

7536681-filing-discipline-regulatory-disclosure-behavior-and-the-cross-section-of-stock-

SUMMARY: This study examines the impact of filing discipline, as measured by the SEC's EDGAR system, on the cross-section of stock returns. The authors build three characteristics from filing timestamps: `q_lag`, the number of days from fiscal quarter-end to SEC acceptance of the most recent 10-Q; `nt_12m`, the count of Form NT 10-K and NT 10-Q filings in the trailing year; and `burst`, the coefficient of variation of the gaps between consecutive 8-K acceptance timestamps. The study finds that the level of a firm's filing discipline, rather than its innovation, predicts the monthly cross-section of stock returns. The authors construct a portfolio based on these characteristics, resulting in a net Sharpe ratio of 1.27 with a six-factor alpha of 0.39% per month. The findings suggest that the predictive content lies in the cross-sectional level of filing discipline, not in its trailing-mean-adjusted within-firm deviation. The effect is concentrated in smaller and more volatile stocks, limiting its capacity.

METHOD: The authors construct three characteristics from EDGAR timestamps: `q_lag`, `nt_12m`, and `burst`. These characteristics are ranked cross-sectionally, smoothed over three months, and signed using in-sample information only. The equally weighted composite, denoted as `disc`, is oriented so that a high score means good discipline. Portfolios are long the top quintile and short the bottom quintile, equally weighted within each leg, neutralised against size, short-term reversal, and lottery demand, and held for six months with one-sixth of the book rolled each month. Trading costs are charged at a flat 10 basis points one way and borrow fees at a flat 0.6% per year, anchored to Frazzini et al. (2018), Novy-Marx and Velikov (2016), and D'Avolio (2002), with a sensitivity table. The main universe includes companies with market capitalisation above \$100M, dollar volume above \$1M, and price above \$3, averaging 2,253 names per month. The net Sharpe ratio is 1.27 (1.29 in-sample, 1.34 held-out) with an alpha of 0.39% per month against the five-factor model of Fama and French (2015) augmented with momentum, and a market beta of -0.03.

KEY FINDINGS:

- The predictive content lies in the cross-sectional level of filing discipline, not in its trailing-mean-adjusted within-firm deviation.
- The effect is concentrated in smaller and more volatile stocks, limiting its capacity.
- The three characteristics are nearly uncorrelated with one another, combining largely distinct variation rather than three measurements of one quantity.
- Filing discipline is persistent: the twelve-month rank autocorrelation of `disc` is 0.55 and 47% of its cross-sectional variance is a permanent firm attribute.
- Predictive power rises monotonically as the universe is widened, from an information coefficient of 0.021 among large liquid names to 0.061 across all common stock.

IMPLICATION FOR TRADING: The findings suggest that the level of filing discipline, rather than its innovation, is the key driver of stock returns. This insight can be used by traders to identify stocks with strong filing discipline, which may be more likely to have positive returns. However, the effect is concentrated in smaller and more volatile stocks, limiting the potential for high returns. Additionally, the effect is not explained by mispricing, but rather by the persistent nature of filing discipline. Traders can use this information to construct portfolios that exploit the predictive power of filing discipline, but they should be aware of the limitations of the effect.

Filing Discipline: Regulatory Disclosure Behavior and the Cross-Section of Stock Returns

Lyuge Zhu

London School of Economics and Political Science

L.Zhu27@lse.ac.uk

September 28, 2026

Abstract

Every SEC filing carries a timestamp. We build three characteristics from those timestamps alone — how late a firm files its quarterly report, how often it files a notification of late filing, and how irregularly its 8-K filings arrive. That filing timeliness predicts the cross-section is known; our question is which component of it does. Over 2010–2025 a dollar-neutral quintile portfolio in a universe of roughly 2,250 US stocks earns a net Sharpe ratio of 1.27 with a six-factor alpha of 0.39% per month ($t = 5.62$) and a market beta of -0.03 . The economically informative result is a decomposition: the predictive content lies in the cross-sectional *level* of filing discipline, not in its trailing-mean-adjusted within-firm deviation. Rebuilt as deviations from each firm’s own history, the same three characteristics produce a gross alpha of 0.08% per month ($t = 1.56$) and no held-out performance. Because the existing evidence measures the level without decomposing it, our finding is best read as a statement about which component carries the signal rather than as an explanation of why the effect survives. The effect is concentrated in smaller and more volatile stocks, which limits its capacity.

JEL classification: G12, G14, M41

Keywords: filing timeliness, disclosure behaviour, cross-section of stock returns, SEC EDGAR, filing metadata, reproducibility

1 Introduction

Research on corporate disclosure reads what firms say. The SEC’s EDGAR system also records, for every filing, exactly when it arrived, to the second. That metadata is free, covers every listed issuer, and admits almost no construction choices *at the level of the characteristic itself*: one timestamp minus another. The portfolio built on top of it involves many choices, which Section 4 states and Section 7 counts. This paper asks whether it prices.

The question is not whether late filing carries bad news. Givoly and Palmon (1982) established four decades ago that good news is released early and bad news late, and a substantial literature has since studied the announcement effects of filing delays and of strategic announcement timing. Our contribution is a decomposition rather than a discovery. We show that what predicts the monthly cross-section is the *level* of a firm’s filing discipline — the fact that a firm is chronically eight days slower than its peers — and not the innovation in that discipline. Rebuilt as deviations from each firm’s own trailing history, the identical characteristics run through an identical portfolio pipeline produce essentially nothing, gross or net. This matters because the prior literature is organised around events: the filing of a Form NT, the announcement window around a delayed report, the

departure of an announcement date from its expected slot. The level is a persistent firm attribute, and it is the attribute that carries the predictive content. We do not identify mispricing, and we use *predicts* throughout in its statistical sense.

We form three characteristics from EDGAR timestamps: `q_lag`, the number of days from fiscal quarter-end to SEC acceptance of the most recent 10-Q; `nt_12m`, the count of Form NT 10-K and NT 10-Q filings in the trailing year; and `burst`, the coefficient of variation of the gaps between consecutive 8-K acceptance timestamps. Each is ranked cross-sectionally, smoothed over three months, and signed using in-sample information only. The equally weighted composite is denoted `DISC`, oriented so that a high score means good discipline. Portfolios are long the top quintile and short the bottom quintile, equally weighted within each leg, neutralised against size, short-term reversal and lottery demand, and held for six months with one-sixth of the book rolled each month. Trading costs are charged at a flat 10 bp one way and borrow fees at a flat 0.6% a year, anchored to Frazzini et al. (2018), Novy-Marx and Velikov (2016) and D’Avolio (2002), with a sensitivity table.

In our main universe — market capitalisation above \$100M, dollar volume above \$1M, price above \$3, averaging 2,253 names per month — the net Sharpe ratio is 1.27 (1.29 in-sample, 1.34 held-out), with an alpha of 0.39% per month ($t = 5.62$) against the five-factor model of Fama and French (2015) augmented with momentum, and a market beta of -0.03 . Both legs contribute: the long basket beats an equally weighted universe by 2.51% per year and the short basket underperforms it by 3.15% per year. The sort rather than the screen does the work: the equally weighted universe earns a small negative alpha of -0.12% per month against the same model, while the long basket earns $+0.11\%$.

Three further results shape the interpretation. First, the three characteristics are nearly uncorrelated with one another (median $|\rho| = 0.05$, maximum 0.16), so the composite combines largely distinct variation rather than three measurements of one quantity. Second, filing discipline is persistent: the twelve-month rank autocorrelation of `DISC` is 0.55 and 47% of its cross-sectional variance is a permanent firm attribute, against 0.95 and 86% for log size and essentially zero for short-term reversal. Third, predictive power rises monotonically as the universe is widened, from an information coefficient of 0.021 among large liquid names to 0.061 across all common stock, which is consistent with the information-environment mechanism of Merton (1987) but is also the most common pattern in the anomaly literature and should not be read as discriminating evidence.

We are deliberately narrow about what the decomposition explains. It is tempting to argue that the effect survives because the literature traded the event component while the level component was never treated as tradeable information. We cannot test that: our sample begins in 2010, so a publication-date design in the manner of McLean and Pontiff (2016) is unavailable, and we never observe whether the event component was arbitrated or simply never had monthly cross-sectional content. What we can show is that the two components are different objects and that only one of them prices.

Section 2 places the paper relative to the disclosure-timing literature; Section 3 describes the sample and three data corrections that materially affect the results; Section 4 defines the characteristics and the portfolio protocol; Section 5 reports the evidence; Section 6 interprets it; Section 7 states what the paper does not claim; and Section 8 concludes.

2 Related Literature

The idea that reporting timeliness carries information is old and well established. Givoly and Palmon (1982) document that firms release good news earlier than bad news. Alford et al. (1994) study violations of the statutory 10-K deadline and the Form NT mechanism that accompanies

them. Bagnoli et al. (2002) show that an earnings announcement arriving later than its expected date predicts a negative surprise, and Doyle and Magilke (2009) and deHaan et al. (2015) examine the strategic scheduling of announcements into low-attention windows. A separate strand treats 8-K filings as a measure of information intensity (Zhao, 2017).

