirect signals are also tricky. Most of our indirect signals are modifications of other characteristics, and whether the modification should increase or decrease the t-stat requires judgment. Complicating the issue is the fact that some indirect signals are modifications of likely or not-predictors, implying that multiple steps of judgment are required to determine which side of t-stat = 1.96 our long-short portfolio should land on.

Each characteristic's predictability category can be found in Tables 2-4, and the assignments are discussed in some more detail in Section 4.1.

Table 1 briefly summarizes our predictability category assignments. Of our 319 characteristics, 161 are clear predictors and 44 are likely predictors. Throughout the text, we describe these two categories as simply "predictors." We also separate these predictors from the other characteristics in our data and code. This separation is helpful because of clear and likely predictors are quite distinct from the other characteristics in how one should analyze them.

14 of our characteristics are not-predictors, and 100 are indirect signals. As seen in the top panel, not-predictors and indirect signals play a minor role in most other meta-studies, with the exception being Hou, Xue, and Zhang (2020). Indeed, the vast majority of our not-predictors and indirect signals are drawn from Hou, Xue, and Zhang (2020).

The bottom panel of Table 1 shows that our data provides more-or-less comprehensive coverage of other meta-studies. We cover all 452 of Hou, Xue, and Zhang's (2020) "anomalies", 97% of the clear predictors from McLean and Pontiff (2016), 88% of the clear predictors from Green, Hand, and Zhang (2017), and 90% of the clear firm-level predictors that use widely-available data from Harvey, Liu, and Zhu (2016). We also cover all of the likely predictors in McLean and Pontiff (2016), Green, Hand, and Zhang (2017), and Hou, Xue, and Zhang (2020). Indeed, most of the clear, widely-available predictors that we are missing are closely related to predictors that we offer (e.g. the industry-adjusted value and momentum predictors of Asness, Porter, and Stevens 2000).

Throughout the paper and code, we separate out the clear and likely predictors. We sometimes refer to these characteristics as just "predictors," for short.

For many purposes, predictors are the only characteristics that should be examined. Moreover, predictors are substantially less redundant than the full dataset. While the full dataset consists of 319 characteristics from 153 studies, the predictors consist of 205 predictors from 137 studies. For 101 studies, we draw only a single predictor from the study.[17]

3. Literature-Level Reproduction Performance

This section contains our main results. We begin by examining literature level reproduction performance for clear and likely predictors (Section 3.1). We then broaden the scope to not-predictors and indirect signals when we compare to

[17] The distribution of predictors per study is right skewed. Most studies have just one predictor, but 20 have 2, and a few have many more. Heston and Sadka (2008) provide 10 strategies related to return seasonality, and we include all 10 as clear predictors to avoid imposing judgment on which is the "right" strategy. A similar philosophy leads us to include many predictors from Richardson et al. (2005) and Daniel and Titman (2006).

other meta-studies (Section 3.2).

3.1. Success of Predictor Reproductions

Our first measure of reproduction success is a simple indicator: does the long-short raw return t-stat exceed 1.96? This measure is the primary focus of Hou, Xue, and Zhang (2020) (HXZ), and provides an easy-to-understand measure of reproduction success.

Figure 2 shows the share of clear predictors that exceed the simple t-stat > 1.96 cutoff, broken down by data focus: accounting, analyst forecasts, corporate events, stock prices, trading data, and a broad "other" category. It shows only clear predictors, as likely predictors had borderline significance in the original papers (Section 2.4). As in all of our reproduction success evaluations, we use the same sample period in the original papers.

The figure shows that reproductions are extremely successful. Our success rates are nearly 100% in every category. 73 out of our 74 accounting focused clear predictor reproductions succeeded. 40 out of 41 price-focused reproductions succeeded. And 43 out of 44 reproductions succeeded among the remaining categories.

[Figure 2 "Reproduction Success Rates for Clear Predictors" about here.]

One might be concerned that our extremely high success rate is sensitive to our definition of a clear predictor. Similarly, one may be concerned that, even if our t-stats exceed 2.0, they may be much smaller than the original t-stats.

Figure 3 should assuage both of those concerns. The figure examines both clear and likely predictors, and evaluates reproduction success more qualitatively, by simply plotting our reproduced t-stats against the t-stats of the original papers. To ensure that t-stats are comparable, the figure excludes predictors that were only examined in regressions or event studies in the original papers. We also drop t-stats from Barber et al. (2001), Frazzini and Pedersen (2014), and Hou and Moskowitz (2005) because our portfolio constructions are very far from the originals for idiosyncratic reasons (see Sections 2.2 and 2.3).

[Figure 3 "Comparison of Reproduced and Original t-stats" about here.]

The figure shows that our extremely high reproduction success rate is not sensitive to our predictor categorizations. Included in this chart are 11 likely predictors (triangles), all of which had t-stats close to 2 in the original papers (horizontal axis). Consistent with the original papers, our reproductions also typically produce t-stats close to 2.0, though roughly half of them fall below the arbitrary 1.96 threshold. More broadly, our extremely high success rate for clear predictors is not due to t-stats just above 2.0. Many reproduced t-stats exceed 5.0, and a few even exceed 10.0.

Perhaps most important, Figure 3 shows that our reproductions match the original results not just qualitatively but quantitatively. A regression of our reproduced t-stats on the original t-stats produces a coefficient of 0.90, not far from the ideal slope of 1.0. The R^2 is 83%, implying that the reproductions do not stray far from the originals. Indeed, much of the remaining deviations may be due to the fact that our t-stats use simple raw mean long-short returns, while many of the original t-stats adjust for characteristics or factor exposures.

As a final demonstration of the reproducibility of the cross-sectional literature, we examine whether mean returns are monotonic in the predictors. From our experience, the vast majority of papers that showed portfolio sorts also showed monotonicity—though we did not hand-collect this information.

Figure 4 examines the monotonicity of decile sorts for clear and likely predictors. The figure limits the data to 170 continuous predictors, as discrete predictors are poorly behaved in decile sorts. It plots the mean return as a function of the decile, with each marker representing one predictor-decile, and boxes summarizing the 25th, 50th, and 75th percentile mean return within each decile. The figure shows that mean returns are indeed monotonically increasing in the predictors. The median across predictors increases monotonically from one decile to the next, as does the 25th percentile and the 75th percentile.

[Figure 4 “Monotonicity” about here.]

The box plots do not show predictor-level monotonicity, however, so we color-code the markers for a closer look. Filled markers indicate predictor-level monotonicity for a given predictor-decile. That is, the marker is filled if the mean return in decile k for predictor i exceeds the mean return for decile $k-1$ for the same predictor i. The figure shows that roughly 80% of the predictor-deciles show an increase, a strong indication of monotonicity at the predictor level. To get a sense of this ratio, note that it suggests the probability that decile k exceeds the return of decile $k-2$ is $1-0.2^2 = 96\%$ in a simple binomial framework, and that this probability is even higher when comparing deciles further apart.

Beyond supporting reproducibility, this monotonicity result demonstrates robustness of the cross-sectional predictability literature. That is, monotonicity shows that the predictability evidence we find is not sensitive to the details of the empirical test. Deciles sorts, cross-sectional regressions, event studies should all produce consistent results. This finding is consistent with early versions of McLean and Pontiff (2016), which found that their results were similar using Fama-Macbeth regressions instead of long-short portfolios.

Interestingly, Figure 4 also provides evidence against the p-hacking explanation for predictability. For p-hacking to explain predictability, it would have to operate across the entire predictor distribution, incrementally increasing mean returns from one level of the predictor to the next. This fine-tuning is an issue that is not examined in previous studies that model p-hacking (Harvey, Liu, and Zhu (2016); Chen and Zimmermann 2020; Chen 2020), nor is it examined in Hou, Xue, and Zhang (2020).

This concludes our main results. Nearly 100% of the literature on cross-sectional stock return predictability can be replicated. This finding implies not only the veracity of the literature, but defends the credibility of the asset pricing community more generally. Finally, these results support the quality of our open source dataset, publically available at https://github.com/OpenSourceAP/CrossSection.

3.2. Not-Predictors, Indirect Signals, and Comparison with Other Meta-Studies

The overwhelming success of our reproductions may appear to be at odds with the literature. McLean and Pontiff (2016) (MP) find that 12% of their 97 predictors produce t-stats < 1.5, suggesting a far higher “failure rate” than to our main results. More strikingly, Hou, Xue, and Zhang (2020) (HXZ) find that roughly 50% of their 452 long-short portfolios produce t-stats < 1.96 in absolute value, even with equal-weighting and when limiting to the original papers’ sample periods.

Our results differ from MP and HXZ primarily because we carefully examine the original papers to check if reproduced long-short t-stats should exceed 1.96 before testing this threshold in our reproductions. In contrast, MP have a somewhat more lenient criterion, limiting themselves to papers "in which the null of no return predictability is rejected at the 5% level." HXZ are much more lenient. HXZ state their "list encompasses the bulk of the published anomalies literature in finance and accounting," but do not specify any more requirements for inclusion that we could identify.

Our careful examination is illustrated in Figure 5, which shows a jitter plot of reproduced t-statistics for all of our characteristics, including those in the "not-predictor" and "indirect signal" categories. The first two rows echo Figures 2 and 3: Clear predictors almost uniformly have t-stats > 1.96 and many t-stats are much larger. Likely predictors are roughly evenly distributed around 1.96. Notably, these rows include predictors with evidence based on regressions and event studies, and thus cover far more likely predictors than Figure 3.

[Figure 5 "Predictive Significance in the Extended Dataset" about here.]

Unlike the clear and likely predictors, not-predictors and indirect signals generally failed to achieve statistical significance in our reproductions. A large mass of indirect signals lies to the left of t-stat = 1.6, and almost all of the not-predictors fall in the same region.

Figure 6 shows the analogous plot, restricted to the predictors that also appear in MP (top panel) or HXZ (bottom panel).

[Figure 6 "Performance vs Other Meta-Studies" about here.]

The top panel echos MP's finding that about 12% their predictors have small t-stats. 14 of our 85 characteristics that overlap with MP have t-stats < 1.96 (dotted line). Most of these we judged as likely predictors, though a handful were judged as indirect signals and one was judged as a not-predictor. In particular, we judged Dichev's (1998) Z-Score as a not-predictor, as it finds a t-stat of 1.59 in a univariate regression (see Dichev's Table 3A). It's likely that MP included Z-Score due to its t-stat of 3.37 in a multivariate regression with size and B/M controls, and reasonable people can disagree on the proper classification of this predictor. Nevertheless, we argue that our reproduced t-stat of 1.20 for Z-Score (Table 4) should not be judged as a reproduction failure, as it is quite consistent with Dichev's univariate results.

The bottom panel of Figure 6 shows the breakdown for characteristics that overlap with HXZ's. The data looks quite similar to our full dataset (Figure 5): Clear predictors are almost entirely above 1.96, likely predictors center around 1.96, not-predictors are below, and indirect signals are dispersed but many fall below the 1.96 cutoff. Thus, HXZ's failure rate of around 50% seems to be due to the misclassification of the "anomalies." In our reproductions, almost all apparent failures are linked to studies that never demonstrated predictability in the first place.

Importantly, MP's choice to be more lenient in determining original-study-predictability is entirely valid for the goals of their study. MP seek to measure the decline in the magnitude of predictability out-of-sample, and one may be interested in this decline for predictors that don't quite meet the 1.96 cutoff.

Indeed, we find that we can closely replicate MP's results, even if we limit our sample to clear predictors. This robustness is seen in Figure 7, which replicates (MP's) Figure 1 using only clear predictors. The top panel plots the in-sample return vs the post-publication return decay, and the bottom panel swaps out the in-sample return for the in-sample t-stat. In both panels, we subset the data to clear predictors in MP (dark circle) and clear predictors that are not in MP (light triangles).

[Figure 7 "McLean and Pontiff (2016) Replication" about here.]

The figure replicates and extends three important facts documented by MP about post-publication decay: (1) decay increases in the in-sample return, (2) decay increases in the in-sample t-stat, and (3) the decay is not large enough to wipe out the in-sample return.

The first two facts can be seen in the upward slope of the regression lines in Figure 7. Strikingly, the upward slope does not depend on whether we fit the regression to predictors that are studied by MP (solid) or missing from MP (dashed). Indeed, the two standard error confidence bands largely overlap, indicating that MP's findings hold quantitatively in an out-of-sample test (or perhaps out-of-out-of-sample test). Visual inspection of MP's Figure 1 suggests our results are also quantitatively consistent with MP's, even though we limit our sample to clear predictors. We also find that using of all of MP's predictors leads to similar results.

The third fact can be seen by comparing the predictor markers to the 45 degree dotted line in the top panel. If post-publication decay is strong enough to wipe out in-sample returns, then the predictors would be evenly distributed around this 45 degree line. However, the majority of predictors lie above the 45 degree line, showing that predictability survives post-publication. It's important to note, however, that these results do not account for trading costs. Indeed, Chen and Velikov (2019) find that the remaining predictability is eliminated by effective bid-ask spreads.

Overall, our results appear to differ from other meta-studies because we carefully categorize predictors based on the original results. Only with this careful categorization can a meta-study accurately evaluate whether replications are successful on such a large scale. These categorizations do not affect McLean and Pontiff's (2016) analysis and indeed we replicate their results among predictors with a higher standard for "original significance."

In contrast, Hou, Xue, and Zhang's (2020) finding of widespread replication failure for equal-weighted portfolios does not survive a more careful inspection. Our results call into question HXZ's main findings regarding value-weighted portfolios, though it is certainly true that anomalies are weaker when value-weighted (See Section 5.2).

4. Characteristic-Level Reproduction Performance

Behind the literature-level results are 319 firm-level characteristics, each of which has its own story, original statistics, and reproduction quality. Tables 2-4 explore this rich data, listing each individual characteristic, the reproduced mean return and t-stat, and the evidence for predictability in the original papers. Table 2 lists clear predictors, Table 3 lists likely predictors, and Table 4 lists not-predictors and indirect signals.

At the surface level, these tables provide a quick reference guide to our dataset. We sort characteristics by author names and provide the acronym used in our code, so readers can easily look up characteristics of interest.

At a deeper level, the tables provide a detailed characterization of our judgments of predictor categories (Section 4.1), our reproduction failures and struggles (Section 4.2), and the reproductions that led to extremely large t-stats (Section 4.3).

[Table 2 "Individual Clear Predictors" about here.]

[Table 3 "Individual Likely Predictors" about here.]

[Table 4 "Additional Characteristics" about here.]

4.1. Which Predictors Are "Clear," "Likely," or "Not?" Which are "Indirect Signals?"

We explained our predictor categorizations in Section 2.4 and showed that the categorizations do not affect our assessment of the literature's reproducibility for long-short portfolios t-stats in Section 3. However, readers may still have questions about why certain predictors are relegated to the "likely," "not," or "indirect signal" categories. Tables 2-4 should answer this question.

4.1.1. Categorization Details for Clear and Likely Predictors

Table 2 shows that most clear predictors produced t-stats that exceed 2.5 in long-short portfolios in the original papers (e.g. Ang et al.'s (2006) idiovol, Belo and Lin's (2012) inventory growth, Dichev's (1998) long-short O-Score strategy). Most of the remaining ones were shown to generate t-stats of 4 or more in regressions (e.g. Fama and French's (1992) book-to-market, Pontiff and Woodgate's (2008) share issuance). As with portfolio sorts, we only consider regressions to be predictive if they forecast returns in period $t+1$ using data available at time t.

In contrast, most of the likely predictors (Table 3) had marginal t-stats in the original papers. Several predictors had t-stats very close to 1.96 in long-short portfolios (Ball et al.'s (2016) operating profitability, George and Hwang's (2004) 52-week high). Others have t-stats between 2 and 3 in regressions (Abarbanell and Bushee's (1998) sales growth over inventory growth, Fama and MacBeth's (1973) CAPM beta). It's worth noting that the t-statistic cutoff of 1.96 is fairly arbitrary, and for some questions regressions are more relevant than long-short portfolios.

Other likely predictors were more ambiguous. Amihud and Mendelson's (1986) bid-ask spread predictor showed strong portfolio sorts, but the original paper did not provide a long-short t-stat. Moreover, they use Fitch's Stock Quotations on the NYSE, while we use Corwin and Schultz's (2012) effective spread based on daily CRSP data (also used in McLean and Pontiff (2016)). Chan, Lakonishok, and Sougiannis's (2001) advertising expense to market produced a 50 bps spread in portfolio sorts, but they did not provide a t-stat. Haugen and Baker (1996) suggest several predictors based on the average t-stat across 180 different multiple regressions, but it's hard to tell say if this procedure should result in t-stats > 2.0 in simple long-short portfolios.

Overall, the individual reproduced t-stats in Table 3 are almost uniformly consistent with the original study's predictability evidence. t-stats much less than 2.0 are in almost every case associated with either middling t-stats in multi-variate regressions (Abarbanell and Bushee's (1998) sales gross to overhead growth), specialized data that we did not employ (Amihud and Mendelson (1986)), or nonstandard methodologies that we did not use (Haugen and Baker (1996)).

4.1.2. Categorization of Not-Predictors and Indirect Signals

Table 4 shows that the not-predictors have a straightforward definition. Most not-predictors had t-stats < 1.96 in long-short portfolios in the original studies (Ang, Chen, and Xing's (2006) past downside beta; Anderson, Ghysels, and Juergens's (2005) long-term forecast dispersion; Whited and Wu's (2006) financial constraints index). It's important to mention that this lack of significance should not be considered a criticism of the original papers. The 5% significance cutoff is arbitrary, and some of these predictors fall just below the cutoff.

Many indirect signals simply did not come with predictability evidence in the original papers. Accrual quality, earnings conservatism, and earnings value relevance all come from Francis et al. (2004), which studies characteristics that are related to an implied cost of capital estimate based on Value Line's price targets. This paper does not, however, examine return prediction. Belo, Lin, and Bazdresch's (2014) model features a variable called brand capital, but the paper does not examine the predictive power of this variable.

Other indirect signals came with predictability-related information in the original papers, but we judged this evidence as too weak allow us to judge statistically significant predictability in portfolio sorts. Several of these weak evidence predictors come from Acharya and Pedersen's (2005) study of liquidity betas. Acharya and Pedersen estimated market prices of risk for these betas in a GMM framework, which would imply predictability if the parameters are very stable. But since betas tend to be unstable (Ang, Chen, and Xing 2006, for example) we judge this GMM result as close to no evidence regarding the results of portfolio sorts. Similarly, we judged the multi-variate regressions of Abarbanell and Bushee (1998) and Soliman (2008) provide insufficient information for our purposes when the coefficient on a regressor is insignificant.

