Rewriting CRSP's History: Impact of Altered Monthly Returns on Asset Pricing

Patrick Schwarz [*] Dominik Walter [‡] Patrick Weiss [†]

February 24, 2026

Abstract

In January 2025, CRSP discontinued the existing stock tape used in many published papers. This transition rewrites 9.62% of monthly returns by more than 1 basis point, primarily due to a change in the dividend reinvestment assumption. Analyzing the impact for a comprehensive set of premia in several thousand sorting specifications reveals that, on average, 11.43% of all monthly long-short returns differ by more than 10 basis points — especially in early periods, NBER recessions, and return-based sorts. Reassuringly, average premia and their significance remain largely unaffected, suggesting CRSP changes mainly introduce unsystematic variation without altering key asset pricing conclusions.

Attachments: Internet Appendix (46 pages)

[*] University of Liège HEC Liège, rue Louvrex 14, 4000 Liège, Belgium. E-mail: patrick.schwarz@uliege.be

[‡] University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany. E-mail: d.walter@uni-konstanz.de

[†] Reykjavik University, Menntavegur 1, 102 Reykjavik, Iceland. E-mail: patrickw@ru.is

We thank the editor Hendrik Bessembinder and an anonymous referee for their helpful comments and constructive suggestions. Moreover, we thank (in alphabetical order) Jules van Binsbergen, Ralf Brüggemann, the Center for Research in Security Prices, Freda Song Drechsler, Marcel Fischer, Heiko Jacobs, Monia Magnani (discussant), Ian Marsh, Alan Myers, Alexander Pasler, Otto Randl, Mihail Velikov, Christopher Walsh, Tom Zimmermann, and Taras Zlupko as well as participants at the 41st International Conference of the French Finance Association (AFFI Best Paper Award), VGSF 20th Anniversary Conference, AWG 2025, and seminar participants at the University of Konstanz, for their comments, insights, and discussions. We also thank the University of Konstanz for supporting open access for this article. Dominik Walter and Patrick Weiss gratefully acknowledge financial support from the Austrian Science Fund (grant number DOC 23-G16 and 10.55776/DOC202), since this project's empirical analysis and data collection were conducted at the Vienna Graduate School of Finance (WU Vienna).

I. Introduction

An important notice by CRSP makes users aware that the release of the Flat File Format 1.0 (SIZ) in January 2025 was effectively the last release of the "old" CRSP tape. From 2025 onward, CRSP will only update the "new" CRSP tape, that is, the Flat File Format 2.0 (CIZ), which was first released in 2022. Although the observations for most data items match exactly between both tapes, we find one striking exception: Monthly holding period returns for stocks, which CRSP describes as follows:

> *There is no direct equivalent to SFZ_MTH's [SIZ] holding period returns, but SFZ_MTH's [SIZ] Mret and Mretx data items map most closely to CIZ MthRet and MthRetx, respectively. The reason for this is that the calculation of legacy [SIZ] Mret and Mretx uses different assumptions than the calculation of MthRet and MthRetx [CIZ]. For example, MRet is a month-to-month holding period return with dividends reinvested at month-end. MthRet is a compound daily return with dividends reinvested on their ex-dates.*

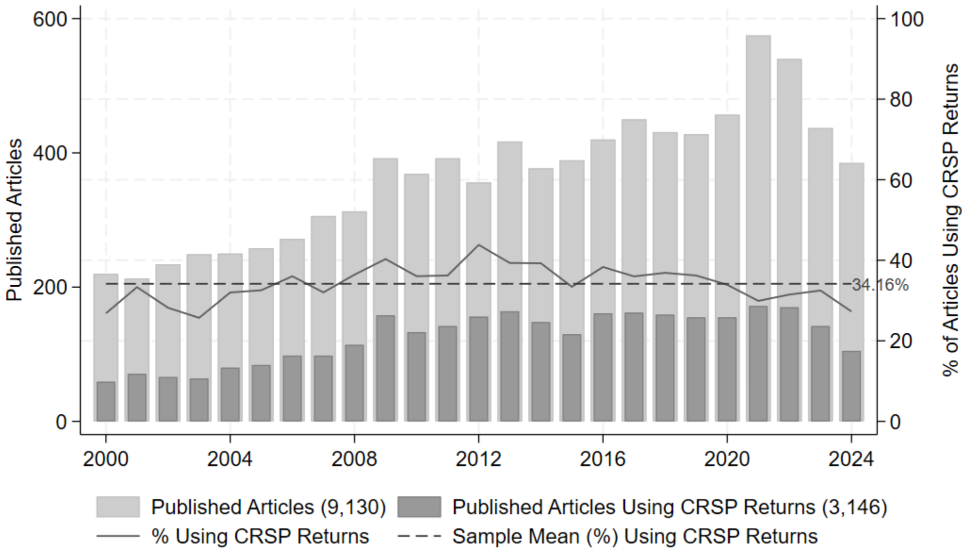
Thus, CRSP computes monthly stock returns in the new tape by reinvesting payouts on the ex-date as opposed to the end of the ex-date month in the old tape. This shift in reinvestment timing and other minor changes alter 9.6% of all monthly returns by at least one basis point between both tapes. The mean absolute difference for these altered returns is 22 basis points. While data providers such as WRDS point to these changes, it is neither clear nor obvious how these changes affect asset pricing studies that directly use monthly CRSP returns. Moreover, it is unclear whether these CRSP-induced changes lead to systematic differences in monthly returns or, rather, constitute unsystematic variation.

The share of asset pricing studies in top finance journals that might be affected by these changes is substantial, as shown in Figure 1. Based on textual analysis, we find that 34% of all articles published in 5 leading finance journals (Top5) over the last 25 years include a CRSP return-based analysis.[1] Additionally, published open-source code (see, e.g., Chen and Zimmermann (2022), Jensen, Kelly, and Pedersen (2023), Scheuch, Voigt, Weiss, and Frey (2024)) suggests that most researchers likely use monthly stock returns directly, irrespective of the underlying payout reinvestment assumption. Consequently, any study using monthly returns from the new tape uses conceptually different monthly returns than studies conducted with the old monthly CRSP tape.[2] Notably, no article in our sample has so far mentioned which CRSP tape (CIZ or SIZ) has been used.

FIGURE 1

Articles Conducting a CRSP-Based Return Analysis

Figure 1 shows the total number of published articles in 5 leading finance journals (Top5, i.e., JF, JFE, RFS, JFQA, and RoF) along with the absolute and relative number of published articles using CRSP return data. The sample period is from 2000 to 2024. An article is counted as conducting a CRSP-based return analysis if the article contains words (e.g., CRSP, return, portfolio sort, among others as described in Internet Appendix IA.I) indicating that return data from CRSP has been used in an empirical analysis.



In this paper, we examine whether shifting the reinvestment timing from the month-end to the ex-date and other changes between the old and new tape translate into systematic differences in asset pricing studies that rely on monthly returns from CRSP. As reinvesting

[1]Internet Appendix IA.I describes the textual analysis methodology, motivated by Berninger, Kiesel, Schiereck, and Gaar (2021) and Berninger, Kiesel, and Schnitzler (2024), we use to derive a reasonable approximation of the relevance of (monthly) CRSP returns in the academic finance literature.

[2]Even before the new tape's release in 2022, the new monthly returns could have been constructed from daily returns. Thus, the return (item "MthRet") in the new CRSP tape is not an entirely new data point. However, we found no evidence for the necessary compounding of daily returns in the literature.

payouts on the ex-date might not be feasible for all investors, we also assess the effect on asset pricing studies if payouts are reinvested on the effective payment date rather than on the preceding ex-date. While an exhaustive replication of years of research is not feasible, our analysis focuses on a comprehensive set of portfolio sorts, a widely used tool found to be sensitive to methodological changes (Walter, Weber, and Weiss (2024) and Soebhag, Van Vliet, and Verwijmeren (2024)).

The changes in monthly CRSP returns might lead to material differences in portfolio sorts for the following reasons: First, by sorting the cross section of stocks into portfolios based on characteristics, one might end up sorting stocks with high (low) payouts into the long (short) portfolio. For instance, Gormsen and Lazarus (2023) indicate that stocks in the long portfolio of common factors have higher payouts compared to the short portfolio. If these stocks with larger payouts also tend to have systematically more positive or negative return differences, then CRSP-induced changes might systematically alter long-minus-short portfolio returns.[3] Second, changes in monthly returns can also directly affect the construction of sorting variables that depend on monthly returns, e.g., momentum. Thus, CRSP changes can also change the cross-sectional ranking of stocks.

We start by investigating why around 9.6% of monthly returns differ between the old and new CRSP tape. We find that almost all documented differences (97.7%) are due to the reinvestment of payouts on the ex-date in the new tape compared to the end of the ex-date month in the old tape. Intuitively, absolute return differences are exacerbated by higher payout yields and higher absolute returns from the day after the ex-date until the end of the month. Moreover, as the average compounded return from the ex-date to the month-end is, on average, positive, we find that monthly returns in the new CRSP tape

[3] We find that stocks with negative compared to positive return differences between the old and new tape significantly differ in their size, book-to-market ratios, and momentum (see Panel A of Table IA.6).

are, on average, slightly but significantly more positive than in the old tape. Additionally, we observe 2.2% of the return differences for months in which the security has a missing monthly price at the end of the previous month (trading gaps) or was first listed during the month (IPO). Naturally, compounding daily returns *within each month* in the new tape leads to different monthly returns for these months compared to using month-end prices in the old tape, which can be stale in the case of trading gaps. Although infrequent, the mean absolute differences in monthly returns for months with trading gaps and IPOs are large at 7%. Lastly, we find a relatively small share (0.1%) of differences for delisting months. Compared to a standard delisting from Shumway (1997), these differences for delisting months arise because CRSP does not set missing delisting returns to -30% to adjust for a delisting bias. Moreover, CRSP requires the payment date of the delisting amount to be close to the delisting date (around 30 days). Thus, to adjust for a delisting bias mentioned by *Shumway* (1997), it is still necessary to construct monthly returns by compounding daily returns and including specific delisting assumptions. In short, trading gaps, new listings, and delistings can also induce differences in portfolio sorts that go beyond those induced by altering the reinvestment timing.

In the new tape, CRSP reinvests payouts on the ex-date to be consistent with the methodology of common total return indices, where payouts are also reinvested on the ex-date. However, reinvesting payouts on the ex-date might not be realistic for capital-constrained investors, as it requires borrowing on the ex-date until the payment date. Thus, we construct a third version of monthly returns based on the new tape, for which we only shift the reinvestment of payouts from the ex-date (new tape) to the payment date (modified new tape). This third reinvestment assumption rewrites around 14.9% of all monthly returns between the new and modified new tape. However, the mean of these absolute return differences is less than a third in relation to the mean absolute return difference between monthly returns from the old and new tape.

Turning to portfolio sorts, we sort the cross section of stock returns into portfolios based on 68 sorting variables commonly used in the literature (see, e.g., Hou, Xue, and Zhang (2020)). We control for the large variation in methodological choices for portfolio sorts in the literature by constructing for each sorting variable several thousand premia[4] that differ in common methodological choices for portfolio sorts (Walter et al., 2024).[5] Based on this methodological uncertainty procedure, we find that, on average, 11.43% of all monthly long-short portfolio returns differ by more than 10 basis points when constructed with the new instead of the old CRSP tape. The corresponding mean absolute difference is 4 basis points per month. These differences are especially pronounced for return-based sorting variables (i.e., which are directly constructed from monthly returns) and in NBER recessions. For those return-based sorting variables and during NBER recessions, on average, 28.07% and 18.95% of all monthly long-short portfolio returns differ at the 10 bp level. The corresponding mean absolute differences are 9 and 7 basis points per month, respectively.

Furthermore, we investigate whether these sizable monthly portfolio return differences translate into changes of the *time-series averages* of these long-short portfolio returns, i.e., premia estimates. Put differently, we investigate whether premia systematically differ when using the new instead of the old CRSP tape. Surprisingly, this comparison reveals that the average return premia differ by less than a basis point (i.e., 0.34 bp) for all sorting variables and by 0.59 bp for return-based sorting variables, respectively. Even looking

[4]We do not claim that all used sorting variables have a risk-based mechanism. For simplicity, we use the term "premia" for the time-series average of long-short returns between characteristics-sorted portfolios.

[5]For instance, we construct 69,120 estimates for the value premium that differ in the methodological choices for 14 common portfolio construction decisions.

across the many portfolio sorts generated by varying methodological choices suggests only small differences in premia. Our results show that the changes in monthly portfolio returns also have a limited influence on the significance of premia. We find that only 0.52% and 0.54% of all premia change their significance at the 1% and 5% significance level, when switching from the old to the new tape. Thus, reinvesting payouts on the ex-date instead of the end of the ex-date month and other minor changes between the new and old tape do not significantly impact premia. Additionally, we find that shifting the reinvestment timing of payouts from the ex-date (new tape) to the payment date (modified new tape) also does not materially change the size and significance of premia.

With these findings, we provide reassuring evidence that neither the recent CRSP changes nor altering the reinvestment timing of payouts from the ex-date, to the corresponding month-end, or to the payment date systematically alter unconditional asset pricing premia. Moreover, while the differences in monthly returns between the old and new tape can lead to considerable portfolio return differences in every month, we do not find any persistence in the direction of these differences in the time series. Thus, differences in long-short portfolio returns tend to cancel out over sample periods of many decades, leaving the time-series average, i.e., premia, virtually unchanged.

While we show that CRSP changes do not significantly impact premia estimates, we cannot generalize this result to all related asset pricing applications. Motivated by the pronounced differences of monthly long-short portfolio returns in NBER recessions, we also investigate whether the CRSP changes impact the cyclicality of premia, i.e., premia conditional on the business cycle. For the vast majority of sorting variables, the CRSP changes do not alter coefficients from regressing long-short portfolio returns on business cycle indicators, i.e., the NBER recession dummy. However, we do find a few exceptions: Across the various methodological choices for portfolio sorts around 24.3% of all significant NBER regression coefficients for portfolios sorted on the real estate ratio (Tuzel, 2010) change their significance at the 5% level.[6] Notably, for a specific common methodological path motivated by Hou et al. (2020), the t-statistic of the NBER regression coefficient loses statistical significance at the 5% significance level as it drops from 2.00 (based on the old tape) to 1.82 (based on the new tape). Thus, CRSP changes induce larger differences in portfolio returns during NBER recessions that can – for specific sorting variables and construction choices – impact the conclusion about their cyclicality.

Related Literature. We relate to the literature that investigates data issues in commonly used financial databases. For instance, Ljungqvist, Malloy, and Marston (2009), Patton, Ramadorai, and Streatfield (2015), Gillan, Hartzell, Koch, and Starks (2018), and Berg, Fabisik, and Sautner (2021), document that historical data rewriting due to revisions, backfilling, or changes in methodology, is widespread in analyst (I/B/E/S), hedge fund, executive compensation (ExecuComp), and ESG (Refinitiv) data sets, respectively.[7] Specifically, we relate to the literature that focuses on data provided by CRSP. The first studies on CRSP data quality (Rosenberg and Houglet (1974) and Bennin (1980)) study error rates in CRSP and Compustat by comparing data points between data sets.[8] Further studies investigate errors and inconsistencies in specific CRSP variables (e.g., for shares outstanding and industry classification codes, see Courtenay and Keller (1994), Guenther and Rosman (1994), and Kahle and Walkling (1996), respectively). Particularly related to our paper are the studies by Shumway (1997) and Shumway and Warther (1999), which document a delisting bias in CRSP returns. We add to this literature on CRSP data by

[6]Tuzel (2010) shows that expected returns of firms with a high share of buildings are counter-cyclical, i.e., higher during recessions using, among others, the NBER recession definition (see their footnote 25).

[7]For an extensive review of the literature, we refer to Liu (2020).

[8]Notably, Beedles and Simkowitz (1978) show that after correcting the data errors found in Rosenberg and Houglet (1974), some results of McEnally (1974) were significantly affected.

showing that considerable differences in monthly stock returns between the old and new CRSP tape do not materially affect unconditional asset pricing premia.

While these studies focus on errors and biases in CRSP data, we investigate differences between the old (SIZ) and new (CIZ) CRSP tapes that are, to a large extent, due to methodological changes (i.e., change in the reinvestment assumption), but also due to revisions (i.e., treatment of trading gaps), and backfilling (i.e., IPOs). In line with the relevance of methodological changes, Akey, Robertson, and Simutin (2024) show that the construction changes in the Fama and French (1993) factors have a greater impact than revisions to the underlying data. Similarly, we raise awareness of methodological changes and their impact by dissecting CRSP changes between the old and new monthly tapes.

Moreover, we contribute to the literature on return computation. For instance, Bessembinder, Chen, Choi, and Wei (2025) study how investors should measure long-term returns and underline the relevance of reinvesting interim cash-flows for long-term returns. We add to this literature by investigating whether reinvesting payouts on the ex-date or payment date matters for return premia. Although reinvesting dividends on the effective payment date rather than on the ex-date rewrites around 14.9% monthly returns by at least 1 bp, it does not materially alter unconditional asset pricing premia.

Lastly, as we study the reinvestment of payouts, we also relate to studies analyzing returns around the day when stocks trade without the dividend for the first time, i.e., Hartzmark and Solomon (2019) or Lakonishok and Vermaelen (1986). However, we can hardly infer the return difference between the old and the new CRSP tape from this literature. Intuitively, the return differences between the old and new CRSP tapes depend on the payout yield and the compounded *raw* return from the ex-date to month-end, once we reinvest payouts on the ex-date (CIZ) instead of the month-end (SIZ). While

Hartzmark and Solomon (2019) and Lakonishok and Vermaelen (1986) study returns after the ex-date, both focus on *abnormal* returns, subtracting expected returns from raw returns.[9] Additionally, Lakonishok and Vermaelen (1986) focus on five days after the ex-date, while we observe, on average, 14 days between the ex-date and the month-end.

II. Differences in Monthly Stock Returns

A. What Changed Between the CRSP Tapes?

From January 2025 onward, CRSP no longer updates the old tape (Flat File Format 1.0, SIZ). Instead, CRSP released a new tape (Flat File Format 2.0, CIZ) in July 2022, covering the U.S. stock and index databases. This update led to the following changes.

Structural Changes. Immediately visible, CRSP renamed data items. For instance, CRSP renamed the item "RET" for monthly stock returns (monthly tape) to "MthRet" and the item "RET" for daily stock returns (daily tape) to "DlyRet" to streamline abbreviations across data tables. Moreover, the data structure and the information content of various data items have changed. For example, CRSP altered the structure of the data items that capture exchanges and share codes by introducing additional variables that must be combined to achieve the same information content as the old data items.[10]

Data Changes. Unlike structural changes that only require a precise mapping between old and new data items, changes in the data can lead to differences between past studies using the discontinued old tape versus future studies based on the new tape. Most notably, CRSP changed the computation of monthly holding period returns for stocks.

[9] We also find negative returns from the ex-date to the month-end once we adjust raw returns for factor exposures. However, we find on average positive compounded raw returns (not adjusted for expected returns) from the ex-date to the month-end. Thus, shifting the reinvestment of payouts from the month-end to the ex-date will create positive return differences of monthly CIZ minus monthly SIZ returns.

[10] WRDS and CRSP provide detailed descriptions of these structural changes in an executive summary.

In the old tape, CRSP computes monthly returns based on month-end stock prices and assumes that payouts are reinvested at the month-end.[11] In contrast, CRSP computes monthly returns in the new tape as: "Daily total return compounded for the period. Will include delisting returns if appropriate".[12] Thus, while CRSP computed monthly returns from monthly price data using month-to-month-end prices in the old tape, CRSP compounds daily returns within each month to compute monthly returns in the new tape.

This change leads to return differences between both tapes for the following reasons: First, CRSP changed the reinvestment assumption of payouts. In contrast to the old tape, where payouts are reinvested at the end of the ex-date month, CRSP assumes that payouts are reinvested on the ex-date in the new tape.[13] This is a direct consequence of using daily data for the computation of monthly returns, as daily returns account for payouts on the ex-date. Second, unlike in the old tape, the new tape includes delisting returns "if appropriate" in the computation of monthly returns. Third, compounding daily returns within each month ensures that stale prices do not impact return computations. For instance, in the old tape, CRSP used the first available previous month-end price, which could be a stale price dating back several months, to compute the monthly return of the current month.[14] Thus, compounding daily returns within each month leads arguably to more precise monthly returns in the new tape compared to the old tape if the previous month-end price is missing due to trading gaps or new listings during the month (IPOs).

Reasons for the Changes. In line with this reasoning, CRSP stated that they changed the computation of monthly returns to improve their precision. Furthermore, CRSP

[11] With payouts, we refer to all distributions to shareholders from the CRSP data table "Distribution". These payouts cover mostly cash dividends, but also special dividends or liquidation payments. Note that dividends do not accrue (risk-free) interest income in the old tape.

[12] See the user guide for the "CRSP U.S. Stock & Indexes data sets, Flat File Format 2.0 (CIZ)".

[13] We refer to the "ex-date" as the first day when the security trades at a price excluding the payout.

[14] See Drechsler (2025) for a detailed discussion of how different forms of trading gaps lead to return differences between the old and new CRSP tapes.

reasoned that incorporating payouts on the ex-date (new CIZ tape), rather than at the end of the ex-date month (old SIZ tape), ensures methodological consistency with total return indices. For instance, total return indices from S&P Dow Jones or MSCI assume that payouts are reinvested on the ex-date. Also note that CRSP already accounts for payouts on the ex-date in the existing daily stock tape when computing daily returns. Lastly, since the monthly CRSP tape predated the daily tape, compounding daily returns was initially not feasible. With available daily data dating back to 1926, CRSP decided that it was time to reevaluate the computation of monthly returns in the new CIZ tape.[15]

Although price pressure around ex-dates suggests that some investors reinvest payouts on the ex-date (see, e.g., Campbell and Beranek (1955), Barclay (1987), and Bali and Hite (1998)), this might not be realistic for all types of investors, such as private investors. For cash-constrained investors, reinvesting on the ex-date requires borrowing on the ex-date until the subsequent payment date. In line with this reasoning, Berkman and Koch (2017) show that there is temporary price pressure on payment dates, especially for stocks with dividend reinvestment plans. This suggests that a significant portion of investors at least partially reinvest their dividends on or around the payment date. Consequently, we study not only the differences in monthly returns between the old (SIZ) and new (CIZ) tapes but also investigate the differences for monthly returns when reinvesting payouts on the payment date (modified new tape) as opposed to the ex-date (new tape). Hence, we construct a version of monthly returns based on the new tape, for which we only shift the reinvestment of payouts from the ex-date to the payment date. We call these returns "modified new returns" with payouts reinvested on the payment date, explain the construction in Appendix IA.II, and provide implementation code on GitHub.[16]

[15]CRSP kindly provided this information during a video call and e-mail exchanges.

[16]On average, there are 22.6 days between the ex-date (new tape) and the payment date (modified new tape) compared to 14.4 days between the ex-date (new tape) and the end of the ex-date month

B. How Large Are the Differences in Monthly Returns?

To quantify the differences due to the CRSP changes, we start by comparing the number of observations and values for commonly used data items between both CRSP tapes (Table B.1 in Appendix B). We observe that the number of observations is almost identical for most monthly and daily data items. The new tape neither loses observations from the old tape nor adds new observations for nearly all data items. Thus, the new CRSP tape covers virtually the same stocks and time horizon as the old tape.

Although CRSP states that "the vast majority of the data match up exactly", we see one striking exception: Monthly stock returns.[17] Thus, we compare monthly stock returns from the new CRSP tape (item "MthRet") with monthly stock returns from the old CRSP tape (item "RET"). We adjust the returns from the old tape for delisting returns from Shumway (1997) as CRSP implements a delisting adjustment in the new tape. Thereby, we avoid counting observations as different, which might not be different for studies that include delisting returns. Thus, besides differences due to altering the reinvestment timing, other differences should only derive from trading gaps, new listings, or different delisting assumptions between CRSP and Shumway (1997).

Applying a tolerance level of 1 basis point (bp), we find that almost 10% of the monthly returns have absolute differences exceeding 1 bp between the old and the new CRSP tape, as reported in Panel A of Table 1. The mean of these absolute return differences that exceed 1 bp is roughly 22 bp (column "MAD"). Moreover, around 1.6% of the monthly returns have absolute differences of at least 10 bp across both tapes. We

(old tape), as reported in Table A.1. Moreover, around 73.6% of all dividends have payment dates in the month(s) after the ex-date month. Thus, shifting the payout from the ex-date to the payment date might alter a larger share of observations than moving the reinvestment from the month-end (old tape) to the ex-date (new tape) within the same month.

[17]Notably, we also find that around 46% of monthly observations have different trading volumes between both tapes (Panel A of Table B.1). However, these differences are due to the rounding to the nearest thousandth, which makes these differences small relative to the average trading volume.

also investigate whether monthly returns are more positive or negative in the new versus the old tape. Although many raw return differences cancel out once we take the average, monthly returns in the new tape are still significantly more positive than in the old tape (column "Diff").[18] Thus, we observe sizable and significant return differences between both CRSP tapes that can potentially impact studies using monthly CRSP returns.

Next, we analyze the potential impact of a more realistic third reinvestment assumption. Specifically, we investigate differences between monthly returns from the new CRSP tape with payouts reinvested on the ex-date and monthly returns based on the new tape with payouts reinvested on the payment date (modified new tape) in Panel B of Table 1. Moving the reinvestment of payouts from the ex-date to the payment date alters around 14.9% of all monthly returns by at least 1 bp. Although this switch from the ex-date to the payment date alters around 50% more observations, the mean absolute difference is, with 6 bp, less than a third compared to the change from the old to the new tape in Panel A. Intuitively, reinvesting payouts on the payment date rather than on the ex-date affects more monthly observations than the change from the old to the new tape. The reason is that 73% of all payout are effectively paid out in the month(s) after the ex-date month. In contrast, shifting the reinvestment from the month-end (old tape) to the ex-date of the same month (new tape) only alters one monthly return for a given payout. Moreover, it is natural to find lower mean absolute differences relative to the comparison of the old vs. the new tape, as we only shift the reinvestment of payouts. In contrast, the differences between the old and new tapes are additionally influenced by potentially large differences based on trading gaps, new listings, and delistings.[19]

[18] Based on a two-sided and paired t-test, we also find that the difference of all (changed + unchanged) new minus old monthly returns is positive and statistically significant with a t-statistic of 5.96.

[19] We provide the distribution of monthly return differences between the new and the old (modified new) tape in Internet Appendix Table IA.1. The share of return differences and their mean absolute differences generalize to subsets of NYSE firms (Tables IA.2) and stocks priced above $5 (Table IA.3).

TABLE 1

Differences in Monthly Stock Returns

Table 1 reports summary statistics for differences between monthly stock returns from the new (CIZ) and old (SIZ) CRSP tape in Panel A. Moreover, Panel B reports statistics for return differences between monthly returns from the new CIZ tape (with payouts reinvested on the ex-date) and modified monthly returns based on the new tape with payouts reinvested on the payment date. We only include common stocks from AMEX, NYSE, and NASDAQ. A return difference exists if its absolute value exceeds the tolerance ("Tol.") or the returns are missing in either of the monthly return versions. We consider 4 tolerance thresholds: 0.1 basis points (bp), 1 bp, 10 bp, and 100 bp (or 1%). We count the absolute frequency of these differences across stocks and time ("Obs.") and the frequency relative to all observations in percent ("Rel."). Then, we compute for each monthly return observation exceeding the respective tolerance level the raw and absolute difference based on monthly returns from the new tape minus monthly returns from the old (modified new) tape. We average these raw and absolute differences across the return observations that exceed the respective tolerance level to obtain mean raw differences ("Diff.") and mean absolute differences ("MAD"), both in percent per month. Finally, we compute whether the raw differences are different from zero with a two-sided paired t-test and indicate the statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. For all statistics, we consider 3 time periods: The full sample period in columns 2 to 5, an early sample from July 1926 until December 1967 in columns 6 to 8, and a late sample from January 1968 until December 2023 in columns 9 to 11.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	474,580	12.78	0.16	0.023***	21.53	0.21	0.037***	11.41	0.15	0.019***
1 bp	357,445	9.62	0.22	0.031***	18.11	0.24	0.044***	8.30	0.21	0.026***
10 bp	60,684	1.63	1.17	0.194***	3.72	1.10	0.240***	1.31	1.20	0.173***
1%	8,312	0.22	15.43	2.776***	0.47	14.33	3.964***	0.18	15.92	2.252***

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	754,384	20.31	0.05	0.003***	33.52	0.06	−0.003***	18.25	0.04	0.005***
1 bp	552,954	14.89	0.06	0.004***	27.21	0.07	−0.004***	12.96	0.06	0.007***
10 bp	77,697	2.09	0.22	0.023***	4.41	0.24	−0.013***	1.73	0.21	0.038***
1%	2,041	0.05	2.14	−0.031	0.11	2.02	−0.584***	0.05	2.23	0.365

The share of return differences vary over time. The time split in Table 1 indicates around twice as many return differences in early sample periods before 1967, relative to periods thereafter.[20]

The differences between returns from the new tape and modified returns from the

[20]We choose 1967 as the cutoff, as Compustat covers most firms from this time onward. Moreover, we plot the absolute return differences and the share of monthly returns that differ between the old, new, and modified new tape in Figure IA.2 and Figure IA.3 in Internet Appendix IA.III. Consistent with the time split in Table 1, the share of return differences decreased from 1960 until 1985.

new tape with payouts reinvested on the payment date are, by construction, only due to changing the reinvestment timing. However, it is not clear to what extent reinvestment timing, trading gaps, new listings, or delistings determine the documented differences in monthly returns between the old and new CRSP tape. Thus, we decompose the return differences between the old and new tape in the following to quantify the relative importance of these determinants. Investigating these determinants will also guide our analysis of how these return differences impact asset pricing studies in Section III below.