Most of these studies are organised around an *event*: a filing date, an announcement window, a departure from an expected date, with a short-horizon return around that date as the dependent variable. Two papers, however, already do what this paper’s first result does. Bannouh et al. (2021) define a reporting lag as the number of days from period end to filing and show that faster filers earn a premium in the cross-section, including within the subsample of firms that meet their deadlines. Brooks et al. (2024) use EDGAR to show that firms filing their 10-K early have better subsequent earnings and higher subsequent stock returns, in both large- and small-capitalisation subsamples. Zhao (2017) studies 8-K filing frequency as a measure of information intensity and reports a long-short spread. That filing timeliness predicts the cross-section is therefore established, not new.

We should be explicit about what is and is not new here. Our `q_lag` closely follows the quarterly-report component of Bannouh et al.’s reporting-lag measure, and its economic intuition is Givoly and Palmon’s: a firm holding bad news delays. We replicate that result rather than discover it, and we regard the replication — on an independently constructed panel, under a frozen protocol, net of costs — as useful but not as a contribution.

What we believe is new is the decomposition. Section 5.4 rebuilds each characteristic as a deviation from the firm’s own history and pushes it through the identical pipeline. The level carries the predictive content for `q_lag` (0.41%, $t = 5.17$, against 0.10%, $t = 1.41$, for its deviation) and for `burst` (0.24%, $t = 3.41$, against 0.04%, $t = 0.60$), whereas `nt_12m` is uninformative in either form (0.21%, $t = 1.50$, and 0.21%, $t = 1.47$). The existing papers measure the level and do not ask which component of it carries the signal; the answer bears directly on how the effect should be interpreted, because a stable between-firm difference and a firm’s own change in behaviour imply different economics. Our second contribution is to place three filing-metadata characteristics, not one, in a common framework and to show that the three measures contain largely distinct variation. Our third is methodological: the entire paper reproduces from five raw inputs in a single notebook whose assertions fail if any reported number changes.

A second literature motivates our treatment of the metadata as a low-degrees-of-freedom laboratory. Text-based disclosure research (Tetlock, 2007; Loughran and McDonald, 2011) requires dictionary and parsing choices; accounting characteristics require restatement and point-in-time decisions. A filing timestamp requires neither. Given the replication concerns raised by Harvey et al. (2016) and Hou et al. (2020), a domain in which the *measurement* of the characteristic carries almost no latitude is a useful place to look. The latitude does not disappear — it moves downstream, into the smoothing window, the holding period, the neutralisation set, the quantile cut and the screens — and we report that search rather than claim it away.

3 Data

Sources. Filing metadata come from the SEC EDGAR submissions API. We download the complete filing history for every issuer, including the historical shards referenced by the `filings.files` field rather than only the truncated `filings.recent` block; the raw download contains 7,596,942 filings, of which 6,637,458 map to a CRSP PERMNO through a CIK–PERMNO link covering 7,379 securities. The link is not a survivor snapshot: 1,211 of the linked securities last filed before 2015. Returns, prices, capitalisations and volumes are from CRSP daily and monthly files; factor returns

are the official monthly series from Kenneth French’s data library. The sample runs from January 2010 to November 2025.

Return construction. Daily returns are compounded to monthly, $\prod_d(1 + r_d) - 1$, rather than summed. Summation overstates monthly returns by 28.1 basis points on average and produces 1,444 stock-months below -100% . CRSP delisting returns `dlret_ciz` are chained explicitly onto the ordinary return.

Universe screens. The main universe requires market capitalisation above \$100M, dollar volume above \$1M and a price above \$3. Each cut is tied to a measurement boundary rather than to performance. The \$1M volume floor keeps the sample inside the range where the cost anchors are actually estimated: Frazzini et al. (2018) measure live institutional executions and Novy-Marx and Velikov (2016) measure TAQ effective spreads, and below roughly \$0.5M of daily volume neither source constrains the cost assumption, so any cost charged there is extrapolation. The \$100M capitalisation floor removes the range in which stock loan is frequently unavailable at any price, which matters because the strategy is dollar-neutral and its short leg is the harder side to borrow. The \$3 price floor removes bid-ask bounce, which inflates measured volatility without carrying signal; Table 9 shows the effect directly, in that a \$1 price screen dominates the fully unscreened universe.

We report six universes rather than one because the informative object is the gradient, not any single cut, and each row answers a different question. *All common stock* applies no screen at all and is the reference point: it shows what the characteristic does when the researcher makes no choice. *Price above \$1* is the minimal defensible screen, removing only the sub-dollar range in which quoted prices bounce between bid and ask by a material fraction of the price; it is also the conventional penny-stock line. *\$50M and \$0.5M* marks the boundary below which neither cost source constrains the assumption. *\$100M and \$1M* is the headline. *\$250M and \$2M* is a common institutional minimum. *\$1B and \$10M* isolates the mega-cap, highly liquid range, where implementation is easiest and where an information-environment mechanism predicts the effect should be weakest — which is what Table 9 finds.

No cut point was selected on performance, and the monotonicity does not depend on where the intermediate lines are drawn. We report the \$100M universe as the headline because it is the widest screen whose implementation assumptions — measured trading costs, available borrow — remain defensible; the wider rows are reported so that the gradient is visible, not because we would trade them.

Security screen. CRSP is distributed in CIZ format, in which `shrcd` is a derived field taking the single value 11; the conventional instruction to retain share codes 10 and 11 therefore cannot be executed literally. The equivalent screen is `SecurityType=EQTY`, `SecuritySubType=COM` and `ShareType=NS`, which excludes ADRs, restricted to NYSE, AMEX and NASDAQ. Where one CIK maps to several PERMNOs (0.97% of stock-months) we flag the largest by capitalisation as primary and report a robustness check restricted to primary classes.

A screen that silently discards delisting returns. The CRSP names file carries a terminal stub record on the delisting date with `SecurityType` null and `SecuritySubType` equal to UNK. Because the monthly common-stock flag is taken from the last trading day of the month, and the delisting day is that last day, every delisting month is classified non-common and dropped — discarding all 8,961 delisting returns even though the merge that attaches them executes correctly.

We forward-fill the classification from the last valid name record, which restores 5,833 of the 8,961 delisting months (the remainder are funds and ADRs and remain excluded). Their delisting returns average -5.6% . The correction leaves the headline results unchanged, because the universe screen admits few securities that survive to a distressed delisting; delisting typically occurs after a security has already fallen out of the universe. We report the correction because a reader cannot otherwise tell whether it was made.

Timing. A signal formed on the last trading day of month t earns the return of month $t + 1$. Because a security can be temporarily absent from the screened panel, we record the month each forward return is actually drawn from and verify the alignment directly: 99.98% of panel observations draw the return from $t + 1$, and 112 rows reach further forward.

Transaction costs. We charge two flat constants rather than a graded ladder, because a ladder requires cut points that no published source supplies while a constant carries a one-line justification and a transparent sensitivity.

One-way trading costs are 10 bp. Frazzini et al. (2018) measure a median execution cost of 6.2 bp across \$1.7 trillion of live institutional trades in 21 markets over 19 years, and report that live costs are an order of magnitude below the effective spreads used in earlier work; Novy-Marx and Velikov (2016) put TAQ effective spreads at 20–57 bp for medium-turnover anomalies. These measure different objects: an effective spread is an upper bound on execution cost, since an institution need not cross it. We take the realised-cost concept as the relevant one and set the charge above the measured median.

Borrow fees are 0.6% per year. Muravyev et al. (2025) report the distribution of realised borrow fees from Markit securities-lending data: the median US stock borrows at 0.38% per year, the 90th percentile is 0.63%, and only about 7% of stock-days exceed 1%. They show that the impact of borrow fees on any strategy comes almost entirely from short positions above that 90th percentile. Our 0.6% sits between their median and their 90th percentile, and is therefore conservative for a short leg that we show below is not in the expensive tail. The older distribution in D’Avolio (2002) — 91% of stocks below 1% per year, the remaining 9% special and averaging 4.3% — implies a comparable range. One caveat on the comparison: the Muravyev et al. (2025) sample covers optionable stocks with a median capitalisation of \$2.1bn, larger than our universe, and smaller firms borrow at higher fees. Short-interest data (Section 7) show the short leg sitting modestly above the universe median rather than in the special tail, which supports a fee inside this range rather than a punitive one; we do not scale the fee by that ratio, since borrow fees are highly non-linear in utilisation. Table 15 reports the sensitivity from 0.5% to 3%: across the 0.38%–0.63% interquartile range of Muravyev et al. (2025) the Sharpe ratio moves by less than 0.08, and the strategy survives a fee of 2% per year — more than three times our assumption — at a Sharpe ratio of 0.89.

All reported returns are net of both charges. Annual turnover is 216%, so the trading charge costs about 43 bp a year while the borrow charge costs 60 bp on the dollar of short exposure; the borrow assumption is therefore the more consequential of the two.

Frozen decisions, and what that does and does not mean. The sign of each characteristic and every screening threshold are determined on 2010–2017 alone. The reported specification is the in-sample optimal subset of the candidate set, so no information from 2018–2025 is required to select it. Reported Sharpe ratios are those of the six-sleeve overlay described in Section 4, never those of a single hand-picked rebalancing schedule.