But the bulk of the indirect signals are Hou, Xue, and Zhang's (2020) variations on characteristics in other studies. These characteristics are noted as "HXZ variant" in the rightmost column of Table 4. Most of these modifications use quarterly versions of annual accounting variables. A few involve arbitrary lags of the denominator or using alternative factor model adjustments when generating return residuals (as in idiosyncratic volatility).

We refrain from assessing the predictor category of HXZ's variants, because some of them require subtle judgments. For example, HXZ's produce a quarterly version of Whited and Wu's (2006), which produced a t-stat of 1.2 in the original paper. It's very hard to say whether a more timely version of this variable would lead to statistical significance. Similarly, it's hard to say if a quarterly version of an annual accounting-based variable will have too much seasonality to be predictive. The seasonality would depend on the precise details, and we did not want to exercise this much judgment. We decided, therefore, to simply label all of HXZ's variants as indirect signals.

Nevertheless, inspection of Table 4 shows that almost all of HXZ's variants demonstrate the robustness of the original results. Chan, Lakonishok, and Sougiannis's (2001) find that R&D to sales fails to predict returns, and our reproduction of HXZ's quarterly version also fails to achieve statistical significance. Meanwhile, our reproductions of HXZ's variations on Anderson and Garcia-Feijoo's (2006) capx growth, Ball et al.'s (2016) cash-based operating profitability, and Lakonishok, Shleifer, and Vishny's (1994) cash flow to market are all statistically significant, consistent with the original constructions.

4.2. Reproduction Failures and Struggles

Despite the overwhelming success in aggregate, Tables 2 and 3 illustrate how our reproductions struggle or even fail in a few instances. As emphasized in the introduction, these failures should not be taken as criticisms of the original papers, as it is quite likely that there are coding errors or remaining deviations among our hundreds of reproduced characteristics.

The smallest reproduced t-stat among our clear predictors (Table 2) is Cremers and Nair's (2005) (CN) takeover vulnerability. Our t-stat is just 1.00 with a mean return of 25 bps per month, despite the fact that the original paper found a t-stat of 3.1 and an alpha of 90 bps per month. Our reproduction uses raw mean returns, while CN use the Carhart (1997) four-factor alpha, so perhaps this deviation is driving the difference in results. However, we typically find that factor adjustments have relatively minor effects (Figure 3), and we also had trouble reproducing CN's active shareholders predictor, which we categorized as a "likely predictor." As seen in Table 3, our reproduction produced a t-stat of 1.02, compared to the original t-stat of 2.04.

The second smallest t-stat comes from our reproduction of Cohen, Diether, and Malloy's (2013) R&D ability. Our reproduction achieves a positive but insignificant t-stat of 1.50, compared to the original paper's t-stat of 2.6. Cohen, Diether, and Malloy measure R&D ability by a rolling estimation of a sales forecasting model that involves several lags of R&D. As R&D is prone to missing and zero values, it is quite possible that we failed to follow the exact same procedures as the original authors.

Among our clear predictors, the only other t-stat < 1.96 is Pástor and Stambaugh's (2003) liquidity beta. However, our reproduced t-stat is 1.93, just a hair below the arbitrary cutoff of 1.96, and not far from the original CAPM-adjusted t-stat of 2.5. We should note that we only aim to reproduce Pastor and Stambaugh, and other replication papers find that this predictor is sensitive to construction details (Li, Novy-Marx, and Velikov 2019; Pontiff and Singla 2019).

A few clear predictors have reproduced t-stats that are notably smaller than the originals, despite being larger than the 1.96 cutoff. The acronyms for these predictors are clearly seen in Figure 3, and the details of these predictors can be found in Table 2.

Several of these predictors come from accounting papers that lag annual accounting data by only 3 or 4 months rather than the 6 months used in the finance literature (Piotroski 2000; Xie (2001); Mohanram 2005). Similarly, we deviate from Johnson and So (2012) in rebalancing our portfolios monthly rather than weekly. Intuitively, these more timely signals would produce notably higher returns, and we do not judge these deviations as reproduction failures.

Another underperforming reproduction is Elgers, Lo, and Pfeiffer's (2001) earnings forecast to price, which in our data has a t-stat of 2.6, far lower than the original t-stat of 5. The original t-stat was size-adjusted, however, and other tables in this paper also show substantial size effects in their data.

The last clear predictor struggle worth mentioning comes from our reproduction of Boudoukh et al.'s (2007) payout yield portfolio. For this predictor, we chose to deviate from the original paper in a subtle way. At the very end of the original paper, the authors offer a long-short strategy that produces a t-stat of 3.92 from a tercile sort with NYSE breakpoints. We found our reproduction of this strategy was sensitive to how we lagged the signal before merging with return data, and that the more robust lagging method generated a very small t-stat. However, NYSE terciles seemed unnecessarily conservative and moreover, Boudoukh et al. (2007) show a decile sort as their first predictability table, which we can reproduce quite well. Thus, our implementation follows this decile sort, though they do not provide a long-short t-stat for this procedure. All told, we judge this predictor to have replicated reasonably well.

Among likely predictors (Table 3), most of the apparent reproduction failures are simply due to deviations between our reproduction attempts and the original papers. Amihud and Mendelson (1986), Barber et al. (2001), and Barry and Brown (1984) all use specialized datasets that we do not have access to (see Section 2). We also did not employ the multivariate regressions of Abarbanell and Bushee (1998), the aggregation of 180 multiple regressions used by Haugen and Baker (1996), or the specialized portfolio sort used by Frazzini and Pedersen (2014).

The most notable likely predictors in terms of reproduction difficulty come from Frankel and Lee (1998). This paper uses analyst forecasts and a present value model to generate three trading strategies that we reproduce. Our reproductions lead to t-stats between 0.96 and 2.01, despite the fact that the original paper finds 1% statistical significance across the board. However, this high statistical significance was "derived using Monte Carlo simulation," and is hard to square with their small return spreads and short 15-year sample. Indeed, we find that B/M is much less significant in their sample than the high significance they show using their Monte Carlo test. In short, we attribute our smaller t-stats to deviations in methodology rather than failed reproductions, but reasonable people can disagree on how to evaluate these reproductions.

4.3. Extremely Strong Predictors

Figure 3 showed that many reproductions achieve huge t-stats of 6.0 or more. The corresponding p-value is 0.000000002, implying that it is absurdly unlikely that these predictors are drawn from the null of no predictability. Indeed, Chen (Forthcoming) argues that it would take in expectation at least 400 years to generate these predictors from p-hacking alone. Consistent with this argument, Harvey, Liu, and Zhu's (2016) SMM estimates and Chordia, Goyal, and Saretto's (2020) calibrations imply that t-stats in excess of 4.0 are almost guaranteed to be true discoveries.

Table 2 takes a closer look at these predictors. Almost all of these outstanding predictors focus on accounting data, analyst forecasts, or stock prices. Stated differently, almost none of them come from the more exotic data categories.

These outstanding performers are quite diverse. They include earnings surprise streaks (Loh and Warachka 2012); net external financing (Bradshaw, Richardson, and Sloan 2006), change in recommendation (Jegadeesh et al. 2004), return seasonality (Heston and Sadka 2008), conglomerate return (Cohen and Lou 2012), dividend seasonality (Hartzmark and Solomon 2013), employment growth (Belo, Lin, and Bazdresch 2014), asset growth (Cooper, Gulen, and Schill 2008), change in taxes (Thomas and Zhang 2002), and enterprise multiple (Loughran and Wellman 2011). These predictors lack any obvious economic connection, consistent with the near zero median correlation we find among clear predictors (Section 5.1). However, these predictors do have in common extremely large t-stats in the original papers, as seen in Table 2 and Figure 3.

5. Additional Evidence of Dataset Quality

This section provides additional results on our dataset's quality. We limit this analysis to the 205 clear and likely predictors. We refer to clear and likely predictors as just "predictors," for short. We focus on predictors here because evaluating the quality of not-predictors and indirect signals is a much more complicated.

The section shows that the dataset contains many distinct predictors (Section 5.1) and that the portfolio returns decline if we impose different liquidity adjustments (Section 5.2) or decrease the rebalancing frequency (Section 5.3).

These results also provide useful benchmark numbers regarding liquidity effects. Namely, imposing value-weighting or market equity screens reduces mean returns by roughly a factor of 1/3. The results also illustrate the flexibility of our portfolio generation code.

5.1. Distinct Predictors

In selecting characteristics, we aim primarily for complete coverage of previous meta-studies (Section 2.1). We make no attempt to eliminate predictors due to subjective similarities.

Thus, we include several profitability-related predictors including those from Fama and French (2006); Balakrishnan, Bartov, and Faurel (2010); and Novy-Marx (2013). Being liberal about distinct predictors is necessary as there is, as of yet, no established methodology for determining distinct predictors. By including all predictors, we allow future users of our code and data to make their own determination on which version of profitability is the "right" one.

Despite this potential redundancy, a simple analysis suggests that this dataset is very high-dimensional. Figure 8 examines this question by showing distributions of correlations.

[Figure 8 "Correlations Between Pairs of Predictors or Portfolio Returns" about here.]

Panel (a) shows correlations at the characteristic level. It shows the distribution of pairwise rank correlations between stock-level predictors (characteristics). Before computing correlations, we sign all predictors so that a higher predictor value implies higher mean returns based on the original papers. The panel shows that predictor-level pairwise correlations are generally close to zero, suggesting that the predictors contain distinct information. These results are consistent with Green, Hand, and Zhang (2013) who also find correlations close to zero among their set of 39 readily programmed predictors.

Panel (b) shows this high dimensionality extends to the portfolio level. It also shows the distribution of pairwise correlations, this time using pairs of long-short portfolio returns. As with all of our long-short predictor portfolios, portfolios are signed to have positive mean returns following the original papers. Similar to the predictor correlations, portfolio return correlations are close to zero for the bulk of the distribution. Indeed the vast majority of correlations lie between -0.5 and +0.5.

Panels (c) and (d) examine whether the HML and momentum factors subsume our long-short portfolios. HML and momentum factors are both downloaded from Ken French's website and constructed from 2x3 sorts following Fama and French (1993). That is, (1) stocks are independently assigned to "S" or "B" based on the NYSE median size and "H" or "L" based on the 30th and 70th percentiles of either B/M or the past year's return within NYSE stocks, (2) value-weighted portfolios are formed for S/L, B/L, S/H, and B/H intersections, and (3) factor returns computed as 0.5(S/H+B/H) - 0.5(S/L+B/L). In contrast, our B/M and momentum portfolios are just single sorts. We follow Rosenberg, Reid, and Lanstein (1985), Fama and French (1992), and Jegadeesh and Titman (1993) in constructing these portfolios, and like most of our predictors these original papers do not employ the Fama and French (1993) 2x3 factor construction for their anomaly strategies.[18]

Panel (c) shows that a handful of our portfolios have correlations of 0.6 of more with HML, but the bulk of the correlation distribution remains close to zero. Panel (d) shows a similar result for the momentum factor. Overall, our long-short portfolios contain many distinct strategies, consistent with McLean and Pontiff's (2016) finding of a near zero average correlation among their reproducted returns.

5.2. Performance by Liquidity Adjustment

Following our philosophy of "reproduction," our baseline portfolios follow the original papers as much as possible (Sections 2.2-2.3). These portfolios likely overstate the profits traders could have earned from these predictors, however, as most of them are equal-weighted (Green, Hand, and Zhang 2013). Equal-weighted portfolios require the trading of illiquid stocks and huge transaction costs (Novy-Marx and Velikov 2016, for example).

Figure 9 examines how simple liquidity adjustments affect predictor performance. The figure shows jitter plots of in-sample mean returns, comparing the

[18]Rosenberg, Reid, and Lanstein (1985) use a complicated procedure to remove exposure to a variety of "risk indexes," which is similar in spirit to the Fama and French (1993) 2x3 approach in that they both remove exposure to size. However, Rosenberg et al's procedure is not very transparent and difficult to implement, so we just use our default quintile sort for simplicity and transparency.

original paper's adjustments (if any) to various liquidity screens as well as the enforcement of value-weighting. The original liquidity adjustments can be found in our hand-collected data at https://github.com/OpenSourceAP/CrossSection.

[Figure 9 "Performance by Liquidity Screen" about here.]

Intuitively, all liquidity adjustments lead to lower mean returns. The price screen (limiting to stocks with share price > $5) appears to be the softest adjustment, producing the smallest decline in performance. The other liquidity adjustments have relatively similar effects.

Overall, simple liquidity adjustments reduce mean returns by a factor of about 1/3, on average. The typical mean return drops from around 60 bps per month to about 40 bps per month regardless of whether the adjustment is an NYSE only screen, a market equity screen, or the enforcement of value-weighting. These results are quantitatively similar to Chen and Velikov (2019), who find that effective bid-ask spreads eliminate about 1/3 of mean returns in-sample, even after cost-mitigation.

Figure 9 also illustrates the flexibility of our code. These various screens are made possible by the fact that we try to delay imposing screens until the portfolio generation step. As a result, the user can choose whether he or she wishes to take signal from all stocks, or just the more liquid ones.

5.3. Performance by Rebalancing Frequency

Our code also allows for a flexible choice of the rebalancing frequency. More precisely, the code allows the user to choose how often stocks are re-assigned to portfolios. We refer to this as "rebalancing," following the cross-sectional literature.[19] This flexibility may be important, for example, when accounting for trading costs.

Figure 10 shows that our code leads to intuitive results when we alter the rebalancing frequency. This figure plots the distribution of mean returns across

[19] Following the cross-sectional literature, our portfolios are always rebalanced monthly in the sense that stock weights are adjusted every month to provide equal- or value-weighting. Most papers do not provide precise explanations of these details, but in our experience this procedure is required for replicating papers. For an explicit example, see https://wrds-www.wharton.upenn.edu/pages/support/applications/risk-factors-and-industry-benchmarks/fama-french-factors/.

predictors for 1-, 3-, 6-, and 12-month rebalancing. For comparison, we also show results using the rebalancing frequency in the original papers.

[Figure 10 "Performance by Rebalancing Frequency" about here.]

Rebalancing at a monthly frequency leads to slightly higher mean returns compared to the original specifications. This is due to the fact that many of the original papers follow Fama and French (1992) and rebalance annually (every June).

Performance declines monotonically as the rebalancing frequency decreases from 1- to 12-months. This pattern is intuitive as less frequent rebalancing implies less exposure to the predictive signal.

6. Conclusion

The credibility of cross-sectional asset pricing is in doubt. Several papers argue that much of the literature is false do to a reliance on statistical methods that are no longer valid (Harvey, Liu, and Zhu 2016; Linnainmaa and Roberts 2018; Chordia, Goyal, and Saretto 2020). One heavily-cited study claims that, not only are the statistical methods invalid, but the numbers cannot be replicated, even when following the original methodologies (Hou, Xue, and Zhang 2020).

We provide data and code that takes a key step toward restoring the credibility of the literature. The data shows that nearly 100% of the literature's predictability results can be reproduced, and our code shows how this surprising result is achieved. This reliability adds to the evidence that the cross-sectional predictability literature is quite credible (McLean and Pontiff 2016; Jacobs and Müller 2020; Chen and Zimmermann 2020; Jensen, Kelly, and Pedersen 2021).

Our code is written explicitly with the user in mind. The structure is modular and parallel, so that pieces of the code can be easily fixed or improved despite the massive size of the entire package. We welcome users to examine and build on our code by visiting https://github.com/OpenSourceAP/CrossSection. We hope this demonstration of open collaboration inspires others to open up their analyses. In our view, a shift toward openness is not only important for the profession's understanding of risk and return, it is also important for protecting the credibility of academic finance in the eyes of the broader public.

References

Abarbanell, Jeffery S and Brian J Bushee. “Abnormal returns to a fundamental analysis strategy”. *Accounting Review* (1998), pp. 19–45.

Acharya, Viral V and Lasse Heje Pedersen. “Asset pricing with liquidity risk”. *Journal of financial Economics* 77.2 (2005), pp. 375–410.

Adrian, Tobias, Erkko Etula, and Tyler Muir. “Financial intermediaries and the cross-section of asset returns”. *The Journal of Finance* 69.6 (2014), pp. 2557–2596.

Ali, Ashiq, Lee-Seok Hwang, and Mark A Trombley. “Arbitrage risk and the book-to-market anomaly”. *Journal of Financial Economics* 69.2 (2003), pp. 355–373.

Alwathainani, Abdulaziz M. “Consistency of firms’ past financial performance measures and future returns”. *The British Accounting Review* 41.3 (2009), pp. 184–196.

Amihud, Yakov. “Illiquidity and stock returns: cross-section and time-series effects”. *Journal of financial markets* 5.1 (2002), pp. 31–56.

Amihud, Yakov and Haim Mendelson. “Asset pricing and the bid-ask spread”. *Journal of financial Economics* 17.2 (1986), pp. 223–249.

Anderson, Christopher W and Luis Garcia-Feijoo. “Empirical evidence on capital investment, growth options, and security returns”. *The Journal of Finance* 61.1 (2006), pp. 171–194.

Anderson, Evan W, Eric Ghysels, and Jennifer L Juergens. “Do heterogeneous beliefs matter for asset pricing?” *The Review of Financial Studies* 18.3 (2005), pp. 875–924.

Ang, Andrew, Joseph Chen, and Yuhang Xing. “Downside risk”. *The review of financial studies* 19.4 (2006), pp. 1191–1239.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. “The cross-section of volatility and expected returns”. *The Journal of Finance* 61.1 (2006), pp. 259–299.

Asness, Clifford S, R Burt Porter, and Ross L Stevens. “Predicting stock returns using industry-relative firm characteristics”. *Available at SSRN 213872* (2000).

Asquith, Paul, Parag A Pathak, and Jay R Ritter. “Short interest, institutional ownership, and stock returns”. *Journal of Financial Economics* 78.2 (2005), pp. 243–276.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov. “Momentum and credit rating”. *The Journal of Finance* 62.5 (2007), pp. 2503–2520.

Balakrishnan, Karthik, Eli Bartov, and Lucile Faurel. “Post loss/profit announcement drift”. *Journal of Accounting and Economics* 50.1 (2010), pp. 20–41.

Bali, Turan G, Nusret Cakici, and Robert F Whitelaw. “Maxing out: Stocks as lotteries and the cross-section of expected returns”. *Journal of Financial Economics* 99.2 (2011), pp. 427–446.

Bali, Turan G, Robert F Engle, and Scott Murray. *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons, 2016.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri Nikolaev. “Accruals, cash flows, and operating profitability in the cross section of stock returns”. *Journal of Financial Economics* 121.1 (2016), pp. 28–45.

Balvers, Ronald J and Dayong Huang. “Productivity-based asset pricing: Theory and evidence”. *Journal of Financial Economics* 86.2 (2007), pp. 405–445.