C. What Drives the Changes in Monthly Returns?

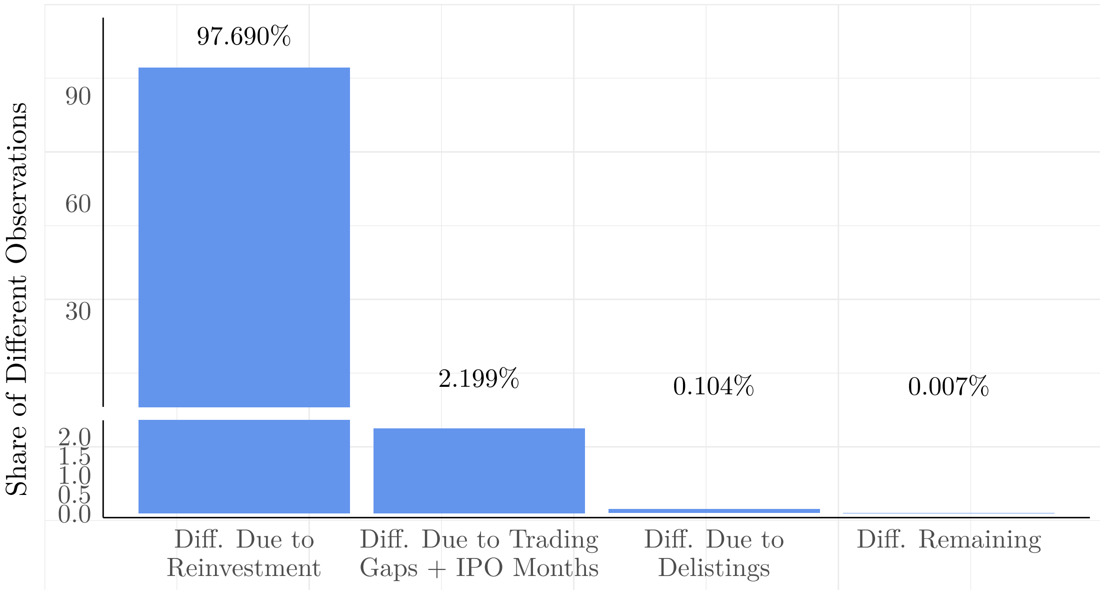
As the daily returns did not change materially in the new (CIZ) tape, we can replicate the new monthly returns using the daily stock returns from the old (SIZ) tape. This allows us to test how many of the monthly return differences are due to the altered reinvestment assumption, trading gaps, new listings, and delistings. First, we start by compounding daily returns within months for which we observe no trading gaps or delistings. Thereby, we ensure that payouts are reinvested on the ex-date as daily returns in the old tape account for payouts on the ex-date. Adjusting for this altered reinvestment assumption explains around 97.7% of all absolute return differences between both tapes that exceed a tolerance level of 0.1 bp, as shown in Figure 2. Thus, the vast majority of monthly return differences are due to changing the reinvestment timing of payouts.

Second, we also compound daily returns for those months that have trading gaps but no recorded delistings. Compounding daily returns within each month can lead to markedly different monthly returns compared to using month-to-month-end prices in the old CRSP tape. That is, CRSP used stale month-end prices that could date back several months, depending on the trading gap. Additionally, CRSP was unable to compute monthly returns for stocks first listed during a respective month because the previous month-end price was unavailable. These observations are no longer missing in the new tape, as daily compounding within a month does not require previous month-end prices. Taking into account these trading gaps and IPO months allows us to explain additionally 2.2% of the return differences between the old and new tape, as shown in Figure 2.

FIGURE 2

Decomposing Monthly Return Differences Between CRSP Tapes

Figure 2 shows the share of absolute return differences between the old (SIZ) and new (CIZ) CRSP tape exceeding 0.1 basis points that can be explained by changing the reinvestment assumption, trading gaps/IPO months, and adjusting delisting returns. We compute these shares as follows: First, we compute how many of the monthly return differences can be explained by compounding the daily returns of the old CRSP tape for months without missing previous month-end prices and no recorded delistings ("Diff. Due to Reinvestment"). Second, we calculate how many of the monthly return differences can be reconciled by compounding daily returns also for months with missing month-end prices from the previous month and no recorded delistings ("Diff. Due to Trading Gaps + IPO Months"). Third, we compute how many of the monthly return differences can be explained by additionally adjusting delisting returns in the old tape ("Diff. Due to Delistings"). Lastly, we show the share of monthly return differences that we cannot explain ("Diff Remaining") after compounding daily returns within all months and adjusting delisting returns. The sample covers all common stocks from AMEX, NYSE, and NASDAQ from 1926 to 2023. All relative frequencies are in percent, and the vertical-axis scales differ for illustrative purposes.



Third, we can explain an additional share of 0.1% of all monthly returns that differ between the old and new tape if we alter two delisting assumptions from Shumway (1997) for the old tape: First, we do not include delisting returns of -30% if the delisting return is missing and the delisting code is between 400 and 591. Second, we set delisting returns to zero if there are more than 30 days between the payment date of the delisting amount and the delisting date. For the remaining delisting dates, we adjust the daily returns by multiplying them by the delisting return. Based on this procedure, we no longer observe differences between the old and the new tape in delisting months. Thus, our results suggest that the delisting procedure included in monthly returns in the new CRSP tape deviates from Shumway (1997) and does not adjust for missing delisting returns. Moreover, CRSP requires the payment date of the delisting amount to be close to the delisting date to ensure that this part of the return is paid out around the delisting date. These delisting adjustments explain what CRSP refers to as delisting "if appropriate".

Although only around 2.3% of all return differences are due to trading gaps, new listings (IPO month), and delistings, their absolute return differences are large: The mean absolute difference for all monthly returns that differ due to trading gaps and new listings is around 7.2%, and even around 14% for monthly returns that differ due to delistings between both tapes (see Internet Appendix Figure IA.4). However, trading gaps, IPO months, or delistings cannot explain the large share of return differences. Moreover, even if we exclude months with trading gaps, new listings, and delisting, the mean of all absolute return differences between the new and old tape is still around 7 bp (see Internet Appendix Table IA.4). These differences are similar to the mean absolute difference that we obtain if we only shift the reinvestment of payouts in the new tape from the ex-date to the payment date (6 bp in Table 1). The remaining differences (32 observations) in Figure 2 have missing returns in the new tape, but valid returns in the old tape.

Next, we analyze the drivers of the differences that arise from reinvesting payouts on the ex-date instead of the month-end (97.7%): First, larger payout yields intuitively lead to larger differences once we shift the reinvestment of payouts. Second, larger absolute compounded returns from the ex-date to the month-end correspond to the reinvestment return of payouts within the month and will increase return differences between the tapes. Consistent with this intuition, the heatmap in Figure B.1 of the Appendix shows that return differences between both CRSP tapes increase with the payout yield and the absolute compounded return from the ex-date to the month-end.[21] In line with these results, stocks with return differences are older, larger, more profitable, and have higher payout yields (see Table IA.5). Also, stocks in traditional dividend-paying industries (e.g., financials or utilities) have a larger share of absolute return differences (see Table IA.7).[22]

Dependency on the Business Cycle. Both components – the payout yield and the absolute compounded return from the ex-date to the month-end – are larger during NBER recessions (see Figure IA.7). This might be intuitive, as payout yields depend on prices that drop at the beginning of recessions. Furthermore, the cross-sectional return dispersion increases during recessions, implying, on average, more extreme returns from the ex-date to the month-end.[23] Thus, both channels that drive the monthly return differences are more pronounced in recessions. Hence, we expect larger impacts of these return differences on asset pricing studies during economic downturns.

To sum up, changing from the old (SIZ) to the new (CIZ) tape rewrites 9.6% of monthly returns by at least 1 bp, which is mostly due to reinvesting payouts on the ex-date instead of the month-end. Moreover, these differences are more pronounced during recessions. Thus, the change from the old to the new tape is potentially impactful, as the code of open source asset pricing studies (see, e.g., Chen and Zimmermann (2022),

[21] We find similar results for return differences that arise from reinvesting payouts on the payment date as opposed to the ex-date in Figure IA.5. The return differences shown in Figure B.1 do not strongly cluster in specific areas of the heatmap, as shown by Figure IA.6 in the Internet Appendix.

[22] We find similar results for return differences that arise from reinvesting payouts on the payment date instead of the ex-date in Panel B. Moreover, even when we exclude stocks in the financial or utility sector, as these are often omitted in empirical studies, 8.5% of all monthly returns still differ by at least 1 bp between the old and new CRSP tape (see Table IA.8 in the Internet Appendix).

[23] We confirm these patterns in Table IA.9 by regressing each component and the mean absolute return differences on the NBER recession indicator. Panel B also shows that return differences that arise from moving the reinvestment of payouts from the ex-date to the payment date are larger in recessions.

Jensen et al. (2023), and Scheuch, Voigt, and Weiss (2023)) suggests that these authors and, thus, likely the majority of original studies directly used monthly returns from the old CRSP tape. We find no indications that, instead of using the monthly return item, studies compute monthly returns by compounding daily returns with dividends reinvested on the ex-date or the payment date. Altering the reinvestment timing in the new CIZ tape from the ex-date to payment date rewrites even 14.9% of monthly returns by at least 1 bp. However, the corresponding mean absolute difference of these monthly returns is less than a third relative to the comparison between the old and new CRSP tape.

III. Differences in Return Premia

It is ex ante unclear how these significant changes in monthly returns between the old and the new tape affect studies that use monthly stock returns from CRSP. The impact of these documented changes is arguably relevant, as 34% of the articles published in the Top5 finance journals in the last 25 years use monthly CRSP returns. While there are many possible applications, we focus on a widely used asset pricing method that relies on monthly CRSP returns, that is, portfolio sorts. By sorting the cross section of stocks into portfolios, researchers uncover (risk) factors and estimate their premia.

A. Data and Methodology

Data. Our premia analysis relies on standard asset pricing data from 1968 until 2023.[24] In particular, we use CRSP data on prices, returns, shares outstanding, industry classifications, and trading volume of common stocks traded on the NYSE, AMEX, and NASDAQ from the old (SIZ) and new (CIZ) tapes. For the old tape, we use delisting returns according to Shumway (1997) and set missing delisting returns with delisting codes 400–591 to -30%.

[24] Note that we include all CRSP observations from 1926 onward in our return analysis in Section II.

Monthly returns in the new tape already include delisting returns. Accounting data are from Compustat, and we require for all analyses a valid primary link between CRSP's permno and Compustat's gvkey in the respective linking table.

Moreover, we follow Akey et al. (2024) and replicate the market factor with their publicly available code, which is based on a fixed methodology following the details in Fama and French (2023). This allows us to construct a market factor with data from the old SIZ and new CIZ tapes separately. The code from Akey et al. (2024) ensures replicability and consistency as the Fama and French (1993) factors are maintained by Dimensional Fund Advisers (DFA) and are subject to methodological changes (Akey et al. (2024)).[25] The risk-free rate comes from DFA data available on Kenneth French's website. Since future database updates might affect the reproducibility of our study, we state the vintage of these data sets, which is from July 2024. Lastly, we obtain the NBER recession indicator and consumption growth from the FRED (vintage from November 2025).

Sorting Variables. We construct 68 sorting variables covering a wide range of economic rationales, which previous studies have suggested to predict the cross section of stock returns. These sorting variables are our test characteristics to assess the impact of the return changes. We provide a list of all sorting variables in Table C.1. Moreover, we follow Hou et al. (2020) in the sorting variable construction (see Internet Appendix IA.VIII) and assign all sorting variables to groups to facilitate comparisons: Financing, intangibles, investment, momentum, profitability, size, trading frictions, and valuation.

Portfolio Sorts. We estimate the premia for each sorting variable by sorting stocks into portfolios based on each of the 68 sorting variables. While portfolio sorts require many choices, e.g., equal- or value-weighted portfolio returns, we control for this methodological

[25] The fixed-factor code for the Fama and French (1993) factors is available on the website of The University of Chicago Law Review, as mentioned in Robertson, Akey, and Simutin (2025).

variation: We follow Walter et al. (2024) and identify 14 common methodological decisions and their choices for portfolio sorts in Figure 3 based on how often published articles implemented these choices, which we indicate by color saturation. As shown in Figure 3, we consider seven sample construction decisions, for example, whether to include or exclude small stocks, financials, utilities, or stocks with negative book equity values. Moreover, we assess seven common portfolio construction decisions, such as different lags between information arrival and portfolio formation, monthly vs. annual rebalancing, five vs. ten portfolios, or equal- vs. value-weighted portfolio returns. To implement these portfolio sorts, we adjust open-source code on GitHub from Walter et al. (2024) for the (modified) new tape with updates from www.tidy-finance.org (Scheuch et al. (2023)).

The choices in Figure 3 imply for the same sorting variable various methodological paths to estimate a premium. We define a path $p = (c_1, ..., c_F)$ as a vector of choices c_f made at each of the 14 decision forks $f = 1, ..., 14$. To account for this methodological variation, we estimate for each sorting variable the premia based on all paths from Figure 3, e.g., 69,120 value premium estimates that differ in the choices made at each decision fork. Thus, we acknowledge the existence of methodological uncertainty (Menkveld, Dreber, Holzmeister, Huber, Johanneson, Kirchler, Razen, Weitzel, Abad, Abudy, and Others (2024)) and ensure that differences between premia constructed with the new, old, or modified new tape are not only prevalent for one potentially arbitrary path.

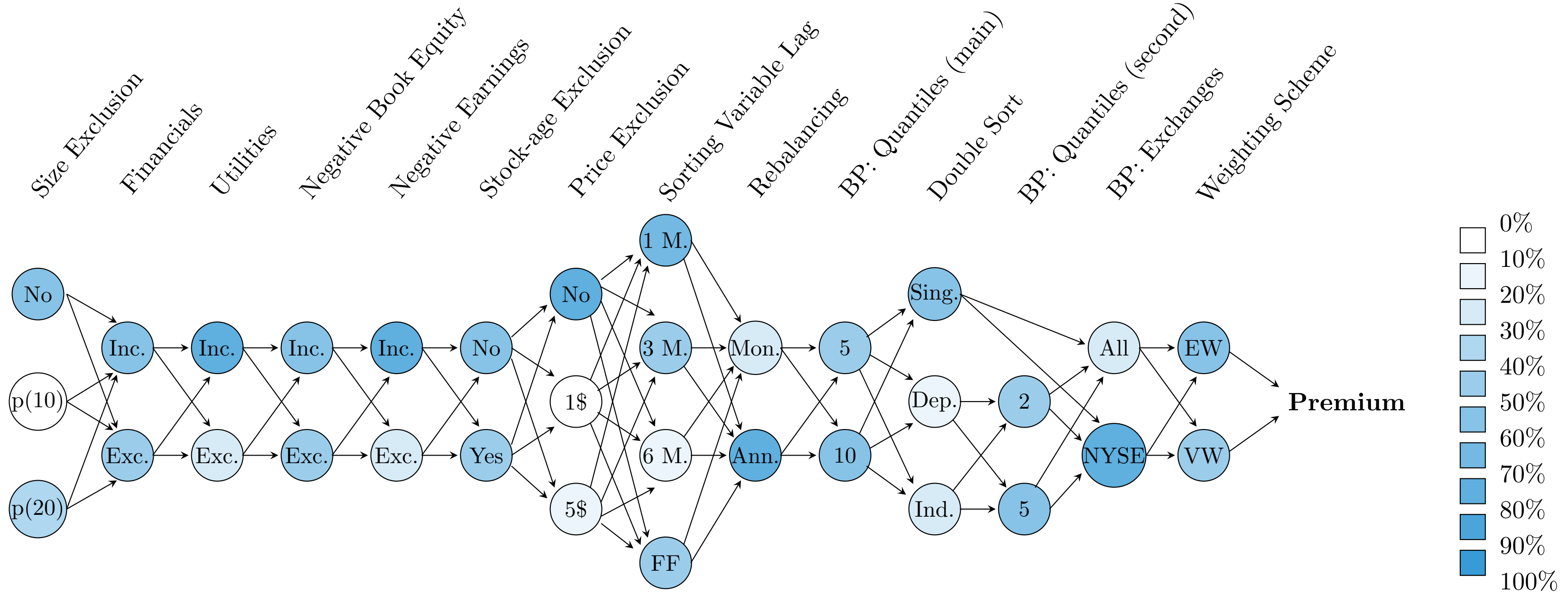
After sorting stocks into portfolios, we compute in every month t the long-minus-short portfolio return for each sorting variable v and all paths $p = 1, ..., P$. We estimate these long-short returns $(r_{t,p}^{v,d})$ with data $d \in \{\text{old}, \text{new}, \text{modified new}\}$ from the old, the new, and the modified new tape (with payouts reinvested on the payout date) separately.

$$r_{t,p}^{v,d} = r_{t,p}^{v,d,\text{Long}} - r_{t,p}^{v,d,\text{Short}}, \tag{1}$$

FIGURE 3

Flowchart of Construction Decisions in Portfolio Sorts

Figure 3 shows the paths based on 14 construction decisions (forks) for a portfolio sort until the premium is estimated. The first seven forks concern the sample construction: Excluding small stocks dependent on market equity quantiles (No, smaller than p(10) or p(20)), financials (included (Inc.) or excluded (Exc.)), utilities (included or excluded), firm-months with negative book equity (included or excluded), firm-months with negative earnings (included or excluded), requiring two years of minimum listing as in Fama and French (1992) (Yes or No), and excluding stock prices smaller than \$1, \$5, or none. The subsequent seven forks belong to the portfolio construction: The lag of the sorting variables (one month (1 M.), three months (3 M.), six months (6 M.), or a Fama and French (1992) lag (FF)), the portfolio rebalancing (monthly (Mon.) or annually (Ann.)), the number of main portfolios (5 or 10), the sorting method (single sorts (Sing.), dependent (Dep.), or independent double sorts (Ind.)), the number of secondary portfolios for double sorts (2 or 5), the exchanges for breakpoints (all stocks (All) or NYSE listed stocks (NYSE)), and the weighting scheme (equal-weights (EW) or value-weights based on the market capitalization (VW)). Note that we only allow for a sorting variable lag of 3 months, 6 months, and as in Fama and French (1992) for sorting variables updated yearly. For sorting variables updated monthly, we allow for sorting variable lags of 1, 3, and 6 months. Sorting variables updated quarterly can have a sorting variable lag of 3 or 6 months. Also, the choice to rebalance portfolios annually is naturally only available for sorting variables updated yearly and not for those updated monthly or quarterly. The color saturation indicates how often the 109 papers analyzed by Hou et al. (2020) implemented each choice.



where $r_{t,p}^{v,d,\text{Long}}$ ($r_{t,p}^{v,d,\text{Short}}$) is the long (short) portfolio return. We compute the premium for each sorting variable v and each path p as the time-series average of long-short portfolio returns. We do this separately for the old, new, and modified new tape:

$$\pi_p^{v,d} = \frac{1}{T}\sum_{t=1}^{T}\left(r_{t,p}^{v,d,\text{Long}} - r_{t,p}^{v,d,\text{Short}}\right). \tag{2}$$

This allows us to compare long-short portfolio returns and premia between different CRSP tapes. Specifically, we compute for each path the absolute difference in monthly long-short portfolio returns between the old and new tape, as well as between the new and modified new tape. Then, we average the absolute differences across all paths and months to obtain mean absolute differences (MADs) for each sorting variable:

$$MAD^v = \frac{1}{T}\frac{1}{P}\sum_{t=1}^{T}\sum_{p=1}^{P}\left|\, r_{t,p}^{v,\text{new}} - r_{t,p}^{v,\text{old (modified new)}} \,\right| \tag{3}$$

Correspondingly, we compute absolute differences for each sorting variable between premia based on the new and old (modified new) tape by averaging across paths:

$$AD^v = \frac{1}{P}\sum_{p=1}^{P}\left|\, \pi_p^{v,\text{new}} - \pi_p^{v,\text{old (modified new)}} \,\right| \tag{4}$$

Additionally, we test whether the long-short portfolio return differences between the old and new tape ("Sig." in Table 2) are significant. Portfolio sorts that share the same choice of a decision fork, but differ in other (potentially less relevant) choices, might not be independent. To control for these dependencies we set up clusters for all portfolio sorting choices from Figure 3 and assign portfolio sorts that make the respective choice to these clusters. Then, we repeatedly sample differences in long-short portfolio returns based on these clusters for every sorting variable and month. Finally, we determine for each sorting variable how many of the bootstrap-implied 99% confidence intervals for the differences in long-short portfolio returns do not contain zero (see Appendix IA.VI).[26]

[26]As an additional robustness check, we analyze a subset of methodological paths for which the

How do CRSP Changes Impact Portfolio Returns? The return changes across the old, new, and modified new tape affect portfolio sorts in two ways: First, monthly portfolio returns can change when aggregating monthly stock returns with differences between tapes into portfolio returns. However, it is unclear how the documented return changes affect portfolio returns as we aggregate potentially positive and negative stock return differences. For example, one might systematically sort stocks with positive (negative) return differences between tapes predominantly into one portfolio, e.g., the long portfolio. If this is the case, CRSP return changes would systematically alter premia.[27]

Second, the impact of CRSP return changes can go beyond the impact induced by aggregating monthly stock returns into portfolio returns for sorting variables that are constructed based on monthly stock returns, e.g., 11-month return momentum. For these sorting variables, the CRSP return changes directly alter the sorting variable before sorting stocks into portfolios. These construction changes can potentially change the cross-sectional rank of affected observations and, in turn, lead to differences in estimated premia. Hence, we flag all sorting variables that directly use *monthly returns* from CRSP with a superscript R.[28] For these predictors constructed with monthly returns, it is even more unclear how the CRSP changes between both tapes impact estimated portfolio returns and premia. Changes in the cross-sectional ranking of stocks can lead to differences along many steps of the portfolio sorting procedure, i.e., when we compute portfolio breakpoints. Thus, we expect a potentially larger impact of CRSP changes for sorting

corresponding long-short portfolio returns have among each other low absolute time-series correlations below a cutoff. Across all sorting variables, we select, on average, 2.1% of all methodological paths and document the path selection in Internet Appendix IA.VII. These methodological paths are less dependent on each other and offer insights into whether a potentially large fraction of highly dependent paths drives our results regarding the share and size of differences in long-short portfolio returns or premia.

[27]For instance, in Panel A of Table IA.6, we find that stocks with positive differences between monthly returns from the new tape minus monthly returns from the old tape tend to have lower book-to-market ratios, higher momentum, and tend to be larger compared to stocks with negative return differences.

[28]Moreover, we mark predictors that use CRSP data in their construction with a superscript C, as also other CRSP variables such as shares outstanding and, thus, market capitalization slightly changed.

variables that are directly constructed from monthly return data.

B. Differences in Return Premia Between Tapes

After sorting stocks into portfolios based on a sorting variable, we compute for all 68 sorting variables monthly long-short portfolio returns and their time-series averages (premia) over various methodological paths. For example, we compute 69,120 value premium estimates that emerge from all possible combinations of the methodological choices for portfolio sorts in Figure 3. We start by comparing monthly long-short portfolio returns computed with data from the old tape with those based on data from the new tape. Table 2 reports the share of these long-short portfolio returns that differ by more than 1 bp ("$N_{1\text{ bp}}$") and by more than 10 bp ("$N_{10\text{ bp}}$") when constructing them based on both tapes separately. We compute these shares of monthly returns that differ between both tapes by averaging across all methodological paths, months, and finally across all sorting variables that belong to the respective group in the first rows of Table 2 in Panel A.

We find the largest share of long-short portfolio returns that differ by at least 10 bp between both CRSP tapes for momentum variables and their methodological paths (14.58%). In contrast, we find the lowest share for profitability-related variables, as 8.97% of their monthly long-short portfolio returns differ by at least 10 bp between both CRSP tapes. Across all groups of sorting variables, 11.43% (45.78%) of all monthly long-short portfolio returns differ by more than 10 (1) basis points.[29] Notably, the share of long-short returns that differ by at least 10 bp when using data from the old vs. the new tape is at least 5.6% for all sorting variables separately, as shown in Table IA.10 of the Internet

[29] We also investigate how large these differences in monthly long-short portfolio returns are if we only allow for differences in monthly stock returns due to altering the reinvestment assumption (97.7% of the differences) and exclude differences due to trading gaps, IPO months, or delistings. The results in Table IA.12 suggest that around 72% (48%) of the differences in monthly portfolio returns exceeding 1 bp (10 bp) from Table 2 – and, thus, a considerably large fraction – are driven by altering the reinvestment assumption from the old to the new CRSP tape.

Appendix. To get a perspective on the size of these differences, we compute the mean absolute difference of monthly long-short returns between the tapes averaged across all methodological paths and months. These mean absolute differences are around 4 bp per month (column "MAD") across all groups of sorting variables. Moreover, the bootstrap procedure indicates that in 59% of the months the differences in long-short portfolio returns across all methodological paths and groups are statistically different from zero at the 1% level (column "Sig.").[30]

Panel B of Table 2 reports that the differences in long-short portfolio returns are more pronounced for sorting variables that are directly constructed from monthly returns. For those return-based sorting variables, we find that more than a fourth (28.07%) of all monthly long-short returns across all methodological paths exceed a difference of 10 bp when comparing them across the old and new CRSP tape. The mean absolute difference of these long-short returns for all return-based sorting variables is around 9 bp per month.

Next, we investigate differences in monthly long-short portfolio returns across all sorting variables and their methodological paths during early sample periods and recessions in Panel C of Table 2. For samples prior to 2000 and during NBER recessions, we find, on average, almost twice as many monthly long-short returns that differ by more than 10 bp compared to the period after 2000 and compared to non-NBER recession months. In particular, prior to 2000 and during NBER recessions, around 15.04% and 18.95% of all monthly long-short returns differ by at least 10 bp. The corresponding mean absolute differences are about 6 and 7 basis points per month, respectively.

[30]We control for dependencies of long-short portfolio returns across methodological paths with a bootstrap procedure (see Appendix IA.VI). Moreover, we find a similar share and size of portfolio return differences when we only include the 2.1% of long-short portfolio returns with the least absolute correlations (see Table IA.14).

TABLE 2

Differences in Monthly Long-Short Portfolio Returns

Table 2 reports summary statistics of the differences between monthly long-short portfolio returns estimated from the old (SIZ) and the new (CIZ) CRSP tape from 1968 to 2023. To account for methodological uncertainty, we estimate long–short portfolio returns for all 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. We compute the share of monthly long-short portfolio returns that differ by at least 1 bp ("$N_{1\text{ bp}}$") and 10 bp ("$N_{10\text{ bp}}$") across all methodological paths and months for each sorting variable. The respective rows report these shares of long-short differences across sorting variables for specific groups (Panel A), types (Panel B), and averaged across specific time periods for all sorting variables (Panel C). Based on the same aggregation, we also report the share of monthly long-short portfolio returns with positive ("Pos.") and negative ("Neg.") return differences between the old and the new tape. Moreover, we compute for each sorting variable the mean absolute difference ("MAD") of long-short portfolio returns from the old vs. the new tape (average across methodological paths and months). The rows report these MADs averaged across groups (Panel A), types (Panel B), and across specific time periods for all sorting variables (Panel C). We also test for each sorting variable and each month whether the long-short portfolio returns from both tapes are significantly different across methodological paths. We control for dependencies across portfolio sorts that share similar choices. Specifically, we bootstrap the differences in long-short portfolio returns for each sorting variable and month with clusters for portfolio sorts that share the same choice of a decision fork (see Internet Appendix IA.VI). Column "Sig." reports the share of monthly long-short portfolio returns for which the bootstrap-implied 99% confidence intervals do not contain zero, averaged over sorting variables for each group. All statistics are in percent. The sorting variables belonging to each group or type can be found in Table C.1 of the Appendix. We classify sorting variables that require CRSP data (monthly CRSP returns) for their construction as CRSP-based (return-based).

Panel A. Sorting Variable Groups

Group	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
Financing	46.72	12.87	0.04	65.80	52.00	48.00
Intangibles	40.60	9.60	0.04	64.22	50.54	49.46
Investment	41.93	9.28	0.03	66.51	51.42	48.58
Momentum	51.27	14.58	0.05	57.69	52.09	47.91
Profitability	42.58	8.97	0.04	62.70	49.06	50.94
Size	44.90	12.04	0.06	35.28	46.79	53.21
Trading Frictions	51.53	12.85	0.05	50.07	52.26	47.74
Valuation	46.71	11.25	0.04	69.66	50.87	49.13
Overall	45.78	11.43	0.04	58.99	50.63	49.37

Panel B. Sorting Variable Types

Type	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
Non-CRSP-Based SVs	41.33	8.82	0.03	65.02	50.48	49.52
CRSP-Based SVs	50.04	13.62	0.05	59.47	51.55	48.45
Return-Based SVs	74.79	28.07	0.09	52.53	51.75	48.25

Panel C. Time Splits

Period	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
Jan. 1968 – Dec. 1999	56.22	15.04	0.06	60.28	51.33	48.67
Jan. 2000 – Dec. 2023	31.91	6.30	0.03	65.05	50.57	49.43
NBER Recessions	61.92	18.95	0.07	61.98	52.37	47.63
No Recessions	43.17	10.08	0.04	62.66	50.91	49.09

Overall, the results of Table 2 suggest that the CRSP changes in the new tape alter monthly long-short portfolio returns by far more than they change monthly stock returns. While about 10% of the monthly stock returns differ at the 1 bp threshold level between the old and new CRSP tape, our results show that a similar share of monthly long-short portfolio returns (11.43%) differ even at the 10 bp level. This suggests that portfolio sorts amplify the CRSP return changes throughout the sorting procedure.

Unconditional Return Premia. In the next step, we investigate whether these significantly altered monthly long-short returns lead to different return premia, i.e., differences in the *time-series averages* of the monthly long-short returns. Surprisingly, our results reveal that changing from the old to the new tape does not materially change premia. Table 3 reports the share of premia that differ between both tapes, averaged across all methodological paths for groups of sorting variables in Panel A and for types of sorting variables in Panel B. On average, only 0.02% (4.67%) of all premia across all groups and their methodological paths differ between the old and new tape at the 10 bp (1 bp) level.

Moreover, for each sorting variable, we compute the average absolute difference between premia constructed with the old versus the new tape by averaging across all methodological paths. These absolute differences (column "AD") between the premia from the old vs. the new tape range from 0.24 bp for all investment related variables to 0.83 bp for the size group. Across all groups of sorting variables and their methodological paths, return premia differ on average by 0.34 bp, which is only about 5.95% (column "Ratio") of the average absolute premium. Panel B reports similar results for the return-based sorting variables with premia differences of, on average, 0.59 bp.