We call 2018–2025 *held out* rather than out of sample, and the distinction is not cosmetic. The research was carried out in 2026. There is no time-stamped pre-registration, and no third party can verify that the protocol was fixed before those returns were examined. The freeze is a self-imposed discipline, and it should be given the weight due to a self-imposed discipline rather than the weight due to a real-time track record. The same applies to our multiple-testing correction: the $M = 260$ of Table 14 counts roughly 130 candidate characteristics in two signs, which is the search over *characteristics*. It does not count the search over portfolio construction — the smoothing window, the holding period, the neutralisation set, the quantile cut, the screens and the cost model — and no Bonferroni threshold we could quote would.

4 Empirical Methodology

4.1 Three characteristics

All three are measured as of the last trading day of the signal month. Filings accepted on the signal day itself are excluded, a deliberately conservative convention.

- **q_lag**: days from the fiscal quarter-end (**reportDate**) to SEC acceptance of the most recent 10-Q, restricted to $[0, 200]$ and treated as missing if the latest 10-Q is more than 200 days stale.
- **nt_12m**: the count of Form NT 10-K and NT 10-Q filings, including their amendments, in the trailing 365 days. Other NT forms are excluded; NT 11-K, for instance, concerns an employee benefit plan rather than the firm’s own periodic report. A firm with no record scores zero rather than missing.
- **burst**: the coefficient of variation of the gaps between consecutive 8-K acceptance timestamps in the trailing 365 days, requiring at least three filings in the window.

burst deserves a comment because it mixes two distinct signals. Item 2.02 filings — the earnings release — are 29.5% of all 8-Ks and arrive on a quarterly schedule; their inter-arrival coefficient of variation averages 0.22 against 0.82 for all other 8-Ks. The measure therefore captures both how reliably a firm keeps its earnings calendar and how irregular its event-driven disclosure flow is. The two components correlate only 0.17, and the blend outperforms either alone in portfolio terms.

4.2 The composite

Each characteristic is converted to a cross-sectional percentile rank, smoothed with a three-month trailing mean, and oriented by the sign of its in-sample rank correlation with forward returns. All three raw characteristics carry a negative sign, so all three are reversed; following the convention of Fama and French (2015) in naming SMB and CMA, and the open-source convention of Chen and Zimmermann (2022), we orient every component so that a high score corresponds to a high expected return. The composite DISC is the equally weighted average, with missing components filled at the median rank.

4.3 Portfolios

At each rebalance we rank on DISC, regress the rank cross-sectionally on the ranks of log size, five-day reversal and the 21-day maximum daily return (Bali et al., 2011), and sort on the residual.

The portfolio is long the top quintile and short the bottom quintile, equally weighted within each leg, with $\sum_i |w_i| = 2$.

Positions are held for six months. Rather than committing to one of the six possible rebalancing schedules, we run six overlapping sleeves offset by one month and allocate one-sixth of the book to each, so that one-sixth of the portfolio turns over per month. This is the overlapping-portfolio construction of Jegadeesh and Titman (1993). It leaves the holding period (six months) and the annual turnover (216%) unchanged, removes the arbitrary choice of start month, and avoids concentrating all execution into two dates per year. The dispersion across the six individual schedules is reported in Table 14.

Between rebalances weights drift with returns and are not reset. A position that leaves the universe mid-holding-period is carried at its unscreened CRSP return, so that delisting returns are earned, and is charged an exit cost at the next rebalance.

Alphas are estimated against the Fama and French (2015) five factors plus momentum (Carhart, 1997), with Newey and West (1987) standard errors at six lags. Cross-sectional regressions follow Fama and MacBeth (1973) on percentile ranks, again with Newey-West errors.

5 Empirical Results

5.1 The panel and the three dimensions

The estimation panel contains 430,337 stock-months over 191 months and 5,916 firms, averaging 2,253 names per month, with a median market capitalisation of \$1.90 billion and median dollar volume of \$17.3 million. Table 1 reports the characteristics. The median firm files its 10-Q 36 days after quarter-end, with the ninetieth percentile at 40 days — the distribution piles up against the statutory deadline. Late-filing notifications are rare: 96.6% of stock-months carry none.

Table 1: Characteristics of the three filing measures

	Coverage	Mean	S.D.	p10	Median	p90	Share zero
q_lag (days)	0.987	34.93	6.51	27.00	36.00	40.00	0.002
nt_12m (count)	1.000	0.045	0.270	0.00	0.00	0.00	0.966
burst (ratio)	0.984	0.829	0.237	0.543	0.818	1.130	0.000

Notes: 430,337 stock-months, January 2010 to November 2025, in the \$100M/\$1M universe. Coverage is the share of stock-months for which the characteristic is observed.

Table 2 reports the average monthly rank correlation between the three. The median absolute off-diagonal correlation is 0.05 and the maximum is 0.16; **q_lag** and **burst** are essentially uncorrelated at 0.001. The three measures therefore contain largely distinct variation: whether a firm files its 10-Q late says almost nothing about whether it has filed a late-filing notice, and nothing at all about whether its 8-K flow is bursty. We stop short of concluding that no common latent factor exists. Low rank correlation is also consistent with substantial measurement error, with the 96.6% of **nt_12m** observations that are exactly zero, and with three measures responding to one latent variable at different thresholds; establishing the absence of a common factor would require a model fitted to zero-inflated data, which we do not attempt. As Table 6 shows, only two of the three carry standalone predictive power in any case.

Table 2: The three dimensions are nearly orthogonal

	q_lag	nt_12m	burst
q_lag	1.000	0.161	0.001
nt_12m		1.000	0.053
burst			1.000

Notes: Cross-sectional Spearman correlations computed each month and averaged over 191 months. Median absolute off-diagonal correlation 0.053; maximum 0.161.

5.2 Main results

Table 3 reports the portfolio. The long-short spread earns 4.70% per year net of costs on a volatility of 3.70%, a Sharpe ratio of 1.27, with a maximum drawdown of 7.1% and a monthly hit rate of 63%. The six-factor alpha is 0.39% per month with $t = 5.62$ and the market beta is -0.03 . In-sample and held-out Sharpe ratios are 1.29 and 1.34.

Table 3: Long-short portfolio and its components

	Return (% p.a.)	Vol. (%)	Sharpe	In-s.	Held-out	Drawdown (%)	Alpha (%/mo)	Beta
Long-short	4.70	3.70	1.27	1.29	1.34	-7.1	0.39*** (5.62)	-0.03
Long basket (gross)	14.68	18.98	0.70	1.00	0.49	-28.8	0.11*** (3.20)	1.00
Universe, equal-wtd.	12.17	19.73	0.55	0.89	0.31	-32.7	-0.12*** (-2.94)	1.04
Short basket (gross)	9.02	20.13	0.38	0.72	0.15	-43.9	-0.38*** (-5.30)	1.05

Notes: \$100M/\$1M universe, January 2010 to November 2025, 191 months. The long-short line is net of trading costs and borrow fees; the basket lines are gross. The long-short book is dollar-neutral and self-financing, so its return series is already an excess return; the three basket lines carry one dollar of net exposure, so the risk-free rate (1.33% per year over the sample) is netted out before computing their Sharpe ratios and alphas. Alpha is the intercept from a regression on the Fama and French (2015) five factors plus momentum, with Newey and West (1987) t -statistics at six lags in parentheses. *** denotes $|t| > 2.58$. Average holdings 899; annual turnover 216%; monthly hit rate 63%.

Two features of Table 3 matter for interpretation. First, both legs contribute. The long basket beats the equally weighted universe by 2.51% per year and the short basket underperforms it by 3.15% per year; decomposing the spread's alpha, the long side supplies 0.13% per month ($t = 2.73$) and the short side 0.26% ($t = 3.39$), which sum to the 0.39% of the combined book. The short side is both the larger and the better identified of the two. Second, the sort does the work rather than the screen: the equally weighted universe earns a small negative alpha of -0.12% per month against the six-factor model, while the long basket earns $+0.11\%$, so the 2.51% annual gap between them is not an artefact of holding an unusual set of securities.

A six-month holding period admits six rebalancing schedules. We hold all six as overlapping sleeves rather than choosing one, so Table 3 reports the combined portfolio. For reference, the six individual schedules deliver Sharpe ratios of 1.01 to 1.25 with a mean of 1.15; the combined book reaches 1.27 because six imperfectly correlated sleeves diversify. The overlay is not a selected

specification — it is the construction that removes the choice of start month, and it is what an implementer holds. Table 14 reports the full dispersion.

The long-short return is small in absolute terms because the portfolio carries no leverage: it is one dollar long and one dollar short, and its volatility is 3.70%. The realised P&L of the two legs is +14.60% and −9.90% per year respectively, which sum to the reported 4.70%. The short leg loses money in level terms throughout a sixteen-year bull market; its contribution is that it loses less than the universe gains. Figure 1 shows both.