Bandyopadhyay, Sati P, Alan G Huang, and Tony S Wirjanto. “The accrual volatility anomaly”. *Unpublished Manuscript, University of Waterloo* (2010).

Banz, Rolf W. “The relationship between return and market value of common stocks”. *Journal of financial economics* 9.1 (1981), pp. 3–18.

Barbee Jr, William C, Sandip Mukherji, and Gary A Raines. “Do sales–price and debt–equity explain stock returns better than book–market and firm size?” *Financial Analysts Journal* 52.2 (1996), pp. 56–60.

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman. “Can investors profit from the prophets? Security analyst recommendations and stock returns”. *The Journal of Finance* 56.2 (2001), pp. 531–563.

Barry, Christopher B and Stephen J Brown. “Differential information and the small firm effect”. *Journal of financial economics* 13.2 (1984), pp. 283–294.

Barth, Mary E and Amy P Hutton. “Analyst earnings forecast revisions and the pricing of accruals”. *Review of accounting studies* 9.1 (2004), pp. 59–96.

Bartov, Eli and Myungsun Kim. “Risk, mispricing, and value investing”. *Review of Quantitative Finance and Accounting* 23.4 (2004), pp. 353–376.

Basu, Sanjoy. “Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis”. *The journal of Finance* 32.3 (1977), pp. 663–682.

Belo, Frederico and Xiaoji Lin. “The inventory growth spread”. *The Review of Financial Studies* 25.1 (2012), pp. 278–313.

Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch. “Labor hiring, investment, and stock return predictability in the cross section”. *Journal of Political Economy* 122.1 (2014), pp. 129–177.

Belo, Frederico, Xiaoji Lin, and Maria Ana Vitorino. "Brand capital and firm value". *Review of Economic Dynamics* 17.1 (2014), pp. 150–169.

Bettis, Richard A. "The search for asterisks: Compromised statistical tests and flawed theories". *Strategic Management Journal* 33.1 (2012), pp. 108–113.

Bhandari, Laxmi Chand. "Debt/equity ratio and expected common stock returns: Empirical evidence". *The journal of finance* 43.2 (1988), pp. 507–528.

Blitz, David, Joop Huij, and Martin Martens. "Residual momentum". *Journal of Empirical Finance* 18.3 (2011), pp. 506–521.

Blume, Marshall E. and Frank Husic. "Price, beta, and exchange listing". *Journal of Finance* 28.2 (1972), pp. 283–299.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts. "On the importance of measuring payout yield: Implications for empirical asset pricing". *The Journal of Finance* 62.2 (2007), pp. 877–915.

Bradshaw, Mark T, Scott A Richardson, and Richard G Sloan. "The relation between corporate financing activities, analysts' forecasts and stock returns". *Journal of accounting and economics* 42.1-2 (2006), pp. 53–85.

Brennan, Michael J, Tarun Chordia, and Avanidhar Subrahmanyam. "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns". *Journal of financial Economics* 49.3 (1998), pp. 345–373.

Brown, David P and Bradford Rowe. "The productivity premium in equity returns". *Available at SSRN 993467* (2007).

Callen, Jeffrey L, Mozaffar Khan, and Hai Lu. "Accounting quality, stock price delay, and future stock returns". *Contemporary Accounting Research* 30.1 (2013), pp. 269–295.

Campbell, John Y, Jens Hilscher, and Jan Szilagyi. "In search of distress risk". *The Journal of Finance* 63.6 (2008), pp. 2899–2939.

Chan, Kevin C, Nai-fu Chen, and David A Hsieh. "An exploratory investigation of the firm size effect". *Journal of Financial Economics* 14.3 (1985), pp. 451–471.

Chan, Louis KC, Narasimhan Jegadeesh, and Josef Lakonishok. "Momentum strategies". *The Journal of Finance* 51.5 (1996), pp. 1681–1713.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis. "The stock market valuation of research and development expenditures". *The Journal of finance* 56.6 (2001), pp. 2431–2456.

Chandrashekar, Satyajit and Ramesh K. S. Rao. “The Productivity of Corporate Cash Holdings and the Cross-Section of Expected Stock Returns”. *Available at SSRN 1334162* (2007).

Chang, Andrew C and Phillip Li. “Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say “Often Not””. *Critical Finance Review* 7 (2018).

Chen, Andrew Y. “Do T-Stat Hurdles Need to Be Raised?” *Available at SSRN 3254995* (2020).

— “The Limits of p-Hacking: Some Thought Experiments”. *the Journal of Finance* (Forthcoming).

Chen, Andrew Y and Mihail Velikov. “Zeroing in on the Expected Returns of Anomalies”. *Working Paper* (2019).

Chen, Andrew Y and Tom Zimmermann. “Publication bias and the cross-section of stock returns”. *The Review of Asset Pricing Studies* 10.2 (2020), pp. 249–289.

Chen, Joseph, Harrison Hong, and Jeremy C Stein. “Breadth of ownership and stock returns”. *Journal of financial Economics* 66.2-3 (2002), pp. 171–205.

Chen, Long and Lu Zhang. “A better three-factor model that explains more anomalies”. *Journal of Finance* 65.2 (2010), pp. 563–595.

Chen, Nai-Fu, Richard Roll, and Stephen A Ross. “Economic forces and the stock market”. *Journal of business* (1986), pp. 383–403.

Chordia, Tarun, Amit Goyal, and Alessio Saretto. “Anomalies and false rejections”. *The Review of Financial Studies* 33.5 (2020), pp. 2134–2179.

Chordia, Tarun, Avanidhar Subrahmanyam, and V Ravi Anshuman. “Trading activity and expected stock returns”. *Journal of financial Economics* 59.1 (2001), pp. 3–32.

Cohen, Lauren, Karl Diether, and Christopher Malloy. “Misvaluing innovation”. *The Review of Financial Studies* 26.3 (2013), pp. 635–666.

Cohen, Lauren and Andrea Frazzini. “Economic links and predictable returns”. *The Journal of Finance* 63.4 (2008), pp. 1977–2011.

Cohen, Lauren and Dong Lou. “Complicated firms”. *Journal of financial economics* 104.2 (2012), pp. 383–400.

Cooper, Michael J, Huseyin Gulen, and Michael J Schill. “Asset growth and the cross-section of stock returns”. *the Journal of Finance* 63.4 (2008), pp. 1609–1651.

Corwin, Shane A and Paul Schultz. “A simple way to estimate bid-ask spreads from daily high and low prices”. *The Journal of Finance* 67.2 (2012), pp. 719–760.

Cremers, KJ Martijn and Vinay B Nair. “Governance mechanisms and equity prices”. *the Journal of Finance* 60.6 (2005), pp. 2859–2894.

Cusatis, Patrick J, James A Miles, and J Randall Woolridge. “Restructuring through spinoffs: The stock market evidence”. *Journal of Financial Economics* 33.3 (1993), pp. 293–311.

Da, Zhi. “Cash flow, consumption risk, and the cross-section of stock returns”. *The Journal of Finance* 64.2 (2009), pp. 923–956.

Da, Zhi and Mitch Warachka. “The disparity between long-term and short-term forecasted earnings growth”. *Journal of Financial Economics* 100.2 (2011), pp. 424–442.

Daniel, Kent and Sheridan Titman. “Market reactions to tangible and intangible information”. *The Journal of Finance* 61.4 (2006), pp. 1605–1643.

Datar, Vinay T, Narayan Y Naik, and Robert Radcliffe. “Liquidity and stock returns: An alternative test”. *Journal of financial markets* 1.2 (1998), pp. 203–219.

De Bondt, Werner FM and Richard Thaler. “Does the stock market overreact?” *The Journal of finance* 40.3 (1985), pp. 793–805.

Dechow, Patricia M, Amy P Hutton, Lisa Meulbroek, and Richard G Sloan. “Short-sellers, fundamental analysis, and stock returns”. *Journal of financial Economics* 61.1 (2001), pp. 77–106.

Dechow, Patricia M, Richard G Sloan, and Mark T Soliman. “Implied equity duration: A new measure of equity risk”. *Review of Accounting Studies* 9.2 (2004), pp. 197–228.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal. “A transaction-cost perspective on the multitude of firm characteristics”. *The Review of Financial Studies* 33.5 (2020), pp. 2180–2222.

Desai, Hemang, Shivaram Rajgopal, and Mohan Venkatachalam. “Value-glamour and accruals mispricing: One anomaly or two?” *The Accounting Review* 79.2 (2004), pp. 355–385.

Detzel, Andrew, Robert Novy-Marx, and Mihail Velikov. “Model selection with transaction costs”. *available at SSRN* 3088828 (2019).

Dewald, William G, Jerry G Thursby, and Richard G Anderson. "Replication in empirical economics: The journal of money, credit and banking project". *The American Economic Review* (1986), pp. 587–603.

Dharan, Bala G and David L Ikenberry. "The long-run negative drift of post-listing stock returns". *The Journal of Finance* 50.5 (1995), pp. 1547–1574.

Dichev, Ilia D. "Is the risk of bankruptcy a systematic risk?" *the Journal of Finance* 53.3 (1998), pp. 1131–1147.

Dichev, Ilia D and Joseph D Piotroski. "The long-run stock returns following bond ratings changes". *The Journal of Finance* 56.1 (2001), pp. 173–203.

Diether, Karl B, Christopher J Malloy, and Anna Scherbina. "Differences of opinion and the cross section of stock returns". *The Journal of Finance* 57.5 (2002), pp. 2113–2141.

Dimson, Elroy. "Risk measurement when shares are subject to infrequent trading". *Journal of Financial Economics* 7.2 (1979), pp. 197–226.

Doyle, Jeffrey T, Russell J Lundholm, and Mark T Soliman. "The predictive value of expenses excluded from pro forma earnings". *Review of Accounting Studies* 8.2 (2003), pp. 145–174.

Easley, David, Soeren Hvidkjaer, and Maureen O'hara. "Is information risk a determinant of asset returns?" *The journal of finance* 57.5 (2002), pp. 2185–2221.

Eberhart, Allan C, William F Maxwell, and Akhtar R Siddique. "An examination of long-term abnormal stock returns and operating performance following R&D increases". *The Journal of Finance* 59.2 (2004), pp. 623–650.

Eisfeldt, Andrea L and Dimitris Papanikolaou. "Organization capital and the cross-section of expected returns". *The Journal of Finance* 68.4 (2013), pp. 1365–1406.

Elgers, Pieter T, May H Lo, and Ray J Pfeiffer. "Delayed security price adjustments to financial analysts' forecasts of annual earnings". *The Accounting Review* 76.4 (2001), pp. 613–632.

Fairfield, Patricia M, Scott Whisenant, and Teri Lombardi Yohn. "Accrued earnings and growth: Implications for earnings persistence and market mispricing". *Accounting Review* 78.1 (2003).

Fama, E.F. and K.R. French. "Common risk factors in the returns on stocks and bonds". *Journal of financial economics* 33.1 (1993), pp. 3–56.

Fama, Eugene F and Kenneth R French. "Profitability, investment and average returns". *Journal of financial economics* 82.3 (2006), pp. 491–518.

Fama, Eugene F and Kenneth R French. "The cross-section of expected stock returns". *the Journal of Finance* 47.2 (1992), pp. 427–465.

Fama, Eugene F and James D MacBeth. "Risk, return, and equilibrium: Empirical tests". *Journal of political economy* 81.3 (1973), pp. 607–636.

Foster, George, Chris Olsen, and Terry Shevlin. "Earnings releases, anomalies, and the behavior of security returns". *Accounting Review* (1984), pp. 574–603.

Francis, Jennifer, Ryan LaFond, Per M Olsson, and Katherine Schipper. "Costs of equity and earnings attributes". *The accounting review* 79.4 (2004), pp. 967–1010.

Francis, Jennifer, Ryan LaFond, Per Olsson, and Katherine Schipper. "The market pricing of accruals quality". *Journal of accounting and economics* 39.2 (2005), pp. 295–327.

Frankel, Richard and Charles MC Lee. "Accounting valuation, market expectation, and cross-sectional stock returns". *Journal of Accounting and economics* 25.3 (1998), pp. 283–319.

Franzoni, Francesco and Jose M Marin. "Pension plan funding and stock market efficiency". *the Journal of Finance* 61.2 (2006), pp. 921–956.

Frazzini, Andrea and Lasse Heje Pedersen. "Betting against beta". *Journal of Financial Economics* 111.1 (2014), pp. 1–25.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. "Dissecting characteristics nonparametrically". *The Review of Financial Studies* 33.5 (2020), pp. 2326–2377.

Gebhardt, William R, Charles MC Lee, and Bhaskaran Swaminathan. "Toward an implied cost of capital". *Journal of accounting research* 39.1 (2001), pp. 135–176.

George, Thomas J and Chuan-Yang Hwang. "The 52-week high and momentum investing". *The Journal of Finance* 59.5 (2004), pp. 2145–2176.

Gettleman, Eric and Joseph M Marks. "Acceleration strategies". *SSRN Electronic Journal* (2006).

Gompers, Paul, Joy Ishii, and Andrew Metrick. "Corporate governance and equity prices". *The quarterly journal of economics* 118.1 (2003), pp. 107–156.

Green, Jeremiah, John RM Hand, and X Frank Zhang. "The characteristics that provide independent information about average us monthly stock returns". *The Review of Financial Studies* (2017), hhx019.

— "The supraview of return predictive signals". *Review of Accounting Studies* 18.3 (2013), pp. 692–730.

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

Hafzalla, Nader, Russell Lundholm, and E Matthew Van Winkle. "Percent accruals". *The Accounting Review* 86.1 (2011), pp. 209–236.

Hahn, Jaehoon and Hangyong Lee. "Financial constraints, debt capacity, and the cross-section of stock returns". *The Journal of Finance* 64.2 (2009), pp. 891–921.

Hartzmark, Samuel M and David H Solomon. "The dividend month premium". *Journal of Financial Economics* 109.3 (2013), pp. 640–660.

Harvey, Campbell R and Yan Liu. "A census of the factor zoo". *Available at SSRN 3341728* (2019).

Harvey, Campbell R, Yan Liu, and Heqing Zhu. "... and the cross-section of expected returns". *The Review of Financial Studies* 29.1 (2016), pp. 5–68.

Harvey, Campbell R and Akhtar Siddique. "Conditional skewness in asset pricing tests". *The Journal of finance* 55.3 (2000), pp. 1263–1295.

Haugen, Robert A and Nardin L Baker. "Commonality in the determinants of expected stock returns". *Journal of financial economics* 41.3 (1996), pp. 401–439.

Heston, Steven L and Ronnie Sadka. "Seasonality in the cross-section of stock returns". *Journal of Financial Economics* 87.2 (2008), pp. 418–445.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang. "Do investors overvalue firms with bloated balance sheets?" *Journal of Accounting and Economics* 38 (2004), pp. 297–331.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li. "Innovative efficiency and stock returns". *Journal of Financial Economics* 107.3 (2013), pp. 632–654.

Holthausen, Robert W and David F Larcker. "The prediction of stock returns using financial statement information". *Journal of accounting and economics* 15.2-3 (1992), pp. 373–411.

Hong, Harrison and Marcin Kacperczyk. "The price of sin: The effects of social norms on markets". *Journal of financial economics* 93.1 (2009), pp. 15–36.

Hou, Kewei. "Industry information diffusion and the lead-lag effect in stock returns". *The Review of Financial Studies* 20.4 (2007), pp. 1113–1138.

Hou, Kewei and Roger K Loh. "Have we solved the idiosyncratic volatility puzzle?" *Journal of Financial Economics* 121.1 (2016), pp. 167–194.

Hou, Kewei and Tobias J Moskowitz. “Market frictions, price delay, and the cross-section of expected returns”. *The Review of Financial Studies* 18.3 (2005), pp. 981–1020.

Hou, Kewei and David T Robinson. “Industry concentration and average stock returns”. *The Journal of Finance* 61.4 (2006), pp. 1927–1956.

Hou, Kewei, Chen Xue, and Lu Zhang. “Digesting anomalies: An investment approach”. *The Review of Financial Studies* 28.3 (2015), pp. 650–705.

— “Replicating anomalies”. *The Review of Financial Studies* 33.5 (2020), pp. 2019–2133.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen. “Market underreaction to open market share repurchases”. *Journal of financial economics* 39.2-3 (1995), pp. 181–208.

Ioannidis, John PA. “Why most published research findings are false”. *PLoS medicine* 2.8 (2005), e124.

Jacobs, Heiko and Sebastian Müller. “Anomalies across the globe: Once public, no longer existent?” *Journal of Financial Economics* 135.1 (2020), pp. 213–230.

Jegadeesh, Narasimhan. “Evidence of predictable behavior of security returns”. *The Journal of finance* 45.3 (1990), pp. 881–898.

Jegadeesh, Narasimhan, Joonghyuk Kim, Susan D Krische, and Charles MC Lee. “Analyzing the analysts: When do recommendations add value?” *The journal of finance* 59.3 (2004), pp. 1083–1124.

Jegadeesh, Narasimhan and Joshua Livnat. “Revenue surprises and stock returns”. *Journal of Accounting and Economics* 41.1-2 (2006), pp. 147–171.

Jegadeesh, Narasimhan and Sheridan Titman. “Returns to buying winners and selling losers: Implications for stock market efficiency”. *The Journal of finance* 48.1 (1993), pp. 65–91.

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen. *Is There A Replication Crisis In Finance?* Tech. rep. National Bureau of Economic Research, 2021.

Johnson, Travis L and Eric C So. “The option to stock volume ratio and future returns”. *Journal of Financial Economics* 106.2 (2012), pp. 262–286.

Kelly, Bryan and Hao Jiang. “Tail risk and asset prices”. *The Review of Financial Studies* 27.10 (2014), pp. 2841–2871.

Kot, Hung Wan and Kalok Chan. “Can contrarian strategies improve momentum profits”. *Journal of Investment Management* 4.1 (2006).

La Porta, Rafael. “Expectations and the cross-section of stock returns”. *The Journal of Finance* 51.5 (1996), pp. 1715–1742.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny. “Contrarian investment, extrapolation, and risk”. *The journal of finance* 49.5 (1994), pp. 1541–1578.

Lamont, Owen, Christopher Polk, and Jesus Saa-Requejo. “Financial Constraints and Stock Returns”. *The Review of Financial Studies* 14.2 (2001), pp. 529–554.

Landsman, Wayne R, Bruce L Miller, Ken Peasnell, and Shu Yeh. “Do investors understand really dirty surplus?” *The Accounting Review* 86.1 (2011), pp. 237–258.

Lee, Charles MC and Bhaskaran Swaminathan. “Price momentum and trading volume”. *the Journal of Finance* 55.5 (2000), pp. 2017–2069.