TABLE 3

Differences in Return Premia

Table 3 reports summary statistics of the differences between premia estimated from the old (SIZ) and the new (CIZ) CRSP tape from 1968 to 2023. To account for methodological uncertainty, we estimate the time-series average of monthly long-short portfolio returns (premium) for each of the 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. We compute the share of premia that differ by at least 1 bp ("$N_{1\ bp}$") and 10 bp ("$N_{10\ bp}$") across all methodological paths for each sorting variable. The respective rows report these shares of premia averaged across sorting variables for groups (Panel A) and types (Panel B) in percent. Moreover, we compute for each sorting variable the absolute difference ("AD") of premia estimates from the old vs. the new tape averaged across methodological paths. The respective rows report these ADs in basis points per month across groups (Panel A) and types (Panel B) of sorting variables. We also scale these absolute differences by the average absolute premium within each group (column "Ratio", in percent). Finally, we report the 90- ("Q90") and 95-percentiles ("Q95") of the absolute premia differences among all methodological paths and sorting variables of each group (in basis points per month). The sorting variables belonging to each group or type can be found in Table C.1 in the Appendix. We classify sorting variables that require CRSP data (monthly CRSP returns) for their construction as CRSP-based (return-based).

Panel A. Sorting Variable Groups

Group	N_{1bp}	N_{10bp}	AD	Ratio	Q90	Q95
Financing	7.01	0.00	0.36	1.18	0.73	0.90
Intangibles	2.94	0.01	0.31	5.64	0.66	0.85
Investment	0.90	0.00	0.24	0.63	0.52	0.65
Momentum	8.15	0.00	0.41	0.86	0.87	1.10
Profitability	2.98	0.00	0.30	2.25	0.70	0.93
Size	16.36	0.62	0.83	10.75	1.80	4.22
Trading Frictions	9.74	0.04	0.51	30.07	1.21	1.66
Valuation	2.38	0.00	0.28	1.33	0.59	0.74
Overall	4.67	0.02	0.34	5.95	0.75	1.00

Panel B. Sorting Variable Types

Type	N_{1bp}	N_{10bp}	AD	Ratio	Q90	Q95
Non-CRSP-Based SVs	1.73	0.00	0.27	1.34	0.59	0.76
CRSP-Based SVs	7.97	0.03	0.42	11.13	0.94	1.28
Return-Based SVs	17.45	0.00	0.59	39.87	1.19	1.47

Even looking at the many methodological paths suggests only small differences in premia for sorting variables belonging to each group: For the methodological paths that generate the 5% largest differences in premia (column "Q95") we find only absolute differences of at least 1 bp across all sorting variables and 1.47 bp across return-based sorting variables. These results also hold for all sorting variables separately (Table IA.11 in the Internet Appendix) and for CAPM alphas relative to the market factor (Table IA.13 in the Internet Appendix).

In line with the small absolute differences in premia, we find that only a few premia estimates change their significance once we switch from the old (SIZ) to the new (CIZ) CRSP tape. Specifically, Table 4 reports that only 0.52% and 0.54% of all methodological paths generate premia estimates across all sorting variables that change their significance at the 1% and 5% significance levels, respectively.

TABLE 4

Significance of Premia Between Tapes

Table 4 reports the share of premia that change their significance once we switch from the old (SIZ) to the new (CIZ) CRSP tape. To account for methodological uncertainty, we estimate the time-series average of long-short portfolio returns (premium) for each of the 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. For each sorting variable and methodological path, we compute the t-statistic whether the corresponding premium is different from zero based on Newey and West (1987) corrected standard errors. Then, we compute the share of premia estimates that change their significance at the 1% and 5% level across all methodological paths for each sorting variable. The respective rows report these shares of premia that change their significance averaged across sorting variables for specific groups (column "Change"). Moreover, we report for each group the share of premia estimates across all methodological paths that are significant ("Sig.") or insignificant ("Insig.") in both tapes. The sample period is from 1968 to 2023. All relative frequencies are in percent, and the sorting variables of each group can be found in Table C.1 in the Appendix.

	1%			5%		
Group	Change	Sig.	Insig.	Change	Sig.	Insig.
Financing	0.83	59.37	39.80	0.44	74.89	24.67
Intangibles	0.54	19.31	80.15	0.89	35.14	63.96
Investment	0.48	85.25	14.28	0.23	94.98	4.79
Momentum	0.71	67.55	31.74	0.46	80.62	18.92
Profitability	0.37	29.27	70.36	0.36	40.49	59.15
Size	0.37	7.17	92.46	0.64	13.28	86.08
Trading Frictions	0.37	11.90	87.74	0.69	22.28	77.03
Valuation	0.49	15.52	83.99	0.64	38.50	60.85
Overall	0.52	38.52	60.96	0.54	52.69	46.78

The changes are similar among groups of sorting variables, ranging from 0.37% to 0.83% for the 1% significance level. We find that none of the sorting variables has more than 2.9% of methodological paths that change the significance of premia at the 1% level between the old (SIZ) and the new (CIZ) CRSP tape (Table D.1 of the Appendix).[31]

While we show that changing from the old to the new CRSP tape does not systematically

[31]Note that the share of differences in return premia, their size, and the changes in the significance of premia are also robust to excluding the vast amount of strongly correlated methodological paths, as shown in the respective Panel B of Table F.1 and Table F.2.

alter premia, the question remains whether reinvesting payouts on the payment date as opposed to the ex-date might substantially alter premia. Thus, we repeat the same analysis and compare premia based on data from the new CRSP tape, in which payouts are reinvested on the ex-date, with premia based on the same tape but with payouts reinvested on the payment date. In line with the change from the old to the new tape, we find that reinvesting payouts on the payment date as opposed to the ex-date also does not materially change premia (Panel C of Table F.1) nor the significance of these premia estimates (Panel C Table F.2) across all groups of sorting variables.

Conditional Return Premia. While unconditional premia are largely unaffected by the CRSP changes, we also investigate whether the CRSP changes affect the cyclicality of premia. Investigating premia conditional on the business cycle might be a natural starting point where CRSP changes potentially matter, as differences in monthly long-short portfolio returns are almost twice as frequent and large during NBER recessions relative to all other months (see Table 2). To provide comprehensive evidence, we identify cyclical premia in our sample and regress the respective monthly long-short portfolio returns on the NBER recession dummy as a realized business cycle indicator. We estimate these NBER regression coefficients for the same sorting variable across various long-short portfolio returns, which we estimate by varying methodological choices in portfolio sorts.[32]

For the vast majority of sorting variables and their methodological paths, we do not find that the regression coefficients for the NBER recession dummy change their significance in Table E.1 when we change from the old to the new CRSP tape. However, we note one exception: the real estate ratio from Tuzel (2010). Around 4.6% (3%) of

[32]Note that we focus on premia that are likely cyclical based on the NBER recession dummy: For each sorting variable, we regress monthly long-short portfolio returns on the NBER recession dummy. We repeat these regressions for all time series of monthly long-short returns that we obtain by varying methodological choices in portfolio sorts. Then, we classify a premium as significant if at least 25% of the NBER regression coefficients are significant across all methodological paths.

the NBER recession coefficients across all long-short portfolio returns, which we estimate by varying methodological sorting choices, change the significance at the 10% (5%) level once we switch from the old to the new tape. These shares are large if we consider that only 28% (12%) of all real estate ratio-sorted portfolio returns generate significant NBER recession coefficients at the 10% (5%) level when varying methodological sorting choices.

To reduce complexity, we also fix a common methodological path for risk-based rationales motivated by the choices in Hou et al. (2020)[33]: We include all stocks except for financials, stocks with negative book equity values, and stocks with less than 2 years of listing. Then, we form monthly decile portfolios based on NYSE breakpoints and value-weighted portfolio returns. In line with the evidence across all methodological paths, we note that the t-statistic of the NBER recession coefficient for long-short portfolio returns sorted on the real estate ratio changes from 2.00 to 1.82 once we switch from the old to the new CRSP tape (see Panel A in Table E.2). For this sorting variable and methodological path, the CRSP changes potentially reverse our judgment that the real estate ratio premium is counter-cyclical at the 5% significance level.[34]

The result that CRSP changes can potentially affect specific conditional premia is not specific to the NBER recession dummy but also pertains to alternative business cycle predictors. We follow the literature, i.e., Cooper and Maio (2019), and *predict* various time series of long-short portfolio returns for the same sorting variable that differ in their methodological sorting choices made to estimate them (see Panel B of Table E.1). Using

[33]Note that Tuzel (2010) provides a risk-based explanation for the real estate ratio premium as stocks with high real estate ratios are more vulnerable to productivity shocks and, thus, have higher expected returns. Intuitively, productivity shocks correlate with the business cycle, such that the real-estate premium is counter-cyclical.

[34]The real estate ratio covers, on average, only around 30% of all observations in our sample, which is one of the lowest coverages across all yearly accounting variables. This low coverage leaves fewer stocks in the long-short portfolio compared to other sorting variables, such that we observe more pronounced differences. However, we still observe, on average, with 466 stocks, a large number of stocks in each monthly long-minus-short portfolio sorted on the real estate ratio.

consumption growth as a cyclical predictor, we note for the composite share issuance variable from Pontiff and Woodgate (2008) that 4.7% (3.7%) of consumption growth coefficients for all portfolio returns generated by varying methodological choices change the significance at the 5% (1%) level. These changed coefficients make up around 9% (21%) of all significant consumption growth coefficients for all composite share issuance-sorted portfolios, which differ in the methodological sorting choices.[35]

Overall, our results suggest that the changes in monthly CRSP returns alter monthly portfolio returns by far more (10% at the 10 bp level) than they change monthly stock returns between both CRSP tapes (around 10% at the 1 bp level). However, these changes in monthly long-short portfolio returns do not lead to meaningful differences in the size and significance of return premia (the time-series average of these monthly long-short portfolio returns). While the changes from the old to the new CRSP tape are mostly due to reinvesting payouts on the ex-date instead of the month-end, we also confirm that reinvesting payouts on the effective payment date does not materially alter premia. Thus, our results provide reassurance for asset pricing studies using the old CRSP tape that their unconditional premia should be almost the same when using the new CRSP tape (with payouts reinvested on the ex-date) or when shifting the reinvestment timing from the ex-date to the payment date. However, we cannot generalize these findings to all asset pricing applications, as we observe meaningful differences in cyclicality estimates for specific sorting variables, e.g., for the real estate ratio. Therefore, we cannot rule out that specific asset pricing applications with specific methodological choices might yield significantly different conclusions about the cyclicality of premia.

[35] A potential reason for these changes is that the composite share issuance ratio uses, among all return-based sorting variables, the longest time series of monthly returns for its construction (five years). This might alter the ranks of stocks sorted on the composite share issuance ratio considerably, leading to differences in portfolio returns beyond aggregating monthly stock returns into portfolios.

IV. Discussion

Our results from Section III raise the question of why monthly return changes between the old and new tape lead to considerable differences for monthly long-short portfolio returns but not for their time-series averages (i.e., premia). Intuitively, the difference in monthly portfolio returns from the new vs. the old tape is positive (negative) if the compounded returns from the ex-date to the end of the same month are positive (negative). These compounded returns represent the differences in reinvestment returns of payouts between the old and the new tape for each month. Furthermore, conditional on the reinvestment return differences, return differences between the old and the new tape increase with the payout yield. However, if the average reinvestment return differences are 0, the return differences for these stocks might cancel out once we aggregate them into portfolios. Thus, the direction of the differences in portfolio returns depends on whether stocks with large payouts have payout yields that correlate positively or negatively with the reinvestment return differences of payouts. Thus, we inspect in each month the cross-sectional correlation between the payout yield and the compounded return from the ex-date to the end of the ex-date month.

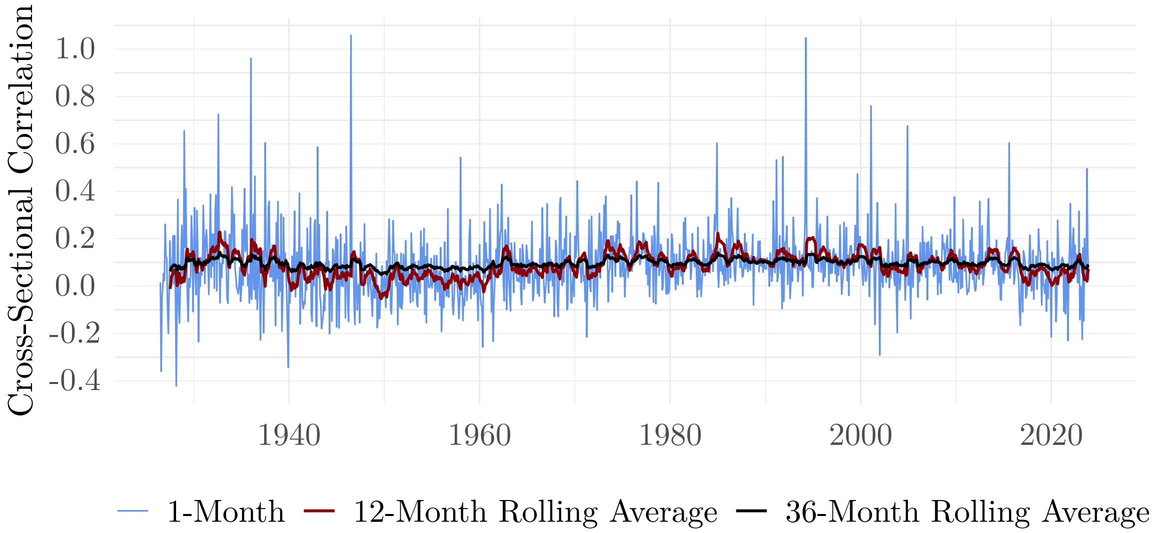
Figure 4 shows that these cross-sectional correlations between the payout yield and the compounded return from the ex-date to the end of the ex-date month range in 90% of the months from -21% to 21% around the mean cross-sectional correlation. However, there is little persistence in the direction of these rank correlations over time. Although these cross-sectional correlations can be either positive or negative over a year, we observe, on average, little variation around a 3-year rolling window. Thus, while cross-sectional correlations between both determinants of CRSP changes - the payout yield and the compounded return from the ex-dates to the end of the ex-date month -

can be large in each month, they are not systematic over time. This might explain why we observe considerable differences in monthly long-short returns that do not translate into differences in their time-series averages (i.e., premia).

FIGURE 4

Cross-Sectional Correlations for Components of Return Changes

Figure 4 shows the time series of cross-sectional correlations between the payout yield and the compounded return from the ex-date to the end of the ex-date month. We compute these cross-sectional correlations in each month based on the observations that have absolute return differences exceeding 0.1 bp between monthly returns from the new (CIZ) and the old (SIZ) tape of CRSP. The red line shows the 12-month rolling average of these cross-sectional correlations, and the black line shows the 36-month rolling average.



We observe a similar picture for monthly long-short portfolio returns in Section III. Across all groups of sorting variables and their methodological paths, we find almost an equal share of monthly long-short returns with positive and negative differences when constructed based on the new versus the old tape (Table 2). Thus, considerable differences in specific months tend to cancel out once we average over many decades.

While we show that the CRSP changes from the old to the new tape do not systematically alter unconditional premia, we cannot generalize this result to all asset pricing applications based on monthly returns. For example, we find that changing from the old to the new CRSP tape alters our assessment of whether the real estate ratio is counter-cyclical based on a 5% significance level. While only a few conditional premia change, those that are sensitive to the CRSP changes have a low data coverage for their sorting variable (real-estate ratio) or require cumulative monthly returns over many years for their construction (composite share issuance).

Furthermore, the CRSP changes imply the following: First, CRSP changes impact the common procedure to compute dividends from the difference between monthly returns including payouts (old CRSP item "ret") and monthly returns excluding payouts (old CRSP item "retx").[36] Inferring dividends directly from monthly returns in the new tape is no longer possible as dividends are reinvested on the ex-date. Thus, differences between the new CRSP item "MthRet" and "MthRetx" reflect payouts and reinvestment proceeds. Second, although monthly returns in the new tape include delisting returns, CRSP does not adjust for a delisting bias following *Shumway* (1997). Thus, to adjust for a delisting bias, it might still be necessary to construct monthly returns by compounding daily returns and including specific delisting assumptions.

V. Conclusion

From January 2025, CRSP no longer updates the Flat File Format 1.0 (SIZ) commonly used as a data source for the returns of U.S. listed stocks. Instead, CRSP started to provide its new Flat File Format 2.0 (CIZ) in 2022. This transition primarily changes monthly stock returns, potentially affecting the results of approximately 34% of all articles published by the Top5 finance journals over the last 25 years.

With the new release, CRSP overwrote the history of 9.6% of monthly stock return observations, which significantly differ by at least 1 bp between the old and the new tape.

[36]See, for instance, Boudoukh, Michaely, Richardson, and Roberts (2007) or, more recently, van Binsbergen, Ma, and Schwert (2022).

The vast share of these differences arises because payouts are reinvested on the ex-date in the new tape as opposed to the end of the ex-date month in the old tape. Although infrequent, we find exceptionally large differences for observations with trading gaps, new listings, and delistings. While reinvesting payouts on the ex-date might not be realistic for some investors, we find that shifting the reinvestment of payouts from the ex-date to the subsequent payment date even rewrites 14.9% of monthly returns by at least 1 bp.

We investigate how these changes for monthly returns affect asset pricing studies by examining a widespread procedure using monthly CRSP returns, that is, portfolio sorts. Changing from the old to the new tape alters around 11.43% of monthly long-short portfolio returns by at least 10 bp with larger differences during recession months. Reassuringly, these sizable positive or negative differences in monthly long-short portfolio returns cancel out over longer horizons, leaving the size and significance of their time-series averages, i.e., premia, virtually unchanged. This also holds if we switch from the new tape to a modified new tape for which we only shift the reinvestment of payouts from the ex-date (new tape) to the payment date.

Although changing from the old to the new tape does not systematically alter unconditional premia, we find that switching from the old to the new tape can alter our conclusion about the cyclicality for a few specific premia. Thus, while changes of monthly CRSP returns unlikely affect studies investigating average effects over long time period, they might lead to differences for a few premia conditional on specific time periods, such as during economic downturns.

References

Abarbanell, J. S., and B. J. Bushee. "Abnormal Returns to a Fundamental Analysis Strategy." *The Accounting Review*, 73 (1998), 19–45.

Akey, P.; A. Z. Robertson; and M. Simutin. "The Noisy Factors? The Retroactive Impact of Methodological Changes on the Fama–French Factors." *Review of Finance,* forthcoming (2024).

Altman, E. I. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *The Journal of Finance*, 23 (1968), 589–609.

Amihud, Y. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5 (2002), 31–56.

Ang, A.; R. J. Hodrick; Y. Xing; and X. Zhang. "The Cross-Section of Volatility and Expected Returns." *The Journal of Finance*, 61 (2006), 259–299.

Balakrishnan, K.; E. Bartov; and L. Faurel. "Post Loss/Profit Announcement Drift." *Journal of Accounting and Economics*, 50 (2010), 20–41.

Bali, R., and G. L. Hite. "Ex Dividend Day Stock Price Behavior: Discreteness or Tax-Induced Clienteles?" *Journal of Financial Economics*, 47 (1998), 127–159.

Bali, T. G.; N. Cakici; and R. F. Whitelaw. "Maxing Out: Stocks as Lotteries and the Cross-Section of Expected Returns." *Journal of Financial Economics*, 99 (2011), 427–446.

Bali, T. G.; R. F. Engle; and S. Murray. *Empirical Asset Pricing: The Cross Section of Stock Returns.* John Wiley & Sons (2016).

Ball, R.; J. Gerakos; J. T. Linnainmaa; and V. Nikolaev. "Accruals, Cash Flows, and Operating Profitability in the Cross Section of Stock Returns." *Journal of Financial Economics*, 121 (2016), 28–45.

Banz, R. W. "The Relationship Between Return and Market Value of Common Stocks." *Journal of Financial Economics*, 9 (1981), 3–18.

Barbee Jr, W. C.; S. Mukherji; and G. A. Raines. "Do Sales–Price and Debt–Equity Explain Stock Returns Better Than Book–Market and Firm Size?" *Financial Analysts Journal*, 52 (1996), 56–60.

Barclay, M. J. "Dividends, Taxes, and Common Stock Prices: The Ex-Dividend Day Behavior of Common Stock Prices Before the Income Tax." *Journal of Financial Economics*, 19 (1987), 31–44.

Basu, S. "The Relationship Between Earnings' Yield, Market Value and Return for NYSE Common Stocks: Further Evidence." *Journal of Financial Economics*, 12 (1983), 129–156.

Beedles, W. L., and M. A. Simkowitz. "A Note on Skewness and Data Errors." *The Journal of Finance*, 33 (1978), 288–292.

Belo, F.; X. Lin; and S. Bazdresch. "Labor Hiring, Investment, and Stock Return Predictability in the Cross Section." *Journal of Political Economy*, 122 (2014), 129–177.

Bennin, R. "Error Rates in CRSP and COMPUSTAT: A Second Look." *The Journal of Finance*, 35 (1980), 1267–1271.

Berg, F.; K. Fabisik; and Z. Sautner. "Is History Repeating Itself? The (Un)Predictable

Past of ESG Ratings." Working Paper, Massachusetts Institute of Technology, Frankfurt School of Finance and Management (2021).

Berkman, H., and P. D. Koch. "DRIPs and the Dividend Pay Date Effect." *Journal of Financial and Quantitative Analysis*, 52 (2017), 1765–1795.

Berninger, M.; F. Kiesel; D. Schiereck; and E. Gaar. "Citations and the Readers' Information-Extracting Costs of Finance Articles." *Journal of Banking and Finance*, 131 (2021), 106188.

Berninger, M.; F. Kiesel; and J. Schnitzler. "Commercial Data in Financial Research." *Review of Corporate Finance*, 4 (2024), 293–335.

Bessembinder, H.; T.-F. Chen; G. Choi; and K. C. J. Wei. "How Should Investors' Long-Term Returns Be Measured?" *Financial Analysts Journal*, 81 (2025), 33–62.

Bhandari, L. C. "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence." *The Journal of Finance*, 43 (1988), 507–528.

Blitz, D.; J. Huij; and M. Martens. "Residual Momentum." *Journal of Empirical Finance*, 18 (2011), 506–521.

Boudoukh, J.; R. Michaely; M. Richardson; and M. R. Roberts. "On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing." *The Journal of Finance*, 62 (2007), 877–915.

Bradshaw, M. T.; S. A. Richardson; and R. G. Sloan. "The Relation Between Corporate Financing Activities, Analysts' Forecasts and Stock Returns." *Journal of Accounting and Economics*, 42 (2006), 53–85.

Brennan, M. J.; T. Chordia; and A. Subrahmanyam. “Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns.” *Journal of Financial Economics*, 49 (1998), 345–373.

Campbell, J. A., and W. Beranek. “Stock Price Behavior on Ex-Dividend Dates.” *The Journal of Finance*, 10 (1955), 425–429.

Chan, L. K.; N. Jegadeesh; and J. Lakonishok. “Momentum Strategies.” *The Journal of Finance*, 51 (1996), 1681–1713.

Chan, L. K.; J. Lakonishok; and T. Sougiannis. “The Stock Market Valuation of Research and Development Expenditures.” *The Journal of Finance*, 56 (2001), 2431–2456.

Chay, J. B.; D. Choi; and J. Pontiff. “Market Valuation of Tax-Timing Options: Evidence from Capital Gains Distributions.” *The Journal of Finance*, 61 (2006), 837–865.

Chen, A. Y., and T. Zimmermann. “Open Source Cross-Sectional Asset Pricing.” *Critical Finance Review*, 27 (2022), 207–264.

Cooper, I., and P. Maio. “New Evidence on Conditional Factor Models.” *Journal of Financial and Quantitative Analysis*, 54 (2019), 1975–2016.

Cooper, M. J.; H. Gulen; and M. J. Schill. “Asset Growth and the Cross-Section of Stock Returns.” *The Journal of Finance*, 63 (2008), 1609–1651.

Courtenay, S. M., and S. B. Keller. “Errors in Databases Revisited: An Examination of the CRSP Shares-Outstanding Data.” *The Accounting Review*, 69 (1994), 285–291.

Daniel, K., and S. Titman. “Market Reactions to Tangible and Intangible Information.” *The Journal of Finance*, 61 (2006), 1605–1643.

Datar, V. T.; N. Y. Naik; and R. Radcliffe. "Liquidity and Stock Returns: An Alternative Test." *Journal of Financial Markets*, 1 (1998), 203–219.

Davis, J. L.; E. F. Fama; and K. R. French. "Characteristics, Covariances, and Average Returns: 1929 to 1997." *The Journal of Finance*, 55 (2000), 389–406.

De Bondt, W. F., and R. Thaler. "Does the Stock Market Overreact?" *The Journal of Finance*, 40 (1985), 793–805.

Desai, H.; S. Rajgopal; and M. Venkatachalam. "Value-Glamour and Accruals Mispricing: One Anomaly or Two?" *The Accounting Review*, 79 (2004), 355–385.

Dichev, I. D. "Is the Risk of Bankruptcy a Systematic Risk?" *The Journal of Finance*, 53 (1998), 1131–1147.

Drechsler, Q. F. S. "Comment on CRSP CIZ Monthly Return." Working Paper, University of Pennsylvania Wharton Research Data Services (2025).

Fama, E. F., and K. R. French. "The Cross-Section of Expected Stock Returns." *The Journal of Finance*, 47 (1992), 427–465.

Fama, E. F., and K. R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Fama, E. F., and K. R. French. "Multifactor Explanations of Asset Pricing Anomalies." *The Journal of Finance*, 51 (1996), 55–84.

Fama, E. F., and K. R. French. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics*, 116 (2015), 1–22.

Fama, E. F., and K. R. French. "Production of U.S. Rm-Rf, SMB, and HML in the Fama–French Data Library." Working Paper, University of Chicago, Dartmouth College Tuck School of Business (2023).

Fama, E. F., and J. D. MacBeth. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy*, 81 (1973), 607–636.

Foster, G.; C. Olsen; and T. Shevlin. "Earnings Releases, Anomalies, and the Behavior of Security Returns." *The Accounting Review*, 59 (1984), 574–603.

Francis, J.; R. LaFond; P. M. Olsson; and K. Schipper. "Costs of Equity and Earnings Attributes." *The Accounting Review*, 79 (2004), 967–1010.

Frazzini, A., and L. H. Pedersen. "Betting Against Beta." *Journal of Financial Economics*, 111 (2014), 1–25.

Gao, X., and J. R. Ritter. "The Marketing of Seasoned Equity Offerings." *Journal of Financial Economics*, 97 (2010), 33–52.

George, T. J., and C.-Y. Hwang. "The 52-Week High and Momentum Investing." *The Journal of Finance*, 59 (2004), 2145–2176.

Gillan, S. L.; J. C. Hartzell; A. Koch; and L. T. Starks. "Getting the Incentives Right: Backfilling and Biases in Executive Compensation Data." *The Review of Financial Studies*, 31 (2018), 1460–1498.

Gormsen, N. J., and E. Lazarus. "Duration-Driven Returns." *The Journal of Finance*, 78 (2023), 1393–1447.

Green, J.; J. R. Hand; and X. F. Zhang. "The Supraview of Return Predictive Signals." *Review of Accounting Studies*, 18 (2013), 692–730.

Guenther, D. A., and A. J. Rosman. “Differences Between COMPUSTAT and CRSP SIC Codes and Related Effects on Research.” *Journal of Accounting and Economics*, 18 (1994), 115–128.

Hafzalla, N.; R. Lundholm; and E. Matthew Van Winkle. “Percent Accruals.” *The Accounting Review*, 86 (2011), 209–236.

Hahn, J., and H. Lee. “Financial Constraints, Debt Capacity, and the Cross-Section of Stock Returns.” *The Journal of Finance*, 64 (2009), 891–921.

Hartzmark, S. M., and D. H. Solomon. “The Dividend Disconnect.” *The Journal of Finance*, 74 (2019), 2153–2199.

Haugen, R. A., and N. L. Baker. “Commonality in the Determinants of Expected Stock Returns.” *Journal of Financial Economics*, 41 (1996), 401–439.

Hirshleifer, D.; K. Hou; S. H. Teoh; and Y. Zhang. “Do Investors Overvalue Firms with Bloated Balance Sheets?” *Journal of Accounting and Economics*, 38 (2004), 297–331.

Hou, K.; C. Xue; and L. Zhang. “Digesting Anomalies: An Investment Approach.” *The Review of Financial Studies*, 28 (2015), 650–705.

Hou, K.; C. Xue; and L. Zhang. “Replicating Anomalies.” *The Review of Financial Studies*, 33 (2020), 2019–2133.

Hribar, P., and D. W. Collins. “Errors in Estimating Accruals: Implications for Empirical Research.” *Journal of Accounting Research*, 40 (2002), 105–134.

Huang, A. G. “The Cross Section of Cashflow Volatility and Expected Stock Returns.” *Journal of Empirical Finance*, 16 (2009), 409–429.

Jegadeesh, N. "Evidence of Predictable Behavior of Security Returns." *The Journal of Finance*, 45 (1990), 881–898.

Jegadeesh, N., and J. Livnat. "Revenue Surprises and Stock Returns." *Journal of Accounting and Economics*, 41 (2006), 147–171.

Jensen, T. I.; B. Kelly; and L. H. Pedersen. "Is There a Replication Crisis in Finance?" *The Journal of Finance*, 78 (2023), 2465–2518.

Kahle, K. M., and R. A. Walkling. "The Impact of Industry Classifications on Financial Research." *Journal of Financial and Quantitative Analysis*, 31 (1996), 309–335.

Kaplan, S. N., and L. Zingales. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *The Quarterly Journal of Economics*, 112 (1997), 169–215.

Lakonishok, J.; A. Shleifer; and R. W. Vishny. "Contrarian Investment, Extrapolation, and Risk." *The Journal of Finance*, 49 (1994), 1541–1578.

Lakonishok, J., and T. Vermaelen. "Tax-Induced Trading Around Ex-Dividend Days." *Journal of Financial Economics*, 16 (1986), 287–319.

Lamont, O.; C. Polk; and J. Saá-Requejo. "New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index." *The RAND Journal of Economics*, 32 (2001), 101–128.

Lev, B., and D. Nissim. "Taxable Income, Future Earnings, and Equity Values." *The Accounting Review*, 79 (2004), 1039–1074.