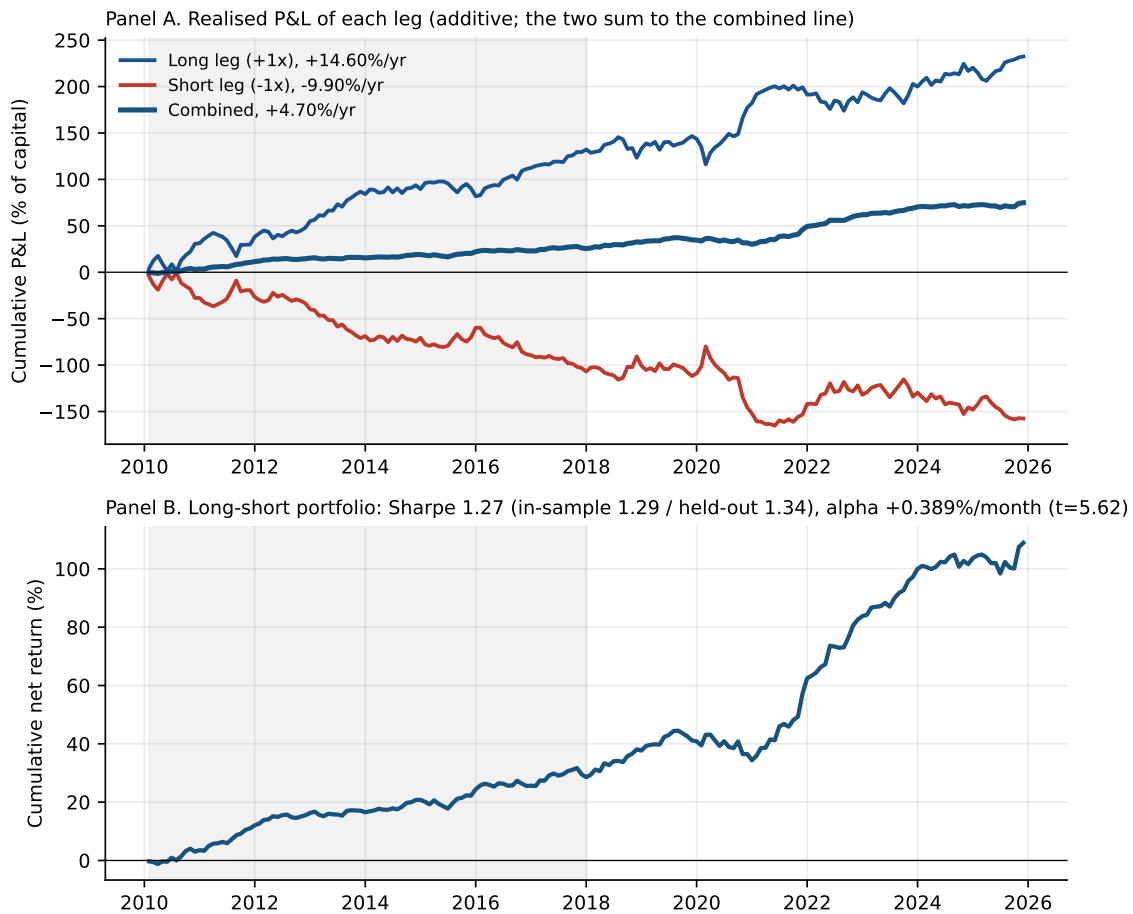


Figure 1: Realised P&L of each leg, and the combined portfolio

Notes: Panel A plots cumulative P&L as a percentage of capital for the long leg (+1 $\times$), the short leg (−1 $\times$) and their sum; the accounting is additive, so the two legs sum exactly to the combined line. Panel B plots the compounded net return of the long-short portfolio. The shaded region is the in-sample period.

Figure 2 plots the cumulative demeaned return by DISC rank. Because returns are demeaned each month, the curve begins and ends at zero by construction; the informative feature is its shape between. The curve is smooth and single-troughed, with the trough at the 37th percentile, indicating that the relationship is monotone rather than driven by either tail. The bottom quintile earns −0.38% per month relative to the cross-sectional mean and the top quintile +0.23%. That the trough sits left of the midpoint says the short side is the shorter and steeper of the two, which is consistent with the alpha decomposition in Table 3. This monotonicity is why we sort into quintiles rather than deciles; narrower cuts add idiosyncratic risk without adding signal.

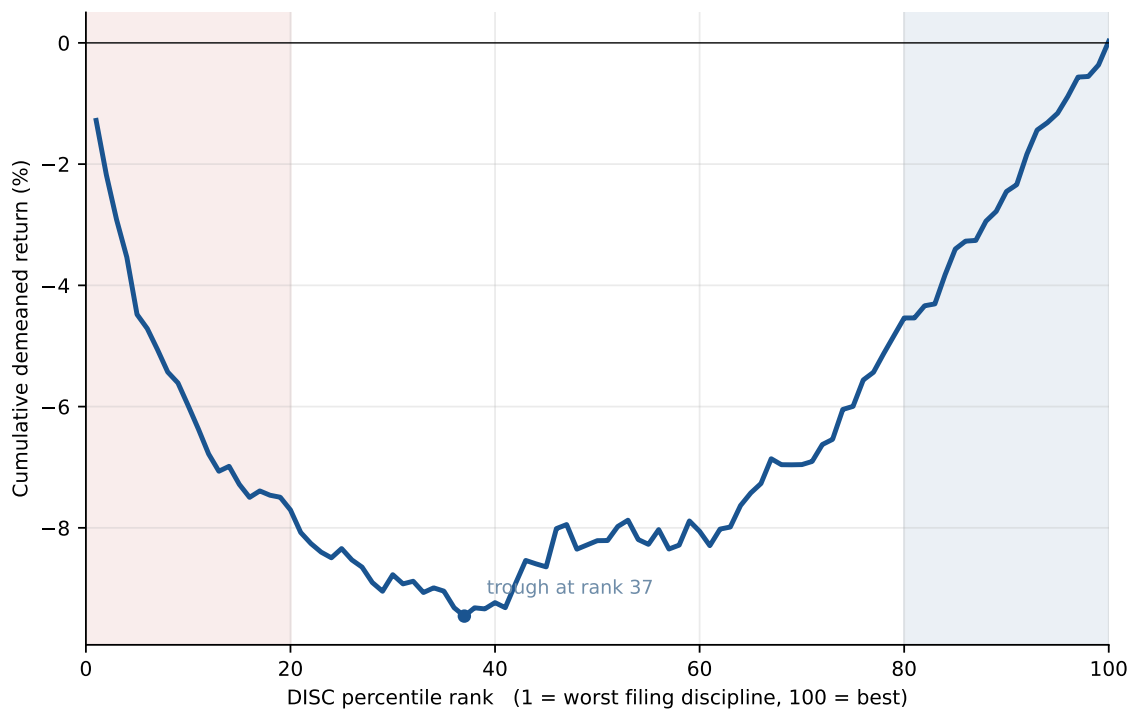


Figure 2: Cumulative demeaned return by filing-discipline rank

Notes: Stocks are sorted into 100 buckets on DISC each month; the plotted series is the cumulative sum of average cross-sectionally demeaned next-month returns over the full sample. Shading marks the quintiles that are traded. We do not split the curve by sub-period: the depth of a sub-period curve reflects that period's cross-sectional dispersion as much as the strength of the signal, and the sub-period spreads appear with proper statistics in Table 13.

5.3 What the choice of construction changes

The alpha is a property of the sort. What changes across implementations is the uncompensated risk carried alongside it. Table 4 trades the identical score four ways.

Table 4: The same signal, four constructions

	Names	Return (% p.a.)	Vol. (%)	Sharpe	Held-out	D'down (%)	Alpha (%/mo)	Beta
Long-only, top quintile	449	14.67	17.73	0.75	0.51	-26.7	0.11 (2.40)	0.97
Decile spread	449	4.43	4.58	0.97	0.92	-8.5	0.41 (4.69)	-0.05
Quintile spread (main)	899	4.70	3.70	1.27	1.34	-7.1	0.39 (5.62)	-0.03
Rank-weighted, all names	2248	3.34	3.12	1.07	1.03	-8.2	0.28 (4.81)	-0.03

Notes: All four use the identical DISC score, signing rule, smoothing, neutralisation, holding period, rebalancing overlay and cost model. The long-only line carries one dollar of gross exposure; the three spreads carry two. Alpha is the six-factor intercept with Newey-West t -statistics at six lags.

Three points follow. First, the signal is not a short-selling artefact: held long-only in the top quintile it earns an alpha of 0.11% per month ($t = 2.41$) and beats the equally weighted universe by 2.50% per year. But it carries a market beta of 0.97, a volatility of 17.7% and a 27% drawdown, so its Sharpe ratio is 0.75 — the alpha is real and the risk around it is almost entirely uncompensated.

Second, concentration raises alpha per dollar and lowers the Sharpe ratio. The decile spread earns the largest alpha in the table, 0.41% per month, on half as many names as the quintile spread; its volatility is 24% higher and its Sharpe ratio falls from 1.27 to 0.97. Sorting harder finds more signal per position and more idiosyncratic risk per position, and beyond the quintile the second effect dominates.

This has a direct implication for what an implementable version looks like. Per dollar of gross exposure the long-only book earns 0.11% per month against 0.20% for the quintile spread, so roughly half the alpha per unit of capital survives in a book that requires no stock loan at all. Hedged against the equally weighted universe it earns 2.50% per year with an alpha of 0.22% ($t = 4.65$). Given that the short leg consists of exactly the population that is hard to borrow (Section 7), a long book hedged with a liquid index proxy is the more plausible large-scale implementation, and the literal 899-name long-short portfolio is better read as the cleanest measurement of the effect than as a portfolio one would run at size.