Lerman, Alina, Joshua Livnat, and Richard R Mendenhall. “The high-volume return premium and post-earnings announcement drift”. *Available at SSRN 1122463* (2008).

Lettau, M. and S. Ludvigson. “Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying”. *Journal of Political Economy* (2001), pp. 1238–1287.

Lev, Baruch and Doron Nissim. “Taxable income, future earnings, and equity values”. *The accounting review* 79.4 (2004), pp. 1039–1074.

Li, Dongmei. “Financial constraints, R&D investment, and stock returns”. *The Review of Financial Studies* 24.9 (2011), pp. 2974–3007.

Li, Hongtao, Robert Novy-Marx, and Mihail Velikov. “Liquidity Risk and Asset Pricing”. *Critical Finance Review* 8.1-2 (2019), pp. 223–255.

Linnainmaa, Juhani T and Michael R Roberts. “The history of the cross-section of stock returns”. *The Review of Financial Studies* 31.7 (2018), pp. 2606–2649.

Liu, Weimin. “A liquidity-augmented capital asset pricing model”. *Journal of financial Economics* 82.3 (2006), pp. 631–671.

Lockwood, Larry and Wikrom Prombutr. “Sustainable growth and stock returns”. *Journal of Financial Research* 33.4 (2010), pp. 519–538.

Loh, Roger K and Mitch Warachka. “Streaks in earnings surprises and the cross-section of stock returns”. *Management Science* 58.7 (2012), pp. 1305–1321.

Lou, Dong. “Attracting investor attention through advertising”. *The Review of Financial Studies* 27.6 (2014), pp. 1797–1829.

Loughran, Tim and Jay W Wellman. “New evidence on the relation between the enterprise multiple and average stock returns”. *Journal of Financial and Quantitative Analysis* (2011), pp. 1629–1650.

Lyandres, Evgeny, Le Sun, and Lu Zhang. “The new issues puzzle: Testing the investment-based explanation”. *The Review of Financial Studies* 21.6 (2008), pp. 2825–2855.

McCullough, Bruce D, Kerry Anne McGeary, and Teresa D Harrison. “Lessons from the JMCB Archive”. *Journal of Money, Credit and Banking* (2006), pp. 1093–1107.

McLean, R David and Jeffrey Pontiff. “Does academic research destroy stock return predictability?” *The Journal of Finance* 71.1 (2016), pp. 5–32.

Menzly, Lior and Oguzhan Ozbas. “Market segmentation and cross-predictability of returns”. *The Journal of Finance* 65.4 (2010), pp. 1555–1580.

Michaely, Roni, Richard H Thaler, and Kent L Womack. “Price reactions to dividend initiations and omissions: Overreaction or drift?” *the Journal of Finance* 50.2 (1995), pp. 573–608.

Mohanram, Partha S. “Separating winners from losers among lowbook-to-market stocks using financial statement analysis”. *Review of accounting studies* 10.2-3 (2005), pp. 133–170.

Moskowitz, Tobias J and Mark Grinblatt. “Do industries explain momentum?” *The Journal of finance* 54.4 (1999), pp. 1249–1290.

Nagel, Stefan. “Short sales, institutional investors and the cross-section of stock returns”. *Journal of financial economics* 78.2 (2005), pp. 277–309.

Naranjo, Andy, M Nimalendran, and Mike Ryngaert. “Stock returns, dividend yields, and taxes”. *The Journal of Finance* 53.6 (1998), pp. 2029–2057.

Nguyen, Giao X and Peggy E Swanson. “Firm characteristics, relative efficiency, and equity returns”. *Journal of Financial and Quantitative Analysis* (2009), pp. 213–236.

Nosek, Brian A, Jeffrey R Spies, and Matt Motyl. “Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability”. *Perspectives on Psychological Science* 7.6 (2012), pp. 615–631.

Novy-Marx, Robert. “Is momentum really momentum?” *Journal of Financial Economics* 103.3 (2012), pp. 429–453.

— “Operating leverage”. *Review of Finance* 15.1 (2011), pp. 103–134.

— “The other side of value: The gross profitability premium”. *Journal of financial economics* 108.1 (2013), pp. 1–28.

Novy-Marx, Robert and Mihail Velikov. “A taxonomy of anomalies and their trading costs”. *The Review of Financial Studies* 29.1 (2016), pp. 104–147.

— “Betting Against Betting Against Beta”. *Journal of Financial Economics* (Forthcoming).

Ortiz-Molina, Hernán and Gordon M Phillips. “Real asset illiquidity and the cost of capital”. *Journal of Financial and Quantitative Analysis* (2014), pp. 1–32.

Ou, Jane A and Stephen H Penman. “Accounting measurement, price-earnings ratio, and the information content of security prices”. *Journal of accounting research* 27 (1989), pp. 111–144.

— “Financial statement analysis and the prediction of stock returns”. *Journal of accounting and economics* 11.4 (1989), pp. 295–329.

Palazzo, Berardino. “Cash holdings, risk, and expected returns”. *Journal of Financial Economics* 104.1 (2012), pp. 162–185.

Pástor, L’uboš and Robert F Stambaugh. “Liquidity risk and expected stock returns”. *Journal of Political economy* 111.3 (2003), pp. 642–685.

Penman, Stephen H, Scott A Richardson, and Irem Tuna. “The book-to-price effect in stock returns: accounting for leverage”. *Journal of accounting research* 45.2 (2007), pp. 427–467.

Piotroski, Joseph D. “Value investing: The use of historical financial statement information to separate winners from losers”. *Journal of Accounting Research* (2000), pp. 1–41.

Pontiff, Jeffrey and Rohit Singla. “Liquidity Risk?” *Critical Finance Review* 8.1-2 (2019), pp. 257–276.

Pontiff, Jeffrey and Artemiza Woodgate. “Share issuance and cross-sectional returns”. *The Journal of Finance* 63.2 (2008), pp. 921–945.

Prakash, Rachna and Nishi Sinha. “Deferred revenues and the matching of revenues and expenses”. *Contemporary Accounting Research* 30.2 (2013), pp. 517–548.

Rajgopal, Shivaram, Terry Shevlin, and Mohan Venkatachalam. “Does the stock market fully appreciate the implications of leading indicators for future earnings? Evidence from order backlog”. *Review of Accounting Studies* 8.4 (2003), pp. 461–492.

Richardson, Scott A, Richard G Sloan, Mark T Soliman, and Irem Tuna. “Accrual reliability, earnings persistence and stock prices”. *Journal of accounting and economics* 39.3 (2005), pp. 437–485.

Ritter, Jay R. “The long-run performance of initial public offerings”. *The journal of finance* 46.1 (1991), pp. 3–27.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein. “Persuasive evidence of market inefficiency”. *The Journal of Portfolio Management* 11.3 (1985), pp. 9–16.

Scherbina, Anna. “Suppressed negative information and future underperformance”. *Review of Finance* 12.3 (2008), pp. 533–565.

Sloan, Richard G. “Do stock prices fully reflect information in accruals and cash flows about future earnings?” *Accounting review* (1996), pp. 289–315.

Soliman, Mark T. “The use of DuPont analysis by market participants”. *The Accounting Review* 83.3 (2008), pp. 823–853.

Spiess, D Katherine and John Affleck-Graves. “The long-run performance of stock returns following debt offerings”. *Journal of Financial Economics* 54.1 (1999), pp. 45–73.

Thomas, Jacob K and Huai Zhang. “Inventory changes and future returns”. *Review of Accounting Studies* 7.2 (2002), pp. 163–187.

Thomas, Jacob and Frank X Zhang. “Tax expense momentum”. *Journal of Accounting Research* 49.3 (2011), pp. 791–821.

Titman, Sheridan, KC John Wei, and Feixue Xie. “Capital investments and stock returns”. *Journal of financial and Quantitative Analysis* 39.4 (2004), pp. 677–700.

Tuzel, Selale. “Corporate real estate holdings and the cross-section of stock returns”. *The Review of Financial Studies* 23.6 (2010), pp. 2268–2302.

Valta, Philip. “Strategic default, debt structure, and stock returns”. *Journal of Financial and Quantitative Analysis* (2016), pp. 197–229.

Vassalou, Maria. “News related to future GDP growth as a risk factor in equity returns”. *Journal of financial economics* 68.1 (2003), pp. 47–73.

Vora, Premal and Luis Palacios. *Fama-French Factors*. Tech. rep. Wharton Research Data Services, 2010.

Watkins, Boyce. “Riding the wave of sentiment: An analysis of return consistency as a predictor of future returns”. *The Journal of Behavioral Finance* 4.4 (2003), pp. 191–200.

Welch, Ivo. “Reproducing, Extending, Updating, Replicating, Reexamining, and Reconciling”. *Critical Finance Review* 8.1-2 (2019), pp. 301–304. ISSN: 2164-5744. URL: `http://dx.doi.org/10.1561/104.00000082`.

Whited, Toni M and Guojun Wu. “Financial constraints risk”. *The Review of Financial Studies* 19.2 (2006), pp. 531–559.

Xie, Hong. “The mispricing of abnormal accruals”. *The accounting review* 76.3 (2001), pp. 357–373.

Xing, Yuhang, Xiaoyan Zhang, and Rui Zhao. “What does the individual option volatility smirk tell us about future equity returns?” *Journal of Financial and Quantitative Analysis* (2010), pp. 641–662.

Yan, Shu. “Jump risk, stock returns, and slope of implied volatility smile”. *Journal of Financial Economics* 99.1 (2011), pp. 216–233.

Figure 1: Illustration of Hand-Collection from Original Papers. Description: The figure shows an excerpt from the spreadsheet we used to hand-collect predictability evidence from the original papers. The full spreadsheet can be found at https://github.com/OpenSourceAP/CrossSection. Interpretation: Evidence required for evaluating reproductions is documented at the characteristic level and readers can easily trace our predictor categorizations to the original papers' exhibits.

Acronym	Authors	Predictability in OP	Key Table in OP	Test in OP	Sign	Return	T-Stat	Stock Weight
GrLTNOA	Fairfield, Whisenant a	2_likely	5A and B	port sort but no LS FF3 adjusted	1	0.61	NA	EW
AM	Fama and French	1_clear	3 Ln(A/ME)	univariate reg	1	NA	5.69	EW
AMq	Fama and French	indirect	NA	NA	NA	NA	NA	NA
BMdec	Fama and French	1_clear	3 Ln(BE/ME)	univariate reg	1	0.50	5.71	EW
BookLeverage	Fama and French	1_clear	3 Ln(A/BE)	mv reg	-1	NA	5.34	EW
BookLeverageQuarterly	Fama and French	indirect	NA	NA	NA	NA	NA	NA
OperProf	Fama and French	2_likely	3 Y_t/B_t	mv reg	1	NA	2.55	EW
OperProfLag	Fama and French	indirect	NA	NA	NA	NA	NA	NA
OperProfLag_q	Fama and French	indirect	NA	NA	NA	NA	NA	NA
Beta	Fama and MacBeth	2_likely	3A t(gamma_1)	univariate reg	1	NA	2.57	EW
BetaSquared	Fama and MacBeth	4_not	3A t(gamma2)	mv reg	-1	NA	0.29	EW
EarningsSurprise	Foster, Olsen and She	1_clear	4 Days +1 to +60	event study 2 months	1	2.98	NA	EW

Figure 2: Reproduction Success Rates for Clear Predictors. Description: We construct one long-short portfolio from each clear predictor following the original papers' results and examine the t-stat for the hypothesis that the mean return is zero in the original papers' sample periods. Clear predictors are those where the original papers clearly demonstrate that our portfolios should be statistically significant (see Section 2.4). Interpretation: Almost 100% of the cross-sectional predictability literature can reproduced and our code and data show how.

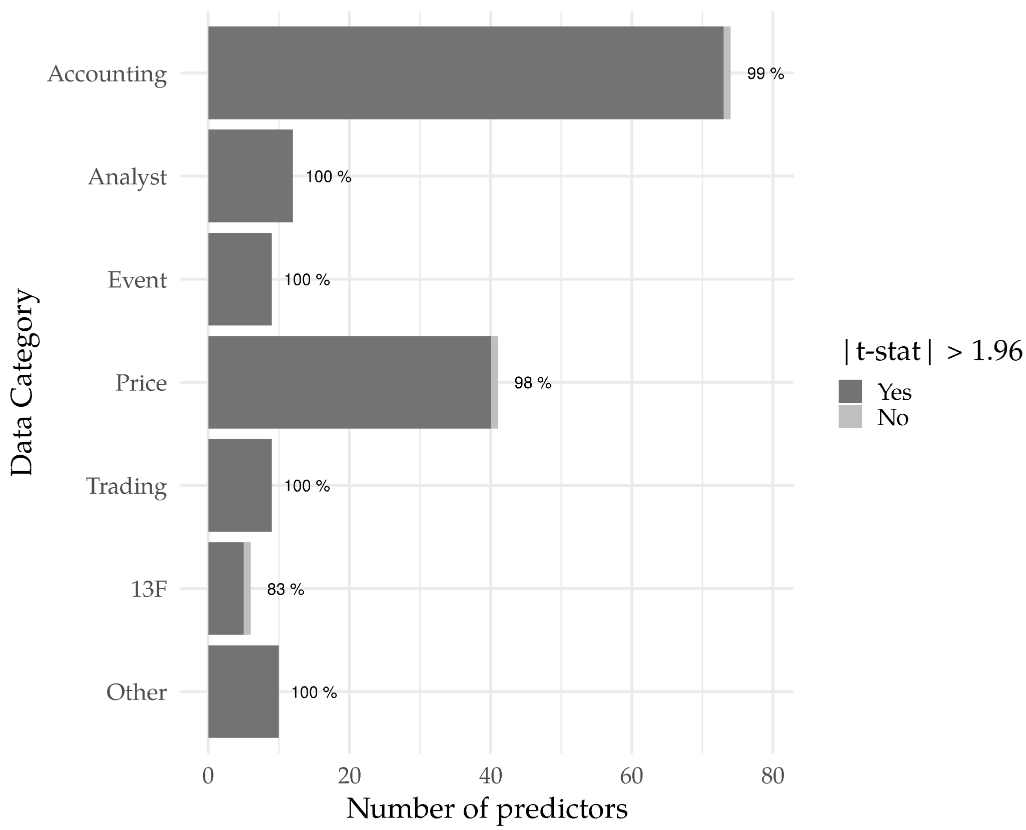


Figure 3: Comparison of Reproduced Long-Short t-stats with Original Long-Short t-stats for All Predictors. Description: We compare reproduced long-short t-stats to hand-collected t-stats from the original studies. We include both clear and likely predictors (defined in Section 2.4). We exclude t-stats from regressions and event studies, and also drop t-stats from three studies because our portfolios deviate significantly from the originals for idiosyncratic reasons (Barber et al. 2001, Frazzini and Pedersen 2014, Hou and Moskowitz 2005, see Section 2). Reproduced t-stats use raw returns but original t-stats may include factor or characteristic adjustments. Axes are log-scale to make the predictor acronyms easier to read. Full references are found in Tables 2 and 3. Regression fit uses OLS in levels, though logs leads to similar results. Interpretation: The overwhelming reproduction success in Figure 2 is not due to predictor categorizations or marginally significant reproductions. Our reproduced t-stats quantitatively match the original papers' for both clear and likely predictors.

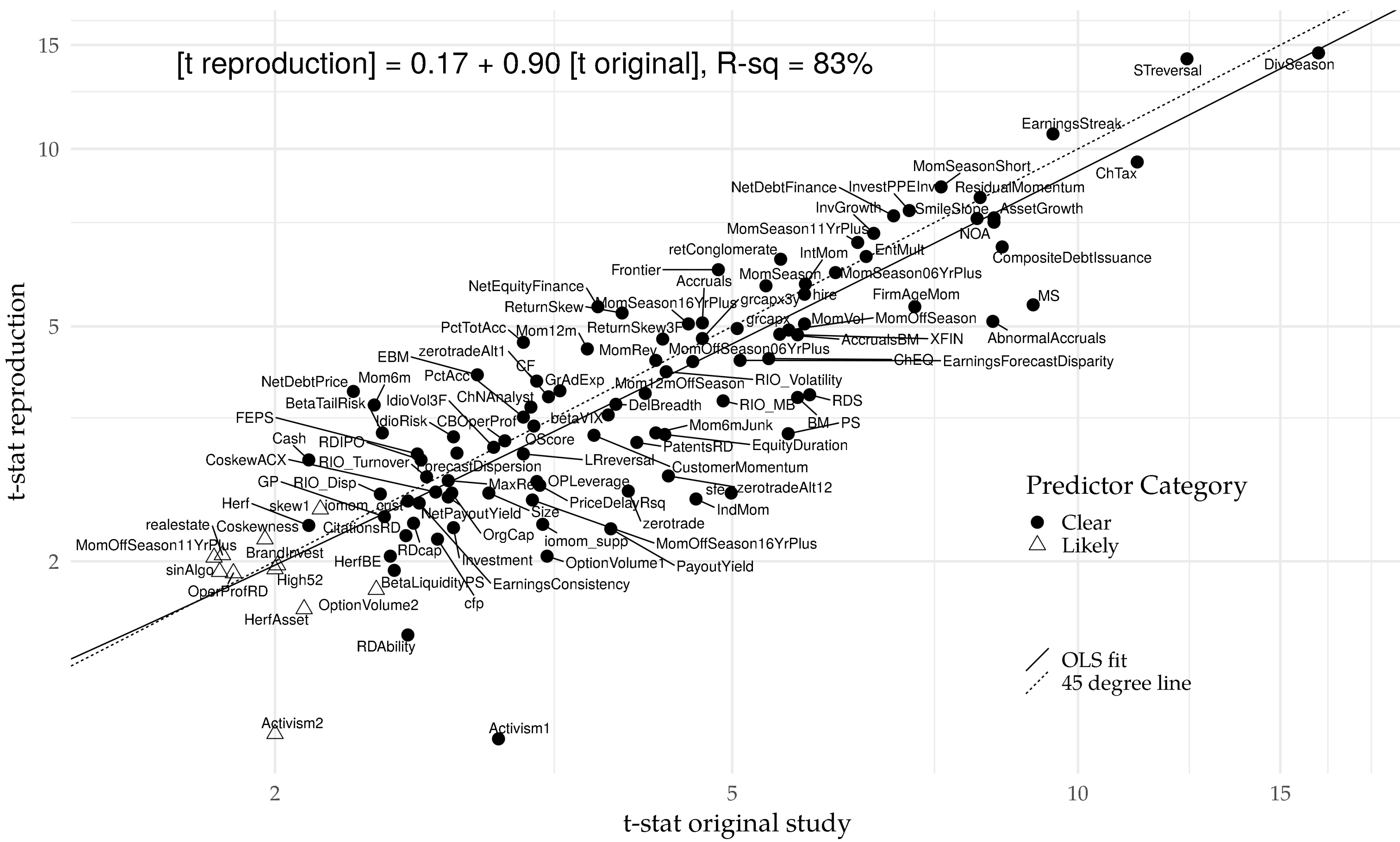


Figure 4: Monotonicity of Mean Returns. Description: We form decile portfolios and examine the mean monthly return in each decile. Data is limited to the 170 continuous predictors. Each marker is one predictor-decile. Filled circles indicate that the mean return in decile k exceeds the mean return for decile $k-1$ for the same predictor. Interpretation: Mean returns increase monotonically in the signal, reproducing the fact that most original portfolio sorts show monotonicity, and showing predictability is robust to implementation details. Monotonicity supports the idea that p-hacking does not explain anomalies, as p-hacking would have to operate across the entire predictor distribution.