Liu, G. "Data Quality Problems Troubling Business and Financial Researchers:

A Literature Review and Synthetic Analysis." *Journal of Business & Finance Librarianship*, 25 (2020), 315–371.

Ljungqvist, A.; C. Malloy; and F. Marston. "Rewriting History." *The Journal of Finance*, 64 (2009), 1935–1960.

Lyandres, E.; L. Sun; and L. Zhang. "The New Issues Puzzle: Testing the Investment-Based Explanation." *The Review of Financial Studies*, 21 (2008), 2825–2855.

McEnally, R. W. "A Note on the Return Behavior of High Risk Common Stocks." *The Journal of Finance*, 29 (1974), 199–202.

Menkveld, A. J.; A. Dreber; F. Holzmeister; J. Huber; M. Johanneson; M. Kirchler; M. Razen; U. Weitzel; D. Abad; M. M. Abudy; and Others. "Nonstandard Errors." *The Journal of Finance*, 79 (2024), 2339–2390.

Newey, W. K., and K. D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (1987), 703–708.

Novy-Marx, R. "Operating Leverage." *Review of Finance*, 15 (2011), 103–134.

Novy-Marx, R. "The Other Side of Value: The Gross Profitability Premium." *Journal of Financial Economics*, 108 (2013), 1–28.

Ohlson, J. A. "Financial Ratios and the Probabilistic Prediction of Bankruptcy." *Journal of Accounting Research*, 18 (1980), 109–131.

Patton, A. J.; T. Ramadorai; and M. Streatfield. "Change You Can Believe in? Hedge Fund Data Revisions." *The Journal of Finance*, 70 (2015), 963–999.

Penman, S. H.; S. A. Richardson; and I. Tuna. "The Book-to-Price Effect in Stock Returns: Accounting for Leverage." *Journal of Accounting Research*, 45 (2007), 427–467.

Pontiff, J., and A. Woodgate. "Share Issuance and Cross-Sectional Returns." *The Journal of Finance*, 63 (2008), 921–945.

Richardson, S. A.; R. G. Sloan; M. T. Soliman; and I. Tuna. "Accrual Reliability, Earnings Persistence and Stock Prices." *Journal of Accounting and Economics*, 39 (2005), 437–485.

Robertson, A. Z.; P. Akey; and M. Simutin. "Noisy Factors in Law." *The University of Chicago Law Review*, 92 (2025), 769–832.

Rosenberg, B., and M. Houglet. "Error Rates in CRSP and Compustat Data Bases and Their Implications." *The Journal of Finance*, 29 (1974), 1303–1310.

Scheuch, C.; S. Voigt; and P. Weiss. *Tidy Finance with R*. Chapman and Hall/CRC (2023).

Scheuch, C.; S. Voigt; P. Weiss; and C. Frey. *Tidy Finance with Python*. Chapman and Hall/CRC (2024).

Shumway, T. "The Delisting Bias in CRSP Data." *The Journal of Finance*, 52 (1997), 327–340.

Shumway, T., and V. A. Warther. "The Delisting Bias in CRSP's Nasdaq Data and Its Implications for the Size Effect." *The Journal of Finance*, 54 (1999), 2361–2379.

Sloan, R. G. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings?" *The Accounting Review*, 71 (1996), 289–315.

Soebhag, A.; B. Van Vliet; and P. Verwijmeren. "Non-Standard Errors in Asset Pricing: Mind Your Sorts." *Journal of Empirical Finance*, 78 (2024), 101517.

Soliman, M. T. "The Use of DuPont Analysis by Market Participants." *The Accounting Review*, 83 (2008), 823–853.

Thomas, J., and F. X. Zhang. "Tax Expense Momentum." *Journal of Accounting Research*, 49 (2011), 791–821.

Thomas, J. K., and H. Zhang. "Inventory Changes and Future Returns." *Review of Accounting Studies*, 7 (2002), 163–187.

Titman, S.; K. J. Wei; and F. Xie. "Capital Investments and Stock Returns." *Journal of Financial and Quantitative Analysis*, 39 (2004), 677–700.

Tuzel, S. "Corporate Real Estate Holdings and the Cross-Section of Stock Returns." *The Review of Financial Studies*, 23 (2010), 2268–2302.

van Binsbergen, J. H.; L. Ma; and M. Schwert. "The Factor Multiverse: The Role of Interest Rates in Factor Return Measurement and Discovery." Working Paper, University of Pennsylvania Wharton School, University of South Carolina Darla Moore School of Business, AQR Arbitrage (2022).

Walter, D.; R. Weber; and P. Weiss. "Methodological Uncertainty in Portfolio Sorts." Working Paper, University of Konstanz, Goethe University Frankfurt, Reykjavik University (2024).

Whited, T. M., and G. Wu. "Financial Constraints Risk." *The Review of Financial Studies*, 19 (2006), 531–559.

Xing, Y. "Interpreting the Value Effect Through the Q-Theory: An Empirical Investigation." *The Review of Financial Studies*, 21 (2008), 1767–1795.

Appendix

A. Summary Statistics

TABLE A.1

Distribution of the Trading Days Between the Payment Date, the Ex-Date, and the End of the Ex-Date Month

Table A.1 reports summary statistics for the distribution of trading days between the last trading date of the ex-date month and the ex-date for all payouts in Panel A. Panel B reports summary statistics for the distribution of trading days between the payment date and the ex-date for all payouts. Since the payment date can be in the month after the ex-date month, we split all payouts based on whether the ex-date month matches the payment date month. Thus, Panel B1 reports summary statistics for the distribution of trading days between the payment date and the ex-date for all payouts for which the payment date is in the same month as the ex-date. In contrast, Panel B2 reports summary statistics for the distribution of trading days between the payment date and the ex-date for those payouts where the payment date is in the next month(s) compared to the ex-date. Finally, we report the distribution of trading days between the payment date and the last trading date of the ex-date month for all payouts in Panel C. For each of these distributions we report the observations ("Obs."), the 5th, 10th, 25th, 50th, 75th, 90th, 95th quantile ("Q95"), the mean ("Mean"), and the standard deviation ("St.Dev."). The sample includes only the 555,361 dividends listed in CRSP with distribution codes (DISTCD) starting with 1 from July 1926 until December 2023.

Panel A. Trading Days Between the End of the Ex-Date Month and the Ex-Date

Obs.	Q5	Q10	Q25	Q50	Q75	Q90	Q95	Mean	St.Dev.
555,361	1.00	2.00	6.00	16.00	22.00	26.00	28.00	14.40	8.98

Panel B. Trading Days Between the Payment Date and the Ex-Date

Obs.	Q5	Q10	Q25	Q50	Q75	Q90	Q95	Mean	St.Dev.
555,361	13.00	15.00	17.00	21.00	27.00	33.00	36.00	22.58	13.99

Panel B1. Trading Days Between the Payment Date and the Ex-Date for Payment Dates in the Same Month as the Ex-Date

Obs.	Q5	Q10	Q25	Q50	Q75	Q90	Q95	Mean	St.Dev.
146,682	10.00	12.00	15.00	17.00	21.00	24.00	26.00	17.66	4.67

Panel B2. Trading Days Between the Payment Date and the Ex-Date for Payment Dates in the Month After the Ex-Date Month

Obs.	Q5	Q10	Q25	Q50	Q75	Q90	Q95	Mean	St.Dev.
408,679	14.00	16.00	19.00	23.00	29.00	35.00	37.00	24.34	15.69

Panel C. Trading Days Between the Payment Date and the End of the Ex-Date Month

Obs.	Q5	Q10	Q25	Q50	Q75	Q90	Q95	Mean	St.Dev.
555,361	-10.00	-5.00	0.00	6.00	15.00	22.00	31.00	8.18	16.52

B. Differences Between the CRSP Tapes

TABLE B.1

Differences in CRSP Items for Monthly and Daily Stock Data

Table B.1 reports differences in data items reported by CRSP in the old (SIZ) tape relative to the new (CIZ) tape. We restrict our sample to common stocks listed on the NYSE, AMEX, and NASDAQ from July 1926 until December 2023. Panel A reports these differences for monthly data and Panel B for daily data. We report the number of observations ("Obs.") in each tape, the amount and the share of observations (in percent) that are different between both CRSP tapes ("Obs. Diff" and "Share Diff."), and the number of observations that are new or lost in the new tape ("Obs. new in CIZ" and "Obs. not in CIZ"). We count observations as different if they either exceed our tolerance level (no exact match for non-numeric variables) or are missing in one and non-missing in the other tape. We use a tolerance level of 0.1 bp for all return variables and stock prices. For trading volume and shares outstanding, we use a tolerance level of 1. Moreover, we calculate the average difference across all observations ("Mean diff.") if the absolute difference exceeds the tolerance level for numeric items. We scale the mean absolute differences with the mean level of each item to put these differences into perspective ("Ratio").

Panel A. Monthly CRSP Data

CRSP Variable	SIZ Obs.	CIZ Obs.	Obs. Diff.	Share Diff.	Mean Diff.	Ratio	Obs. new in CIZ	Obs. not in CIZ
Trading Volume	3,455,439	3,455,929	1,743,577	46.4	7213.1	0.083	490	0
Stock Return	3,710,761	3,713,871	474,580	12.8	0.1638	20.57	3,922	812
Stock Return exc. Div.	3,710,760	3,713,871	10,035	0.27	10.84	1078.6	3,923	812
Stock Price	3,741,978	3,746,133	4,168	0.11	0.0114	0.030	4,155	0
Shares Outstanding	3,760,171	3,760,051	587	0.02	1.0000	0.002	0	120
CUSIP	3,251,081	3,251,081	0	0.00			0	0
SIC Code	3,760,171	3,760,171	7	0.00			0	0
Ticker	3,374,813	3,374,813	0	0.00			0	0
Company Name	3,760,171	3,760,171	1	0.00			0	0
Share Class	216,380	216,380	0	0.00			0	0
Delisting Return	2,660	2,579	105	0.00			12	93
VW-index incl. Div.	1,176	1,176	0	0.00			0	0
VW-index excl. Div.	1,176	1,176	0	0.00			0	0
EW-index incl. Div.	1,176	1,176	0	0.00			0	0
EW-index excl. Div.	1,176	1,176	0	0.00			0	0

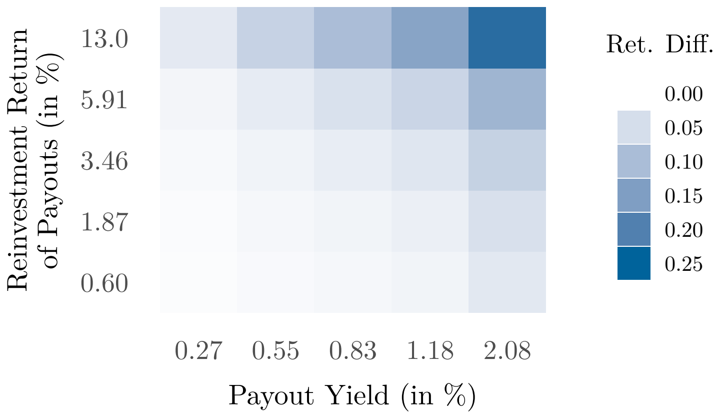
Panel B. Daily CRSP Data

CRSP Variable	SIZ Obs.	CIZ Obs.	Obs. Diff.	Share Diff.	Mean Diff.	Ratio	Obs. new in CIZ	Obs. not in CIZ
Shares Outstanding	79,554,008	79,551,681	12,573	0.02	1.0000	0.002	1	2,328
Trading Volume	73,123,769	73,123,773	4	0.00			4	0
Share Price	79,152,909	79,152,909	334	0.00	0.0120	0.031	0	0
Stock Return	79,123,110	79,123,110	1	0.00	0.0012	1.477	0	0
Stock Return exc. Div.	79,123,110	79,123,110	1	0.00	0.0012	1.616	0	0
VW-index incl. Div.	25,798	25,798	0	0.00			0	0
VW-index excl. Div.	25,798	25,798	0	0.00			0	0
EW-index incl. Div.	25,798	25,798	0	0.00			0	0
EW-index excl. Div.	25,798	25,798	0	0.00			0	0

FIGURE B.1

Dissecting Reinvestment Return Differences

Figure B.1 shows the average absolute return differences in percent ("Ret. Diff.") in each of the 25 double-sorted buckets with increasing color intensity. We assign observations with monthly return differences exceeding 0.1 bp between the new (CIZ) and old (SIZ) CRSP tape into quintiles based on their payout yield and the stock's absolute compounded return from the ex-date to the month-end ("Reinvestment Return of Payouts"). We replace missing daily returns for the computation of the reinvestment return with zero. Based on these independent double sorts, we assign stocks into 25 buckets and depict the absolute return differences for each bucket by color intensity. We show the mean values for the payout yield (horizontal axis) and reinvestment return of payouts (vertical axis) for each bucket in percent. We include all common stocks listed on AMEX, NYSE, and NASDAQ from 1926 until 2023.



C. Sorting Variables

TABLE C.1

List of 68 Sorting Variables

Table C.1 reports the 68 sorting variables used throughout the paper. We document the group, data frequency ("Freq."), abbreviation ("Abb."), description, original reference paper, and data availability for all 68 sorting variables. Additionally, a superscript C indicates that the variable construction requires CRSP data, and a superscript R indicates that the construction requires monthly returns from CRSP. We provide the construction details for all sorting variables in the Internet Appendix IA.VIII.

Group	Freq.	Abb.	Description	Original Reference Paper
	yearly	CDI	Composite Debt Issuance	Lyandres, Sun, and Zhang (2008)
	monthly	CSI^{C,R}	Composite Share Issuance	Daniel and Titman (2006)
	yearly	DBE	Change in Common Equity	Richardson, Sloan, Soliman, and Tuna (2005)
Financing	yearly	DCOL	Change in Current Operating Liabilities	Richardson et al. (2005)
	yearly	DFNL	Change in Financial Liabilities	Richardson et al. (2005)
	yearly	NDF	Net Debt Financing	Bradshaw, Richardson, and Sloan (2006)
	yearly	NEF	Net Equity Financing	Bradshaw et al. (2006)
	yearly	NXF	Net External Financing	Bradshaw et al. (2006)
	yearly	ADM^{C}	Advertisement Expenses to Market Equity	Chan, Lakonishok, and Sougiannis (2001)
	quarterly	CFV	Cash-Flow Volatility	Huang (2009)
	yearly	EPRD	Earnings' Predictability	Francis, LaFond, Olsson, and Schipper (2004)
	yearly	HR	Hiring Rate	Belo, Lin, and Bazdresch (2014)

Continued on the next page

Table C.1: (continued)

Group	Freq.	Abb.	Description	Original Reference Paper
Intangibles	yearly	KZI^{C}	Kaplan and Zingales (1997) Index	Lamont, Polk, and Saá-Requejo (2001)
	yearly	LFE	Labor Force Efficiency	Abarbanell and Bushee (1998)
	yearly	OL	Operating Leverage	Novy-Marx (2011)
	yearly	RDM^{C}	R&D Expenses to Market Equity	Chan et al. (2001)
	yearly	RER^{C}	Real Estate Ratio	Tuzel (2010)
	yearly	TAN	Tangibility	Hahn and Lee (2009)
	yearly	WW^{C}	Whited and Wu Index	Whited and Wu (2006)
Investment	yearly	ACI	Abnormal Corporate Investment	Titman, Wei, and Xie (2004)
	yearly	AG	Asset Growth	Cooper, Gulen, and Schill (2008)
	yearly	DNOA	Change in Net Operating Assets	Hirshleifer, Hou, Teoh, and Zhang (2004)
	yearly	DPIA	Change in Property, Plant, Equipment and Inventory to Assets	Lyandres et al. (2008)
	yearly	DWC	Change in Net Non-Cash Working Capital	Richardson et al. (2005)
	yearly	IG	Investment Growth	Xing (2008)
	yearly	DINV	Inventory Changes	Thomas and Zhang (2002)
	yearly	NOA	Net Operating Assets	Hirshleifer et al. (2004)
	yearly	OA	Operating Accruals	Sloan (1996)
	yearly	PTA	Percent Total Accruals	Hafzalla, Lundholm, and Matthew Van Winkle (2011)
Momentum	monthly	ABR^{C}	Abnormal Returns on Earnings' Announcements	Chan, Jegadeesh, and Lakonishok (1996)
	monthly	MOM^{C,R}	Return Momentum	Fama and French (1996)
	monthly	RMOM^{C,R}	Residual Momentum	Blitz, Huij, and Martens (2011)
	quarterly	RS	Revenue Surprise	Jegadeesh and Livnat (2006)
	quarterly	SUE	Standardized Unexpected Earnings	Foster, Olsen, and Shevlin (1984)
	quarterly	TES	Tax Expense Surprise	Thomas and Zhang (2011)
	monthly	52W^{C}	52-Week High	George and Hwang (2004)
Profitability	yearly	ATO	Asset Turnover	Soliman (2008)
	yearly	BL	Book Leverage	Fama and French (1992)
	yearly	CBOP	Cash-Based Operating Profitability	Ball, Gerakos, Linnainmaa, and Nikolaev (2016)
	yearly	CTO	Capital Turnover	Haugen and Baker (1996)
	yearly	GPA	Gross Profits to Assets	Novy-Marx (2013)
	yearly	O	Ohlson (1980) O-Score	Dichev (1998)
	yearly	OPE	Operating Profits to Book Equity	Fama and French (2015)
	quarterly	ROA	Return on Assets	Balakrishnan, Bartov, and Faurel (2010)
	quarterly	ROE	Return on Equity	Hou et al. (2015)
	yearly	TBI	Taxable Income to Book Income	Lev and Nissim (2004)
	yearly	Z^{C}	Altman (1968) Z-Score	Dichev (1998)
Size	monthly	ME^{C}	Logarithm of Market Equity	Banz (1981)
Trading-	monthly	AMI^{C}	Amihud Illiquidity Measure	Amihud (2002)
	monthly	BETA^{C,R}	Beta Relative to the Market	Fama and MacBeth (1973)
	monthly	BFP^{C}	Frazzini and Pedersen Beta	Frazzini and Pedersen (2014)
	monthly	DTV^{C}	Dollar Trading Volume	Brennan, Chordia, and Subrahmanyam (1998)

Continued on the next page

Table C.1: (continued)

Group	Freq.	Abb.	Description	Original Reference Paper
Frictions	monthly	ISKEW^{C}	Idiosyncratic Skewness	Bali, Engle, and Murray (2016)
	monthly	IVOL^{C}	Idiosyncratic Volatility	Ang, Hodrick, Xing, and Zhang (2006)
	monthly	MDR^{C}	Maximum Daily Return	Bali, Cakici, and Whitelaw (2011)
	monthly	SREV^{C,R}	Short-Term Reversal	Jegadeesh (1990)
	monthly	TUR^{C}	Share Turnover	Datar, Naik, and Radcliffe (1998)
Valuation	yearly	AM^{C}	Assets to Market Equity	Fama and French (1992)
	yearly	BM^{C}	Book Equity to Market Equity	Davis, Fama, and French (2000)
	yearly	CFM^{C}	Cash Flow to Market Equity	Lakonishok, Shleifer, and Vishny (1994)
	yearly	DM^{C}	Debt to Market Equity	Bhandari (1988)
	yearly	EBM^{C}	Enterprise Book Equity to Market Equity	Penman, Richardson, and Tuna (2007)
	yearly	EM^{C}	Earnings to Market Equity	Basu (1983)
	yearly	NDM^{C}	Net Debt to Market Equity	Penman et al. (2007)
	yearly	NPY^{C}	Net Payout Yield	Boudoukh, Michaely, Richardson, and Roberts (2007)
	yearly	OCM^{C}	Operating Cash Flow to Market Equity	Desai, Rajgopal, and Venkatachalam (2004)
	monthly	REV^{C,R}	Long-Term Reversal	De Bondt and Thaler (1985)
	yearly	SM^{C}	Sales to Market Equity	Barbee Jr, Mukherji, and Raines (1996)

D. Differences in the Significance of Premia

TABLE D.1

Significance of Premia Between Tapes for each Sorting Variable

Table D.1 reports the share of premia that change their significance once we switch from the old (SIZ) to the new (CIZ) CRSP tape. We report these shares for each sorting variable ("SV"). To account for methodological uncertainty, we estimate the time-series average of long-short portfolio returns (premium) for each of the 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. For each sorting variable and methodological path, we compute the t-statistic whether the corresponding premium is different from zero based on Newey and West (1987) corrected standard errors. Then, we compute the share of premia estimates that change their significance at the 1% and 5% level across all methodological paths for each sorting variable (column "Change"). Moreover, we report the share of premia estimates across all methodological paths that are significant ("Sig.") or insignificant ("Insig."), averaged over premia from the old and the new tape. All relative frequencies are in percent, and the variables are arranged in descending order by the change at the 1% significance level. We label sorting variables that require CRSP data (monthly CRSP returns) for their construction with a superscript "C" ("R"). Moreover, we provide an overview of all sorting variables in Table C.1 and construction details in the Internet Appendix IA.VIII. The sample period is from 1968 to 2023.

		1%			5%		
SV	Group	Change	Sig.	Insig.	Change	Sig.	Insig.
RER^{C}	Intangibles	2.86	32.46	64.68	2.51	72.70	24.78
CSI^{C,R}	Financing	2.22	93.27	4.51	0.04	99.82	0.14

Continued on the next page

Table D.1. (continued)

SV	Group	1% Change	1% Sig.	1% Insig.	5% Change	5% Sig.	5% Insig.
NPY[C]	Valuation	1.39	25.68	72.93	1.20	75.26	23.53
OCM[C]	Valuation	1.10	41.90	57.00	0.76	82.83	16.40
OPE	Profitability	1.04	49.02	49.94	0.71	80.86	18.42
EM[C]	Valuation	1.00	26.79	72.21	1.05	75.60	23.34
ACI	Investment	0.94	75.40	23.66	0.39	92.97	6.64
MOM[C,R]	Momentum	0.92	47.97	51.11	0.56	63.56	35.88
HR	Intangibles	0.91	45.42	53.67	0.71	66.93	32.36
PTA	Investment	0.89	74.75	24.36	0.49	90.28	9.23
SREV[C,R]	Trading frictions	0.89	53.23	45.87	0.98	78.19	20.83
TES	Momentum	0.89	84.69	14.42	0.35	97.13	2.52
52W[C]	Momentum	0.86	48.48	50.67	0.27	61.60	38.14
NEF	Financing	0.80	30.46	68.74	0.65	62.39	36.96
GPA	Profitability	0.79	35.77	63.44	0.59	59.09	40.31
ROE	Profitability	0.78	77.37	21.84	0.17	93.00	6.83
NDF	Financing	0.78	88.70	10.52	0.24	96.68	3.08
DBE	Financing	0.76	60.92	38.32	0.43	84.87	14.71
OL	Intangibles	0.76	16.08	83.16	1.39	52.09	46.52
DINV	Investment	0.75	80.53	18.72	0.24	96.47	3.29
CFM[C]	Valuation	0.75	27.27	71.98	1.10	64.25	34.65
CDI	Financing	0.69	19.84	79.47	0.92	39.52	59.56
ISKEW[C]	Trading frictions	0.69	11.12	88.19	1.32	20.07	78.61
RMOM[C,R]	Momentum	0.66	67.96	31.38	1.06	81.12	17.82
DTV[C]	Trading frictions	0.66	10.85	88.49	1.77	29.85	68.38
SUE	Momentum	0.65	73.24	26.11	0.34	86.34	13.32
RS	Momentum	0.64	65.76	33.60	0.38	83.77	15.85
ROA	Profitability	0.62	62.48	36.90	0.56	83.96	15.48
NXF	Financing	0.55	81.14	18.31	0.23	92.65	7.12
OA	Investment	0.55	73.23	26.22	0.44	87.59	11.96
DFNL	Financing	0.52	83.96	15.53	0.32	92.90	6.78
SM[C]	Valuation	0.50	25.52	73.98	0.84	61.36	37.80
CFV	Intangibles	0.49	9.77	89.74	1.91	37.93	60.16
AG	Investment	0.48	81.19	18.33	0.27	93.42	6.32
IG	Investment	0.46	87.69	11.84	0.16	95.90	3.94
TAN	Intangibles	0.40	4.49	95.11	0.63	13.80	85.57
CBOP	Profitability	0.39	92.23	7.38	0.07	96.83	3.11
DPIA	Investment	0.38	84.03	15.60	0.22	93.51	6.27
ME[C]	Size	0.37	7.17	92.46	0.64	13.28	86.08
ATO	Profitability	0.37	4.12	95.51	0.89	18.90	80.20
ABR[C]	Momentum	0.35	84.76	14.89	0.27	90.80	8.93
MDR[C]	Trading frictions	0.35	14.55	85.10	0.29	24.17	75.54
IVOL[C]	Trading frictions	0.33	8.67	91.00	0.76	23.87	75.38
DCOL	Financing	0.33	16.64	83.03	0.68	30.28	69.04
AMI[C]	Trading frictions	0.29	5.21	94.50	0.80	15.07	84.13
RDM[C]	Intangibles	0.28	2.59	97.13	1.01	21.44	77.55
REV[C,R]	Valuation	0.27	4.25	95.49	1.11	14.55	84.35
BM[C]	Valuation	0.25	13.83	85.92	0.51	29.89	69.60
ADM[C]	Intangibles	0.17	1.22	98.61	1.18	18.54	80.27
DWC	Investment	0.16	97.64	2.20	0.02	99.71	0.26
NOA	Investment	0.11	98.52	1.37	0.00	99.99	0.01
TUR[C]	Trading frictions	0.09	3.42	96.50	0.27	9.34	90.39

Continued on the next page

Table D.1. (continued)

SV	Group	1%			5%		
		Change	Sig.	Insig.	Change	Sig.	Insig.
EBM^{C}	Valuation	0.08	3.21	96.71	0.21	10.17	89.62
TBI	Profitability	0.06	0.56	99.38	0.47	6.77	92.76
AM^{C}	Valuation	0.06	2.30	97.65	0.26	9.42	90.32
WW^{C}	Intangibles	0.04	0.28	99.68	0.23	1.77	98.00
O	Profitability	0.03	0.15	99.82	0.16	1.97	97.87
DNOA	Investment	0.03	99.50	0.47	0.01	99.98	0.02
LFE	Intangibles	0.02	0.04	99.93	0.23	1.17	98.60
CTO	Profitability	0.02	0.12	99.85	0.25	3.03	96.73
Z^{C}	Profitability	0.01	0.09	99.90	0.05	1.02	98.93
KZI^{C}	Intangibles	0.00	0.01	99.99	0.02	0.20	99.77
EPRD	Intangibles	0.00	100.00	0.00	0.00	100.00	0.00
BL	Profitability	0.00	0.00	100.00	0.00	0.00	100.00
BETA^{C,R}	Trading frictions	0.00	0.00	100.00	0.00	0.00	100.00
BFP^{C}	Trading frictions	0.00	0.00	100.00	0.00	0.00	100.00
DM^{C}	Valuation	0.00	0.00	100.00	0.01	0.08	99.91
NDM^{C}	Valuation	0.00	0.00	100.00	0.02	0.13	99.85

E. Differences in Cyclicality Coefficients

TABLE E.1

Significance of Cyclicality Coefficients Across All Paths

Table E.1 compares cyclicality coefficients for long-short portfolio returns based on the old (SIZ) and new (CIZ) CRSP tape. To account for methodological uncertainty, we estimate long–short portfolio returns for each of the 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. We regress these long-short portfolio returns from SIZ and CIZ data on the contemporaneous NBER recession dummy (Panel A) and the one-month lagged consumption growth (Panel B). Then, we compute the share of regression coefficients that change their significance at the 10%, 5%, and 1% level across all methodological paths for each sorting variable (column “Change”) in percent. Note that we use Newey and West (1987) corrected standard errors. Moreover, we report the average share of methodological paths across the old and the new tape for which we observe significant regression coefficients (“Sig.”) in percent. Lastly, we report the ratio (“Ratio”) of the share of coefficients that change their sign relative to the share of all significant coefficients for each sorting variable across both tapes (in percent). We include only sorting variables for which we observe in at least 25% of the methodological paths t-statistics larger than 1.65 for the respective cyclicality proxy. We label sorting variables that require CRSP data (monthly CRSP returns) for their construction with a superscript “C” (“R”). Table C.1 reports all sorting variables and construction details are in the Internet Appendix IA.VIII. The variables are arranged in descending order by the change at the 10% significance level. The sample period is from 1968 to 2023.