Third, spreading weight across the entire cross-section goes too far in the other direction. The rank-weighted book holds 2,248 names, achieves the lowest volatility in the table at 3.12%, and dilutes the alpha to 0.29%; its Sharpe ratio of 1.14 is below the quintile spread's. The quintile cut is where the two effects balance, which is what the shape of Figure 2 already suggested: the relationship is monotone across the middle of the distribution, so the extra positions a rank-weighted book takes in the middle deciles add turnover and variance without adding expected return.

5.4 Level, not innovation

Table 5 is the paper's central result. Each of the three characteristics is rebuilt in four ways and pushed through an identical portfolio pipeline. The level is the cross-sectional rank of the raw characteristic. The cross-sectional residual removes size within each month. The within-firm deviation subtracts the firm's own trailing 24-month mean — the natural construction if the informative quantity is the surprise in filing behaviour rather than its level. The two-way residual applies both.

Table 5: The level prices; the innovation does not

	Gross		Net				Turn.
	Sharpe	Alpha	Sharpe	In-s.	Held-out	Alpha	(%)
Level	1.56	0.48 (6.92)	1.27	1.29	1.34	0.39 (5.62)	216
Cross-sectional residual	1.26	0.32 (5.76)	0.93	0.74	1.08	0.23 (4.20)	217
Within-firm deviation	0.34	0.08 (1.56)	-0.04	-0.01	-0.07	-0.01 (-0.17)	262
Two-way residual	0.16	0.04 (0.69)	-0.18	-0.16	-0.20	-0.05 (-0.88)	262

Notes: Alphas are monthly percentages against the six-factor model with Newey-West t -statistics in parentheses. All four versions use the same signing rule, the same smoothing, the same neutralisation, the same quintile cut and the same cost model; only the input characteristic differs. The portfolio is already neutralised against size, so the cross-sectional-residual row removes size twice and is reported for completeness; the informative contrast is the first row against the third.

The within-firm deviation produces a gross alpha of 0.08% per month with $t = 1.56$ and an held-out Sharpe ratio of 0.01. It is not that the innovation carries a signal that trading costs consume: its turnover is higher, at 262% against 216%, but the cost differential is 0.09% per year and the gap in Table 5 is in the gross numbers. The innovation simply carries little information at this horizon.

Figure 3 plots the four cumulative return series; the level and the deviation are not close. The single-characteristic version of the same contrast is in Table 7. For **q_lag** the level earns a gross alpha of 0.41% ($t = 5.17$) against 0.10% ($t = 1.41$) for the deviation; for **burst**, 0.24% ($t = 3.41$) against 0.04% ($t = 0.60$). The exception is **nt_12m**, for which the two are indistinguishable and both are insignificant.

Table 6 takes this further and asks what each characteristic contributes to the composite. **nt_12m** traded on its own loses money: a Sharpe ratio of -0.16 and an alpha of -0.04% per month ($t = -0.27$). Its entire contribution is diversification, and that contribution is small — dropping it lowers the Sharpe ratio from 1.27 to 1.19 and the held-out Sharpe ratio from 1.34 to 1.31, and the two return series correlate 0.98. We retain it because the three-characteristic set is the in-sample optimum (1.29 against 1.11) and was fixed before the held-out period was opened; dropping it now, after observing that it is weak in ancillary tables, would itself be a specification chosen with hindsight. Nothing in the paper’s argument turns on the choice. The two-characteristic composite delivers the same decomposition: the level earns a gross alpha of 0.44% per month ($t = 6.62$) and a net Sharpe ratio of 1.19, while the within-firm deviation earns 0.07% ($t = 1.31$) gross and a net Sharpe ratio of -0.13 .

Table 6 also carries a methodological warning. The information coefficient ranks these specifications almost backwards. The two highest-IC rows, **q_lag** alone and **q_lag + nt_12m**, at 0.037 and 0.037, have the lowest Sharpe ratios of any specification containing **q_lag**, at 0.81 and 0.86. **burst** has the lowest IC in the table at 0.003, yet adding it to **q_lag** raises the Sharpe ratio from 0.81 to 1.19 and cuts volatility from 5.64% to 3.71%. The IC measures how accurately a signal orders the cross-section; it says nothing about whether a characteristic reduces the variance of the resulting portfolio. Selecting components on IC alone would have discarded the one that contributes most.

5.5 Filing discipline is persistent

Table 8 reports rank autocorrelations and a between-firm variance share. Because all three characteristics use 365-day windows or a most-recent-filing value, a twelve-month lag involves essentially non-overlapping information, so the autocorrelation at that horizon is not mechanical.

Table 6: What each characteristic contributes

Specification	IC	Return (% p.a.)	Vol. (%)	Sharpe	In-s.	Held-out	Alpha (%/mo)
q_lag + nt_12m + burst (main)	0.029	4.70	3.70	1.27	1.29	1.34	0.39 (5.62)
q_lag + burst	0.028	4.41	3.71	1.19	1.11	1.31	0.35 (5.29)
q_lag + nt_12m	0.037	4.64	5.43	0.86	1.18	0.74	0.34 (4.45)
nt_12m + burst	0.006	1.90	4.34	0.44	0.46	0.44	0.21 (2.59)
q_lag alone	0.037	4.58	5.64	0.81	0.99	0.77	0.33 (4.18)
burst alone	0.003	1.20	4.19	0.29	0.23	0.34	0.15 (2.10)
nt_12m alone	0.008	-0.96	6.15	-0.16	-0.22	-0.09	-0.04 (-0.27)

Notes: Each row rebuilds the composite from the listed characteristics and runs the identical pipeline. IC is the average monthly Spearman correlation with next-month return.

Table 7: Level versus deviation, one characteristic at a time

	Level Gross alpha (%/mo)	Within-firm deviation Gross alpha (%/mo)
q_lag	0.41 (5.17)	0.10 (1.41)
nt_12m	0.21 (1.50)	0.21 (1.47)
burst	0.24 (3.41)	0.04 (0.60)

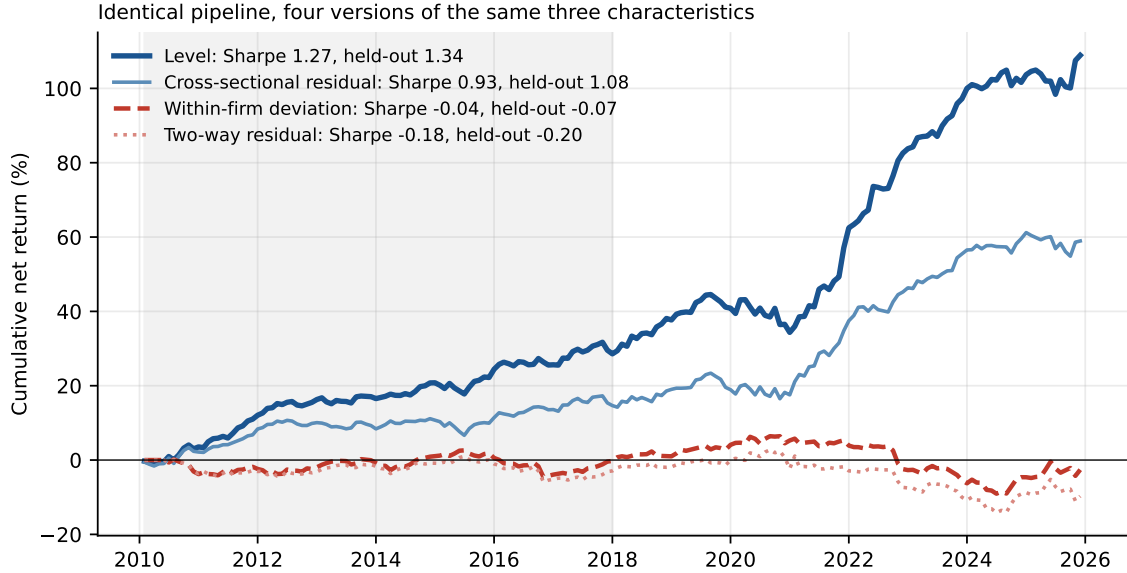


Figure 3: Level versus innovation: identical pipeline, four versions of the same three characteristics

Notes: Cumulative net return of the long-short portfolio when the three characteristics enter as levels, as cross-sectional residuals on size, as deviations from each firm's trailing 24-month mean, and as two-way residuals. Every other choice — signing, smoothing, neutralisation, quintile cut, holding period, cost model — is held fixed. The shaded region is the in-sample period.

Table 8: Persistence of filing discipline

	Rank autocorrelation at lag (months)					Between-firm variance share
	1	12	36	60	120	
DISC	0.975	0.549	0.422	0.373	0.291	0.475
q_lag	0.942	0.733	0.598	0.547	0.428	0.585
nt_12m	0.932	0.215	0.064	0.068	0.017	0.325
burst	0.922	0.339	0.293	0.260	0.208	0.337
Log size (benchmark)	0.996	0.953	0.886	0.842	0.754	0.862
Five-day reversal (benchmark)	0.000	-0.001	-0.012	0.005	0.009	0.018

DISC sits between the two benchmarks and much closer to size than to reversal: 47% of its cross-sectional variation is a permanent firm attribute and its ten-year autocorrelation is 0.29. Quintile transitions tell the same story. One year ahead the diagonal of the transition matrix averages 36% against a random benchmark of 20%, and five years ahead 30%; the probability of moving from the worst quintile to the best is 4% at one year and 10% at five. The matrix is asymmetric: five years on, 47% of the best-disciplined quintile remains in place against 28% of the worst. Good filing discipline is a durable organisational capability; poor discipline is more transient, because firms either fix it or cease to exist.