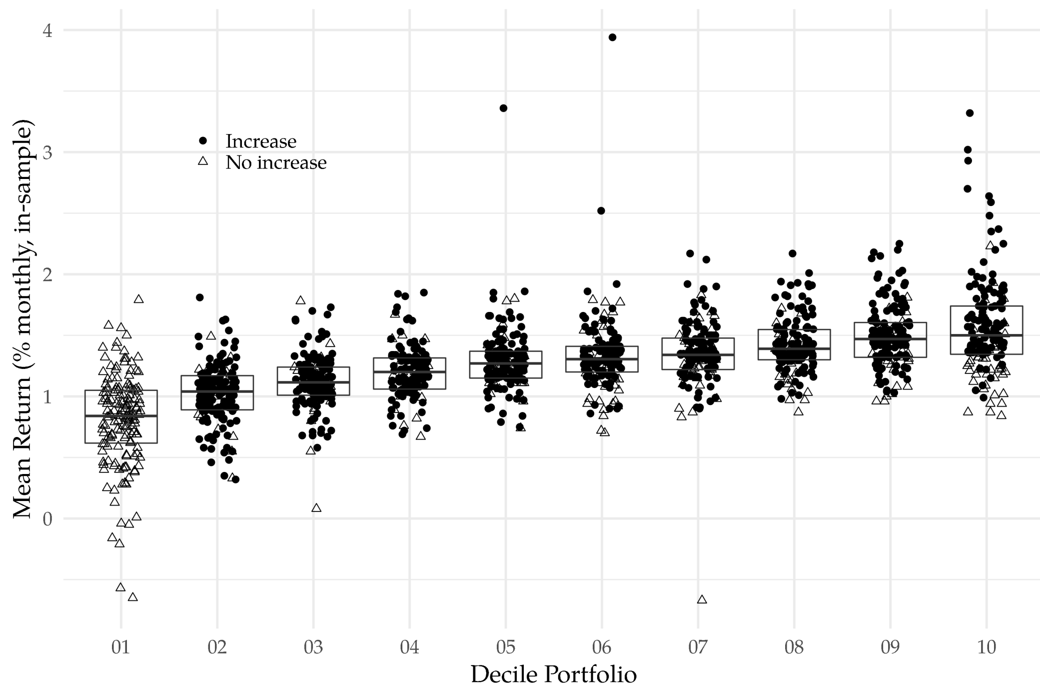


Figure 5: Reproduction Performance for All Characteristics. Description: We construct one long-short portfolio from each characteristic following the original papers and examine the t-stat for the hypothesis that the mean return is zero in the original papers' sample periods. Predictor categories use results from the original papers to judgmentally determine whether we should expect to find statistical significance in our portfolio tests (see Section 2.4). Clear predictors provide clear evidence, likely predictors have borderline evidence, and not-predictors imply insignificance. Indirect signals had only suggestive evidence of predictive power. Almost all indirect signals come from Hou, Xue, and Zhang (2020). Interpretation: Likely predictors have reproduced t-stats that average around 2.0, consistent with the original evidence. Not predictors are also reliably reproduced. Indirect signals vary wildly in terms of their performance.

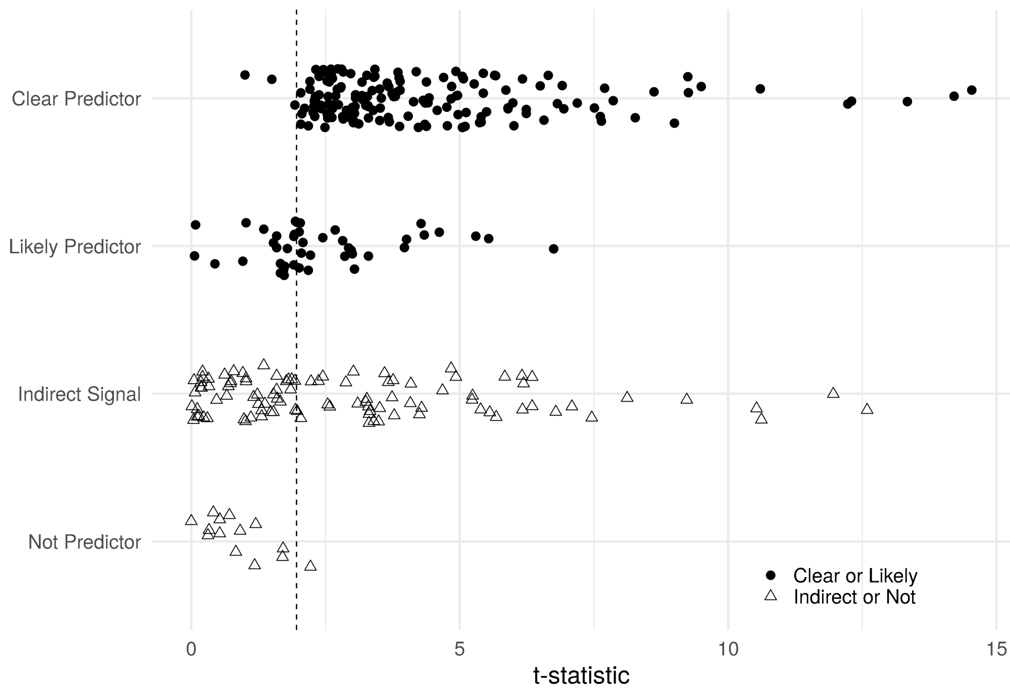
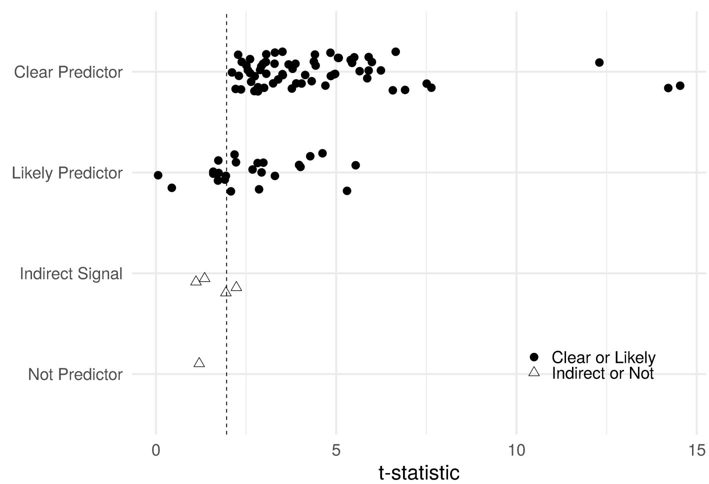


Figure 6: Replication of Other Meta-Study Replication Rates. Description: We examine subsets of our characteristics that are in McLean and Pontiff (2016) (top panel) or Hou, Xue, and Zhang (2020) (bottom panel). Each marker is one characteristic's long-short t-stat. Predictability categories are based on results in the original papers (see Sections 2.4 and 4.1). Vertical line is 1.96. Interpretation: We replicate MP's finding that roughly 12% of their predictors have small t-stats. HXZ's high "failure rate" seems to be driven by indirect signals, which were not tested for predictability in the original papers.

(a) McLean and Pontiff (2016)



(b) Hou, Xue, Zhang (2020)

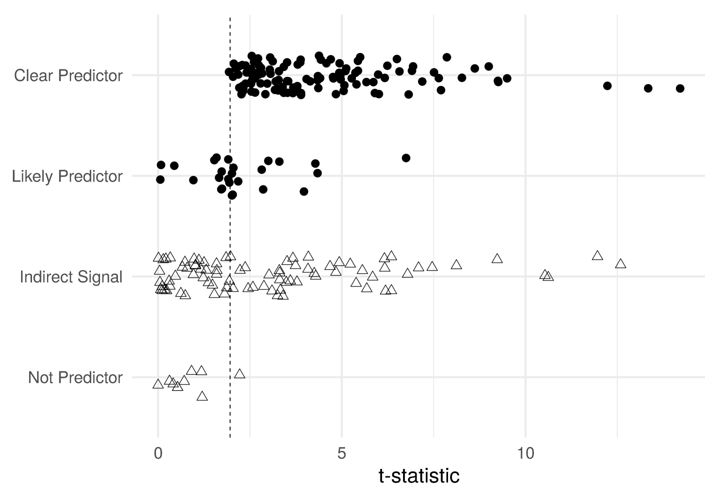


Figure 7: Out-of-Sample Replication of McLean and Pontiff's (2016) Out-of-Sample Tests (MP). Description: we compare the in-sample return (top panel) or in-sample t-stat (bottom panel) with the difference between in-sample mean returns and post-publication mean returns (ppt per month). OLS fit is shown using either predictors in MP (solid line) or predictors not in MP (dashed line). Fits in bottom panel are difficult to see because both fits are nearly identical. Shaded area is 2 S.E. Axis limits are identical to MP's Figure 1. Dotted line is the 45 degree line. Interpretation: MP's findings replicate, even out-of-sample. Returns decay post-publication (markers are right of 0) but remain positive (above 45 deg line in top panel), and the decay is higher for predictors that are stronger in-sample (upward slopes), consistent with investors learning about mispricing from academic studies.

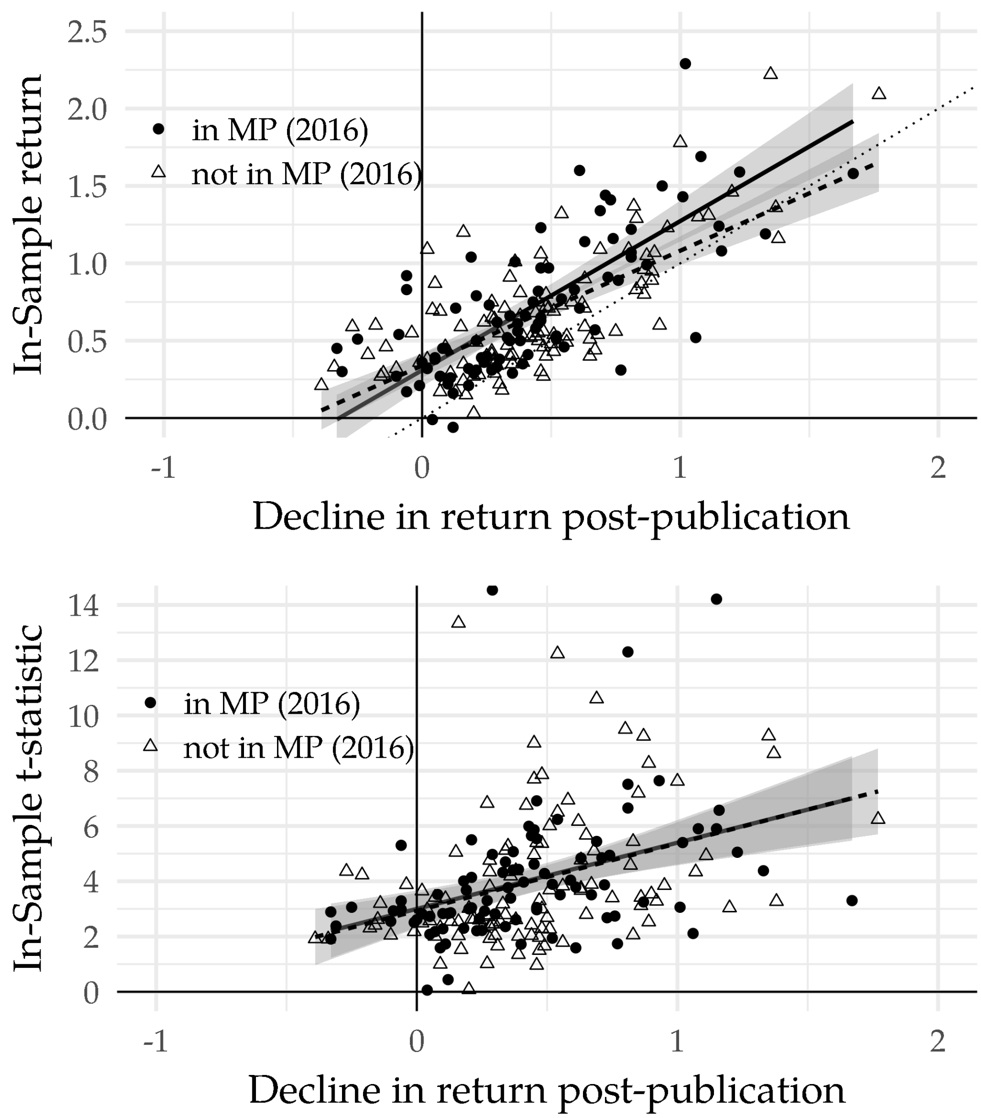
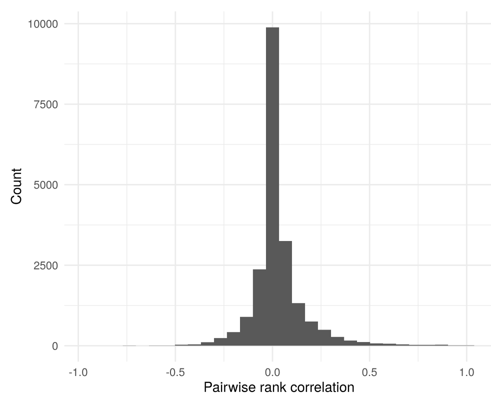
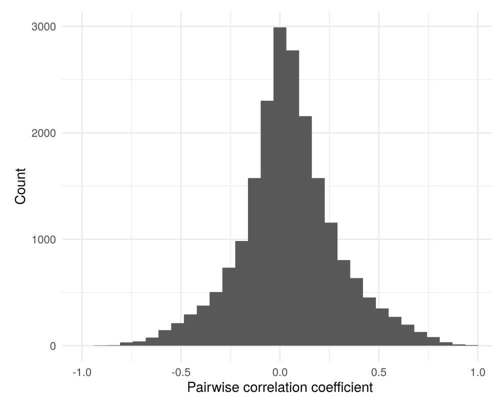


Figure 8: Correlations Between Pairs of Predictive Characteristics or Pairs of Portfolio Returns. Description: We show the distribution of correlations for (a) pairs of characteristics, (b) pairs of long-short portfolio returns, (c) long-short returns paired with HML, and (d) long-short returns paired with a Fama and French (1993)-style momentum factor. HML and FF3-style momentum are downloaded from Ken French's website. Data is limited to clear and likely predictors. Characteristics are signed so that a higher value implies higher mean returns and portfolios are signed to have positive mean returns, both following the original papers. Characteristic are pooled across all firm-months available and returns are pooled across the longest overlapping sample. Interpretation: Our dataset contains many distinct predictors.

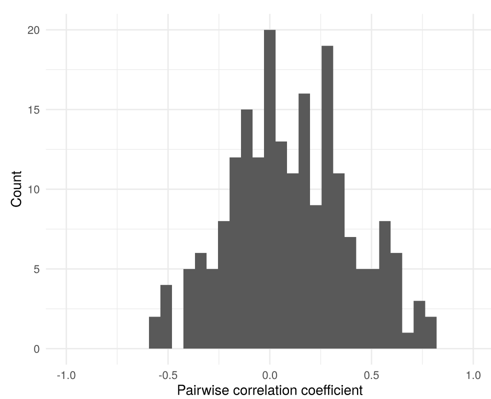
(a) Signed Predictive Characteristics



(b) Signed Predictor Portfolio Returns



(c) Signed Portfolios and HML



(d) Signed Portfolios and FF3 Momentum

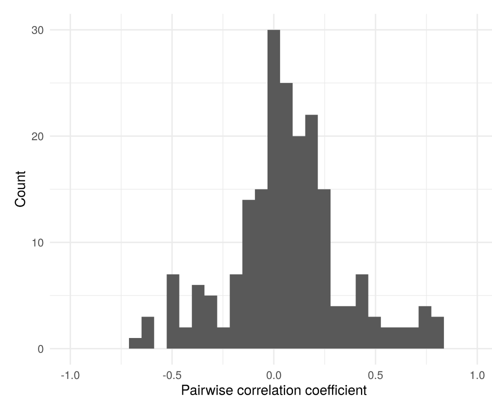


Figure 9: Performance by Liquidity Adjustment. Description: We use predictive characteristics to construct long-short portfolios using various liquidity adjustments. "Original papers" uses adjustments from the original papers (if any); "price > 5," "NYSE only," and "ME > NYSE 20 pct" only take positions in stocks if the share price exceeds \$5, stock is listed on the NYSE, or if market equity exceeds to 20th percentile among NYSE stocks in the month. "VW force" forces value-weighting. Each dot is one portfolio. Middle line is median, boxes are 25 and 75 percentiles, and the whiskers extend to the smallest (largest) value within the 25th (75th) percentile minus (plus) 1.5 times the interquartile range. Interpretation: Our code can impose a variety of liquidity adjustments and produces the intuitive result that liquid stocks are less predictable. Most simple liquidity adjustments reduce mean returns by a factor of about 1/3.