Panel A: NBER Recession Coefficients

SV	Group	10%			5%			1%		
		Change	Sig.	Ratio	Change	Sig.	Ratio	Change	Sig.	Ratio
RER^{C}	Intangibles	4.60	27.98	16.44	3.01	12.42	24.27	0.42	0.93	45.53
CBOP	Profitability	2.32	49.18	4.72	1.82	33.36	5.45	1.16	11.80	9.80

Continued on the next page

Table E.1. (continued)

		10%			5%			1%		
SV	Group	Change	Sig.	Ratio	Change	Sig.	Ratio	Change	Sig.	Ratio
ISKEW[C]	Trading fric.	2.26	49.00	4.61	1.95	34.63	5.63	1.26	13.44	9.36
O	Profitability	2.06	44.49	4.63	1.85	26.32	7.05	0.74	6.33	11.73
BL	Profitability	2.06	78.26	2.63	2.24	57.94	3.87	1.66	17.00	9.73
RS	Momentum	2.04	64.43	3.17	2.19	48.12	4.55	1.84	20.94	8.77
KZI[C]	Intangibles	1.54	56.96	2.71	1.77	34.95	5.06	0.72	9.34	7.70
ATO	Profitability	1.44	51.48	2.80	1.12	33.13	3.37	0.65	7.82	8.29
Z[C]	Profitability	1.37	32.55	4.20	1.32	16.34	8.09	0.21	0.87	24.33
CTO	Profitability	1.18	76.16	1.54	1.31	62.04	2.11	1.10	34.92	3.15
GPA	Profitability	1.06	73.18	1.46	1.34	56.29	2.37	1.16	27.42	4.23
OL	Intangibles	0.74	92.46	0.80	1.24	77.45	1.60	1.76	36.06	4.89
EPRD	Intangibles	0.15	95.66	0.16	0.31	92.28	0.33	1.43	69.06	2.07

Panel B: Consumption Growth Coefficients

		10%			5%			1%		
SV	Group	Change	Sig.	Ratio	Change	Sig.	Ratio	Change	Sig.	Ratio
CSI[C,R]	Financing	3.86	73.27	5.26	4.66	52.48	8.87	3.74	17.49	21.39
IG	Investment	3.00	57.78	5.19	3.21	40.23	7.98	1.62	14.26	11.36
RER[C]	Intangibles	2.33	45.92	5.07	2.24	34.87	6.43	1.56	18.35	8.52
IVOL[C]	Trading fric.	2.23	32.93	6.76	1.74	18.18	9.57	0.90	4.95	18.25
RMOM[C,R]	Momentum	2.13	62.02	3.44	1.64	47.42	3.45	1.22	23.63	5.14
DBE	Financing	2.07	58.48	3.54	2.24	39.72	5.64	1.19	12.80	9.30
NDF	Financing	1.94	79.59	2.44	2.60	68.49	3.79	3.39	41.48	8.17
OPE	Profitability	1.75	40.61	4.31	1.77	29.44	6.01	1.14	12.51	9.08
DNOA	Investment	1.74	59.47	2.92	1.75	43.41	4.03	1.23	15.70	7.83
52W[C]	Momentum	1.67	77.60	2.15	2.13	53.84	3.95	0.92	25.20	3.66
BL	Profitability	1.48	68.49	2.16	1.42	56.68	2.51	1.40	32.05	4.38
WW[C]	Intangibles	1.42	34.79	4.08	1.13	24.61	4.61	0.84	10.83	7.76
CFV	Intangibles	1.38	34.26	4.03	1.02	20.67	4.91	0.44	5.08	8.64
TAN	Intangibles	1.33	44.90	2.95	1.57	34.09	4.61	0.99	15.49	6.42
CBOP	Profitability	1.26	31.98	3.93	1.19	22.86	5.21	0.90	9.28	9.70
DFNL	Financing	1.21	72.47	1.67	1.41	61.60	2.29	1.68	38.43	4.38
MOM[C,R]	Momentum	1.20	30.48	3.95	0.88	15.85	5.55	0.27	2.46	11.06
AMI[C]	Trading fric.	1.17	51.38	2.27	0.96	43.22	2.23	0.96	28.75	3.33
CTO	Profitability	1.06	45.39	2.34	1.17	33.90	3.47	0.73	16.03	4.56
TUR[C]	Trading fric.	1.05	79.00	1.33	1.51	63.61	2.37	1.52	29.00	5.26
ME[C]	Size	1.02	29.63	3.43	0.99	22.19	4.46	0.65	11.02	5.86
DWC	Investment	0.96	33.53	2.87	0.82	22.13	3.69	0.39	7.71	5.09
DTV[C]	Trading fric.	0.89	76.65	1.17	1.02	68.79	1.49	1.33	47.81	2.78
GPA	Profitability	0.87	41.11	2.12	0.98	28.30	3.46	0.48	12.71	3.78
DPIA	Investment	0.87	74.21	1.17	1.50	62.54	2.39	2.11	28.69	7.34
BM[C]	Valuation	0.85	49.24	1.73	0.87	32.42	2.69	0.53	7.69	6.87
DM[C]	Valuation	0.80	42.46	1.89	0.71	30.11	2.36	0.44	10.13	4.34
AG	Investment	0.75	85.39	0.88	1.29	72.72	1.77	2.01	34.98	5.74
DINV	Investment	0.73	74.81	0.97	0.88	64.17	1.37	1.29	42.52	3.03
HR	Intangibles	0.67	82.03	0.82	0.92	72.44	1.27	1.31	50.41	2.60
EBM[C]	Valuation	0.63	45.51	1.39	0.82	29.11	2.81	0.30	7.23	4.10

Continued on the next page

Table E.1. (continued)

SV	Group	10%			5%			1%		
		Change	Sig.	Ratio	Change	Sig.	Ratio	Change	Sig.	Ratio
AM^{C}	Valuation	0.62	71.90	0.86	0.83	60.78	1.37	0.82	34.34	2.39
RS	Momentum	0.59	26.20	2.27	0.43	16.20	2.68	0.26	5.66	4.52
OL	Intangibles	0.57	49.86	1.14	0.94	42.35	2.22	1.22	24.51	4.99
SM^{C}	Valuation	0.28	40.10	0.69	0.38	32.25	1.19	0.45	16.33	2.76

TABLE E.2

Significance of Cyclicality Coefficients for a Specific Path

Table E.2 compares cyclicality coefficients for long-short portfolio returns based on the old (SIZ) and new (CIZ) CRSP tape. We estimate the long-short portfolio returns for all sorting variables ("SV") based on methodological sorting choices from Hou et al. (2020): We include all stocks except for financials and negative book equity stocks, and require two years of minimum listing. Then, we form monthly decile portfolios with NYSE breakpoints and value-weighted returns. We regress these long-short portfolio returns from SIZ and CIZ data on the contemporaneous NBER recession dummy (Panel A) and the one-month lagged consumption growth (Panel B). We report these regression coefficients based on SIZ and CIZ data ("Coef. SIZ" and "Coef. CIZ") and their t-statistics based on Newey and West (1987) ("T-stat SIZ" and "T-stat CIZ"). All coefficients are standardized, except those for the NBER recession dummy. We include only sorting variables with t-statistics larger than 1.65 and label sorting variables that require CRSP data (monthly CRSP returns) for their construction with a superscript "C" ("R"). Table C.1 reports all sorting variables and construction details are in the Internet Appendix IA.VIII. The sample period is from 1968 to 2023.

Panel A: NBER Recession Coefficients

SV	Group	Coef. SIZ	Coef. CIZ	T-stat SIZ	T-stat CIZ
RER^{C}	Intangibles	0.90	0.85	2.00	1.82
$REV^{C,R}$	Valuation	1.49	1.56	1.81	1.91
CTO	Profitability	0.97	0.97	2.06	2.11
EPRD	Intangibles	2.10	2.10	2.58	2.59
OL	Intangibles	1.25	1.25	2.32	2.32
BL	Profitability	1.03	1.04	2.46	2.46

Panel B: Consumption Growth Coefficients

SV	Group	Coef. SIZ	Coef. CIZ	T-stat SIZ	T-stat CIZ
$CSI^{C,R}$	Financing	0.08	0.07	2.78	2.64
HR	Intangibles	0.06	0.06	2.79	2.85
DFNL	Financing	0.05	0.05	1.72	1.76
$MOM^{C,R}$	Momentum	0.05	0.05	1.67	1.67
$52W^{C}$	Momentum	0.09	0.09	3.32	3.32
DM^{C}	Valuation	0.09	0.09	2.75	2.75
$RMOM^{C,R}$	Momentum	0.13	0.13	4.16	4.15
ABR^{C}	Momentum	0.07	0.07	1.68	1.68

F. Overview of Differences in Premia

TABLE F.1

Differences in Return Premia - Overview of Comparisons

Table F.1 reports summary statistics of the differences between return premia constructed from different versions of monthly CRSP returns. Panel A reports differences between premia estimated from the old (SIZ) and the new (CIZ) CRSP tape. Panel B reports differences between premia estimated from the old (SIZ) and the new (CIZ) CRSP tape, only including methodological paths in portfolio sorts that have the least-correlated portfolio returns among each other (2.1% of all methodological paths used in Panel A). Moreover, Panel C reports differences between premia estimated from the new (CIZ) tape (with payouts reinvested on the ex-date) and the modified new tape (with payouts reinvested on the payment date). We start by estimating the time-series averages of monthly long-short portfolio returns (premia) for each of the 68 sorting variables using the respective tape of monthly CRSP returns. We account for methodological uncertainty by estimating these premia across all portfolio-construction paths implied by the 14 methodological decisions illustrated in Figure 3. For each of the three comparisons of monthly CRSP tapes in the respective Panels, we compute the following statistics averaged across methodological paths and groups of sorting variables (row "Overall"): First, the share of premia (in percent) that differ by at least 1 bp ("$N_{1\text{ bp}}$") and 10 bp ("$N_{10\text{ bp}}$"). Second, the absolute difference ("AD") of premia estimates between the two tapes under investigation (in basis points per month). Third, we also scale these absolute differences by the average absolute premium (column "Ratio", in percent). Finally, we report the 90- ("Q90") and 95-percentiles ("Q95") of the absolute premia differences across all methodological paths and groups of sorting variables (in basis points per month). The sample period is from 1968 to 2023.

Panel A. New (CIZ) vs. Old (SIZ) Tape of CRSP

Group	N_{1bp}	N_{10bp}	AD	Ratio	Q90	Q95
Overall	4.67	0.02	0.34	5.95	0.75	1.00

Panel B. New (CIZ) vs. Old (SIZ) Tape of CRSP Only Including the Least-Correlated Paths

Group	N_{1bp}	N_{10bp}	AD	Ratio	Q90	Q95
Overall	7.95	0.04	0.43	3.27	0.96	1.23

Panel C. New (CIZ) vs. Modified New Tape with Payouts Reinvested on the Payment Date

Group	N_{1bp}	N_{10bp}	AD	Ratio	Q90	Q95
Overall	0.32	0.00	0.08	2.30	0.14	0.17

TABLE F.2

Differences in the Significance of Premia - Overview of Comparisons

Table F.2 reports the share of premia that change their significance once we switch between the underlying monthly CRSP tape used to estimate these premia. Panel A reports the share of premia that change their significance when switching from the old (SIZ) to the new (CIZ) CRSP tape. Panel B reports the share of premia that change their significance when switching from the old (SIZ) to the new (CIZ) CRSP tape, only including methodological paths in portfolio sorts that have the least-correlated portfolio returns among each other (2.1% of all methodological paths used in Panel A). Moreover, Panel C reports the share of premia that change their significance when switching from the new (CIZ) tape (with payouts reinvested on the ex-date) to the modified new tape (with payouts reinvested on the payment date). We start by estimating the time-series averages of monthly long-short portfolio returns (premia). For each sorting variable and methodological path, we compute the t-statistic whether the corresponding premium is different from zero based on Newey and West (1987) corrected standard errors. We account for methodological uncertainty by estimating these premia and their significance across all portfolio-construction paths implied by the 14 methodological decisions illustrated in Figure 3. For each of the three comparisons of monthly CRSP tapes in the respective Panels, we compute the following statistics averaged across methodological paths and groups of sorting variables (row "Overall"): First, the share of premia estimates that change their significance at the 1% and 5% level. Second, we report the share of premia estimates that are significant ("Sig.") or insignificant ("Insig.") at the 1% and level. All relative frequencies are in percent. The sample period is from 1968 to 2023.

Panel A. New (CIZ) vs. Old (SIZ) Tape of CRSP

	1%			5%		
Group	Change	Sig.	Insig.	Change	Sig.	Insig.
Overall	0.52	38.52	60.96	0.54	52.69	46.78

Panel B. New (CIZ) vs. Old (SIZ) Tape of CRSP Only Including the Least-Correlated Paths

	1%			5%		
Group	Change	Sig.	Insig.	Change	Sig.	Insig.
Overall	0.49	36.42	63.09	0.50	49.15	50.35

Panel C. New (CIZ) vs. Modified New Tape with Payouts Reinvested on the Payment Date

	1%			5%		
Group	Change	Sig.	Insig.	Change	Sig.	Insig.
Overall	0.12	37.77	58.65	0.12	51.79	44.63

Internet Appendix

Rewriting CRSP's History: Impact of Altered Monthly Returns on Asset Pricing

February 24, 2026

IA.I. CRSP Return Usage in Finance Journals

In this section, we describe the data collection process and analysis for Figure 1 and Figure IA.1 in the Internet Appendix IA.I. The figure motivates the potential relevance of an altered monthly return calculation for the existing empirical finance literature.

In order to see how relevant the CRSP data set is in the empirical finance literature, we downloaded all articles from 5 leading finance journals (Top5), i.e., the Journal of Finance (JF), the Journal of Financial Economics (JFE), the Review of Financial Studies (RFS), the Journal of Financial and Quantitative Analysis (JFQA), and the Review of Finance (RoF) from 2000 until the end of 2024. We do not include Management Science as it is a general interest (Management) journal. We obtained the PDF files of all articles by manually downloading, web-scraping, or downloading via the respective publishers' application programming interfaces (APIs). We downloaded all articles listed on Scopus with a valid digital object identifier (DOI) when searching for the respective journals from 2000 to 2024. This ensures full coverage of all articles, but also leads to downloading various other published non-full article PDFs with a valid DOI, i.e., errata, corrigenda, retraction notices, comments/notes, replies on comments/notes, editorial or publisher notes, editorials, and editorial statistics. We remove all those non-full articles through various screening procedures as common in the literature (see Berninger et al. (2021)). To ensure that the text is readable, we run each PDF through an optical character recognition (OCR) model, compare the quality of the machine-readable OCR text with the regular text import function, and use the better text files for the search tasks. The described procedure leads to an overall sample of 9,130 full articles from the Top5 finance journals from 2000 until 2024.

Then, we search for words and phrases throughout the full-text article to proxy whether the article uses monthly stock return data from CRSP for the empirical analysis. Specifically, we have 4 conditions (3 positive and 1 negative condition) that must be met for an article to count as conducting a CRSP-based return analysis: (1) "CRSP" or "Center for Research in Security Prices" must appear at least once; (2) "monthly stock return(s)", "stock return(s)", or "excess return(s)" needs to appear at least once; (3) any of the following words indicating that an empirical analysis was conducted (similar approach with different keywords as in Berninger et al. (2021) and Berninger et al. (2024)) must be present at least once: "empirical", "standard error(s)", "portfolio sort", "Fama(-)MacBeth regression", "Fama-French factors", "return predictability", "alpha", "factor model(s)", "anomaly", or "anomalies"; (4) articles meeting the first 3 conditions are only considered if the phrase "mutual fund(s) database" in combination with either 1 mention of "mutual fund return(s)" or at least 10 mentions of "mutual fund(s)" are not present in the article. However, this screen can be offset if the article has at the same time at least 10 mentions of "stock return(s)". The last screen ensures that we do not count articles using only monthly mutual fund returns from the CRSP US mutual fund database. All words in the article and all search words are converted to lowercase before searching.

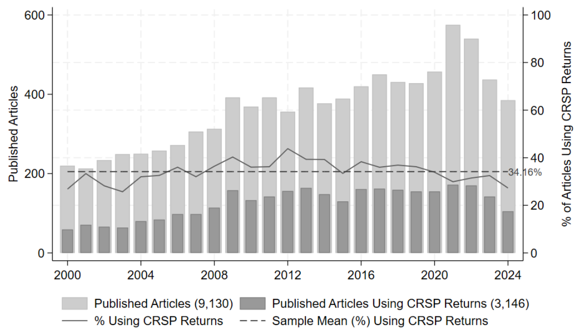
We are aware that this rather simple methodology does not ensure that the percentage of articles conducting a (monthly) CRSP-based returns analysis is fully accurate. However, it is a reasonable approximation of the relevance of (monthly) stock returns from CRSP in the finance literature. Our results show that 34.16% of all Top5 full-text articles published in the last 25 years have performed a (monthly) CRSP-based returns analysis. Our approach is conservative, as we focus on empirical studies using monthly CRSP returns. Otherwise, we find similar numbers as reported in Berninger et al. (2024), who approximate the overall CRSP usage with 51.45% from 2000–2016 in a broader setting. There are only small differences between the 5 journals; the percentages range from 20.24% for the RoF to 39.45% for the JFQA. The Top3 journals show similar percentages, with 31.67% for the RFS, 33.62% for the JF, and 36.60% for the JFE, respectively. Interestingly, no published article after 2022 has so far mentioned which CRSP tape (CIZ or SIZ) has been used, suggesting that most researchers are not aware of the documented differences between the tapes and their potential implications.

FIGURE IA.1

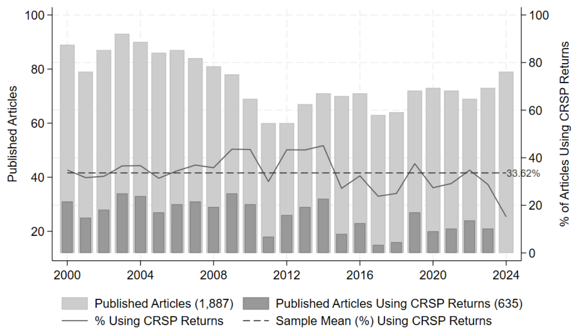
Articles Conducting a CRSP-Based Return Analysis

Figure IA.1 shows the total number of published articles together with the absolute and relative number of published articles using CRSP return data. We consider all articles published in the 5 leading (Top5) finance journals (Graph A). We also report journal-specific results for the JF, JFE, RFS, JFQA, and RoF in Graphs B through F, respectively. The sample period is from 2000 to 2024. An article is counted as conducting a CRSP-based return analysis if the article contains words (e.g., CRSP, return, portfolio sort, etc.) indicating that return data from CRSP has been used in an empirical analysis (see Internet Appendix IA.I).

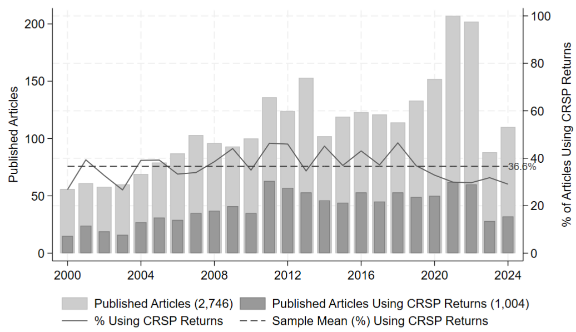
Graph A. Top5



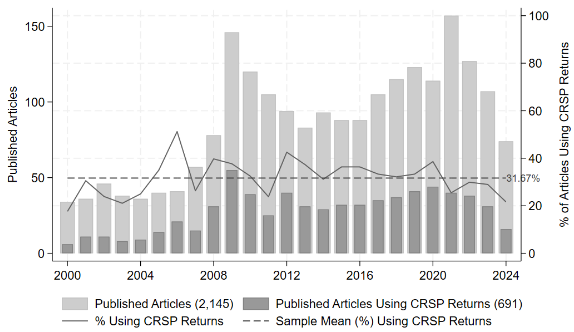
Graph B. JF



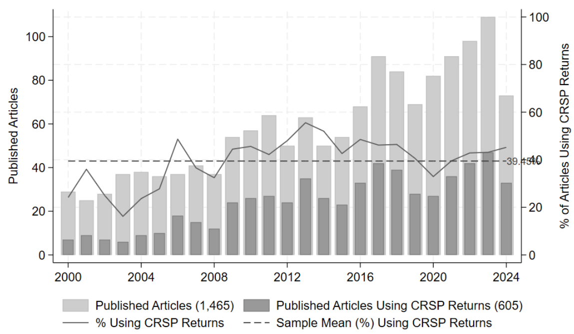
Graph C. JFE



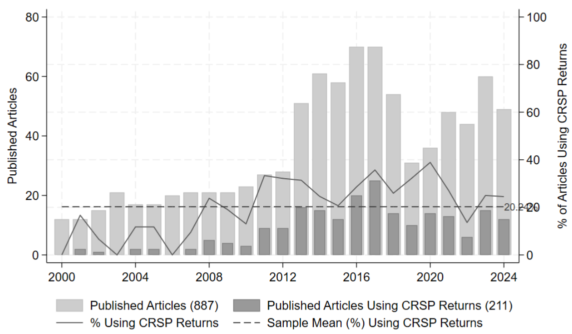
Graph D. RFS



Graph E. JFQA



Graph F. RoF



IA.II. Modified Returns from the New Tape with Payouts Reinvested on the Payment Date

In this section, we describe the methodology to compute monthly returns based on the new CIZ tape, for which we only move the reinvestment of payouts from the ex-date (CIZ) to the payment date. Moreover, we report summary statistics for the number of trading days between the ex-date, the end of the ex-date month, and the payment date for dividends from CRSP.

We start with all payouts for common stocks traded on the NYSE, AMEX, and NASDAQ from July 1926 to December 2023. We obtain these payouts from the CRSP data table "Distribution". We exclude observations that do not represent payouts or do not affect return computations: First, all observations that CRSP does not code in its return computation (distribution codes 1999, 2999, and 4999) and stock splits (distribution code 5523). Second, rights and share offers because CRSP assumes that these offers are always accepted (distribution codes starting with 4, 6, and 7). Third, we exclude stock dividends (distribution codes starting with 5) because the firm distributes only new shares to existing shareholders and does not pay out a specified amount that investors might need to reinvest.

Additionally, we also exclude all payouts for which we observe, on average, only a few trading days between the payment date and the date when the stock trades for the first time without the payout: Liquidation dividends (distribution codes starting with 2) and payouts from M&As or restructurings (distribution codes starting with 3). The median number of trading days between the payment date and a potential ex-date for these payment types is 1 to 2 trading days. These payouts arguably do not affect return computations once we move the reinvestment of payouts from the ex-date to the payment date, as there are almost no trading days between these two dates.

Moreover, we remove dividends with payment dates before or on the ex-date, after the delisting date, and after the end of our sample period. Excluding dividends with payment dates before or on the ex-date is a common procedure and avoids including potentially erroneous records (Chay, Choi, and Pontiff, 2006; Berkman and Koch, 2017). This leaves us essentially with 555,361 dividend observations from CRSP that potentially affect monthly returns once we reinvest payouts on the payment date rather than on the ex-date. We report the distribution of the trading days between the ex-date, the payment date, and the last trading date of the ex-date month for these dividends in Table A.1 below.

For each of these dividends, we collect the ex-date, the payment date, and the payment amount from the CRSP data table "Distribution". We also compute a cumulative factor to adjust share prices based on data from the daily CRSP tape for the factor to adjust share prices (FACPR). We replace payment dates (and a few ex-dates) with the next trading date with an available price in case the payment date (ex-date) is not a trading date with a valid price.

As CRSP does not share its full code, we cannot ensure that we consistently recompute daily returns with payments reinvested on the payment date for dates with multiple payouts, share issuances, or stock splits on one event date. Thus, we focus on a standard set of dividend observations that represents the vast majority of all dividends (around 93% of all 555,361 dividend observations in our sample):

1. We include only dividend observations for which we observe at most one ex-date

and one payment date per stock-month. However, we still allow for multiple payouts on these dates.

2. Moreover, we require that all dividends on the same ex-date have the same payment date.

3. Furthermore, we only investigate dividends with payment dates in the ex-date month or in the month after the ex-date. Thus, we ensure that some outliers with extremely large time lags between the ex-date and payment date do not impact our results, as it is common in the literature (Berkman and Koch, 2017).

4. We require that there are no additional payouts except dividends on the respective ex-date and payment date that could influence our return computation, as we might make slightly different computation assumptions than CRSP for these complex dates.

These steps reduce complexity and ensure that we do not create differences in the monthly return computation that are not due to moving the reinvestment assumption but rather due to us making slightly different methodological assumptions than CRSP, i.e., when there are multiple payments other than common dividends on the same date. These steps still leave us with the vast majority of dividends, namely 515,206 from 555,361 potentially relevant dividends. In the following, we describe how we compute modified returns with payouts reinvested on the payment date based on daily returns. In general, we compute daily returns with payouts reinvested on the payment dates and compound daily returns within each month to obtain modified monthly returns.

Payment Date in the Same Month as the Ex-Date. We start by replacing the daily return on the ex-date with the daily return excluding payouts from CRSP. Then, we recompute the daily return on the payment date based on the dividend amount, the stock price from the end of the previous day, and the stock price from the end of the payment day. Then we compound these daily returns within the month to obtain the modified returns based on the new CIZ tape with payouts reinvested on the payment date.

Payment Date in the Month After the Ex-Date Month. Again, we start by replacing the daily return on the ex-date with the daily return excluding payouts from CRSP. As we move the payment to the next month, we need to account for the fact that the investor has a claim on the dividend from the ex-date onward, but simply cannot reinvest it. Thus, we replace the daily return at the end of the month by adding the dividend (without reinvestment) to the month-end price. In the next payment date month, we recompute the daily return on the payment date based on the dividends distributed on this payment day, the stock price at the end of the previous day, and the stock price from the end of the payment day. If the payment date is not on the first trading date of a month, we also need to recompute the daily return at the beginning of the month, because the investor also had a claim to the dividend but could not reinvest it. Thus, we recompute daily returns on the first trading date in affected months by adding the dividend (without reinvestment) to the stock price at the end of the first trading date of that month. Then, we compound these modified daily returns within both months to obtain the modified returns based on the new CIZ tape with payouts reinvested on the payment date.

Note that we do not recompute all daily returns. Instead, we only change a few daily returns and ensure that the compounded modified daily returns over the month reflect the monthly holding-period returns with payouts reinvested on the payment date. That said, we ensure consistent results by computing the dividend amount plus the reinvestment proceeds from the payment date to the month-end of the payment date month. Thereby, we can recompute the modified monthly returns from month-end prices in months without trading gaps and new listings. We find a precise match between both procedures for stock months without trading gaps or new listings. Additionally, we check whether our modified return in the ex-date months with payment dates in the next month equals the old SIZ monthly returns (which reinvests dividends at the end of the month). Intuitively, there should not be a difference, as we assume that investors cannot reinvest the dividend in the ex-date month if the payment date is in the next month relative to reinvesting at the month-end. Also, for these modified returns, we do not observe differences relative to the monthly returns in the old SIZ tape for the ex-date month.

IA.III. Differences in Monthly Stock Returns

TABLE IA.1

Distribution of Differences in Monthly Stock Returns

Table IA.1 reports summary statistics for the distribution of absolute differences between monthly stock returns from the new (CIZ) and old (SIZ) CRSP tape in Panel A. Moreover, Panel B reports statistics for the distribution of absolute differences between monthly returns (in percent) from the new CIZ tape (with payouts reinvested on the ex-date) and modified monthly returns based on the modified new tape with payouts reinvested on the payment date. We only include common stocks from AMEX, NYSE, and NASDAQ from 1926 until 2023. A return difference exists if its absolute value exceeds the tolerance ("Tol.") or the returns are missing in either of the tapes. We consider 4 tolerance thresholds: 0.1 basis points (bp), 1 bp, 10 bp, and 100 bp (or 1%). We count the absolute frequency of these differences across stocks and time ("Obs.") and the frequency relative to all observations in percent ("Rel."). Then, we report the first ("Q10") and ninth ("Q90") decile, the median ("Median"), the mean ("Mean"), the standard deviation ("SD"), the skewness ("Skew."), and the kurtosis ("Kurt.") for the absolute return differences that pass the corresponding tolerance level. Return differences are displayed in percent per month.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

Tol.	Obs.	Rel.	Q10	Median	Q90	Mean	SD	Skew.	Kurt.
0.1 bp	474,580	12.78	0.00	0.02	0.11	0.16	2.65	66.23	8156.60
1 bp	357,445	9.62	0.01	0.04	0.14	0.22	3.06	57.43	6133.70
10 bp	60,684	1.63	0.11	0.16	0.50	1.17	7.61	23.14	997.95
1%	8,312	0.22	1.34	7.16	36.39	15.43	26.25	7.04	93.32

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

Tol.	Obs.	Rel.	Q10	Median	Q90	Mean	SD	Skew.	Kurt.
0.1 bp	754,384	20.31	0.00	0.02	0.10	0.05	0.22	325.74	152941.27
1 bp	552,954	14.89	0.01	0.03	0.12	0.06	0.26	284.07	114905.72
10 bp	77,697	2.09	0.11	0.15	0.35	0.22	0.67	117.40	18374.47
1%	2,041	0.05	1.06	1.40	2.90	2.14	4.81	18.29	394.11

TABLE IA.2

Differences in Monthly Stock Returns (NYSE-Listed Stocks Only)

Table IA.2 reports summary statistics for differences between monthly stock returns from the new (CIZ) and old (SIZ) CRSP tape in Panel A. Moreover, Panel B reports statistics for return differences between monthly returns from the new CIZ tape (with payouts reinvested on the ex-date) and modified monthly returns based on the new tape with payouts reinvested on the payment date. We only include common stocks listed on the NYSE. A return difference exists if its absolute value exceeds the tolerance ("Tol.") or the returns are missing in either of the monthly return versions. We consider 4 tolerance thresholds: 0.1 basis points (bp), 1 bp, 10 bp, and 100 bp (or 1%). We count the absolute frequency of these differences across stocks and time ("Obs.") and the frequency relative to all observations in percent ("Rel."). Then, we compute for each monthly return observation exceeding the respective tolerance level the raw and absolute difference based on monthly returns from the new tape minus monthly returns from the old (modified new) tape. We average these raw and absolute differences across the return observations that exceed the respective tolerance level to obtain mean raw differences ("Diff.") and mean absolute differences ("MAD"), both in percent per month. Finally, we compute whether the raw differences are different from zero with a two-sided paired t-test and indicate the statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. For all statistics, we consider 3 time periods: The full sample period in columns 2 to 5, an early sample from July 1926 until December 1967 in columns 6 to 8, and a late sample from January 1968 until December 2023 in columns 9 to 11.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	288,439	20.69	0.11	0.015***	22.41	0.20	0.040***	19.88	0.06	0.002
1 bp	217,813	15.62	0.14	0.020***	18.83	0.24	0.048***	14.12	0.08	0.003
10 bp	34,288	2.46	0.68	0.130***	3.70	1.04	0.260***	1.88	0.36	0.009
1%	1,343	0.10	13.36	3.292***	0.22	14.56	4.718***	0.04	9.89	−0.832

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	467,626	33.54	0.05	0.001***	35.33	0.06	−0.004***	32.70	0.04	0.004***
1 bp	343,434	24.63	0.06	0.002***	28.66	0.07	−0.005***	22.75	0.05	0.006***
10 bp	47,385	3.40	0.21	0.007***	4.65	0.23	−0.022***	2.81	0.19	0.030***
1%	609	0.04	1.91	−0.690***	0.10	2.01	−0.961***	0.02	1.65	0.022

TABLE IA.3

Differences in Monthly Stock Returns (Stock Price ≥ \$5)

Table IA.3 reports summary statistics for differences between monthly stock returns from the new (CIZ) and old (SIZ) CRSP tape in Panel A. Moreover, Panel B reports statistics for return differences between monthly returns from the new CIZ tape (with payouts reinvested on the ex-date) and modified monthly returns based on the new tape with payouts reinvested on the payment date. We only include common stocks from AMEX, NYSE, and NASDAQ with share prices larger than or equal to \$5. A return difference exists if its absolute value exceeds the tolerance ("Tol.") or the returns are missing in either of the monthly return versions. We consider 4 tolerance thresholds: 0.1 basis points (bp), 1 bp, 10 bp, and 100 bp (or 1%). We count the absolute frequency of these differences across stocks and time ("Obs.") and the frequency relative to all observations in percent ("Rel."). Then, we compute for each monthly return observation exceeding the respective tolerance level the raw and absolute difference based on monthly returns from the new tape minus monthly returns from the old (modified new) tape. We average these raw and absolute differences across the return observations that exceed the respective tolerance level to obtain mean raw differences ("Diff.") and mean absolute differences ("MAD"), both in percent per month. Finally, we compute whether the raw differences are different from zero with a two-sided paired t-test and indicate the statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. For all statistics, we consider 3 time periods: The full sample period in columns 2 to 5, an early sample from July 1926 until December 1967 in columns 6 to 8, and a late sample from January 1968 until December 2023 in columns 9 to 11.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	458,358	16.28	0.09	0.002	23.42	0.16	0.013*	14.92	0.07	−0.001
1 bp	341,860	12.15	0.12	0.003	19.63	0.19	0.016*	10.72	0.10	−0.002
10 bp	50,893	1.81	0.62	0.018	3.71	0.81	0.099**	1.44	0.52	−0.022
1%	2,537	0.09	10.64	−0.080	0.21	11.74	2.030**	0.07	9.82	−1.651***

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	737,037	26.19	0.04	0.003***	36.93	0.06	−0.003***	24.14	0.04	0.005***
1 bp	536,675	19.07	0.06	0.004***	29.92	0.07	−0.004***	17.00	0.05	0.007***
10 bp	70,793	2.52	0.20	0.023***	4.71	0.23	−0.014***	2.10	0.19	0.039***
1%	1,239	0.04	1.72	−0.452***	0.09	1.85	−0.891***	0.04	1.61	−0.060

TABLE IA.4

Differences in Monthly Stock Returns Without Trading Gaps, New Listings, and Delistings

Table IA.4 reports summary statistics for differences between monthly returns from the new (CIZ) and old (SIZ) CRSP tape, excluding observations with trading gaps, new listings (IPOs), and delistings. Moreover, we only include common stocks from AMEX, NYSE, and NASDAQ. A return difference exists if its absolute value exceeds the tolerance ("Tol.") or the returns are missing in either of the monthly return versions. We consider 4 tolerance thresholds: 0.1 basis points (bp), 1 bp, 10 bp, and 100 bp (or 1%). We count the absolute frequency of these differences across stocks and time ("Obs.") and the frequency relative to all observations in percent ("Rel."). Then, we compute for each monthly return observation exceeding the respective tolerance level the raw and absolute difference based on monthly returns from the new tape minus monthly returns from the old (modified new) tape. We average these raw and absolute differences across the return observations that exceed the respective tolerance level to obtain mean raw differences ("Diff.") and mean absolute differences ("MAD"), both in percent per month. Finally, we compute whether the raw differences are different from zero with a two-sided paired t-test and indicate the statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. For all statistics, we consider 3 time periods: The full sample period in columns 2 to 5, an early sample from July 1926 until December 1967 in columns 6 to 8, and a late sample from January 1968 until December 2023 in columns 9 to 11.