We should be careful not to present Table 8 and Table 5 as independent evidence. Taking within-firm deviations removes the between-firm variation by construction, and 47% of DISC’s variance is between-firm. The two tables are two views of the same fact. Their joint economic content is that the priced object is a stable difference across firms rather than a firm’s own time-series movement.

The exception again is `nt_12m`, whose autocorrelation falls to 0.22 at twelve months and 0.02 at ten years. It is an event indicator, not a trait, and it should not be described as one. The three characteristics operate on three different time scales: a standing capability, a medium-term habit, and an occasional accident.

5.6 Where the effect is priced

Table 9 widens the universe. The information coefficient rises monotonically from 0.021 among large liquid names to 0.061 across all common stock, and the Sharpe ratio from 0.81 to 1.98. This is consistent with Merton (1987): the marginal value of knowing that a firm cannot file on time should be largest where the information environment is thinnest. It is also the single most common pattern in the anomaly literature, and on its own it does not discriminate between an information-environment mechanism and small-stock noise.

5.7 Cross-sectional regressions

Table 10 reports Fama-MacBeth regressions on percentile ranks. The coefficient on DISC falls from 0.76% to 0.60% as controls are added but its t -statistic rises from 3.69 to 5.89: the controls absorb variance rather than signal. In the full specification only momentum carries an independent coefficient of any significance.

Table 9: Predictive power rises as the universe widens

Universe	Names per mo.	IC	Return (% p.a.)	Vol. (%)	Sharpe	In-s.	Held-out	Alpha (%/mo)
All common stock	3588	0.061	9.40	5.44	1.73	1.63	1.92	0.73 (5.93)
Price > \$1	3415	0.052	9.03	4.57	1.98	1.84	2.22	0.70 (7.09)
\$50M / \$0.5M	2518	0.033	5.23	3.93	1.33	1.43	1.38	0.41 (5.23)
\$100M / \$1M	2253	0.029	4.70	3.70	1.27	1.29	1.34	0.39 (5.62)
\$250M / \$2M	1947	0.024	3.94	3.68	1.07	1.10	1.11	0.31 (4.92)
\$1B / \$10M	1273	0.021	3.03	3.76	0.81	0.88	0.78	0.21 (3.66)

Notes: Figure 4 plots the same information. Universes are defined by market capitalisation, dollar volume and, for the tighter screens, a minimum price. IC is the average monthly Spearman correlation between DISC and next-month return. The \$1 price screen dominates the fully unscreened universe because sub-dollar stocks carry bid-ask bounce.

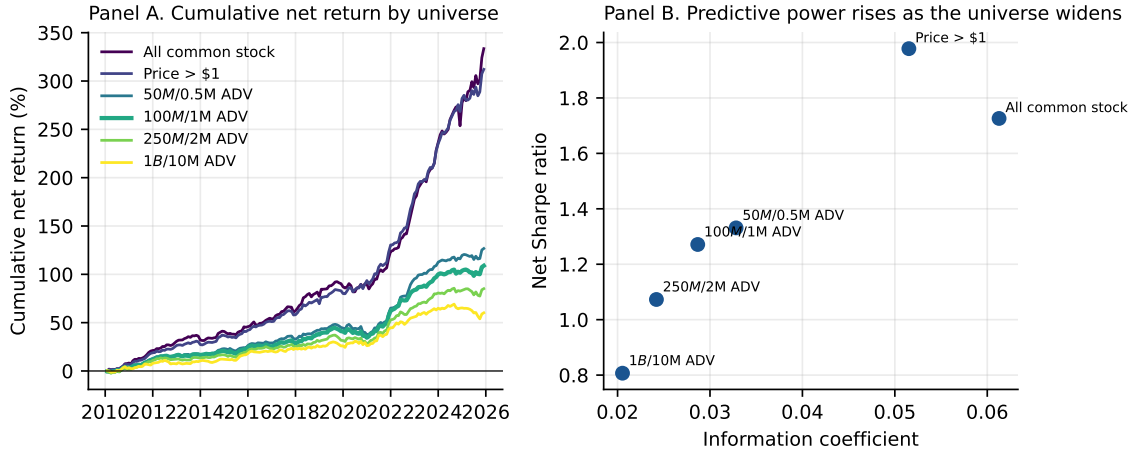


Figure 4: Cumulative net return by universe, and predictive power against universe breadth

Notes: Panel A plots the cumulative net return of the long-short portfolio in each of the six universes of Table 9; the main specification is drawn in bold. Panel B plots the net Sharpe ratio against the information coefficient across the same six universes.

Table 10: Fama-MacBeth regressions of next-month return on DISC

	(1)	(2)	(3)	(4)	(5)
DISC	0.764 (3.69)	0.693 (4.57)	0.647 (4.42)	0.607 (5.74)	0.603 (5.89)
Log size		0.192 (0.58)	0.087 (0.28)	-0.067 (-0.29)	-1.027 (-1.38)
Momentum			0.509 (2.01)	0.644 (2.49)	0.631 (2.44)
Reversal			-0.129 (-0.47)	-0.089 (-0.41)	-0.047 (-0.22)
Volatility				-0.332 (-0.56)	-0.288 (-0.49)
MAX				0.004 (0.03)	0.094 (0.61)
Log Amihud					-1.059 (-1.37)
Turnover					-0.162 (-0.63)
Months	191	191	191	191	191

Notes: Monthly cross-sectional regressions of next-month return on percentile ranks, coefficients in monthly percent, Newey and West (1987) t -statistics at six lags in parentheses. Amihud illiquidity follows Amihud (2002); MAX follows Bali et al. (2011). The coefficient on log size moves from +0.19 in column (2) to -1.03 in column (5); log size and log Amihud illiquidity are close to collinear in the cross-section, so their individual loadings are not separately identified. Neither is significant in any specification.

5.8 Is it just quality?

Firms that file late, file notices of late filing and disclose erratically are plausibly just low-quality firms. The six-factor model in Table 3 controls for *factor exposures* — RMW and CMA — which is not the same as controlling for the *characteristics* themselves. We therefore build accounting characteristics from the SEC Financial Statement Data Sets. Because these come from the same EDGAR filings whose timestamps generate our signal, they are point-in-time by construction: each variable enters the panel on the acceptance timestamp of the 10-K that reports it, and is carried forward until the next one. We use return on assets, operating profitability, accruals in the sense of Sloan (1996), leverage, asset growth and book-to-market. All six are non-missing for 42.8% of the panel.

Table 11 reports the result. Adding the accounting characteristics to the full specification of Table 10 leaves the coefficient on DISC essentially unchanged, at +0.603% in the baseline against +0.618% with all six controls. The t -statistic falls from 5.89 to 4.31, but the estimation sample falls at the same time from 414,538 to 183,940 observations, which accounts for most of the decline. In the full specification not one of the six accounting characteristics is individually significant, the largest t -statistic in absolute value being 1.00. The cross-sectional rank correlation between DISC and an equal-weighted composite of the five quality characteristics is 0.154. Filing discipline is not a repackaging of accounting quality.

The concentration result is different, and we state it plainly. Sorting first into terciles on the quality composite and then into DISC quintiles within each tercile, the spread is +0.806% per month ($t = 3.97$) among the lowest-quality third, +0.467% ($t = 2.45$) in the middle, and +0.197% ($t = 1.58$) — insignificant — among the highest-quality third. The effect is not *explained* by quality but it is *concentrated* where quality is low. This is the same message as the volatility sort in Table 12: junk firms carry high idiosyncratic risk, and the two conditioning variables are not independent.

Running the full six-sleeve strategy on the subsample for which all six accounting variables exist gives a Sharpe ratio of 0.82, with an alpha of 0.37% per month ($t = 4.31$). Most of the decline from the headline 1.27 is a universe effect rather than a quality effect: the quality-complete subsample averages 1012 names a month rather than 2,253, and Table 9 already shows a Sharpe ratio of 0.81 in a universe of comparable size. Residualising DISC on the accounting ranks each month before forming portfolios costs a further 0.26 in Sharpe ratio; the alpha falls to 0.30% ($t = 3.52$) but remains significant, while the held-out Sharpe ratio of 0.34 is the weakest number in the paper. We report it rather than omit it.

5.9 Conditioning variables

Table 12 sorts into terciles on a conditioning variable and reports the DISC quintile spread within each tercile. Three patterns are worth separating.

First, the spread declines in size and dollar volume, and among the most heavily traded tercile it is only marginally significant ($t = 1.83$). We can say that the effect does not vanish among large capitalisations ($t = 3.03$), but not that it is comfortably present among the most liquid names.

Second, the spread *rises* in turnover, from 0.55% among the least traded tercile to 0.72% among the most traded. A pure investor-neglect account predicts the opposite: neglected stocks should show the largest mispricing. The gradient is modest and the high tercile carries the weakest t -statistic of the three, so we treat this as inconsistent with a pure neglect story rather than as a rejection of one.

Third, and most strongly, the spread rises sharply in volatility, from an insignificant 0.16% to 1.16% ($t = 4.42$). This is the classic signature of limits to arbitrage (Shleifer and Vishny, 1997;

Table 11: Does accounting quality explain DISC?