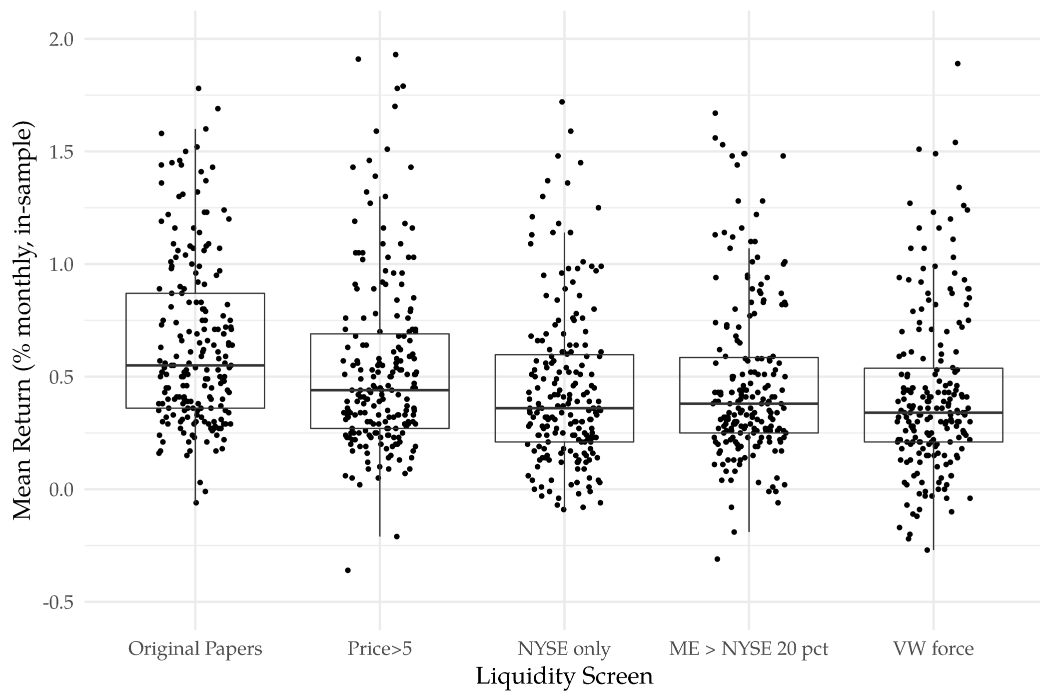


Figure 10: Performance by Rebalancing Frequency. Description: We use predictive characteristics to construct long-short portfolios that take on new signal data every 1-, 3-, 6-, or 12-months and measure mean returns in the original sample periods. "Original papers" uses the rebalancing frequency in the original papers. Middle line is median, boxes are 25 and 75 percentiles, and the whiskers extend to the smallest (largest) value within the 25th (75th) percentile minus (plus) 1.5 times the interquartile range. Interpretation: Our code is flexible in rebalancing frequencies and produces the intuitive result that less frequent rebalancing leads to lower mean returns. Enforcing 12-month holding periods reduces mean returns by roughly 20% compared to the original specifications.

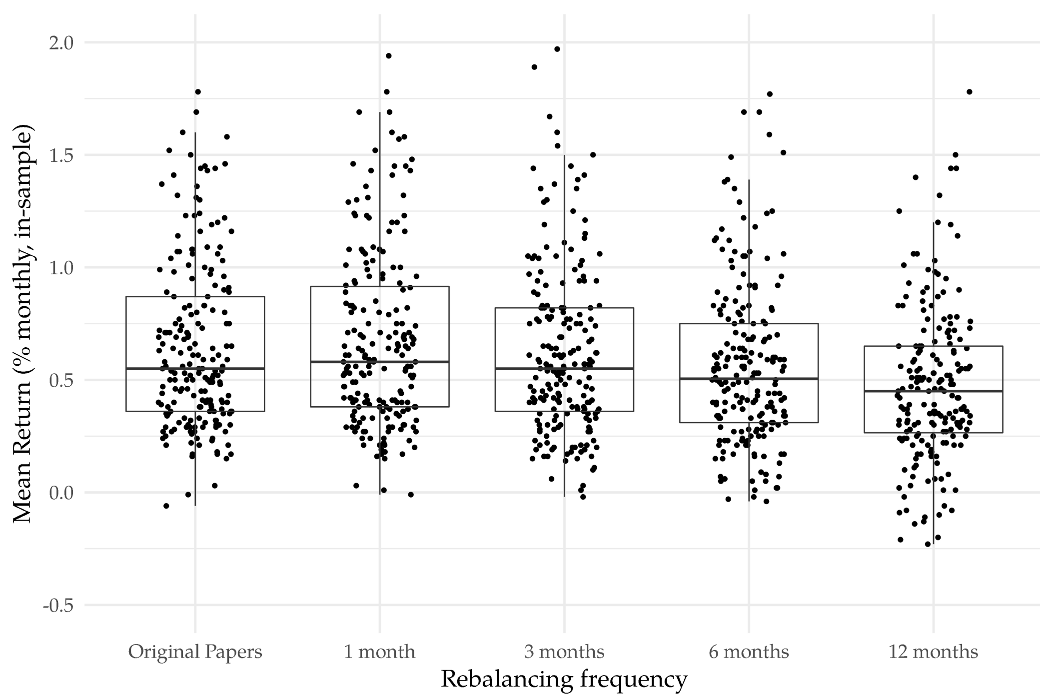


Table 1: An Open Source Dataset for Asset Pricing

Description: This table compares our dataset to McLean and Pontiff (2016) (MP); Green, Hand, and Zhang (2017) (GHZ); Harvey, Liu, and Zhu (2016) (HLZ); and Hou, Xue, and Zhang (2020) (HXZ). Predictor categories use results from the original papers to determine whether we should expect to find statistical significance in our long-short portfolio tests. Clear predictors provide clear evidence (e.g. long-short t-stat > 4), likely predictors have borderline evidence (e.g. multi-variate regression t-stat ≈ 2.5), and not-predictors have t-stats < 1.96. Indirect signals had suggestive evidence of predictive power (e.g. used as an ingredient in a larger model, correlated with B/M). Widely-available data includes CRSP-Compustat, IBES, OptionMetrics, 13F, among others. A detailed list of characteristic definitions is in the Online Appendix. Code and data are available at https://github.com/OpenSourceAP/CrossSection. Interpretation: Our dataset offers comprehensive coverage of firm-level predictors.

Panel A: Variable Counts

	Our Data		Other Metastudies			
	Predictor	Extended	MP	GHZ	HLZ	HXZ
Firm-Level Characteristics from Widely-Available Data						
Clear Predictor	161	161	68	66	229	118
Likely Predictor	44	44	24	13	7	27
Not-Predictor		14	1	3	1	10
Indirect Signals		100	4	20	46	85
Total	205	319	97	102	283	240
Additional Portfolios Made from Alternative Rebalancing Frequencies						
		957				212
Other Variables						
Theory					22	
Not Firm-Level					91	
Non-Widely Available Data					38	
Total	205	1276	97	102	283	452

Panel B: Our Coverage of Other Metastudies (%)

	MP	GHZ	HLZ	HXZ
Firm-Level Characteristics from Widely-Available Data				
Clear Predictor	97	88	90	100
Likely Predictor	100	100	43	100
not-predictor	100	100	100	100
Indirect Signals	100	100	91	100

Table 2: Performance of Individual Clear Predictors. Description: This table lists clear predictors (defined in Section 2.4) in the baseline data along with their original in-sample periods; the in-sample mean return (% monthly) and t-stat in our reproduced portfolio; and the predictability evidence in the original paper. "port sort" is portfolio sort, "LS" is long-short portfolio, "mv reg" is multivariate regression. Interpretation: Reproduced t-stats are close to the original results and support the credibility of the literature as well as the quality of our code and data. The table also provides a quick-reference guide to our code and data.

Original Study	Predictor	Acronym	Sample	Reproduction		Original Study's Predictability Evidence
				Mean Ret	t-Stat	
Abarbanell and Bushee (1998)	Change in capital inv (ind adj)	ChInvIA	1974-1988	0.50	5.50	t=2.9 in mv reg
Ali, Hwang, and Trombley (2003)	Idiosyncratic risk (AHT)	IdioVolAHT	1976-1997	0.89	2.53	t = 2.7 in mv reg
Alwathainani (2009)	Earnings consistency	EarningsConsistency	1971-2002	0.21	2.51	t=2.7 in complicated LS port
Amihud (2002)	Amihud's illiquidity	Illiquidity	1964-1997	0.57	3.51	t=6.6 in univariate reg
Anderson and Garcia-Feijoo (2006)	Change in capex (two years)	grcapx	1976-1999	0.50	4.96	t=5 in port sort
Anderson and Garcia-Feijoo (2006)	Change in capex (three years)	grcapx3y	1976-1999	0.59	4.77	t=4.7 in port sort
Ang et al. (2006)	Systematic volatility	betaVIX	1986-2000	1.07	3.54	t=3.9 in port sort
Ang et al. (2006)	Idiosyncratic risk	IdioRisk	1963-2000	0.99	3.25	t=2.9 in port sort
Ang et al. (2006)	Idiosyncratic risk (3 factor)	IdioVol3F	1963-2000	0.96	3.12	t=3.1 in port sort
Ang, Chen and Xing (2006)	Coskewness using daily returns	CoskewACX	1963-2001	0.29	2.62	t=2.8 in port sort
Avramov et al (2007)	Junk Stock Momentum	Mom6mJunk	1985-2003	1.58	3.30	t=4.3 in port sort
Baik and Ahn (2007)	Change in order backlog	OrderBacklogChg	1971-1999	0.38	2.49	p<0.01 in port sort
Balakrishnan, Bartov and Faurel (2010)	Return on assets (qtrly)	roaq	1976-2005	1.69	5.90	t=6.5 in port sort, nontraditional
Bali, Cakici, and Whitelaw (2010)	Maximum return over month	MaxRet	1962-2005	0.89	2.74	t=2.8 in port sort
Bali, Engle and Murray (2015)	Return skewness	ReturnSkew	1963-2012	0.41	5.27	t=4 in port sort
Bali, Engle and Murray (2015)	Idiosyncratic skewness (3F model)	ReturnSkew3F	1963-2012	0.29	4.76	t=4.4 in port sort
Ball et al. (2016)	Cash-based operating profitability	CBOperProf	1963-2014	0.46	3.20	t=3.2 in port sort
Banz (1981)	Size	Size	1926-1975	0.50	2.61	t=3.1 in long-short
Barth and Hutton (2004)	Change in Forecast and Accrual	ChForecastAccrual	1981-1996	1.22	12.30	p-val < 0.001 in port sort
Bartov and Kim (2004)	Book-to-market and accruals	AccrualsBM	1980-1998	1.44	4.85	t=5.5 in long-short
Basu (1977)	Earnings-to-Price Ratio	EP	1957-1971	0.39	2.21	monotonic port sort but no LS
Bazdresch, Belo and Lin (2014)	Employment growth	hire	1965-2010	0.51	5.67	t=5.8 in port sort
Belo and Lin (2012)	Inventory Growth	InvGrowth	1965-2009	0.87	7.19	t=6.6 in port sort
Bhandari (1988)	Market leverage	Leverage	1952-1981	0.36	2.64	t=3.9 in regression
Blitz, Huij and Martens (2011)	Momentum based on FF3 residuals	ResidualMomentum	1930-2009	0.95	8.27	t=8 in long-short ff3+ alpha
Blume and Husic (1972)	Price	Price	1932-1971	1.43	3.06	t=3 in regressions
Boudoukh et al. (2007)	Net Payout Yield	NetPayoutYield	1984-2003	0.87	2.57	t=2.8 in conservative LS, strong port sort
Boudoukh et al. (2007)	Payout Yield	PayoutYield	1984-2003	0.43	2.27	t=3.9 in conservative LS, strong port sort
Bradshaw, Richardson, Sloan (2006)	Net debt financing	NetDebtFinance	1971-2000	0.75	7.70	t=6.9 in port sort

continued on next page

Table 2: (continued)

Original Study	Predictor	Acronym	Sample	Reproduction Mean Ret	Reproduction t-Stat	Original Study's Predictability Evidence
Bradshaw, Richardson, Sloan (2006)	Net equity financing	NetEquityFinance	1971-2000	1.06	5.40	t=3.8 in port sort
Bradshaw, Richardson, Sloan (2006)	Net external financing	XFIN	1971-2000	1.14	4.84	t=5.7 in port sort
Brennan, Chordia, Subra (1998)	Past trading volume	DolVol	1966-1995	0.75	2.72	t=2.9 in regression
Cen, Wei, and Zhang (2006)	Analyst earnings per share	FEPS	1983-2002	1.46	3.04	t=2.7 in port sort
Chan and Ko (2006)	Momentum and LT Reversal	MomRev	1965-2001	1.19	4.38	t=4.3 in long-short
Chan, Jegadeesh and Lakonishok (1996)	Earnings announcement return	AnnouncementReturn	1977-1992	1.20	13.34	t=9.3 in regression
Chan, Jegadeesh and Lakonishok (1996)	Earnings forecast revisions	REV6	1977-1992	1.29	5.44	t=4.1 in regression
Chan, Lakonishok and Sougiannis (2001)	R&D over market cap	RD	1975-1995	1.23	6.91	strong port sort
Chandrashekar and Rao (2009)	Cash Productivity	CashProd	1963-2003	0.56	3.40	t=3.6 in regression
Chen, Hong and Stein (2002)	Breadth of ownership	DelBreadth	1979-1998	0.69	3.69	t=4.0 in port sort
Chordia, Subra, Anshuman (2001)	Share turnover volatility	std_turn	1966-1995	0.80	3.42	t=3.7 in regression
Chordia, Subra, Anshuman (2001)	Volume Variance	VolSD	1966-1995	0.38	2.82	t=3.6 in regression
Cohen and Frazzini (2008)	Customer momentum	CustomerMomentum	1980-2004	1.16	3.27	t=3.8 in port sort
Cohen and Lou (2012)	Conglomerate return	retConglomerate	1977-2009	1.32	6.50	t=5.5 in port sort
Cohen, Diether and Malloy (2013)	R&D ability	RDAbility	1980-2009	0.27	1.50	t=2.6 in double sort
Cooper, Gulen and Schill (2008)	Asset growth	AssetGrowth	1968-2003	1.50	7.64	t=8.5 in port sort
Cremers and Nair (2005)	Takeover vulnerability	Activism1	1990-2001	0.24	1.00	t=3.1 in port sort
Da and Warachka (2011)	Long-vs-short EPS forecasts	EarningsForecastDisparity	1983-2006	0.66	4.38	t=5.1 in LS port
Daniel and Titman (2006)	Composite equity issuance	CompEquIss	1968-2003	0.27	2.41	t=4.4 in mv reg
Daniel and Titman (2006)	Intangible return using BM	IntanBM	1968-2003	0.40	2.29	t=4.0 in mv reg
Daniel and Titman (2006)	Intangible return using CFtoP	IntanCFP	1968-2003	0.40	2.32	t=4.9 in mv reg
Daniel and Titman (2006)	Intangible return using EP	IntanEP	1968-2003	0.34	2.46	t=4.6 in mv reg
Daniel and Titman (2006)	Intangible return using Sale2P	IntanSP	1968-2003	0.53	2.42	t=4.3 in mv reg
Daniel and Titman (2006)	Share issuance (5 year)	ShareIss5Y	1968-2003	0.52	4.32	t=4.4 in univar reg
Datar, Naik and Radcliffe (1998)	Share Volume	ShareVol	1962-1991	0.91	3.87	t=8.9 in univariate reg
De Bondt and Thaler (1985)	Long-run reversal	LRreversal	1929-1982	0.79	3.04	t=3.3 in long-short
Dechow, Sloan and Soliman (2004)	Equity Duration	EquityDuration	1962-1998	0.60	3.28	t=4.4 in conservative long-short
Desai, Rajgopal, Venkatachalam (2004)	Operating Cash flows to price	cfp	1973-1997	0.36	2.18	t=2.77 in port sort
Dharan and Ikenberry (1995)	Exchange Switch	ExchSwitch	1962-1990	0.45	2.89	t = 3.6 in event study
Dichev (1998)	O Score	OScore	1981-1995	1.01	3.39	t=3.36 in LS port
Dichev and Piotroski (2001)	Credit Rating Downgrade	CredRatDG	1986-1998	0.73	2.92	t=11 in event study w/ special data
Diether, Malloy and Scherbina (2002)	EPS Forecast Dispersion	ForecastDispersion	1976-2000	0.65	3.05	t=2.9 in port sort
Doyle, Lundholm and Soliman (2003)	Excluded Expenses	ExclExp	1988-1999	0.27	3.02	t=5.7 in mv reg
Eberhart, Maxwell and Siddique (2004)	Unexpected R&D increase	SurpriseRD	1974-2001	0.29	3.00	t=3.5 in long-short
Eisfeldt and Papanikolaou (2013)	Organizational capital	OrgCap	1970-2008	0.36	2.61	t=2.9 in port sort

continued on next page

Table 2: (continued)