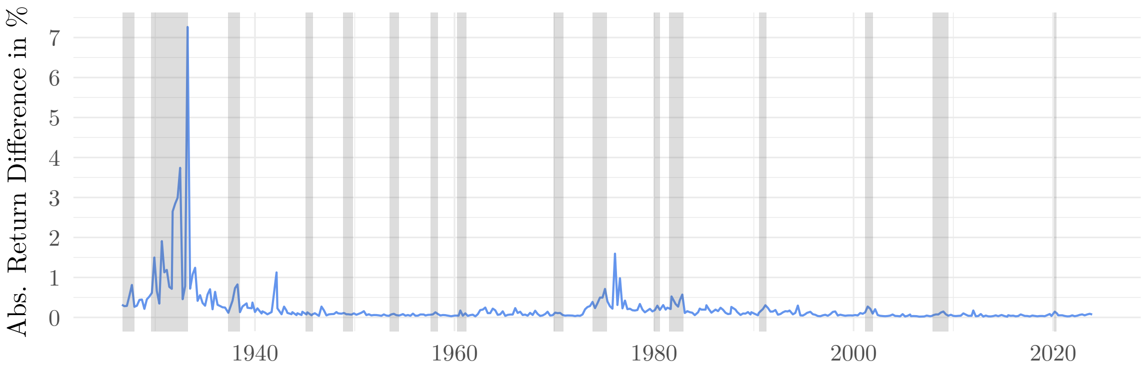
	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	463,640	12.60	0.06	0.003***	21.29	0.07	−0.002	11.29	0.05	0.005***
1 bp	347,240	9.44	0.07	0.004***	17.84	0.08	−0.003	8.17	0.07	0.007***
10 bp	53,108	1.44	0.27	0.029***	3.30	0.26	0.002	1.16	0.28	0.041***
1%	1,622	0.04	4.69	0.671	0.05	3.95	0.816	0.04	4.91	0.627

FIGURE IA.2

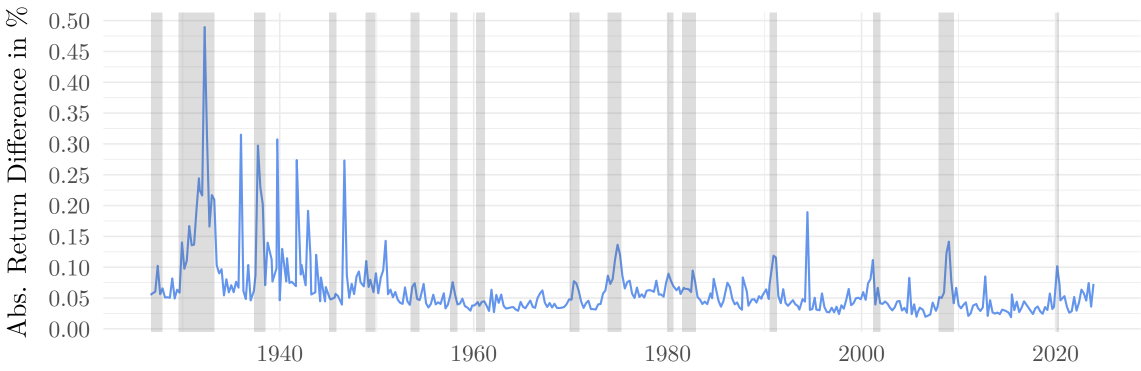
Absolute Return Differences over Time

Figure IA.2 shows the absolute differences in monthly stock returns over time. We average these absolute return differences exceeding 0.1 basis points for all observations within a quarter. Graph A (Graph B) shows these absolute return differences between the new (CIZ) and the old (SIZ) CRSP tape (excluding observations with trading gaps, IPOs, and delistings). Graph C shows these absolute differences between returns from the new CIZ tape and modified new returns from the new CIZ tape with payouts reinvested on the payment date. We only include common stocks from AMEX, NYSE, and NASDAQ. The sample covers the time period from 1926 to 2023. Absolute return differences are in percent per quarter, and gray shaded areas correspond to NBER recession months.

Graph A. New (CIZ) vs. Old (SIZ) Returns from CRSP



Graph B. New (CIZ) vs. Old (SIZ) Returns from CRSP Without Trading Gaps and Delistings



Graph C. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

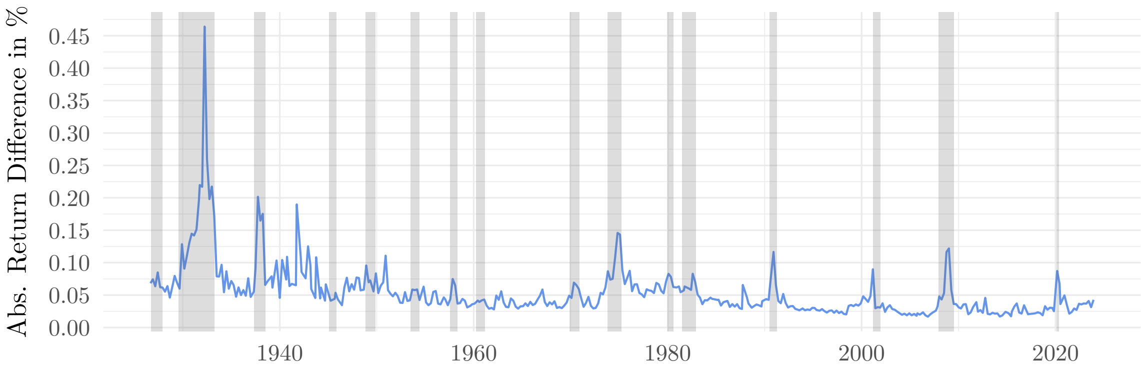
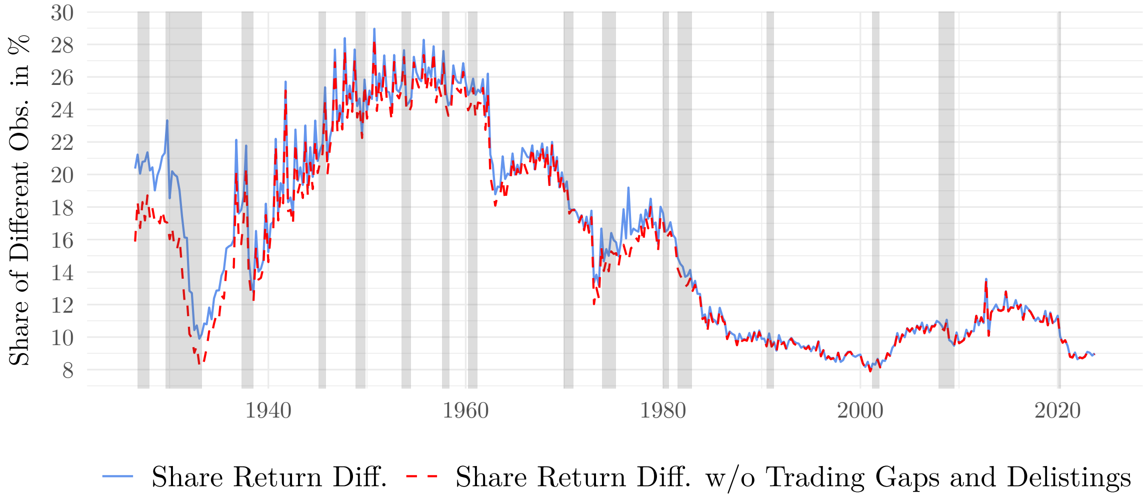


FIGURE IA.3

Share of Monthly Stock Return Differences over Time

Figure IA.3 shows the time-series of the number of return differences between the new (CIZ) and old (SIZ) CRSP tapes relative to all monthly CRSP observations in Graph A. Furthermore, we show the relative frequency of return differences excluding return differences due to trading gaps or delistings (red dashed line). Moreover, Graph B shows the relative frequency of return differences between monthly returns from the new CIZ tape (with payouts reinvested on the ex-date) and modified new returns based on the new CIZ tape with payouts reinvested on the payment date. We only consider observations as different if the absolute return difference exceeds a tolerance level of 0.1 basis points or the returns are missing in one of the monthly return versions. We only include common stocks from AMEX, NYSE, and NASDAQ. The sample covers the time period from 1926 to 2023. The relative frequency is in percent per quarter, and gray shaded areas correspond to NBER recession months.

Graph A. New (CIZ) vs. Old (SIZ) Returns from CRSP



Graph B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

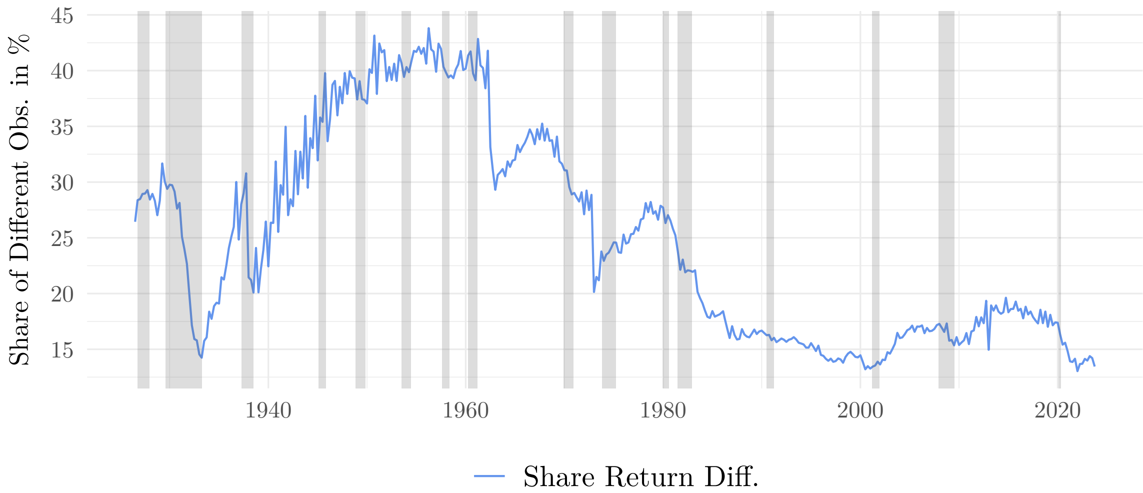


FIGURE IA.4

Mean Absolute Return Differences by Category

Figure IA.4 shows the mean absolute differences between monthly returns from the new (CIZ) and old (SIZ) CRSP tape by category of return changes, i.e., changing the reinvestment assumption, incorporating trading gaps/IPO months, and adjusting delisting returns. We compute these shares as follows: First, we compute the mean absolute differences for the monthly returns that we can explain by compounding the daily returns of the old CRSP tape for months without missing previous month-end prices and no recorded delistings ("Diff. Due to Reinvestment"). Second, we calculate the mean absolute differences for the monthly returns that we can reconcile by compounding daily returns, also for months with missing month-end prices from the previous month and no recorded delistings ("Diff. Due to Trading Gaps + IPO Months"). Third, we compute the mean absolute differences for monthly returns that we can explain by additionally adjusting delisting returns in the old tape ("Diff. Due to Delistings"). Lastly, we also show the mean absolute differences of monthly returns that are still different between both tapes ("Diff Remaining") after compounding daily returns within all months and adjusting delisting returns. We count monthly returns as different if the absolute difference exceeds 0.1 bp. The sample covers all common stocks from the AMEX, NYSE, and NASDAQ from 1926 to 2023. The mean absolute differences are in percent, and the vertical-axis scales differ for illustrative purposes.

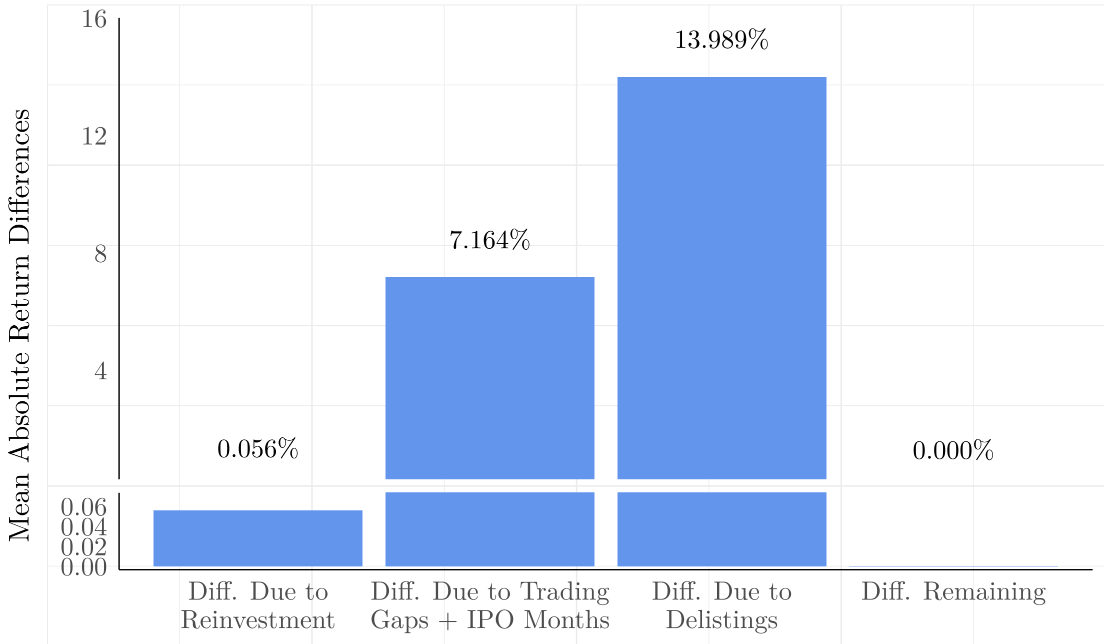


FIGURE IA.5

Dissecting Reinvestment Return Differences for Modified New Returns

Figure IA.5 shows the average absolute return differences in percent ("Ret. Diff.") in each of the 25 double-sorted buckets with increasing color intensity. We assign observations with monthly return differences exceeding 0.1 bp between the new (CIZ) CRSP tape and the modified new (CIZ) tape with payouts reinvested on the payment date into quintiles based on their payout yield and the stock's absolute compounded return from the ex-date to the payment date ("Reinvestment Return of Payouts"). We replace missing daily returns for the computation of the reinvestment return with zero. Based on these independent double sorts, we assign stocks into 25 buckets and depict the absolute return differences by color for each bucket by color intensity. Additionally, we show the mean values for the payout yield (horizontal axis) and reinvestment return of payouts (vertical axis) for each bucket in percent. We include all common stocks listed on AMEX, NYSE, and NASDAQ from 1926 until 2023.

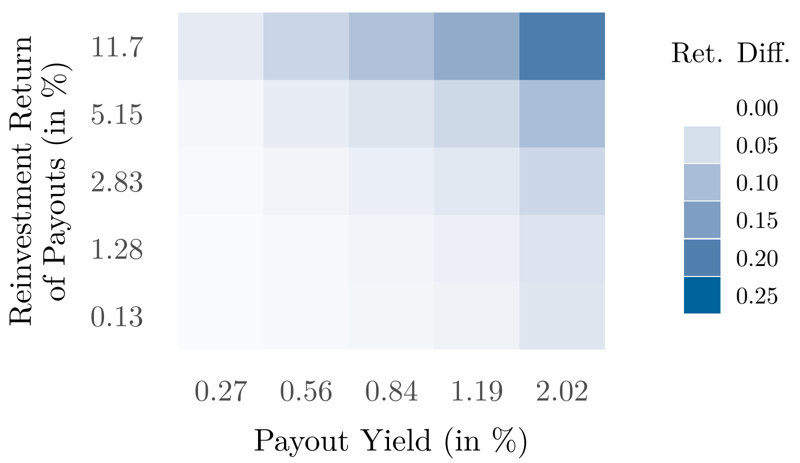
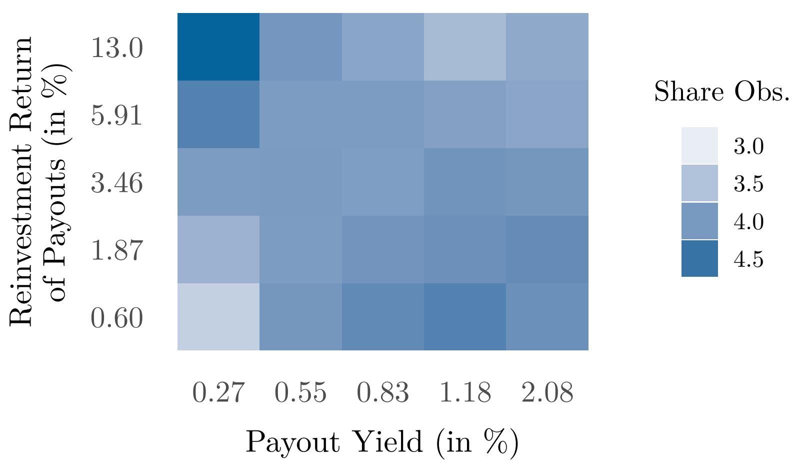


FIGURE IA.6

Determinants of Return Differences - Share of Observations

Figure IA.6 shows the share of observations in each of the 25 double-sorted buckets along with the mean values for each of the buckets. For Graph A, we assign observations with monthly return differences exceeding 0.1 bp between the the new (CIZ) and old (SIZ) CRSP tape into quintiles based on their payout yield and the stock's absolute compounded return from the ex-date to the month-end ("Reinvestment Return of Payouts"). For Graph B, we assign observations with monthly return differences exceeding 0.1 bp between the new (CIZ) CRSP tape and the modified new (CIZ) tape with payouts reinvested on the payment date into quintiles based on their payout yield and the stock's absolute compounded return from the ex-date to the payment date ("Reinvestment Return of Payouts"). We replace missing daily returns for the computation of the reinvestment return with zero. Based on these independent sorts, we assign stocks into 25 buckets and show the relative share of observations ("Share Obs.") for each bucket by color intensity. Additionally, we show the mean values for the payout yield (horizontal axis) and reinvestment return of payouts (vertical axis) for each bucket in percent. We include all common stocks listed on AMEX, NYSE, and NASDAQ from 1926 until 2023.

Graph A. New (CIZ) vs. Old (SIZ) Returns from CRSP



Graph B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

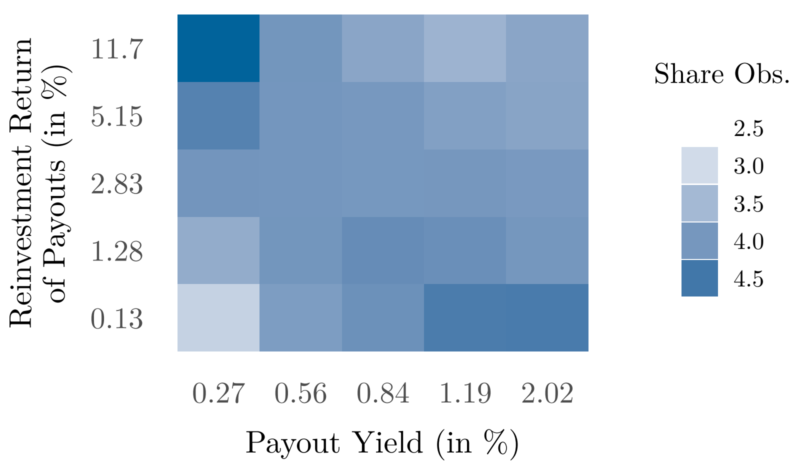


TABLE IA.5

Characteristics of Stocks With Versus Without Return Differences

Table IA.5 reports mean characteristics for stocks that have different monthly returns between the new (CIZ) and old (SIZ) CRSP tapes ("Return Difference") and those that do not have different monthly returns based on this comparison ("No Return Difference") in Panel A. Furthermore, Panel B reports mean characteristics for stocks that have different monthly returns between the new (CIZ) CRSP tape and the modified new (CIZ) tape with payouts reinvested on the payment date ("Return Difference") and those that do not have different monthly returns based on this comparison ("No Return Difference"). We consider a monthly return to be different between two versions of monthly returns if their absolute difference exceeds 0.1 basis points. Moreover, we test with an unpaired, two-sided t-test whether the mean characteristics of stocks with return differences are significantly different from those that do not show return differences ("Δ Mean"). We include all stocks listed on AMEX, NYSE, and NASDAQ. The sample period is from 1926 to 2023. We indicate statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. The construction of the characteristics is laid out in Section IA.IX of the Internet Appendix.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

Characteristic	No Return Difference	Return Difference	Δ Mean
Age	11.38	16.70	5.311***
Beta	1.149	0.956	-0.193***
Size	18.49	19.51	1.018***
Book-to-Market	0.847	0.858	0.011***
Momentum	0.133	0.157	0.023***
Asset Growth	0.328	0.129	-0.199***
Operating Profitability	0.082	0.302	0.220***
Real Estate Ratio	-0.002	0.012	0.015***
Dividend Ratio	0.187	0.446	0.259***
Issuance Ratio	0.155	0.030	-0.126***
Total Payout Ratio	-0.109	0.056	0.165***

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

Characteristic	No Return Difference	Return Difference	Δ Mean
Age	10.83	16.98	6.156***
Beta	1.169	0.954	-0.215***
Size	18.39	19.56	1.168***
Book-to-Market	0.849	0.849	0.000
Momentum	0.132	0.154	0.022***
Asset Growth	0.347	0.127	-0.221***
Operating Profitability	0.060	0.306	0.246***
Real Estate Ratio	-0.004	0.013	0.016***
Dividend Ratio	0.162	0.450	0.289***
Issuance Ratio	0.166	0.029	-0.137***
Total Payout Ratio	-0.123	0.058	0.181***

TABLE IA.6

Characteristics of Stocks with Return Differences

Table IA.6 reports mean characteristics for stocks sorted on the differences between monthly returns from the new (CIZ) minus monthly returns from the old (SIZ) CRSP tape in Panel A. Furthermore, Panel B reports mean characteristics for stocks sorted on differences between monthly returns from the new (CIZ) CRSP tape minus monthly returns from the modified new (CIZ) tape with payouts reinvested on the payment date. We sort stocks into quintiles based on the return difference between these versions of monthly returns and report the average characteristic for each quintile. We consider a monthly return to be different between two versions of monthly returns if the absolute difference exceeds 0.1 basis points. Moreover, we test with an unpaired, two-sided t-test whether the mean characteristics between stocks with the largest positive ("Q5") return differences are different from stocks with the most negative ("Q1") return differences ("Q5 − Q1"). We include all stocks listed on AMEX, NYSE, and NASDAQ. The sample period is from 1926 to 2023. We indicate statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. The construction of the characteristics is laid out in Section IA.IX of the Internet Appendix.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

Characteristic	Q1	Q2	Q3	Q4	Q5	Q5 − Q1
Return Difference	-0.251	-0.018	0.000	0.020	0.324	0.575***
Age	16.84	16.78	16.54	16.86	16.79	-0.051
Beta	0.926	0.976	0.995	0.957	0.922	-0.004
Size	19.25	19.65	19.79	19.67	19.29	0.042***
Book-to-Market	0.939	0.818	0.770	0.814	0.932	-0.007**
Momentum	0.111	0.172	0.197	0.179	0.128	0.017***
Asset Growth	0.120	0.130	0.140	0.130	0.117	-0.003
Operating Profitability	0.294	0.308	0.314	0.312	0.292	-0.002
Real Estate Ratio	0.012	0.014	0.012	0.013	0.011	-0.001
Dividend Ratio	0.519	0.414	0.384	0.420	0.505	-0.014***
Issuance Ratio	0.030	0.029	0.031	0.029	0.028	-0.001
Total Payout Ratio	0.060	0.054	0.052	0.059	0.061	0.000

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

Characteristic	Q1	Q2	Q3	Q4	Q5	Q5 − Q1
Return Difference	-0.072	-0.017	0.001	0.020	0.089	0.161***
Age	17.01	16.84	16.83	17.16	17.13	0.125**
Beta	0.932	0.972	0.990	0.954	0.921	-0.012***
Size	19.28	19.66	19.82	19.70	19.32	0.034***
Book-to-Market	0.924	0.811	0.766	0.815	0.932	0.008***
Momentum	0.111	0.171	0.196	0.172	0.122	0.011***
Asset Growth	0.118	0.131	0.137	0.128	0.116	-0.003*
Operating Profitability	0.297	0.309	0.315	0.313	0.296	-0.001
Real Estate Ratio	0.012	0.013	0.013	0.012	0.011	-0.001
Dividend Ratio	0.519	0.413	0.383	0.425	0.516	-0.003
Issuance Ratio	0.029	0.029	0.031	0.029	0.027	-0.002***
Total Payout Ratio	0.061	0.054	0.054	0.060	0.063	0.002**

TABLE IA.7

Differences in Returns per Industry

Table IA.7 reports summary statistics of the monthly return differences between the new (CIZ) and old (SIZ) CRSP tapes within industries in Panel A. Moreover, Panel B reports statistics within industries for return differences between monthly returns from the new CIZ tape (with payouts reinvested on the ex-date) and modified new returns based on the CIZ tape with payouts reinvested on the payment date. First, this table reports the relative number of firms with payouts ("Div." in percent) in each industry based on all common stocks within each industry listed on AMEX, NYSE, and NASDAQ. For the remaining summary statistics ("Obs", "Rel.", "MAD", "Diff"), we include only common stocks listed on AMEX, NYSE, and NASDAQ that have absolute return differences exceeding 0.1 bp (10 bp) or missing returns in one of the monthly return versions in the respective industry. Then, we count the absolute frequency of these differences across stocks and time within each industry ("Obs.") and the frequency relative to all observations in percent within each industry ("Rel."). Additionally, we compute for each monthly observation the raw and absolute difference based on monthly returns from the CIZ minus monthly returns from the SIZ (modified CIZ) tape. We average these raw and absolute differences across all observations within each industry to obtain mean raw differences ("Diff.") and mean absolute differences ("MAD"), both in percent per month. Finally, we compute for each industry whether the raw differences are different from zero with a two-sided paired t-test and indicate the statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. We consider the whole sample from 1926 until 2023.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

		0.1 bp				10 bp			
Industry	Div.	Obs.	Rel.	MAD	Diff.	Obs.	Rel.	MAD	Diff.
Utilities	86.48	35,107	25.06	0.07	−0.000	4,050	2.89	0.35	−0.003
Finance	69.40	105,859	17.89	0.11	0.020***	10,006	1.69	0.92	0.190***
Manufacturing	51.94	213,035	13.51	0.15	0.009**	27,597	1.75	0.99	0.080**
Retail	51.14	31,537	13.28	0.14	0.021*	3,783	1.59	0.99	0.183*
Transportation	44.17	19,640	10.96	0.31	0.141***	2,734	1.53	2.06	1.023***
Wholesale	40.95	11,744	9.57	0.23	0.024	1,220	0.99	1.98	0.227
Agriculture	43.89	995	9.52	0.20	−0.084	128	1.22	1.92	−0.893
Construction	40.43	3,550	9.37	0.33	0.003	452	1.19	2.44	0.034
Mining	35.65	14,501	8.49	0.30	0.118***	1,998	1.17	2.17	0.913***
Services	25.90	23,425	5.26	0.31	0.026	2,835	0.64	2.46	0.215
Public	14.58	3,228	2.74	0.19	0.006	304	0.26	1.80	0.065