	(1) Baseline (Table 10 col. 5)	(2) + profit- ability	(3) + accruals, leverage	(4) + investment, value
DISC	+0.603 (5.89)	+0.509 (4.56)	+0.517 (4.04)	+0.618 (4.31)
Return on assets		+0.061	+0.246	+0.152
Operating profitability		−0.003	−0.174	−0.150
Accruals			−0.103	+0.031
Leverage			+0.129	+0.031
Asset growth				−0.021
Book-to-market				−0.129
Months	191	190	190	178
Observations	414,538	293,703	203,051	183,940
<i>DISC quintile spread within quality terciles (%/month)</i>				
Low quality (junk)		+0.806	($t = 3.97$)	
Middle		+0.467	($t = 2.45$)	
High quality		+0.197	($t = 1.58$)	

Notes: Monthly cross-sectional regressions of next-month return on percentile ranks, coefficients in monthly percent, Newey and West (1987) t -statistics at six lags. Column (1) repeats the full specification of Table 10; the controls it contains — log size, momentum, reversal, volatility, MAX, Amihud illiquidity and turnover — are included in every column but not reported. Accounting variables are built from the SEC Financial Statement Data Sets, taken from the most recent 10-K whose acceptance timestamp precedes the month end. The quality composite equally weights the percentile ranks of return on assets, operating profitability, and the negatives of accruals, leverage and asset growth.

Table 12: Double sorts: DISC quintile spread within terciles

Conditioning variable	Low	Middle	High
Market capitalisation	0.93 (4.42)	0.51 (2.72)	0.32 (3.03)
Dollar volume	0.91 (5.64)	0.65 (3.00)	0.32 (1.83)
Turnover	0.55 (3.52)	0.61 (4.62)	0.72 (2.80)
Volatility	0.16 (1.58)	0.33 (3.67)	1.16 (4.42)

Notes: Monthly percent, Newey-West t -statistics at six lags.

Pontiff, 2006): idiosyncratic risk deters the arbitrageur who would otherwise close the gap. It is a competing explanation to the information-environment reading of Table 9, it is better identified than that reading, and we report it as such.

5.10 Where in calendar time the return accrues

If the effect were an event-time underreaction, the return should concentrate on information dates. Using earnings announcement dates taken from 8-K item 2.02, the daily long-short return inside a $[-1, +1]$ announcement window is 0.14% against 0.018% on other days, an intensity ratio of 7.9. But announcement windows are 2.1% of trading days and account for only 20% of the monthly spread. Four-fifths of the return accrues on ordinary days on which nothing is disclosed, which is what the gradual re-rating of a standing characteristic looks like and is not what the digestion of a news event looks like.

The announcement date is itself derived from the filing stream that produces **burst**, so signal and event clock would otherwise share timestamps. The figures above use a variant of **burst** that excludes item 2.02 filings entirely, eliminating the overlap; the intensity ratio rises from 7.0 to 7.9 when the overlap is removed. The full strategy built on that variant earns a Sharpe ratio of 0.94 with an alpha of 0.33% ($t = 4.00$) and correlates 0.87 with the headline series, so the main result does not depend on the shared timestamps either.

We describe this evidence as consistent with a gradual correction. It does not by itself reject a risk-based account, since risk premia can also be earned around information events.

5.11 Time series

Table 13 splits the sample. The information coefficient is higher after 2018 than before, so the signal has not decayed. What varies is the realisation of the short leg: its excess over the universe falls from 4.44% per year in 2018–2022 to 1.87% in 2023–2025, while the universe return itself doubles. Shorting poorly disciplined firms does not pay in a strong bull market. The single negative year in the sample, 2020 (−3.0%), has the same character.

Table 13: Sub-period evidence

	IC	t	Quintile spread (%/mo)	Long side (% p.a.)	Short side (% p.a.)	Universe (% p.a.)	Sharpe
2010–2017	0.021	5.25	0.42	1.34	2.82	14.87	1.36
2018–2022	0.039	4.64	0.84	3.78	4.44	7.02	1.45
2023–2025	0.033	3.72	0.78	3.45	1.87	13.65	1.31

Notes: Long and short side are the excess of each basket over the equally weighted universe.

5.12 Robustness

Table 14 collects the diagnostics. A value-weighted market proxy loads 1.004 on the market factor with $R^2 = 0.994$ at zero lag and collapses to $R^2 \approx 0.02$ at ± 1 month, confirming factor alignment. Delaying the signal by one to four months decays the Sharpe ratio smoothly from 1.16 to 0.92 with no cliff, which is the signature of a genuine signal rather than a look-ahead artefact. Winsorising returns at the 1st and 99th percentiles *raises* the Sharpe ratio to 1.31, so extreme returns work against the strategy rather than driving it. Excluding the best single month leaves 1.26. Restricting to primary share classes leaves 1.27. A moving-block bootstrap gives a 95% interval for the Sharpe

ratio of $[0.76, 1.80]$. The held-out Sharpe ratio of 1.34 exceeds the in-sample 1.29, which should not be read as the strategy strengthening after the protocol was frozen. The excess is concentrated in two years: excluding 2021 and 2022 leaves a held-out Sharpe ratio of 0.83 on an annualised return of 3.19%, against 3.31% in sample. What changed in the held-out period is the dispersion environment in which the short leg realises, not the strength of the signal — the information coefficient is 0.021, 0.039 and 0.033 across the three sub-periods of Table 13.

The dispersion across the six rebalancing schedules is 1.01 to 1.25 with a mean of 1.15. The reported overlay achieves 1.27, above every individual schedule, because six imperfectly correlated sleeves diversify; this is a property of the combined portfolio, which is the portfolio an implementer would hold.

Against the multiple-testing thresholds of Harvey et al. (2016), a t -statistic of 5.62 clears the authors’ suggested threshold of 3.0 and a Bonferroni 5% critical value of 3.73. We set $M = 260$ because the search behind this paper screened roughly 130 candidate characteristics built from filing metadata, each admitting two signs. That count covers the search within this study; it does not cover hypotheses entertained in unrelated projects, and no multiple-testing correction does.

Table 14: Robustness diagnostics

Diagnostic	Result
Factor alignment, market beta at lag 0 / ± 1	1.004 ($R^2 = 0.994$) / $R^2 \approx 0.02$
Signal delayed 1 / 2 / 3 / 4 months (Sharpe)	1.16 / 1.03 / 0.94 / 0.92
Rebalancing schedule, six phases (Sharpe)	1.01–1.25, mean 1.15
Returns winsorised at 1% / 99% (Sharpe)	1.31
Excluding best / worst month (Sharpe)	1.26 / 1.37
Primary share class only (Sharpe)	1.27
Delisting returns restored (Sharpe)	unchanged at 1.27
burst excluding item 2.02 (Sharpe, alpha)	0.94, 0.33 (4.00)
Moving-block bootstrap, 95% interval for Sharpe	$[0.76, 1.80]$
Bonferroni 5% critical t at $M = 260$	3.73 vs. realised 5.62

Table 15 varies each cost constant while holding the other at its baseline. The strategy’s exposure to the two charges is mechanical: turnover and short exposure are fixed by construction, so a change in either constant shifts every monthly return by the same amount and leaves volatility unchanged. At the borrow fee the reconstruction supports, 0.5–0.7%, the Sharpe ratio spans 1.30 to 1.24. The strategy survives a borrow fee of 2% per year, more than three times our assumption and above the 1% threshold that D’Avolio (2002) finds 91% of stocks below, at a Sharpe ratio of 0.89; it does not survive a fee near the 4.3% average of his special tranche. On the trading side it survives the 57bp upper bound of Novy-Marx and Velikov (2016)’s effective-spread range at a Sharpe ratio of 0.72, but that figure is an effective spread rather than a realised cost, and applying it to a strategy that trades 216% of capital a year is a deliberately punitive test.

6 Discussion

The evidence supports a narrow claim and does not support a broader one that is tempting to make.

The narrow claim is that the object priced in the monthly cross-section is the level of a firm’s filing discipline, not its innovation. Table 5 shows that the level survives in the held-out period

Table 15: Sensitivity to the two cost constants

	Return (%)	Sharpe	Six-factor alpha	
	p.a.		%/mo	<i>t</i>
<i>Panel A: annual borrow fee (trading cost held at 10 bp)</i>				
0.0%	5.30	1.43	0.44	6.34
0.5%	4.80	1.30	0.40	5.74
0.6% (baseline)	4.70	1.27	0.39	5.62
0.7%	4.60	1.24	0.38	5.50
1.0%	4.30	1.16	0.36	5.14
2.0%	3.30	0.89	0.27	3.94
3.0%	2.30	0.62	0.19	2.73
<i>Panel B: one-way trading cost (borrow fee held at 0.6%)</i>				
5 bp	4.92	1.33	0.41	5.88
10 bp (baseline)	4.70	1.27	0.39	5.62
20 bp	4.27	1.15	0.35	5.10
57 bp	2.68	0.72	0.22	3.18

Long-short quintile spread, 2010–2025, 191 months. Both charges enter linearly: the borrow fee applies to one dollar of short exposure every month and the trading charge to 216% annual turnover across two legs, so each row shifts the monthly return by a constant and volatility is 3.70% throughout. Alpha is against the five-factor model plus momentum; *t*-statistics use Newey–West standard errors with six lags.

and the deviation does not, gross or net. Table 8 shows that the level is a persistent firm attribute rather than a sequence of independent draws. The announcement-window decomposition shows that four-fifths of the return accrues on days when nothing is disclosed. Taken together these describe a standing, freely observable characteristic that is slowly re-rated, rather than a news event that is underreacted to.