Original Study	Predictor	Acronym	Sample	Reproduction Mean Ret	Reproduction t-Stat	Original Study's Predictability Evidence
Elgers, Lo and Pfeiffer (2001)	Earnings Forecast to price	sfe	1982-1998	0.81	2.61	t=5 in long-short size adjusted
Fama and French (1992)	Total assets to market	AM	1963-1990	0.69	3.42	t=5.7 in univar reg
Fama and French (1992)	Book to market using December ME	BMdec	1963-1990	0.98	5.37	t=5.71 in univariate reg
Fama and French (1992)	Book leverage (annual)	BookLeverage	1963-1990	0.29	3.37	t=5.3 in mv reg
Foster, Olsen and Shevlin (1984)	Earnings Surprise	EarningsSurprise	1974-1981	1.16	4.94	huge spread in event study
Gompers, Ishii and Metrick (2003)	Governance Index	Governance	1990-1999	0.52	2.11	t=2.7 in long short FF3 alpha
Gou, Lev and Shi (2006)	IPO and no R&D spending	RDIPO	1980-1995	0.97	2.97	t=2.68 in port sort FF3+Mom alpha
Grinblatt and Moskowitz (1999)	Industry Momentum	IndMom	1963-1995	0.27	2.55	t=4.6 in long-short
Hafzalla, Lundholm, Van Winkle (2011)	Percent Operating Accruals	PctAcc	1989-2008	0.46	3.51	t>2.6 in size-adjusted long-short
Hafzalla, Lundholm, Van Winkle (2011)	Percent Total Accruals	PctTotAcc	1989-2008	0.50	4.70	t>2.6 in size-adjusted long-short
Hahn and Lee (2009)	Tangibility	tang	1973-2001	0.71	3.67	t=3.37 in univariate FMB
Hartzmark and Salomon (2013)	Dividend seasonality	DivSeason	1927-2011	0.33	14.54	t=16 in long-short
Hawkins, Chamberlin, Daniel (1984)	EPS forecast revision	AnalystRevision	1975-1980	0.91	5.12	t=3.2 in long only CAPM alpha
Heston and Sadka (2008)	Momentum without the seasonal part	Mom12mOffSeason	1965-2002	1.23	3.85	t=4 in port sort
Heston and Sadka (2008)	Off season long-term reversal	MomOffSeason	1965-2002	1.31	4.93	t=5.6 in port sort
Heston and Sadka (2008)	Off season reversal years 6 to 10	MomOffSeason06YrPlus	1965-2002	0.59	4.36	t=4.6 in port sort
Heston and Sadka (2008)	Off season reversal years 16 to 20	MomOffSeason16YrPlus	1965-2002	0.35	2.54	t=3.4 in port sort
Heston and Sadka (2008)	Return seasonality years 2 to 5	MomSeason	1965-2002	0.82	5.86	t=5 in port sort
Heston and Sadka (2008)	Return seasonality years 6 to 10	MomSeason06YrPlus	1965-2002	0.74	6.17	t=6.1 in port sort
Heston and Sadka (2008)	Return seasonality years 11 to 15	MomSeason11YrPlus	1965-2002	0.75	6.94	t=6.4 in port sort
Heston and Sadka (2008)	Return seasonality years 16 to 20	MomSeason16YrPlus	1965-2002	0.59	5.05	t=4.5 in port sort
Heston and Sadka (2008)	Return seasonality last year	MomSeasonShort	1965-2002	1.36	8.62	t=7.6 in port sort
Hirschleifer, Hsu and Li (2013)	Citations to RD expenses	CitationsRD	1982-2008	0.21	2.21	t=2.6 in FF3 style long-short
Hirschleifer, Hsu and Li (2013)	Patents to RD expenses	PatentsRD	1982-2008	0.30	3.18	t=4.1 in FF3 style long-short
Hirshleifer et al. (2004)	Net Operating Assets	NOA	1964-2002	1.07	7.51	t=8.5 in long-short
Hirshleifer, Hou, Teoh, Zhang (2004)	change in net operating assets	dNoa	1964-2002	1.05	9.25	t=8.9 in mv reg
Hou (2007)	Earnings surprise of big firms	EarnSupBig	1972-2001	0.37	2.38	t=9 in mv reg weekly
Hou (2007)	Industry return of big firms	IndRetBig	1972-2001	2.22	9.26	t=11 in mv reg
Hou and Moskowitz (2005)	Price delay r square	PriceDelayRsq	1964-2001	0.48	2.69	t =3.4 in port sort char adj
Hou and Robinson (2006)	Industry concentration (sales)	Herf	1963-2001	0.21	2.30	t = 2.14 in port sort
Hou and Robinson (2006)	Industry concentration (equity)	HerfBE	1963-2001	0.22	2.04	t = 2.52 in characteristics-adjusted port sort
Jegadeesh (1989)	Short term reversal	STreversal	1934-1987	2.97	14.21	t=12 in port sort
Jegadeesh and Livnat (2006)	Revenue Surprise	RevenueSurprise	1987-2003	0.75	5.99	t>2.6 in many event studies
Jegadeesh and Titman (1993)	Momentum (12 month)	Mom12m	1964-1989	1.37	4.58	t=3.7 long-short
Jegadeesh and Titman (1993)	Momentum (6 month)	Mom6m	1964-1989	1.04	3.68	t=2.4 long-short

continued on next page

Table 2: (continued)

Original Study	Predictor	Acronym	Sample	Reproduction Mean Ret	t-Stat	Original Study's Predictability Evidence
Jegadeesh et al. (2004)	Change in recommendation	ChangeInRecommendation	1985-1998	1.04	6.65	p<0.01 in LS port, but we lack the data
Johnson and So (2012)	Option to stock volume	OptionVolume1	1996-2010	0.68	2.04	t = 3.45 in port sort CAPM alpha weekly
Kelly and Jiang (2014)	Tail risk beta	BetaTailRisk	1963-2010	0.46	3.30	Tab4A t-stat 2.48
La Porta (1996)	Long-term EPS forecast	fgr5yrLag	1983-1990	0.83	2.06	t=4.9 in regression
Lakonishok, Shleifer, Vishny (1994)	Cash flow to market	CF	1968-1990	0.83	4.04	t=3.4 in port sort
Lakonishok, Shleifer, Vishny (1994)	Revenue Growth Rank	MeanRankRevGrowth	1968-1990	0.53	3.89	t=4.5 in double sort
Landsman et al. (2011)	Real dirty surplus	RDS	1976-2003	0.49	3.83	t=5.8 in port sort
Lee and Swaminathan (2000)	Momentum in high volume stocks	MomVol	1965-1995	1.59	5.05	t=6 in long-short, lots of robustness
Lev and Nissim (2004)	Taxable income to income	Tax	1973-2000	0.45	3.52	t=3.9 in regression
Li (2011)	R&D capital-to-assets	RDcap	1980-2007	0.46	2.32	t=2.6 in long-short
Litzenberger and Ramaswamy (1979)	Predicted div yield next month	DivYieldST	1936-1977	0.41	4.23	t=6 in mv reg
Liu (2006)	Days with zero trades	zerotrade	1960-2003	0.49	2.63	t=4.1 in port sort
Liu (2006)	Days with zero trades	zerotradeAlt1	1960-2003	0.64	3.80	t = 3.46 in port sort (12m holding)
Liu (2006)	Days with zero trades	zerotradeAlt12	1960-2003	0.40	2.79	t > 4 in port sort (diff holding periods)
Lockwood and Prombutr (2010)	Growth in book equity	ChEQ	1964-2007	0.61	4.41	t=5.38 in EW port sort
Loh and Warachka (2012)	Earnings surprise streak	EarningsStreak	1987-2009	1.09	10.60	t=9.5 in port sort ff3 alpha
Lou (2014)	Growth in advertising expenses	GrAdExp	1974-2010	0.44	3.89	t=3.5 in long-short
Loughran and Wellman (2011)	Enterprise Multiple	EntMult	1963-2009	1.08	6.57	t=6.54 in decile sort CAPM alpha
Lyandres, Sun and Zhang (2008)	Composite debt issuance	CompositeDebtIssuance	1970-2005	0.41	6.82	t=8.59 in port sort CAPM alpha
Lyandres, Sun and Zhang (2008)	change in ppe and inv/assets	InvestPPEInv	1970-2005	0.80	7.86	t=7 in long-short port
Menzly and Ozbas (2010)	Customers momentum	iomom_cust	1986-2005	0.71	2.53	t=2.6 in industry port sort
Menzly and Ozbas (2010)	Suppliers momentum	iomom_supp	1986-2005	0.60	2.31	t=3.4 in industry port sort
Michaely, Thaler and Womack (1995)	Dividend Initiation	DivInit	1964-1988	0.58	5.65	t=3.4 in event study
Michaely, Thaler and Womack (1995)	Dividend Omission	DivOmit	1964-1988	0.51	3.06	t=6 in event study
Mohanram (2005)	Mohanram G-score	MS	1978-2001	1.34	5.44	t=9 in port sort nonstandard data lag
Nagel (2005)	Inst Own and Forecast Dispersion	RIO_Disp	1980-2003	0.62	2.60	t = 2.47 in conditional sort
Nagel (2005)	Inst Own and Market to Book	RIO_MB	1980-2003	0.90	3.74	t = 4.91 in conditional sort
Nagel (2005)	Inst Own and Turnover	RIO_Turnover	1980-2003	0.65	2.78	t = 2.71 in conditional sort
Nagel (2005)	Inst Own and Idio Vol	RIO_Volatility	1980-2003	1.01	4.19	t = 4.38 in conditional sort
Nguyen and Swanson (2009)	Efficient frontier index	Frontier	1980-2003	2.09	6.24	t=5 in port sort
Novy-Marx (2010)	Operating leverage	OPLeverage	1963-2008	0.38	2.73	t=3.38 in port sort
Novy-Marx (2012)	Intermediate Momentum	IntMom	1927-2010	1.24	5.90	Tab2 t-stat 5.79
Novy-Marx (2013)	gross profits / total assets	GP	1963-2010	0.30	2.38	t=2.5 in VW LS quint
Palazzo (2012)	Cash to assets	Cash	1972-2009	0.70	2.97	t=2.14 in port sort but strong with adjustments
Pastor and Stambaugh (2003)	Pastor-Stambaugh liquidity beta	BetaLiquidityPS	1968-1999	0.35	1.93	t=2.54 in VW port sort CAPM alpha

continued on next page

Table 2: (continued)

Original Study	Predictor	Acronym	Sample	Reproduction Mean Ret	Reproduction t-Stat	Original Study's Predictability Evidence
Penman, Richardson and Tuna (2007)	Leverage component of BM	BPEBM	1963-2001	0.22	2.83	t=4.1 in univariate reg
Penman, Richardson and Tuna (2007)	Enterprise component of BM	EBM	1963-2001	0.31	4.14	t=3.0 in double sort
Penman, Richardson and Tuna (2007)	Net debt to price	NetDebtPrice	1963-2001	0.55	3.88	t=2.3 in double sort
Piotroski (2000)	Piotroski F-score	PS	1976-1996	0.92	3.29	t=5.59 in port sort nonstandard data lag
Pontiff and Woodgate (2008)	Share issuance (1 year)	ShareIss1Y	1970-2003	0.62	4.97	t=7.08 in univariate reg
Rajgopal, Shevlin, Venkatachalam (2003)	Order backlog	OrderBacklog	1981-1999	0.40	2.80	t=2.38 in univariate size-adjusted FMB
Richardson et al. (2005)	Change in current operating assets	DelCOA	1962-2001	0.54	6.01	t=9 in mv reg
Richardson et al. (2005)	Change in current operating liabilities	DelCOL	1962-2001	0.35	4.35	t=4.5 in mv reg
Richardson et al. (2005)	Change in equity to assets	DelEqu	1963-2001	0.47	3.18	t=6.3 in mv reg
Richardson et al. (2005)	Change in financial liabilities	DelFINL	1962-2001	0.73	12.23	t=8 in univariate reg
Richardson et al. (2005)	Change in long-term investment	DelLTI	1962-2001	0.17	2.55	t=3.4 in mv reg
Richardson et al. (2005)	Change in net financial assets	DelNetFin	1962-2001	0.55	9.00	t=6 in unvivariate reg
Richardson et al. (2005)	Total accruals	TotalAccruals	1962-2001	0.28	2.63	t=6 in mv reg
Ritter (1991)	Initial Public Offerings	IndIPO	1975-1987	0.66	2.36	t=4 in event study
Rosenberg, Reid, and Lanstein (1985)	Book to market using most recent ME	BM	1973-1984	1.60	3.79	t=6 in nonstandard long-short
Scherbina (2008)	Decline in Analyst Coverage	ChNAnalyst	1982-2005	1.09	3.65	t > 3 in port sort FF3 alpha for small stocks
Sloan (1996)	Accruals	Accruals	1962-1991	0.56	5.07	t > 4 in port sort CAPM alpha 12 month holding
Soliman (2008)	Change in Asset Turnover	ChAssetTurnover	1984-2002	0.29	3.77	t=5 in mv reg
Soliman (2008)	Change in Net Noncurrent Op Assets	ChNNCOA	1984-2002	0.35	4.43	t=4.3 in mv reg
Soliman (2008)	Change in Net Working Capital	ChNWC	1984-2002	0.16	2.83	t=4.6 in mv reg
Thomas and Zhang (2002)	Inventory Growth	ChInv	1970-1997	0.77	6.24	t>2.6 in port sort
Thomas and Zhang (2011)	Change in Taxes	ChTax	1977-2006	1.09	9.50	t = 11.26 in decile sort
Titman, Wei and Xie (2004)	Investment to revenue	Investment	1973-1996	0.25	2.28	t=2.86 in VW port sort
Valta (2016)	Convertible debt indicator	ConvDebt	1985-2012	0.38	4.34	t > 2.6 in mv reg
Xie (2001)	Abnormal Accruals	AbnormalAccruals	1971-1992	0.54	5.10	t=8 port sort w/ nonstandard data lag
Yan (2011)	Put volatility minus call volatility	SmileSlope	1996-2005	1.78	7.62	t=8 in port sort
Zhang (2004)	Firm Age - Momentum	FirmAgeMom	1983-2001	2.29	5.40	t = 7.21 in long portfolio

Table 3: Performance of Individual Likely Predictors. Description: This table lists likely predictors (defined in Section 2.4) in the baseline data along with their original in-sample periods; the in-sample mean return (% monthly) and t-stat in our reproduced portfolio; and the predictability evidence in the original paper. “port sort” is portfolio sort, “LS” is long-short portfolio, “mv reg” is multivariate regression. Interpretation: Our categorization of predictors as “likely” can be justified by the original studies. As in Table 2, reproduced t-stats are close to the original results and support the credibility of the literature as well as the quality of our code and data. The table also provides a quick-reference guide to our code and data.

Related Original Study	Predictor	Acronym	Sample	Reproduction Mean Ret	Reproduction t-Stat	Original Study’s Predictability Evidence
Abarbanell and Bushee (1998)	Sales growth over inventory growth	GrSaleToGrInv	1974-1988	0.31	3.30	t=2.4 in mv reg
Abarbanell and Bushee (1998)	Sales growth over overhead growth	GrSaleToGrOverhead	1974-1988	-0.06	-0.44	t=2.1 in mv reg
Amihud and Mendelsohn (1986)	Bid-ask spread	BidAskSpread	1961-1980	0.71	1.59	strong port sorts but no LS special data
Asquith Pathak and Ritter (2005)	Inst own among high short interest	IO_ShortInterest	1980-2002	2.22	3.04	strong port sort but no long-short
Ball et al. (2016)	Operating profitability R&D adjusted	OperProfRD	1963-2014	0.33	1.91	t=1.8 in port sort
Barbee, Mukherji and Raines (1996)	Sales-to-price	SP	1979-1991	0.71	2.86	t=2.5 in mv reg
Barber et al. (2002)	Consensus Recommendation	ConsRecomm	1985-1997	0.53	1.35	t=3.2 in port sort nonstandard data
Barber et al. (2002)	Down forecast EPS	DownRecomm	1985-1997	0.63	5.54	t>8 in 3-day event study
Barber et al. (2002)	Up Forecast	UpRecomm	1985-1997	0.61	4.62	t>8 in 3-day event study
Barry and Brown (1984)	Firm age based on CRSP	FirmAge	1931-1980	-0.01	-0.06	t=2.5 in reg nonstandard data
Belo, Lin and Vitorino (2014)	Brand capital investment	BrandInvest	1975-2010	0.56	1.97	t=2.0 in port sort
Chan, Lakonishok and Sougiannis (2001)	Advertising Expense	AdExp	1975-1996	0.97	4.28	53 bps spread but no t-stat
Cremers and Nair (2005)	Active shareholders	Activism2	1990-2001	0.43	1.02	t=2.0 in port sort
Cusatis, Miles and Woolridge (1993)	Spinoffs	Spinoff	1965-1988	0.40	2.22	t=2.3 in event study
De Bondt and Thaler (1985)	Medium-run reversal	MRreversal	1933-1980	0.39	2.08	large ret in similar long-short
Dechow et al. (2001)	Short Interest	ShortInterest	1976-1993	0.83	5.30	35 bps spread in port sort
Easley, Hvidkjaer and O’Hara (2002)	Probability of Informed Trading	ProbInformedTrading	1984-1998	1.30	4.34	t=2.5 in mv reg
Fairfield, Whisenant and Yohn (2003)	Growth in long term operating assets	GrLTNOA	1964-1993	0.41	3.97	61 bps spread in long-short
Fama and French (2006)	operating profits / book equity	OperProf	1977-2003	0.72	3.00	t=2.6 in mv reg
Fama and MacBeth (1973)	CAPM beta	Beta	1929-1968	0.66	1.72	t=2.6 univar reg
Frankel and Lee (1998)	Analyst Value	AnalystValue	1975-1993	0.26	1.73	p<0.01 in port sort but nonstandard stats
Frankel and Lee (1998)	Analyst Optimism	AOP	1975-1993	0.36	2.01	p<0.01 in port sort but nonstandard stats
Frankel and Lee (1998)	Predicted Analyst forecast error	PredictedFE	1979-1993	0.30	0.96	p<0.01 in reg but nonstandard stats
Franzoni and Marin (2006)	Pension Funding Status	FR	1980-2002	0.31	1.74	49 bps long-short
Frazzini and Pedersen (2014)	Frazzini-Pedersen Beta	BetaFP	1929-2012	0.03	0.08	t=7 in nonstandard port sort
George and Hwang (2004)	52 week high	High52	1963-2001	0.51	1.94	t=2.0 in long-short
Harvey and Siddique (2000)	Coskewness	Coskewness	1964-1993	0.27	2.18	p-val<0.05 in long-short

continued on next page

Table 3: (continued)

Original Study	Predictor	Acronym	Sample	Reproduction Mean Ret	Reproduction t-Stat	Original Study's Predictability Evidence
Haugen and Baker (1996)	net income / book equity	RoE	1979-1993	0.32	2.82	t=4.5 in mv reg nonstandard
Haugen and Baker (1996)	Cash-flow to price variance	VarCF	1979-1993	-0.56	-1.91	t=2.5 in mv reg nonstandard
Haugen and Baker (1996)	Volume to market equity	VolMkt	1979-1993	0.45	1.59	t=4 in mv reg nonstandard
Haugen and Baker (1996)	Volume Trend	VolumeTrend	1979-1993	0.54	2.93	t=3 in mv reg nonstandard
Heston and Sadka (2008)	Off season reversal years 11 to 15	MomOffSeason11YrPlus	1965-2002	0.24	2.03	t=1.8 in port sort, but similar strats do better
Hong and Kacperczyk (2009)	Sin Stock (selection criteria)	sinAlgo	1926-2006	0.21	1.92	t-stat = 1.8 in LS nontraditional
Hou and Moskowitz (2005)	Price delay coeff	PriceDelaySlope	1964-2001	0.17	2.01	t =7.7 in port sort w/ complicated signal
Hou and Moskowitz (2005)	Price delay SE adjusted	PriceDelayTstat	1964-2001	0.15	1.53	t =7.39 in port sort w/ complicated signal
Hou and Robinson (2006)	Industry concentration (assets)	HerfAsset	1963-2001	0.18	1.66	t = 2.12 in characteristics-adjusted port sort
Ikenberry, Lakonishok, Vermaelen (1995)	Share repurchases	ShareRepurchase	1980-1990	0.32	4.01	t=1.85 in long - benchmark port
Johnson and So (2012)	Option volume to average	OptionVolume2	1996-2010	0.53	1.79	t = 2.5 in port sort CAPM alpha weekly data
Loh and Warachka (2012)	Earnings streak length	NumEarnIncrease	1987-2009	0.52	6.75	similar results in port sorts but not exact
Prakash and Sinha (2012)	Deferred Revenue	DelDRC	2002-2007	0.71	1.66	t=3.6 in nonstandard reg 5 year sample
Ritter (1991)	IPO and age	AgeIPO	1981-1984	1.41	2.68	Event study, no t-stat
Spiess and Affleck-Graves (1999)	Debt Issuance	DebtIssuance	1975-1989	0.17	2.98	t = 2.19 FF3 alpha on long port
Tuzel (2010)	Real estate holdings	realestate	1971-2005	0.32	2.05	t=1.8 (VW) and t= 1.28 (EW) in port sort
Xing, Zhang and Zhao (2010)	Volatility smirk near the money	skew1	1996-2005	0.55	2.45	t = 2.19 in port sort

Table 4: Performance of Not-Predictors and Indirect Signals. This table lists not-predictors and indirect signals (defined in Section 2.4). The vast majority of these characteristics are included only to nest Hou, Xue, and Zhang (2020) (HXZ) (see Table 1). We list the original or related study's in-sample periods; the in-sample mean return (% monthly) and t-stat in our reproduced portfolio; and the predictability evidence in the original paper, if available. Unlike clear and likely predictors, we do not sign these portfolios or select portfolio implementations based on the original papers' results. All portfolios are equal-weighted and long-short quintiles, unless the characteristic is discrete. "HXZ variant" indicates our characteristic is based on HXZ's modification of a (not necessarily predictive) characteristic in a previous study. Interpretation: Our categorization of predictors as "not" or "indirect evidence" can be justified by evidence in the original or related studies. Many of HXZ's "replication failures" come from studies that were never shown to produce statistically significant predictability.