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

		0.1 bp				10 bp			
Industry	Div.	Obs.	Rel.	MAD	Diff.	Obs.	Rel.	MAD	Diff.
Utilities	86.48	60,342	43.08	0.05	0.003***	6,488	4.63	0.18	0.016***
Finance	69.40	166,736	28.18	0.04	0.005***	13,867	2.34	0.22	0.036***
Manufacturing	51.94	344,590	21.85	0.05	0.002***	37,223	2.36	0.22	0.019***
Retail	51.14	48,536	20.44	0.05	−0.000	4,907	2.07	0.21	−0.001
Agriculture	43.89	1,804	17.26	0.04	0.004	166	1.59	0.23	0.045
Transportation	44.17	30,467	17.01	0.05	−0.000	3,654	2.04	0.25	0.000
Wholesale	40.95	18,279	14.89	0.04	0.004***	1,576	1.28	0.22	0.036***
Construction	40.43	5,620	14.83	0.05	0.003	638	1.68	0.25	0.033
Mining	35.65	23,338	13.66	0.05	0.001	2,632	1.54	0.24	0.007
Services	25.90	36,976	8.31	0.04	0.003***	3,581	0.80	0.24	0.020**
Public	14.58	4,780	4.06	0.04	0.001	369	0.31	0.23	−0.006

TABLE IA.8

Differences in Monthly Stock Returns Without Finance and Utilities

Table IA.8 reports summary statistics for differences between monthly stock returns from the new (CIZ) and old (SIZ) CRSP tape in Panel A. Moreover, Panel B reports statistics for return differences between monthly returns from the new CIZ tape (with payouts reinvested on the ex-date) and modified monthly returns based on the new tape with payouts reinvested on the payment date. We only include common stocks from AMEX, NYSE, and NASDAQ excluding stocks in the finance and utility sector. A return difference exists if its absolute value exceeds the tolerance ("Tol.") or the returns are missing in either of the monthly return versions. We consider 4 tolerance thresholds: 0.1 basis points (bp), 1 bp, 10 bp, and 100 bp (or 1%). We count the absolute frequency of these differences across stocks and time ("Obs.") and the frequency relative to all observations in percent ("Rel."). Then, we compute for each monthly return observation exceeding the respective tolerance level the raw and absolute difference based on monthly returns from the new tape minus monthly returns from the old (modified new) tape. We average these raw and absolute differences across the return observations that exceed the respective tolerance level to obtain mean raw differences ("Diff.") and mean absolute differences ("MAD"), both in percent per month. Finally, we compute whether the raw differences are different from zero with a two-sided paired t-test and indicate the statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. For all statistics, we consider 3 time periods: The full sample period in columns 2 to 5, an early sample from July 1926 until December 1967 in columns 6 to 8, and a late sample from January 1968 until December 2023 in columns 9 to 11.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	333,614	11.18	0.19	0.027***	21.38	0.21	0.041***	9.38	0.18	0.021***
1 bp	253,326	8.49	0.25	0.035***	18.14	0.25	0.049***	6.79	0.25	0.029***
10 bp	46,628	1.56	1.31	0.213***	3.89	1.07	0.257***	1.15	1.45	0.186***
1%	7,500	0.25	16.02	2.838***	0.50	14.70	4.472***	0.21	16.66	2.040***

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

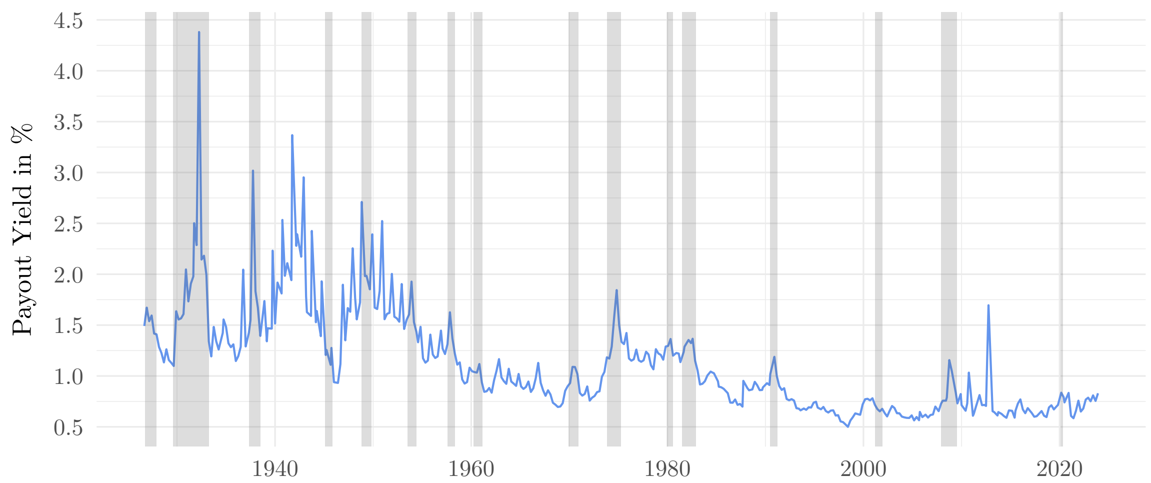
	Jul. 1926 – Dec. 2023				Jul. 1926 – Dec. 1967			Jan. 1968 – Dec. 2023		
Tol.	Obs.	Rel.	MAD	Diff.	Rel.	MAD	Diff.	Rel.	MAD	Diff.
0.1 bp	527,306	17.68	0.05	0.002***	33.03	0.06	−0.003***	14.96	0.04	0.005***
1 bp	387,829	13.00	0.06	0.003***	27.05	0.07	−0.004***	10.52	0.06	0.007***
10 bp	57,342	1.92	0.23	0.021***	4.58	0.24	−0.014***	1.45	0.22	0.041***
1%	1,639	0.05	2.10	−0.092	0.11	2.06	−0.603***	0.04	2.13	0.359

FIGURE IA.7

Time Series of Components for the Return Differences

Figure IA.7 shows the payout yield in Graph A and the absolute compounded return from the ex-date to the end of the ex-date month in Graph B for stock-month observations with differences in monthly returns between the new (CIZ) and the old (SIZ) CRSP tape. In each month we compute the average payout yield and the average absolute compounded return from the ex-date to the end of the ex-date month across all stocks with differences in monthly returns between the old and new CRSP tape that exceed 0.1 bp. We only include common stocks from AMEX, NYSE, and NASDAQ. The sample covers the time period from 1926 until 2023. All measures are shown in percent per quarter and gray shaded areas correspond to NBER recession months.

Graph A. Payout Yield



Graph B. Absolute Compounded Returns from the Ex-Date to the End of the Ex-Date Month

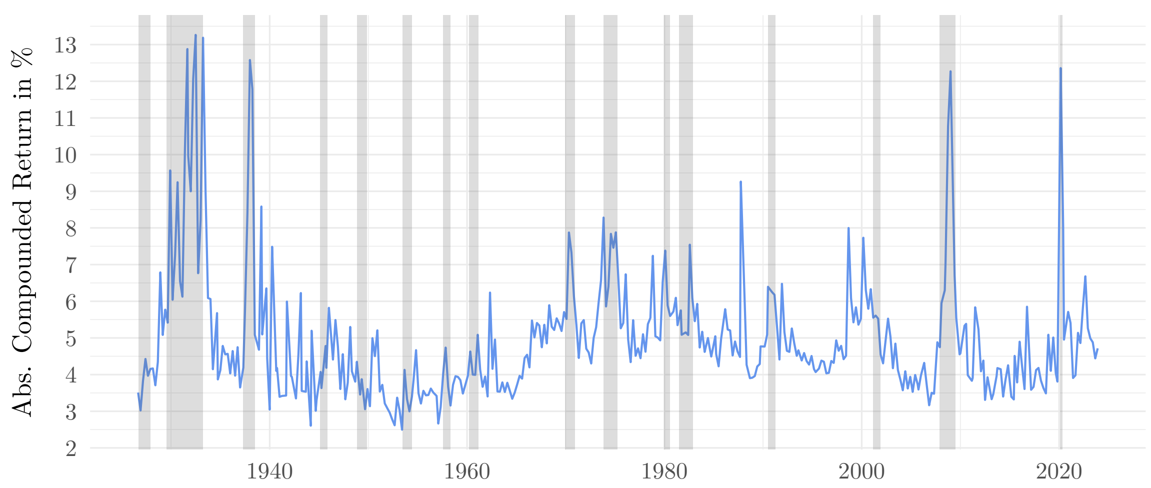


TABLE IA.9

Cyclicality of Monthly Stock Return Differences

Table IA.9 reports in Panel A cyclicality coefficients for the return differences between the new (CIZ) and old (SIZ) CRSP tape and the components of these return differences. Panel B reports the cyclicality coefficients for the return differences between the new (CIZ) tape and the modified new (CIZ) tape with payouts reinvested on the payment date. Specifically, we regress the absolute return differences between these monthly return differences, the payout yield of the corresponding differences, and the absolute reinvestment return of these return differences on the NBER recession indicator and the cross sectional return dispersion. For Panel A the reinvestment return corresponds to the absolute compounded return from the ex-date to the end of the ex-date month and in Panel B to the absolute compounded return from the ex-date to the payment date. We replace missing daily returns for the computation of the reinvestment return with zero. We only include observations if their monthly returns exceed 0.1 bp and average all dependent variables across all stocks in each month. We use Newey and West (1987) corrected standard errors and standardize all variables except for the NBER recession indicator, which is one during recession months and zero otherwise. We indicate statistical significance at the ten-, five-, and one-percent level by *, **, and ***, respectively. The sample covers the time period from 1926 until 2023.

Panel A. New (CIZ) vs. Old (SIZ) Returns from CRSP

	Absolute Return Differences		Payout Yield		Abs. Reinvestment Return of Payouts	
Intercept	-0.10***	-1.05***	1.25***	1.44***	4.73***	2.02***
	(-2.81)	(-3.34)	(24.68)	(11.39)	(55.15)	(7.38)
NBER Recession Ind.	0.60***	0.46***	0.39***	0.41***	1.95***	1.56***
	(3.11)	(2.86)	(3.79)	(4.23)	(5.48)	(6.09)
Return Dispersion		0.11***		-0.02*		0.32***
		(2.79)		(-1.65)		(9.17)
R^2	5.00%	21.10%	2.20%	2.90%	11.90%	41.20%
Nobs.	1,154	1,154	1,154	1,154	1,154	1,154

Panel B. New (CIZ) vs. Modified New Returns with Payouts Reinvested on the Payment Date

	Absolute Return Differences		Payout Yield		Abs. Reinvestment Return of Payouts	
Intercept	-0.21***	-0.79***	1.05***	1.26***	3.87***	1.88***
	(-5.09)	(-3.86)	(22.25)	(8.92)	(52.67)	(12.20)
NBER Recession Ind.	1.23***	1.15***	0.48***	0.51***	1.42***	1.13***
	(4.67)	(5.29)	(4.80)	(5.52)	(5.38)	(6.10)
Return Dispersion		0.07***		-0.02		0.23***
		(2.82)		(-1.57)		(13.05)
R^2	21.50%	27.50%	12.70%	15.70%	11.60%	40.70%
Nobs.	1,154	1,154	1,154	1,154	1,154	1,154

IA.IV. Differences in Portfolio Returns and Premia

TABLE IA.10

Differences in Long-Short Portfolio Returns for each Sorting Variable

Table IA.10 reports summary statistics of the differences between monthly long-short portfolio returns estimated from the old (SIZ) and the new (CIZ) CRSP tape from 1968 to 2023. We report these differences for each sorting variable ("SV"), which we assign to groups in the respective Panels. To account for methodological uncertainty, we estimate long–short portfolio returns for each of the 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. We compute the share of monthly long-short portfolio returns that differ by at least 1 bp ("$N_{1\text{ bp}}$") and 10 bp ("$N_{10\text{ bp}}$") across all methodological paths and months for each sorting variable. Based on the same aggregation, we also report the share of monthly long-short portfolio returns with positive ("Pos.") and negative ("Neg.") return differences between the old and the new tape. Moreover, we compute for each sorting variable the mean absolute difference ("MAD") of long-short returns from the old vs. the new tape and average across methodological paths and months. We also test for each sorting variable and each month whether the long-short returns from both tapes are significantly different across methodological paths. We control for dependencies across portfolio sorts that share similar choices. Specifically, we bootstrap the differences in long-short portfolio returns for each sorting variable and month with clusters for portfolio sorts that share the same choice of a decision fork (see Internet Appendix IA.VI). Column "Sig." reports the share of monthly long-short portfolio returns for which the bootstrap-implied 99% confidence intervals do not contain zero. All statistics are in percent. We label sorting variables that require CRSP data (monthly CRSP returns) for their construction with a superscript "C" ("R"). Moreover, we provide an overview of all sorting variables in Table C.1 and construction details in the Internet Appendix IA.VIII.

Panel A. Financing

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
$CSI^{C,R}$	88.50	41.92	0.13	61.64	52.46	47.54
DCOL	41.62	9.95	0.04	67.58	52.82	47.18
NDF	40.98	9.22	0.03	63.29	49.60	50.40
NXF	42.08	9.16	0.03	65.06	50.40	49.60
DFNL	42.36	9.12	0.03	67.43	51.29	48.71
DBE	43.29	8.96	0.03	67.43	52.97	47.03
CDI	36.14	7.66	0.03	67.16	51.56	48.44
NEF	38.77	6.96	0.03	66.83	54.91	45.09
Mean	46.72	12.87	0.04	65.80	52.00	48.00

Panel B. Intangibles

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
RER^{C}	48.98	15.72	0.05	56.90	52.25	47.75
WW^{C}	47.89	12.14	0.06	44.36	47.71	52.29
ADM^{C}	36.42	10.12	0.04	63.55	47.29	52.71
KZI^{C}	44.01	9.93	0.04	71.39	52.82	47.18
LFE	44.13	9.54	0.04	66.16	47.56	52.44
HR	43.65	9.20	0.03	66.16	51.83	48.17
TAN	42.99	9.15	0.03	68.52	52.02	47.98
OL	44.46	9.09	0.03	69.72	45.58	54.42
CFV	33.76	7.65	0.03	63.44	51.37	48.63
RDM^{C}	31.92	7.45	0.03	66.43	51.75	48.25
EPRD	28.33	5.56	0.02	69.76	55.74	44.26
Mean	40.60	9.60	0.04	64.22	50.54	49.46

Continued on the next page

Table IA.10. (continued)

Panel C. Investment

SV	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Sig.	Pos.	Neg.
DPIA	43.60	10.11	0.04	68.95	50.99	49.01
AG	44.56	10.10	0.04	68.04	52.21	47.79
DNOA	43.93	9.73	0.03	71.99	51.14	48.86
OA	42.53	9.59	0.03	64.84	50.99	49.01
DWC	39.34	9.33	0.03	62.56	50.99	49.01
PTA	41.39	9.15	0.03	65.45	49.77	50.23
NOA	44.09	8.85	0.03	71.08	54.03	45.97
DINV	39.67	8.76	0.03	64.69	51.60	48.40
IG	41.18	8.70	0.03	61.95	52.05	47.95
ACI	38.98	8.46	0.03	65.56	50.39	49.61
Mean	41.93	9.28	0.03	66.51	51.42	48.58

Panel D. Momentum

SV	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Sig.	Pos.	Neg.
$RMOM^{C,R}$	78.06	31.30	0.10	53.69	53.37	46.63
$MOM^{C,R}$	71.72	23.13	0.08	49.77	56.15	43.85
$52W^{C}$	52.73	13.85	0.05	58.03	56.52	43.48
ABR^{C}	37.42	10.24	0.04	57.30	52.65	47.35
SUE	40.53	8.11	0.03	62.84	49.60	50.40
TES	39.73	8.01	0.03	56.92	44.31	55.69
RS	38.67	7.46	0.03	65.27	52.02	47.98
Mean	51.27	14.58	0.05	57.69	52.09	47.91

Panel E. Profitability

SV	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Sig.	Pos.	Neg.
ROE	42.41	10.25	0.04	57.51	53.31	46.69
ATO	44.22	9.99	0.04	71.39	47.03	52.97
OPE	42.91	9.66	0.04	64.17	53.97	46.03
ROA	43.00	9.44	0.04	56.38	51.21	48.79
TBI	41.72	9.18	0.05	34.33	48.43	51.57
CBOP	41.95	9.11	0.03	63.27	48.58	51.42
Z^{C}	41.25	8.63	0.03	67.62	49.93	50.07
O	39.05	8.56	0.03	66.51	51.29	48.71
CTO	45.07	8.34	0.03	70.62	43.68	56.32
BL	43.99	7.79	0.03	70.46	46.18	53.82
GPA	42.76	7.68	0.03	67.47	46.03	53.97
Mean	42.58	8.97	0.04	62.70	49.06	50.94

Panel F. Size

SV	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Sig.	Pos.	Neg.
ME^{C}	44.90	12.04	0.06	35.28	46.79	53.21

Panel G. Trading Frictions

SV	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Sig.	Pos.	Neg.
$BETA^{C,R}$	71.14	20.93	0.07	53.17	50.54	49.46
$SREV^{C,R}$	58.96	15.60	0.05	40.66	47.98	52.02
BFP^{C}	56.97	15.30	0.06	62.61	49.70	50.30
TUR^{C}	52.34	13.35	0.05	62.01	50.60	49.40
DTV^{C}	48.04	11.52	0.07	27.33	53.45	46.55
AMI^{C}	46.96	10.99	0.06	24.92	53.30	46.70

Continued on the next page

Table IA.10. (continued)

$IVOL^{C}$	45.22	10.39	0.04	61.58	58.59	41.41
MDR^{C}	44.40	9.68	0.04	63.23	60.54	39.46
$ISKEW^{C}$	39.72	7.93	0.03	55.16	45.59	54.41
Mean	51.53	12.85	0.05	50.07	52.26	47.74

Panel H. Valuation

SV	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Sig.	Pos.	Neg.
$REV^{C,R}$	80.39	35.53	0.12	56.23	50.00	50.00
OCM^{C}	45.90	10.31	0.04	73.27	49.44	50.56
EBM^{C}	44.81	9.69	0.04	69.72	49.63	50.37
SM^{C}	44.13	9.41	0.04	69.87	49.48	50.52
CFM^{C}	44.39	8.99	0.03	71.66	53.52	46.48
NDM^{C}	43.00	8.87	0.03	70.87	49.85	50.15
BM^{C}	43.02	8.86	0.03	72.11	50.07	49.93
EM^{C}	44.48	8.75	0.03	72.11	52.77	47.23
AM^{C}	44.98	8.27	0.03	72.71	52.32	47.68
NPY^{C}	37.12	7.77	0.03	71.29	52.58	47.42
DM^{C}	41.56	7.34	0.03	66.42	49.93	50.07
Mean	46.71	11.25	0.04	69.66	50.87	49.13

TABLE IA.11

Differences in Return Premia for each Sorting Variable

Table IA.11 reports summary statistics of the differences between return premia estimated from the old (SIZ) and the new (CIZ) CRSP tape from 1968 to 2023. We report these differences for each sorting variable ("SV"), which we assign to groups in the respective Panels. To account for methodological uncertainty, we estimate the time-series average of monthly long-short portfolio returns (premium) for each of the 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. We compute the share of premia that differ by at least 1 bp ("$N_{1\ bp}$") and 10 bp ("$N_{10\ bp}$") across all methodological paths for each sorting variable. Moreover, we compute for each sorting variable the absolute difference ("AD") of premia estimates from the old vs. the new tape averaged across methodological paths. We also scale these absolute differences by the average absolute premium for each sorting variable (column "Ratio", in percent). Finally, we report the 90- ("Q90") and 95-percentiles ("Q95") of the absolute premia differences across all methodological paths for each sorting variable (in basis points per month). We label sorting variables that require CRSP data (monthly CRSP returns) for their construction with a superscript "C" ("R"). Moreover, we provide an overview of all sorting variables in Table C.1 and construction details in the Internet Appendix IA.VIII.

Panel A. Financing

SV	$N_{1\ bp}$	$N_{10\ bp}$	AD	Ratio	Q90	Q95
$CSI^{C,R}$	47.37	0.00	1.05	2.22	1.97	2.35
NDF	3.31	0.00	0.42	1.41	0.78	0.92
NXF	1.29	0.00	0.26	0.57	0.57	0.72
DBE	1.17	0.00	0.24	0.60	0.54	0.70
DCOL	1.07	0.00	0.24	1.48	0.53	0.68
DFNL	1.04	0.00	0.29	0.98	0.60	0.73
CDI	0.45	0.00	0.20	1.62	0.43	0.54
NEF	0.40	0.00	0.19	0.56	0.42	0.53
Mean	7.01	0.00	0.36	1.13	0.73	0.90

Continued on the next page

Table IA.11. (continued)

Panel B. Intangibles

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	AD	Ratio	Q90	Q95
WW[C]	13.44	0.07	0.53	5.83	1.25	1.86
CFV	5.61	0.00	0.41	1.17	0.84	1.03
ADM[C]	2.61	0.00	0.32	1.25	0.68	0.84
RDM[C]	2.39	0.00	0.26	0.88	0.55	0.73
RER[C]	2.32	0.00	0.30	1.95	0.64	0.81
LFE	1.82	0.00	0.32	7.19	0.66	0.81
TAN	1.48	0.00	0.29	1.81	0.62	0.77
HR	0.93	0.00	0.27	0.97	0.56	0.68
OL	0.68	0.00	0.24	0.82	0.51	0.64
EPRD	0.57	0.00	0.23	0.32	0.49	0.60
KZI[C]	0.53	0.00	0.22	39.87	0.47	0.59
Mean	2.94	0.01	0.31	1.43	0.66	0.85

Panel C. Investment

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	AD	Ratio	Q90	Q95
PTA	1.75	0.00	0.27	0.92	0.58	0.74
DPIA	1.28	0.00	0.26	0.57	0.56	0.71
AG	1.27	0.00	0.28	0.57	0.59	0.72
NOA	0.94	0.00	0.24	0.48	0.51	0.66
DNOA	0.87	0.00	0.24	0.42	0.52	0.65
DWC	0.64	0.00	0.23	0.55	0.48	0.60
ACI	0.62	0.00	0.23	0.94	0.49	0.62
DINV	0.59	0.00	0.23	0.61	0.50	0.62
OA	0.56	0.00	0.23	0.64	0.49	0.62
IG	0.48	0.00	0.21	0.61	0.45	0.57
Mean	0.90	0.00	0.24	0.60	0.52	0.65

Panel D. Momentum

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	AD	Ratio	Q90	Q95
MOM[C,R]	17.57	0.00	0.61	1.11	1.25	1.52
ABR[C]	17.41	0.00	0.57	1.03	1.45	2.00
RMOM[C,R]	12.60	0.00	0.52	1.23	1.09	1.34
52W[C]	8.19	0.00	0.47	1.08	0.93	1.16
TES	0.85	0.00	0.25	0.59	0.52	0.65
SUE	0.25	0.00	0.22	0.47	0.46	0.57
RS	0.15	0.00	0.19	0.53	0.41	0.50
Mean	8.15	0.00	0.41	0.88	0.87	1.10

Panel E. Profitability

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	AD	Ratio	Q90	Q95
TBI	21.39	0.01	0.80	6.63	2.36	3.54
CBOP	3.03	0.00	0.33	0.57	0.71	0.88
ROA	1.34	0.00	0.27	0.56	0.58	0.73
ROE	1.31	0.00	0.28	0.56	0.59	0.73
ATO	1.17	0.00	0.24	1.24	0.52	0.67
O	1.00	0.00	0.27	4.38	0.55	0.68
OPE	0.93	0.00	0.22	0.66	0.49	0.63

Continued on the next page

Table IA.11. (continued)

GPA	0.91	0.00	0.24	0.75	0.50	0.63
CTO	0.86	0.00	0.22	1.63	0.48	0.61
Z^C	0.44	0.00	0.22	3.84	0.47	0.58
BL	0.41	0.00	0.21	3.90	0.45	0.56
Mean	2.98	0.00	0.30	1.33	0.70	0.93

Panel F. Size

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	AD	Ratio	Q90	Q95
ME^C	16.36	0.62	0.83	10.75	1.80	4.22

Panel G. Trading Frictions

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	AD	Ratio	Q90	Q95
DTV^C	27.70	0.19	1.03	3.62	2.88	4.21
AMI^C	25.60	0.14	0.95	6.90	2.61	3.86
BFP^C	9.23	0.00	0.50	5.58	0.98	1.19
$IVOL^C$	6.22	0.00	0.40	1.30	0.84	1.07
$SREV^{C,R}$	5.95	0.00	0.39	14.37	0.84	1.05
MDR^C	5.51	0.01	0.36	1.58	0.78	1.04
$BETA^{C,R}$	3.52	0.00	0.35	217.36	0.73	0.91
TUR^C	2.96	0.00	0.31	1.39	0.68	0.86
$ISKEW^C$	0.96	0.00	0.29	18.52	0.59	0.72
Mean	9.74	0.04	0.51	3.57	1.21	1.66

Panel H. Valuation

SV	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	AD	Ratio	Q90	Q95
$REV^{C,R}$	17.69	0.01	0.59	2.90	1.30	1.67
OCM^C	1.66	0.00	0.30	0.70	0.63	0.78
CFM^C	1.55	0.00	0.30	0.76	0.62	0.77
EBM^C	0.99	0.00	0.26	1.06	0.54	0.67
NDM^C	0.82	0.00	0.25	3.68	0.52	0.65
AM^C	0.67	0.00	0.24	0.98	0.49	0.62
BM^C	0.65	0.00	0.23	0.67	0.49	0.61
EM^C	0.57	0.00	0.24	0.68	0.51	0.63
NPY^C	0.57	0.00	0.22	0.83	0.47	0.59
SM^C	0.55	0.00	0.23	0.52	0.48	0.60
DM^C	0.47	0.00	0.22	1.88	0.48	0.60
Mean	2.38	0.00	0.28	0.99	0.59	0.74

TABLE IA.12

Differences in Long-Short Portfolio Returns due to Reinvestment Timing

Table IA.12 reports summary statistics of the differences between monthly long-short portfolio returns estimated from the old (SIZ) and the new (CIZ) CRSP tape from 1968 to 2023. We focus solely on differences between the old and new CRSP tape due to the alteration of the reinvestment assumption. Thus, we replace returns in the CIZ tape that differ due to reasons other than altering the reinvestment assumption, i.e., trading gaps, with the returns in the old SIZ tape. To account for methodological uncertainty, we estimate long–short portfolio returns for all 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. We compute the share of monthly long-short portfolio returns that differ by at least 1 bp ("$N_{1\text{ bp}}$") and 10 bp ("$N_{10\text{ bp}}$") across all methodological paths and months for each sorting variable. The respective rows report these shares of long-short differences across sorting variables for specific groups (Panel A), types (Panel B), and averaged across specific time periods for all sorting variables (Panel C). Based on the same aggregation, we also report the share of monthly long-short portfolio returns with positive ("Pos.") and negative ("Neg.") return differences between the old and the new tape. Moreover, we compute for each sorting variable the mean absolute difference ("MAD") of long-short portfolio returns from the old vs. the new tape (average across methodological paths and months). The rows report these MADs averaged across groups (Panel A), types (Panel B), and across specific time periods for all sorting variables (Panel C). We also test for each sorting variable and each month whether the long-short portfolio returns from both tapes are significantly different across methodological paths. We control for dependencies across portfolio sorts that share similar choices. Specifically, we bootstrap the differences in long-short portfolio returns for each sorting variable and month with clusters for portfolio sorts that share the same choice of a decision fork (see Internet Appendix IA.VI). Column "Sig." reports the share of monthly long-short portfolio returns for which the bootstrap-implied 99% confidence intervals do not contain zero, averaged over sorting variables for each group. All statistics are in percent. The sorting variables belonging to each group or type can be found in Table C.1 of the Appendix. We classify sorting variables that require CRSP data (monthly CRSP returns) for their construction as CRSP-based (return-based).

Panel A. Sorting Variable Groups

Group	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
Financing	34.21	7.14	0.03	71.25	52.68	47.32
Intangibles	27.77	4.15	0.02	73.95	50.88	49.12
Investment	27.32	3.84	0.02	74.62	52.40	47.60
Momentum	41.05	7.97	0.03	60.32	52.92	47.08
Profitability	29.13	4.01	0.02	72.36	49.39	50.61
Size	29.70	4.67	0.03	52.47	47.83	52.17
Trading Frictions	40.18	7.17	0.03	53.02	51.60	48.40
Valuation	33.32	4.89	0.02	76.52	51.38	48.62
Overall	32.84	5.48	0.02	66.81	51.14	48.86

Panel B. Sorting Variable Types

Type	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
Non-CRSP-Based SVs	27.93	3.82	0.02	73.43	51.27	48.73
CRSP-Based SVs	37.89	7.08	0.03	64.87	51.57	48.43
Return-Based SVs	67.28	17.51	0.06	40.41	51.79	48.21

Panel C. Time Splits

Period	$N_{1\text{ bp}}$	$N_{10\text{ bp}}$	MAD	Sig.	Pos.	Neg.
Jan. 1968 – Dec. 1999	34.00	4.93	0.02	69.94	52.16	47.84
Jan. 2000 – Dec. 2023	30.78	5.84	0.02	68.71	50.51	49.49
NBER Recessions	42.58	7.67	0.03	67.76	52.71	47.29
No Recessions	31.07	4.99	0.02	69.94	51.35	48.65

IA.V. Differences in Factor Alphas

TABLE IA.13

Differences in CAPM Alphas

Table IA.13 reports summary statistics of the differences between CAPM alphas estimated from the old (SIZ) and new (CIZ) CRSP tapes from 1968 to 2023. To account for methodological uncertainty, we estimate long–short portfolio returns for all 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. For each sorting variable we compute the CAPM alpha by regressing monthly long-short returns constructed from the SIZ (CIZ) tape on the market factor based on SIZ (CIZ) data. We start by computing the share of CAPM alphas that differ by at least 1 bp ("$N_{1\ bp}$") and 10 bp ("$N_{10\ bp}$") across all methodological paths for each sorting variable. The respective rows report these shares of premia averaged across sorting variables for specific groups (Panel A) and types (Panel B) in percent. Moreover, we compute for each sorting variable the absolute difference ("AD") of CAPM alphas from the SIZ vs. the CIZ tape averaged across methodological paths. The respective rows report these ADs in basis points per month across groups (Panel A) and types (Panel B) of sorting variables. We also scale these absolute differences by the average absolute CAPM alpha within each group (column "Ratio", in percent). Finally, we report the 90- ("Q90") and 95-percentiles ("Q95") of the absolute CAPM alpha differences across all methodological paths and sorting variables for each group and type (in basis points per month). The sorting variables belonging to each group or type can be found in Table C.1 in the Appendix. We classify sorting variables that require CRSP data (monthly CRSP returns) for their construction as CRSP-based (return-based). Lastly, the construction of the market factor is based on the fixed Fama and French (1993) factors proposed by Akey et al. (2024).