That reading has bite relative to the existing literature, which is organised around events: the Form NT filing, the announcement window, the departure of an earnings date from its expected slot. Those studies are about a different object, measured at a different horizon. Our characteristics are constructed from the same filings and are informative about the same firms, but the quantity that prices in the monthly cross-section is not the quantity that literature measures.

The broader claim we decline to make is that this explains why a forty-four-year-old idea still earns a positive alpha. It would be natural to argue that the event component was arbitrated away while the level component was never treated as tradeable information because it never constitutes news. We cannot test it. Our sample begins in 2010, so a publication-date design in the manner of McLean and Pontiff (2016) is unavailable; and the evidence is equally consistent with the innovation never having had monthly cross-sectional content in the first place, in which case the announcement-window literature is studying a short-horizon phenomenon that may well remain alive. We report the decomposition and leave the persistence question open.

Two further caveats bear on interpretation. First, Table 8 and Table 5 are not independent evidence: taking within-firm deviations removes between-firm variation by construction, and that is where 47% of DISC’s variance lives. Second, the strongest pattern in Table 12 is in volatility, not in size or turnover, and it points to limits to arbitrage rather than to information environment. An account in which idiosyncratic risk deters arbitrage fits our cross-sectional evidence at least as well

as an account based on Merton (1987), and it is better identified.

7 Limitations

The sample begins in January 2010, which is when EDGAR metadata coverage becomes complete enough to compute trailing-year filing statistics for the full cross-section. This precludes a publication-date test and limits the sample to 191 months.

The strategy is essentially unlevered, with a volatility of 3.70%. Reaching an institutionally meaningful return requires two to three times gross exposure, and gross exposure scales the borrow requirement on the short leg. That leg consists of firms that file late, file notifications of late filing, and disclose erratically, and it is natural to suppose it is also the population that is expensive to borrow. The data support that supposition only weakly. Matching the portfolio to FINRA semi-monthly short interest, which begins in December 2017, the short leg carries a median short interest of 3.9% of shares outstanding against 3.4% for the universe, and 3.9% of its stock-months sit above 20% short interest against 2.6% for the universe. The leg is tilted towards crowded shorts but sits nowhere near the special-borrow tail, which is why we take a general-collateral fee as the baseline (Table 15). The binding constraint on scale is therefore the volatility of the unlevered book and the capacity of the small and volatile names it holds, rather than borrow availability.

`nt_12m` is the weakest element of the design. It contributes a third of the composite, loses money when traded on its own (Table 6), and is an event indicator rather than a trait (Table 8); it is also the one characteristic for which the level and the deviation cannot be told apart. It survives in the specification because the three-characteristic set was the in-sample optimum and was frozen before the held-out period was opened, and it adds 0.08 to the Sharpe ratio through diversification. A reader who prefers the two-characteristic composite loses 0.08 in Sharpe and 0.03 held out, and every conclusion in the paper is unchanged.

Our measure of the innovation is coarse, and the decomposition should be read with that in mind. We take the deviation of a characteristic from the firm’s own trailing twenty-four-month mean. That is a within-firm deviation, not an unexpected component in any structural sense: it inherits the arbitrary window, the estimation error in the mean, the low signal-to-noise ratio that a highly persistent series implies, and mean reversion. It also mismeasures the timing. `q_lag` updates only when a 10-Q is accepted and is carried forward in between, so in most months the deviation we compute reflects no new filing behaviour at all. The sharper test — predicting each filing’s lag from the firm’s own history, its filer status, the fiscal quarter and the statutory deadline, and asking whether the residual surprise forecasts returns in the months after that filing — is one we do not run here. What we can say is that a between-firm ranking of filing behaviour forecasts returns and a within-firm deviation from a trailing mean does not; whether a properly specified filing surprise would be open.

The effect is concentrated in low-quality firms. Table 11 shows that accounting characteristics do not explain DISC — its coefficient is unchanged when six of them are added — but that the quintile spread falls monotonically from 0.81% per month among the lowest-quality third of the cross-section to 0.20% ($t = 1.58$) among the highest. Filing discipline is not a proxy for quality, but an investor restricted to high-quality names would not have been able to trade it.

Finally, the results are concentrated in smaller and more volatile stocks. Table 9 and Table 12 both make this explicit, and any implementation at scale would face the capacity constraint that implies.

8 Conclusion

What is priced in filing behaviour is its level. A composite of three characteristics built entirely from SEC filing metadata — quarterly report lag, late-filing notifications, and the irregularity of 8-K arrivals — earns a net Sharpe ratio of 1.27 with a six-factor alpha of 0.39% per month ($t = 5.62$) and a market beta of -0.03 over 2010–2025, in a universe of roughly 2,250 US stocks. The three characteristics are nearly uncorrelated with one another: the three measures capture largely distinct dimensions of filing behaviour, two of which price on their own.

The result we regard as informative is the decomposition. Rebuilt as deviations from each firm’s own history, the identical characteristics produce a gross alpha of 0.08% per month with $t = 1.56$ and nothing held out. What is priced is the level — a persistent, freely observable attribute of the firm — rather than the innovation. Since the existing literature on filing delay is organised around announcement events, this is a statement about which object prices, not an explanation of why the effect has survived.

The natural next step is to test the level-versus-innovation distinction on other administrative characteristics that are observable at no cost and change slowly, and to obtain an announcement-date source independent of EDGAR so that the event-window decomposition can be run without any shared timestamps.

References

- Alford, A. W., Jones, J. J., and Zmijewski, M. E. (1994). Extensions and violations of the statutory SEC form 10-k filing requirements. *Journal of Accounting and Economics*, 17(1–2):229–254.
- Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.
- Bagnoli, M., Kross, W., and Watts, S. G. (2002). The information in management’s expected earnings report date: A day late, a penny short. *Journal of Accounting Research*, 40(5):1275–1296.
- Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2):427–446.
- Bannouh, K., Geng, D., and Peeters, B. (2021). Filing, fast and slow: Reporting lag and stock returns. *Working paper, SSRN 3432446*.
- Brooks, L. L. Z., Cheng, Y., Liu, L., and Yu, M. D. (2024). The timeliness of 10-k filings, financial performance, and stock returns. *Journal of Corporate Accounting & Finance*, 35(1):277–307.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.
- Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, 11(2):207–264.
- D’Avolio, G. (2002). The market for borrowing stock. *Journal of Financial Economics*, 66(2–3):271–306.
- deHaan, E., Shevlin, T., and Thornock, J. (2015). Market (in)attention and the strategic scheduling and timing of earnings announcements. *Journal of Accounting and Economics*, 60(1):36–55.

- Doyle, J. T. and Magilke, M. J. (2009). The timing of earnings announcements: An examination of the strategic disclosure hypothesis. *The Accounting Review*, 84(1):157–182.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.
- Frazzini, A., Israel, R., and Moskowitz, T. J. (2018). Trading costs. *SSRN Working Paper 3229719*.
- Givoly, D. and Palmon, D. (1982). Timeliness of annual earnings announcements: Some empirical evidence. *The Accounting Review*, 57(3):486–508.
- Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.
- Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5):2019–2133.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.
- Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65.
- McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.
- Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3):483–510.
- Muravyev, D., Pearson, N. D., and Pollet, J. M. (2025). Anomalies and their short-sale costs. *The Journal of Finance*, 80(5):2519–2566.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.
- Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147.
- Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, 42(1–2):35–52.
- Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1):35–55.
- Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, 71(3):289–315.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.
- Zhao, X. (2017). Does information intensity matter for stock returns? evidence from form 8-k filings. *Management Science*, 63(5):1382–1404.

A AI-Usage Statement

In preparing this manuscript the author used Claude (Anthropic) to write and refactor the data pipeline, to propose diagnostic and robustness tests, and to draft the text. The author directed the research — the question, the choice of characteristics, the frozen in-sample protocol, the universe definition and the cost assumptions — and reviewed, corrected and revised the generated code and prose before submission.

Every number reported here is produced by the reproduction notebook described in Appendix B and is checked against that notebook’s saved output by a script that fails if any figure in the manuscript and the corresponding computed value disagree. The author has reviewed the content of the submitted manuscript and takes full responsibility for it.

B Reproducibility

Every table and figure is generated by `paper_reproduction.ipynb`, which runs end to end in approximately seven minutes from five primary inputs: the EDGAR filing metadata, CRSP daily files, the CRSP names and delisting files, the CIK–PERMNO link, and the Fama-French monthly factors. The three characteristics are constructed from EDGAR within the notebook rather than read from intermediate files. All tunable parameters — the in-sample cutoff, the universe thresholds, the smoothing window, the neutralisation set, the quintile fraction, the holding period and the cost model — appear in a single configuration cell.

The notebook ends with assertions on the headline numbers, on the alignment of signals and returns, on the identity between the two legs and their sum, on the equality of the level row of Table 5 with the headline specification, and on the delisting correction. If any number in this paper ceases to match the code, the notebook fails rather than producing a revised number silently.