Original / Related Study	Predictor	Acronym	Predictor Category	Reproduction Mean Ret	Reproduction t-Stat	Original Study's Predictability Evidence (if Available)
Abarbanell and Bushee (1998)	Effective Tax Rate	ETR	Indirect Evidence	0.01	0.18	t=1.5 in mv reg
Abarbanell and Bushee (1998)	Gross margin growth to sales growth	GrGMToGrSales	Indirect Evidence	0.37	3.30	t=1.9 in mv reg
Abarbanell and Bushee (1998)	Change in sales vs change in receiv	GrSaleToGrReceivables	Indirect Evidence	0.06	0.66	t=1.6 in mv reg
Abarbanell and Bushee (1998)	Laborforce efficiency	LaborforceEfficiency	Indirect Evidence	-0.07	-0.74	t=0.6 in mv reg
Abarbanell and Bushee (1998)	Change in gross margin vs sales	pchgm_pchsale	Indirect Evidence	0.42	3.76	GHZ variant of GrGMToGrSale
Acharya and Pedersen (2005)	Illiquidity-illiquidity beta (beta2i)	betaCC	Indirect Evidence	0.33	1.87	in-sample only
Acharya and Pedersen (2005)	Illiquidity-market return beta (beta4i)	betaCR	Indirect Evidence	-0.09	-0.98	in-sample only
Acharya and Pedersen (2005)	Net liquidity beta (betanet,p)	betaNet	Indirect Evidence	0.35	1.97	in-sample only
Acharya and Pedersen (2005)	Return-market illiquidity beta	betaRC	Indirect Evidence	0.06	0.30	in-sample only
Acharya and Pedersen (2005)	Return-market return illiquidity beta	betaRR	Indirect Evidence	-0.03	-0.14	in-sample only
Adrian, Etula and Muir (2014)	Broker-Dealer Leverage Beta	BetaBDLeverage	Not Predictor	0.41	2.22	t=1 in conservative port sort
Anderson and Garcia-Feijoo (2006)	Investment growth (1 year)	grcapx1y	Indirect Evidence	-0.28	-3.74	HXZ variant
Anderson, Ghysels, and Juergens (2005)	Long-term forecast dispersion	ForecastDispersionLT	Not Predictor	-0.00	-0.00	t=1.0 in conservative long-short
Ang et al. (2006)	Idiosyncratic risk (CAPM)	IdioVolCAPM	Indirect Evidence	-0.31	-1.01	HXZ variant
Ang et al. (2006)	Idiosyncratic risk (q factor)	IdioVolQF	Indirect Evidence	-0.39	-1.16	HXZ variant
Ang, Chen and Xing (2006)	Downside beta	DownsideBeta	Not Predictor	0.07	0.31	t=0.6 in port sort
Balakrishnan, Bartov and Faurel (2010)	Change in Return on assets	ChangeRoA	Indirect Evidence	1.32	12.59	HXZ variant
Balakrishnan, Bartov and Faurel (2010)	Change in Return on equity	ChangeRoE	Indirect Evidence	1.08	10.53	HXZ variant
Bali, Engle and Murray (2015)	Idiosyncratic skewness (CAPM)	ReturnSkewCAPM	Indirect Evidence	-0.36	-5.39	HXZ variant
Bali, Engle and Murray (2015)	Idiosyncratic skewness (Q model)	ReturnSkewQF	Indirect Evidence	-0.26	-4.29	HXZ variant
Ball et al. (2016)	Cash-based oper prof lagged assets	CBOperProfLagAT	Indirect Evidence	0.46	3.33	HXZ variant
Ball et al. (2016)	Cash-based oper prof lagged assets qtrly	CBOperProfLagAT_q	Indirect Evidence	0.86	6.19	HXZ variant
Ball et al. (2016)	Oper prof R&D adj lagged assets	OperProfRDLagAT	Indirect Evidence	0.05	0.33	HXZ variant

continued on next page

Table 4: (continued)

Original / Related Study	Predictor	Acronym	Predictor Category	Reproduction Mean Ret	t-Stat	Original Study's Predictability Evidence (if Available)
Ball et al. (2016)	Oper prof R&D adj lagged assets (qtrly)	OperProfRDLagAT_q	Indirect Evidence	1.17	5.56	HXZ variant
Barbee, Mukherji and Raines (1996)	Sales-to-price quarterly	SP_q	Indirect Evidence	1.18	4.93	HXZ variant
Basu (1977)	Earnings-to-Price Ratio	EPq	Indirect Evidence	1.31	6.79	HXZ variant
Belo, Lin and Vitorino (2014)	Brand capital to assets	BrandCapital	Indirect Evidence	0.24	1.25	not studied for predictability
Bhandari (1988)	Market leverage quarterly	Leverage_q	Indirect Evidence	0.26	0.79	HXZ variant
Blitz, Huij and Martens (2011)	6 month residual momentum	ResidualMomentum6m	Indirect Evidence	0.39	4.08	HXZ variant
Boudoukh et al. (2007)	Net Payout Yield quarterly	NetPayoutYield_q	Indirect Evidence	0.76	2.04	HXZ variant
Boudoukh et al. (2007)	Payout Yield quarterly	PayoutYield_q	Indirect Evidence	0.72	6.16	HXZ variant
Brown and Rowe (2007)	Return on invested capital	roic	Not Predictor	0.08	0.33	t=0.9 in port sort
Callen, Khan and Lu (2013)	Accounting component of price delay	DelayAcct	Not Predictor	-0.16	-0.83	t=1 in long-short
Callen, Khan and Lu (2013)	Non-accounting component of price delay	DelayNonAcct	Not Predictor	0.27	1.70	t=1 in long-short
Campbell, Hilscher and Szilagyi (2008)	Failure probability	FailureProbability	Not Predictor	0.40	0.91	t=1.5 in conservative port sort
Campbell, Hilscher and Szilagyi (2008)	Failure probability	FailureProbabilityJune	Indirect Evidence	0.03	0.07	HXZ variant
Chan, Lakonishok and Sougiannis (2001)	R&D over market cap quarterly	RD_q	Indirect Evidence	1.89	5.23	HXZ variant
Chan, Lakonishok and Sougiannis (2001)	R&D to sales	rd_sale	Not Predictor	0.17	0.71	8 bps spread in port sort
Chan, Lakonishok and Sougiannis (2001)	R&D to sales	rd_sale_q	Indirect Evidence	0.71	1.48	HXZ variant
Cooper, Gulen and Schill (2008)	Asset growth quarterly	AssetGrowth_q	Indirect Evidence	-0.94	-4.84	HXZ variant
Desai, Rajgopal, Venkatachalam (2004)	Operating Cash flows to price quarterly	cfpq	Indirect Evidence	1.07	8.12	HXZ variant
Dichev (1998)	O Score quarterly	OScore_q	Indirect Evidence	-1.10	-3.02	HXZ variant
Dichev (1998)	Altman Z-Score	ZScore	Not Predictor	-0.35	-1.20	t=1.59 in univar reg
Dichev (1998)	Altman Z-Score quarterly	ZScore_q	Indirect Evidence	-0.13	-0.47	HXZ variant
Dimson (1979)	Dimson Beta	BetaDimson	Indirect Evidence	-0.23	-1.53	only shown to forecast beta
Eisfeldt and Papanikolaou (2013)	Org cap w/o industry adjustment	OrgCapNoAdj	Indirect Evidence	0.64	3.31	HXZ variant
Elgers, Lo and Pfeiffer (2001)	Number of analysts	nanalyst	Indirect Evidence	0.19	1.02	spread in median ret each leg size adj
Fama and French (1992)	Total assets to market (quarterly)	AMq	Indirect Evidence	0.78	3.31	HXZ variant
Fama and French (1992)	Book leverage (quarterly)	BookLeverageQuarterly	Indirect Evidence	-0.23	-1.59	HXZ variant
Fama and French (2006)	operating profits / book equity	OperProfLag	Indirect Evidence	0.40	1.81	HXZ variant
Fama and French (2006)	operating profits / book equity	OperProfLag_q	Indirect Evidence	1.02	3.40	HXZ variant
Fama and MacBeth (1973)	CAPM beta squred	BetaSquared	Not Predictor	-0.66	-1.71	t=0.3 in mv reg
Francis, LaFond, Olsson, Schipper (2004)	Earnings conservatism	EarningsConservatism	Indirect Evidence	-0.00	-0.01	correlated with BM and other predictors
Francis, LaFond, Olsson, Schipper (2004)	Earnings persistence	EarningsPersistence	Indirect Evidence	-0.21	-1.59	correlated with BM and other predictors
Francis, LaFond, Olsson, Schipper (2004)	Earnings Predictability	EarningsPredictability	Indirect Evidence	-0.60	-3.49	correlated with BM and other predictors
Francis, LaFond, Olsson, Schipper (2004)	Earnings Smoothness	EarningsSmoothness	Indirect Evidence	0.02	0.11	correlated with BM and other predictors
Francis, LaFond, Olsson, Schipper (2004)	Earnings timeliness	EarningsTimeliness	Indirect Evidence	-0.02	-0.21	correlated with BM and other predictors
Francis, LaFond, Olsson, Schipper (2004)	Value relevance of earnings	EarningsValueRelevance	Indirect Evidence	-0.02	-0.32	correlated with BM and other predictors

continued on next page

Table 4: (continued)

Original / Related Study	Predictor	Acronym	Predictor Category	Reproduction Mean Ret	t-Stat	Original Study's Predictability Evidence (if Available)
Francis, LaFond, Olsson, Schipper (2004)	RoA volatility	roavol	Indirect Evidence	-0.07	-0.19	correlated with BM and other predictors
Francis, LaFond, Olsson, Schipper (2005)	Accrual Quality	AccrualQuality	Indirect Evidence	0.16	0.62	correlated with E/P and factor structure
Francis, LaFond, Olsson, Schipper (2005)	Accrual Quality in June	AccrualQualityJune	Indirect Evidence	0.19	0.73	HXZ variant
Frankel and Lee (1998)	Intrinsic or historical value	IntrinsicValue	Indirect Evidence	0.48	2.45	not studied. Ingredient variable.
Franzoni and Marin (2006)	Pension Funding Status	FRbook	Indirect Evidence	0.32	2.58	HXZ variant
Hafzalla, Lundholm, Van Winkle (2011)	Percent Abnormal Accruals	AbnormalAccrualsPercent	Indirect Evidence	-0.29	-4.09	HXZ variant
Hahn and Lee (2009)	Tangibility quarterly	tang_q	Indirect Evidence	0.85	5.84	HXZ variant
Haugen and Baker (1996)	Capital turnover	CapTurnover	Indirect Evidence	0.25	1.23	t<2 in mv reg nonstandard
Haugen and Baker (1996)	Capital turnover (quarterly)	CapTurnover_q	Indirect Evidence	0.85	4.68	HXZ variant
Holthausen and Larcker (1992)	Depreciation to PPE	depr	Indirect Evidence	0.28	1.02	ingredient in complicated model
Holthausen and Larcker (1992)	Change in depreciation to PPE	pchdepr	Indirect Evidence	0.18	1.66	ingredient in complicated model
Hou and Loh (2016)	Bid-ask spread (TAQ)	BidAskTAQ	Not Predictor	0.12	0.41	t=1.3 in mv reg
La Porta (1996)	Long-term EPS forecast (Monthly)	fgr5yrNoLag	Indirect Evidence	-0.66	-1.60	HXZ variant
Lakonishok, Shleifer, Vishny (1994)	Cash flow to market quarterly	CFq	Indirect Evidence	1.69	10.62	HXZ variant
Lakonishok, Shleifer, Vishny (1994)	Annual sales growth	sgr	Indirect Evidence	-0.60	-4.25	HXZ variant
Lakonishok, Shleifer, Vishny (1994)	Annual sales growth quarterly	sgr_q	Indirect Evidence	0.60	3.51	HXZ variant
Lamont, Polk and Saa-Requejo (2001)	Kaplan Zingales index	KZ	Not Predictor	0.08	0.53	t=1.1 in conservative port sort
Lamont, Polk and Saa-Requejo (2001)	Kaplan Zingales index quarterly	KZ_q	Indirect Evidence	-1.49	-9.23	HXZ variant
Lev and Nissim (2004)	Taxable income to income (qtrly)	Tax_q	Indirect Evidence	0.03	0.23	HXZ variant
Loughran and Wellman (2011)	Enterprise Multiple quarterly	EntMult_q	Indirect Evidence	-1.59	-11.96	HXZ variant
Naranjo, Nimalendran, Ryngaert (1998)	Dividend yield for small stocks	DivYield	Indirect Evidence	0.34	1.11	mixed results, small spread
Naranjo, Nimalendran, Ryngaert (1998)	Last year's dividends over price	DivYieldAnn	Indirect Evidence	0.01	0.11	HXZ variant
Novy-Marx (2010)	Operating leverage (qtrly)	OPLeverage_q	Indirect Evidence	0.39	2.37	HXZ variant
Novy-Marx (2013)	gross profits / total assets	GPlag	Indirect Evidence	0.20	1.85	HXZ variant
Novy-Marx (2013)	gross profits / total assets	GPlag_q	Indirect Evidence	0.88	6.17	HXZ variant
Ortiz-Molina and Phillips (2014)	Asset liquidity over book assets	AssetLiquidityBook	Indirect Evidence	0.35	1.37	no predictability. Correlated with ICC
Ortiz-Molina and Phillips (2014)	Asset liquidity over book (qtrly)	AssetLiquidityBookQuart	Indirect Evidence	0.31	0.96	HXZ variant
Ortiz-Molina and Phillips (2014)	Asset liquidity over market	AssetLiquidityMarket	Indirect Evidence	1.41	7.46	no predictability. Correlated with ICC
Ortiz-Molina and Phillips (2014)	Asset liquidity over market (qtrly)	AssetLiquidityMarketQuart	Indirect Evidence	1.31	6.35	HXZ variant
Ou and Penman (1989)	CF to debt	cashdebt	Indirect Evidence	-0.06	-0.21	ingredient in complicated model
Ou and Penman (1989)	Current Ratio	currat	Indirect Evidence	0.29	1.93	ingredient in complicated model
Ou and Penman (1989)	Change in Current Ratio	pchcurrat	Indirect Evidence	0.21	2.54	ingredient in complicated model
Ou and Penman (1989)	Change in quick ratio	pchquick	Indirect Evidence	0.32	3.27	ingredient in complicated model
Ou and Penman (1989)	Change in sales to inventory	pchsaleinv	Indirect Evidence	0.49	5.24	ingredient in complicated model
Ou and Penman (1989)	Quick ratio	quick	Indirect Evidence	0.30	1.77	ingredient in complicated model

continued on next page

Table 4: (continued)

Original / Related Study	Predictor	Acronym	Predictor Category	Reproduction Mean Ret	Reproduction t-Stat	Original Study's Predictability Evidence (if Available)
Ou and Penman (1989)	Sales to cash ratio	salecash	Indirect Evidence	0.22	1.31	ingredient in complicated model
Ou and Penman (1989)	Sales to inventory	saleinv	Indirect Evidence	0.03	0.18	ingredient in complicated model
Ou and Penman (1989)	Sales to receivables	salerec	Indirect Evidence	0.29	1.53	ingredient in complicated model
Penman, Richardson and Tuna (2007)	Enterprise component of BM	EBM_q	Indirect Evidence	0.81	7.09	HXZ variant
Penman, Richardson and Tuna (2007)	Net debt to price	NetDebtPrice_q	Indirect Evidence	-0.73	-3.78	HXZ variant
Piotroski (2000)	Piotroski F-score	PS_q	Indirect Evidence	1.39	6.35	HXZ variant
Richardson et al. (2005)	Change in short-term investment	DelSTI	Not Predictor	-0.04	-0.53	t=0.4 in mv reg
Rosenberg, Reid, and Lanstein (1985)	Book to market (quarterly)	BMq	Indirect Evidence	1.65	3.67	HXZ variant
Soliman (2008)	Asset Turnover	AssetTurnover	Indirect Evidence	0.40	2.23	t=0.3 in mv reg
Soliman (2008)	Asset Turnover	AssetTurnover_q	Indirect Evidence	0.59	3.10	HXZ variant
Soliman (2008)	Change in Noncurrent Operating Assets	ChNCOA	Indirect Evidence	-1.02	-5.68	No predictability. Ingredient for predictor.
Soliman (2008)	Change in Noncurrent Operating Liab	ChNCOL	Indirect Evidence	-0.54	-3.60	No predictability. Ingredient for predictor.
Soliman (2008)	Change in Profit Margin	ChPM	Indirect Evidence	0.11	1.35	t=0.3 in mv reg
Soliman (2008)	Profit Margin	PM	Indirect Evidence	0.52	1.94	t=1 in mv reg
Soliman (2008)	Profit Margin	PM_q	Indirect Evidence	1.29	2.88	HXZ variant
Soliman (2008)	Return on Net Operating Assets	RetNOA	Indirect Evidence	0.01	0.05	t=1.4 in mv reg
Soliman (2008)	Return on Net Operating Assets	RetNOA_q	Indirect Evidence	1.26	3.25	HXZ variant
Valta (2016)	Secured debt	secured	Indirect Evidence	-0.00	-0.04	t > 1.96 in mv reg
Valta (2016)	Secured debt indicator	securedind	Indirect Evidence	-0.06	-0.69	GHZ variant
Whited and Wu (2006)	Whited-Wu index	WW	Not Predictor	0.33	1.18	t=1.3 in port sort
Whited and Wu (2006)	Whited-Wu index	WW_Q	Indirect Evidence	0.50	1.31	HXZ variant