Panel A. Sorting Variable Groups

Group	N_{1bp}	N_{10bp}	AD	Ratio	Q90	Q95
Financing	7.59	0.00	0.37	0.84	0.73	0.89
Intangibles	2.88	0.01	0.31	1.60	0.65	0.84
Investment	0.78	0.00	0.24	0.52	0.51	0.64
Momentum	7.85	0.00	0.40	0.76	0.84	1.06
Profitability	2.82	0.00	0.29	1.94	0.68	0.91
Size	16.01	0.46	0.81	31.01	1.82	4.20
Trading Frictions	9.41	0.05	0.50	28.60	1.22	1.67
Valuation	2.33	0.00	0.28	1.36	0.59	0.74
Overall	4.60	0.01	0.34	5.29	0.75	0.99

Panel B. Sorting Variable Types

Type	N_{1bp}	N_{10bp}	AD	Ratio	Q90	Q95
Non-CRSP-based SVs	1.66	0.00	0.27	1.11	0.58	0.75
CRSP-based SVs	7.90	0.03	0.42	9.98	0.93	1.27
Return-based SVs	18.70	0.00	0.60	2.19	1.22	1.49

IA.VI. Bootstrapping Differences in Monthly Long-Short Portfolio Returns

In this section, we describe our bootstrapping procedure to control for the dependencies among long-short portfolio returns differing in methodological choices. These dependencies might arise as we conduct several portfolio sorts for the same sorting variable that only differ in the methodological portfolio sorting choices, e.g., whether to equally-weight or value-weight stock returns when aggregating them into portfolios. Intuitively, if we fix a sorting variable, all corresponding series of monthly portfolio returns that are based on value-weighted portfolio returns – regardless of other methodological decisions – might be highly correlated as they are all based on the same choice for this fork. This dependency structure might be particularly relevant for the most impactful forks, i.e., those that create the largest differences in portfolio sorts when altering the choice at this fork. Observing large differences between portfolio sorts that differ in the choice of one impactful fork might imply that the series of portfolio returns that share the same choice for this impactful fork but differ in additional and less impactful choices might be strongly correlated.

Thus, our methodological portfolio sorting procedure suggests that monthly portfolio returns that share the same methodological choice at a specific fork might be dependent on each other. To control for this dependency structure, we assign the portfolio sorts that share the same choice at one decision fork into a cluster. For instance, all portfolio sorts that are based on value-weighted returns constitute one cluster, and all portfolio sorts that are based on equal-weighted returns represent another cluster. We repeat this assignment of portfolio sorts into clusters based on all remaining choices for all 14 decision forks depicted in Figure 3. This implies 32 clusters for sorting variables that are updated on a yearly basis.[1] As a robustness test, we also limit the number of clusters to the choices that correspond to the 7 most impactful forks determined in Walter et al. (2024).

In short, we assume that monthly portfolio returns that share the same choice at a fork under investigation are dependent on each other. These clusters will ensure that we preserve the dependency structure when repeatedly sampling portfolio returns during the following bootstrap procedure. Intuitively, when sampling portfolio returns from these clusters, we can determine the difference in monthly portfolio returns when using the new (CIZ) tape instead of the old (SIZ) tape for their construction. Repeating this with dependent samples will give an idea of how large these differences are, or in other words, whether the difference in the population likely contains zero or not. To test this, we implement the following bootstrap procedure for each of the 68 sorting variables separately:

1. For each cluster, we take a sample of portfolio sorts and their corresponding time series of monthly long-short returns. As our clusters have an overlap, we always take from each cluster a sample of portfolio sorts that have not yet been sampled in other clusters. Thereby, we avoid sampling essentially duplicate observations. That

[1] We only allow for a sorting variable lag of 3 months, 6 months, and at least 6 months, as in Fama and French (1992) for these sorting variables updated yearly. Thus, while Figure 3 depicts one more choice (a sorting variable lag of one month), this choice is only available for sorting variables updated monthly.

said, we still sample with replacement once we repeat this step. We set the number of portfolio sorts sampled from each cluster such that all portfolio sorts taken from all available clusters make up around 1% of all available portfolio sorts. Thus, we effectively sample 1% of the portfolio sorts, assuming that portfolio sorts that share the same choice of a fork under investigation are dependent on each other.

2. Based on this sample, we compute the raw difference between the long-short portfolio returns constructed with the new (CIZ) tape and the long-short portfolio returns constructed with data from the old (SIZ) tape. Then we compute for each portfolio sort and each month the raw differences of long-short portfolio returns when constructed with the old (SIZ) vs. the new (CIZ) tape. Thereafter, we compute the mean difference within each month across all differences in long-short portfolio returns that differ in methodological choices. Thus, we obtain for each month from January 1968 to December 2023 a difference for long-short portfolio returns from the old (SIZ) vs. the new (CIZ) tape.

3. We repeat the first and second steps above 10,000 times and always sample with replacement when moving to the next repetition.

These steps will generate 10,000 raw differences of monthly long-short portfolio returns in each month from January 1968 to December 2023. Then, we aggregate these 10,000 raw differences in each month into distributions and determine the 99% confidence interval. Intuitively, if the confidence interval does not contain zero, we can be almost certain that the difference in the real population – conditional on the dependency structure assumed above – is different from zero and, thus, statistically significant at the 1% level. Hence, we count the months in which the confidence interval does not contain zero and report the corresponding share of months with statistically significant differences in long-short portfolio returns.

After repeating these steps for all sorting variables, we find across all sorting variables that monthly long-short portfolio returns are statistically different from zero in 58.99% of the months (at the 1% significance level). We also investigate how this result changes once we only investigate the clusters based on the choices implied by the 7 most impactful forks determined by Walter et al. (2024). Including only the clusters from 7 out of 14 decision forks from Figure 3 yields quantitatively similar results: In 59.21% of the months, we find significantly different monthly long-short portfolio returns across all groups of sorting variables (at the 1% significance level).

IA.VII. Least-Correlated Methodological Paths

In this section, we explain how we estimate the set of methodological paths for which we observe the lowest time-series correlations among the corresponding long-short portfolio returns.

We identify the methodological paths that yield the least-correlated long-short portfolio returns for each sorting variable. For a relatively small variance-covariance matrix, one could directly estimate the average correlations of long-short portfolio returns of all possible subsets with a specific number of constituents. Then, one can keep those subsets below a certain absolute correlation cutoff, i.e., the 20% with the lowest absolute correlations within each subset.

However, we observe for most sorting variables 69,120 paths, which would require estimating the lower triangular of a very large correlation matrix with dimensions $69,120 \times 69,120$ for each sorting variable. Moreover, it would require analyzing the average correlation within all subsets, i.e., of 10 paths. Thus, we start with the following sampling procedure to determine a reasonable cutoff for the absolute correlation for each sorting variable:

A. Determine the cutoff for the absolute correlation:

1. Randomly select 100 methodological paths for portfolio sorts from Figure 3 for the sorting variable under investigation. Then, compute the correlation matrix capturing the time-series correlation between all long-short portfolio returns across the 100 methodological paths. Compute the average absolute correlation based on the lower triangular of the correlation matrix.

2. Repeat the first step 10,000 times and aggregate the average absolute correlations from all steps to a distribution. Identify the correlation corresponding to the 20th percentile of this distribution. Use the correlation corresponding to the 20th percentile as a cutoff for the next step.

B. Determine the subset of least-correlated paths based on the cutoff for the absolute correlation:

3. Randomly select 100 methodological paths for the sorting variable under investigation. Identify all possible subsets capturing 10 methodological paths. For each of these subsets, compute the correlation matrix based on the time-series correlation of all long-short portfolio returns within the subset. Then, compute the average absolute correlation based on the lower triangular of the correlation matrix. Keep only those subsets of paths that have a lower average absolute correlation among all paths of that subset than the correlation cutoff from Step 2.

4. Repeat the third step 10,000 times for each sorting variable and collect the least-correlated subsets of paths from each iteration. Drop duplicates of paths, as several iterations of Step 3 might select the same methodological path(s).

5. Take the subset of the least-correlated paths from Step 4. These methodological paths are candidates for the subset of least-correlated paths. Repeat Steps 1 - 4 again. Thereby, we further narrow down the set of least-correlated paths and ensure that we investigate the average absolute correlation within almost all subsets of the methodological paths that are left after Steps 1-4.

Thus, we obtain for each sorting variable the subset of methodological paths that result in the lowest absolute correlation between the corresponding long-short portfolio returns. Across all sorting variables, this sampling-based procedure selects around 2.1% of all methodological paths and reduces the absolute correlation between long-short portfolio returns across all methodological paths based on SIZ data from around 70.2% to around 55.7%. A selection based on CIZ data results in almost the same selected paths.

TABLE IA.14

Differences in Monthly Long-Short Portfolio Returns - Least-Correlated Methodological Paths

Table IA.14 reports summary statistics of the differences between monthly long-short portfolio returns estimated from the old (SIZ) and the new (CIZ) CRSP tape from 1968 to 2023. To account for methodological uncertainty, we estimate long–short portfolio returns for all 68 sorting variables across all portfolio-construction paths implied by the 14 methodological decisions from Figure 3. To control for dependencies across methodological paths, we select only a subset of paths (around 2.1%) that lead to the lowest absolute time-series correlations across the long-short portfolio returns of this subset. We compute the share of monthly long-short portfolio returns that differ by at least 1 bp ("$N_{1\ bp}$") and 10 bp ("$N_{10\ bp}$") across the selected paths and months for each sorting variable. The respective rows report these shares of long-short differences across sorting variables for specific groups (Panel A), types (Panel B), and averaged across specific time periods for all sorting variables (Panel C). Based on the same aggregation, we also report the share of monthly long-short portfolio returns with positive ("Pos.") and negative ("Neg.") return differences between the old and the new tape. Moreover, we compute for each sorting variable the mean absolute difference ("MAD") of long-short portfolio returns from the old vs. the new tape (average across selected methodological paths and months). The rows report these MADs averaged across groups (Panel A), types (Panel B), and across specific time periods for all sorting variables (Panel C). All statistics are in percent. The sorting variables belonging to each group or type can be found in Table C.1 of the Appendix. We classify sorting variables that require CRSP data (monthly CRSP returns) for their construction as CRSP-based (return-based).

Panel A. Sorting Variable Groups

Group	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Pos.	Neg.
Financing	47.93	14.47	0.05	51.41	48.59
Intangibles	44.46	10.93	0.04	49.69	50.31
Investment	46.10	11.19	0.04	51.99	48.01
Momentum	52.01	15.57	0.06	52.08	47.92
Profitability	47.16	11.10	0.04	48.68	51.32
Size	48.25	19.11	0.07	51.72	48.28
Trading Frictions	50.98	13.27	0.06	52.02	47.98
Valuation	49.77	13.20	0.05	49.76	50.24
Overall	48.33	13.61	0.05	50.92	49.08

Panel B. Sorting Variable Types

Type	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Pos.	Neg.
Non-CRSP-Based SVs	45.06	10.63	0.04	50.40	49.60
CRSP-Based SVs	51.53	15.01	0.06	50.95	49.05
Return-Based SVs	73.13	30.14	0.11	51.00	49.00

Panel C. Time Splits

Period	$N_{1\ bp}$	$N_{10\ bp}$	MAD	Pos.	Neg.
Jan. 1968 – Dec. 1999	59.08	17.17	0.07	51.01	48.98
Jan. 2000 – Dec. 2023	34.27	7.16	0.03	50.23	49.77
NBER Recessions	63.88	20.57	0.08	51.81	48.19
No Recessions	45.96	11.65	0.05	50.56	49.43

IA.VIII. Sorting Variables

In this section, we provide the details for the sorting variable construction. We analyze 68 sorting variables as reported in Table C.1 in the Appendix, which we group into 8 groups to facilitate comparisons. Note that our construction closely follows Hou et al. (2020). Moreover, we replace any negative values of total assets (AT), sales (SALE), capital expenditures (CAPX), and inventories (INVT) with missing values.

A. Financing

Composite Debt Issuance (CDI). We follow Lyandres et al. (2008) and measure composite debt issuance (CDI) for each firm in each fiscal year t from Compustat annual data as the logarithmic growth rate in the book value of debt from fiscal year $t-5$ to fiscal year t. The book value of debt is measured by the sum of current debt (DLC) and long-term debt (DLTT).

Composite Share Issuance (CSI). Daniel and Titman (2006) propose to measure composite share issuance (CSI) from CRSP data as the difference between the change in market equity and the cumulative log return of a stock. Both, the change in market equity and cumulative log returns are measured in each month from year t to year $t-5$.

Change in Common Equity (DBE). To capture the change in common equity (DBE) according to Richardson et al. (2005), we calculate the following ratio from Compustat annual data:

$$DBE_t = \frac{CEQ_t - CEQ_{t-1}}{AT_{t-1}},$$

where CEQ represents common equity and AT total assets.

Change in Current Operating Liabilities (DCOL). Richardson et al. (2005) measure the change in current operating liabilities (DCOL) for each firm in each fiscal year t from annual Compustat data:

$$DCOL_t = \frac{(LCT_t - DLC_t) - (LCT_{t-1} - DLC_{t-1})}{AT_{t-1}},$$

where LCT are current liabilities, DLC short-term debt, and AT total assets. We replace missing values of DLC with 0.

Change in Financial Liabilities (DFNL). We define the change in financial liabilities (DFNL) similar to Richardson et al. (2005) for each stock in each fiscal year t from annual Compustat data in the following way:

$$DFNL_t = \frac{(DLTT_t + DLC_t + PSTK_t) - (DLTT_{t-1} + DLC_{t-1} + PSTK_{t-1})}{AT_{t-1}},$$

where DLTT is long-term debt, DLC short-term debt, PSTK the value of preferred stocks, and AT total assets. Missing values of DLTT, DLC, and PSTK are set to 0 if at least 1 of the 3 variables is available.

Net Debt Financing (NDF). We follow Bradshaw et al. (2006) and compute net debt financing (NDF) for each stock in each fiscal year t from annual Compustat data as:

$$NDF_t = \frac{DLTIS_t - DLTR_t + DLCCH_t}{\frac{1}{2}\left(AT_t + AT_{t-1}\right)},$$

where DLTIS are cash proceeds from the issuance of long-term debt, DLTR are cash payments for long-term debt reductions, DLCCH are the net changes in current debt, and AT total assets. We replace missing values of DLCCH with 0. Data starts in January 1972 due to data availability of financing variables. Data starts in January 1972 due to data availability of financing variables.

Net Equity Financing (NEF). We measure net equity financing (NEF) similar to Bradshaw et al. (2006) for each stock in each fiscal year t from annual Compustat data:

$$NEF_t = \frac{SSTK_t - PRSTKC_t - DV_t}{\frac{1}{2}\big(AT_t + AT_{t-1}\big)},$$

where SSTK are proceeds from the sale of common and preferred stocks, PRSTKC are payments for the repurchase of common and preferred stocks, DV are cash payments for dividends, and AT total assets. Data starts in January 1972 due to data availability of financing variables. Data starts in January 1972 due to data availability of financing variables.

Net External Financing (NXF). We capture net external financing (NXF) for each stock in each fiscal year t similar to Bradshaw et al. (2006) by the sum of net debt financing and net equity financing. Both variables are described above. Data starts in January 1972 due to data availability of financing variables.

B. Intangibles

Advertisement Expenses to Market Equity (ADM). Chan et al. (2001) suggest measuring the advertising expense to market ratio (ADM) as advertising expenses (Compustat item XAD) divided by market equity, which is obtained from CRSP at the end of each fiscal year. We exclude observations with negative advertising expenses. We start our measure in January 1973 to ensure sufficient data coverage.

Cash-Flow Volatility (CFV). We follow Huang (2009) and compute operating cash flows to sales for each stock in each fiscal quarter q from quarterly Compustat data:

$$Operating\ cash\ flows_q = \frac{IBQ_q + DPQ_q + (WCAP_q - WCAP_{q-1})}{SALEQ_q},$$

where IBQ are quarterly income before extraordinary items, DPQ are quarterly depreciation and amortizations, WCAPQ are quarterly working capital, and SALEQ are quarterly sales. Cash-flow volatility (CFV) for each stock in each fiscal quarter corresponds to the standard deviation of operating cash flows during the past 16 quarters. We require a minimum of 8 observations. We start our measure in January 1978 to ensure sufficient data coverage.

Earnings' Predictability (EPRD). Francis et al. (2004) define split-adjusted earnings per share (EPSA) from Compustat data as earnings per share (EPSPX) divided by the adjustment factor (AJEX). We follow Francis et al. (2004) and measure earnings predictability (EPRD) for each stock as the residual volatility (u_t) from the following auto-regressive process:

$$EPSA_t = \alpha + \beta \cdot EPSA_{t-1} + u_t.$$

Moreover, we measure this auto-regressive process over the last 10 years and always require 10 years of non-missing observations.

Hiring Rate (HR). We follow Belo et al. (2014) and obtain the hiring rate (HR) for each stock in each fiscal year t from annual Compustat data:

$$HR_t = \frac{EMP_t - EMP_{t-1}}{\frac{1}{2}\big(EMP_t + EMP_{t-1}\big)},$$

where EMP represents the number of employees. Moreover, we exclude firms with a hiring rate of 0.

Kaplan and Zingales (1997) Index (KZI). We obtain the Kaplan and Zingales index (KZI) for each firm in each fiscal year from annual Compustat data by following Lamont et al. (2001):

$$\begin{aligned} KZI_t = & -1.002 \cdot \frac{IB_t + DP_t}{PPENT_{t-1}} + 0.283 \cdot \frac{AT_t + ME_t - CEQ_t - TXDB_t}{AT_t} + 3.139 \cdot \frac{DLC_t + DLTT_t}{DLC_t + DLTT_t + SEQ_t} \\ & - 39.368 \cdot \frac{DVC_t + DVP_t}{PPENT_{t-1}} - 1.315 \cdot \frac{CHE_t}{PPENT_{t-1}}, \end{aligned}$$

where IB corresponds to income before extraordinary items, DP to depreciation and amortization, PPENT to property, plant, and equipment, AT to total assets, ME to market equity from CRSP at the end of each fiscal year, CEQ to common equity, TXDB to deferred taxes, DLC to current debt, DLTT to long-term debt, SEQ to shareholder equity, DVC to dividends of common stock, DVP to dividends of preferred stocks, and CHE to cash holdings.

Labor Force Efficiency (LFE). We define the labor force efficiency (LFE) as in Abarbanell and Bushee (1998) for each firm in each fiscal year t from annual Compustat data:

$$LFE_t = \left(\frac{SALE_t}{EMP_t} - \frac{SALE_{t-1}}{EMP_{t-1}} \right) \Big/ \frac{SALE_{t-1}}{EMP_{t-1}},$$

where SALE corresponds to sales and EMP to employees.

Operating Leverage (OL). We follow Novy-Marx (2013) and compute operating leverage (OL) from Compustat data as cost of goods sold (COGS) plus selling, general and administrative expenses (XSGA), both scaled by current total assets (AT).

R&D Expenses to Market Equity (RDM). Chan et al. (2001) propose to compute the R&D expense to market ratio (RDM) as R & D expenses (Compustat item XRD) divided by market equity from the end of each fiscal year. We obtain market equity data from CRSP and include only observations with positive R&D expenses. We start our measure in January 1976 because R&D expenses were standardized in 1975.

Real-Estate Ratio (RER). We define the real-estate ratio (RER) similar to Tuzel (2010) with Compustat data. Prior to 1983, it corresponds to the sum of buildings (PPENB) and capital leases (PPENLS) scaled by net property, plant and equipment (PPENT). After the end of 1983, it is measured as the sum of buildings at cost (FATB) and leases at cost (FATL), both divided by gross property, plant and equipment (PPEGT). Subsequently, we winsorize the real estate ratios in each fiscal year at the 1% and 99% percentile. The industry-adjusted real-estate ratio is obtained by subtracting the industry average real-estate ratio from each stock-specific real-estate ratio. We use 2-digit SIC codes to assign stocks to industries. We always require at least 5 observations to calculate the industry average each year. Note that real estate data starts in 1969, limiting the observation period for this specific sorting variable. Data for the real estate ratio starts in January 1970 due to data availability.

Tangibility (TAN). We capture the tangibility (TAN) of each firm in each fiscal year according to Hahn and Lee (2009) from annual Compustat data:

$$TAN_t = \frac{CHE_t}{AT_t} + 0.715 \cdot \frac{RECT_t}{AT_t} + 0.547 \cdot \frac{INVT_t}{AT_t} + 0.535 \cdot \frac{PPEGT_t}{AT_t},$$

where CHE corresponds to cash holdings, RECT to accounts receivable, INVT to inventory, PPEGT to property, plant, and equipment, and AT to total assets.

Whited and Wu Index (WW). We closely follow Whited and Wu (2006) and measure financing constraints for each firm in each fiscal year t from annual Compustat data:

$$\begin{aligned} WW_t = & -0.091 \cdot \frac{IB_t + DP_t}{AT_t} - 0.062 \cdot \mathbb{1}_{DVPSX\_F>0} + 0.021 \cdot \frac{DLTT_t}{AT_t} - 0.044 \cdot \ln(AT_t) + 0.102 \cdot ISG_t \\ & - 0.035 \cdot \frac{SALE_t - SALE_{t-1}}{SALE_{t-1}}, \end{aligned}$$

where IB is income before extraordinary items, DP depreciation and amortization, AT total assets, $\mathbb{1}_{DVPSX>0}$ a dummy variable equal to 1 if the firm pays out cash dividends (DVPSX_F), DLTT long-term debt and SALE sales. Moreover, ISG is the industry growth rate of sales, while industries are defined by 3-digit SIC codes. Industries with less than 2 firms are excluded. Since Whited and Wu (2006) estimate this index with quarterly data, we replace the annual growth rates in industry sales growth and stock-specific sales growth with their implied quarterly compounded growth rates. Lastly, we winsorize the distribution of each sub-variable of the Whited and Wu index at the 1% and 99% quantile.

C. Investment

Abnormal Corporate Investment (ACI). We measure abnormal corporate investments (ACI) from Compustat annual data for each firm in each fiscal year t as in Titman et al. (2004):

$$ACI_t = \frac{CE_t}{\frac{1}{3}\left(CE_{t-1} + CE_{t-2} + CE_{t-3}\right)} - 1,$$

where CE corresponds to capital expenditures (Compustat item CAPX) divided by sales (SALE). We follow Hou et al. (2020) and exclude stocks with sales below 10 million dollars.

Asset Growth (AG). We follow Cooper et al. (2008) and measure asset growth (AG) for each stock in each fiscal year t from Compustat data as the change in total assets (AT) from year t to year $t-1$, divided by total assets from year $t-1$.

Change in Net Operating Assets (DNOA). We follow Hirshleifer et al. (2004) and measure net operating assets for each stock in each fiscal year t from Compustat annual data:

$$Net\ operating\ assets_t = (AT_t - CHE_t) - (AT_t - DLC_t - DLTT_t - MIB_t - PSTK_t - CEQ_t),$$

where AT corresponds to total assets, CHE to cash and short-term investments, DLC to current liabilities, DLTT to long-term debt, MIB to minority interests, PSTK to the value of preferred stocks, and CEQ to common equity. Missing values in DLC, DLTT, MIB, and PSTK are set to 0. The change in net operating assets (DNOA) is then the difference between net operating assets of fiscal year t and fiscal year $t-1$ scaled by total assets of year $t-1$.

Change in Property, Plant, Equipment and Inventory to Assets (DPIA). We add the annual change in gross property, plant and equipment (PPEGT) to the annual change in inventory (INVT) and scale this sum by 1-year-lagged total assets (AT). Thus, we obtain the change in property, plant, equipment and inventories (DPIA) as in Lyandres et al. (2008) from Compustat data.

Change in Net Non-Cash Working Capital (DWC). Following Richardson et al. (2005) we define non-cash working capital from Compustat data as:

$$WC_t = (ACT_t - CHE_t) - (LCT_t - DLC_t),$$

where ACT corresponds to current assets, CHE to cash, LCT to current liabilities, and DLC to short-term debt. We set missing values of DLC to 0. The change in net non-cash working capital (DWC) corresponds to the change of WC from fiscal year t to fiscal year $t-1$ scaled by total assets from fiscal year $t-1$.

Investment Growth (IG). We compute investment growth (IG) from annual Compustat data as the annual change in capital expenditures (CAPX) from fiscal year t to year $t-1$, scaled by capital expenditures from year $t-1$.

Inventory Changes (DINV). Thomas and Zhang (2002) suggest measuring the change in inventory (DINV) from Compustat data as the annual change in inventories (INVT) from fiscal year t to fiscal year $t-1$, divided by average total assets (AT) over the fiscal year t and $t-1$.

Net Operating Assets (NOA). We follow Hirshleifer et al. (2004) and compute net operating assets (NOA) from Compustat data:

$$NOA_t = \frac{(AT_t - CHE_t) - (AT_t - DLC_t - DLTT_t - MIB_t - PSTK_t - CEQ_t)}{AT_{t-1}},$$

where AT is total assets, CHE cash and short-term investments, DLC short-term debt, DLTT long-term debt, MIB minority interest, PSTK preferred stock, and CEQ common equity. We replace missing values of DLC, DLTT, MIB, and PSTK as 0.

Operating Accruals (OA). The definition of operating accruals (OA) before 1988 closely follows Sloan (1996):

$$OA_t = \frac{(\Delta ACT_t - \Delta CHE_t) - (\Delta LCT_t - \Delta DLC_t - \Delta TXP_t) - DP_t}{AT_{t-1}},$$

where ACT is current assets, CHE cash, LCT current liabilities, DLC short-term debt, TXP taxes payable, and DP depreciation and amortization. Moreover, we replace missing values of DLC and TXP with 0. Due to data availability, we follow Hribar and Collins (2002) and compute operating accruals from 1988 and onward as:

$$OA_t = \frac{NI_t - OANCF_t}{AT_{t-1}},$$

where NI is net income and OANCF corresponds to net cash flow from operations. All items are from Compustat data.

Percent Total Accruals (PTA). Hafzalla et al. (2011) suggest measuring percent total accruals (PTA) from Compustat data as total accruals (TA) divided by the absolute value of net income (NI). Before 1988 we follow Hou et al. (2020) and define PTA as:

$$\begin{aligned} PTA_t =& \big(\Delta(ACT_t - CHE_t - LCT_t + DLC_t) + \Delta(AT_t - ACT_t - IVAO_t - LT_t + LCT_t + DLTT_t) \\ & + \Delta(IVST_t + IVAO_t - DLTT_t - DLC_t - PSTK_t)\big)/|NI_t|, \end{aligned}$$

where ACT is current assets, LCT current liabilities, DLC short-term debt, AT total assets, IVAO investments and advances, LT total liabilities, DLTT long-term debt, IVST short-term investments, PSTK preferred stock, and NI net income. Δ refers to the change from fiscal year t to fiscal year $t-1$. Moreover, missing values of IVAO, DLTT, DLC, IVST, and PSTK are set to 0. From 1988 and, thereafter, we follow Hribar and Collins (2002) and measure PTA from Compustat data as

$$PTA_t = \frac{NI_t - OANCF_t - IVNCF_t - FINCF_t + SSTK_t - PRSTKC_t - DV_t}{|NI_t|},$$

where NI corresponds to net income, OANCF to total operating cash flows, IVNCF to total investing cash flows, FINCF to total financing cash flows, SSTK to the sale of stocks, PRSTKC to stock repurchases, and DV to dividends. Moreover, we set missing values of SSTK and DV to 0.

D. Momentum

Abnormal Returns on Earnings' Announcements (ABR). We follow Chan et al. (1996) and estimate abnormal returns around quarterly earnings' announcements in month t as the difference between the individual stock return $r_{i,d}$ and the market index $r_{m,d}$ on day d. We cumulate these abnormal returns around the following 4-day event window including 2 days before the quarterly earnings announcement, the day of the announcement, and 1 day after:

$$ABR_{i,t} = \sum_{d=-2}^{d=1} (r_{i,d} - r_{m,d}).$$

The quarterly earnings announcement date corresponds to Compustat item RDQ and has to be after the fiscal quarter end date to exclude potential recording errors. Data starts in January 1972 because the earnings announcement date RDQ is only available from 1972 onwards.

Return Momentum (MOM). We compute return momentum (MOM) for each stock in each month as the cumulative return from month $t-12$ to month $t-2$ skipping the most recent month as in Fama and French (1996).

Residual Momentum (RMOM). As in Blitz et al. (2011), we define the 11-month residual momentum (RMOM) in each month and for each stock as cumulative residual returns from month $t-2$ to month $t-12$, scaled by the standard deviation of residual returns over the same time horizon. Residual returns are obtained in each month from regressing monthly excess stock returns from month $t-1$ to month $t-36$ on the Fama and French (1993) 3-factor model. Throughout these rolling regressions, we always require 36 monthly returns.

Revenue Surprise (RS). Similar to Jegadeesh and Livnat (2006), we construct revenues per share from Compustat quarterly data for each stock i in each quarter q as:

$$Revenues\,per\,share_q = \frac{SALEQ_q}{CSHPRQ_q \cdot AJEXQ_q},$$

where SALEQ corresponds to quarterly revenues, CHSPRQ to the correction factor for quarterly shares outstanding, and AJEXQ to quarterly shares outstanding. Revenue surprises (RS) correspond then to the change in revenues per share over the last 4 quarters scaled by the standard deviation of the change in revenues per share over the last 8 quarters. We require at least 6 quarterly observations for this rolling standard deviation. Lastly, the earnings announcement date has to be after the fiscal quarter end date. Data starts in January 1972 because the earnings announcement date RDQ is only available from 1972 onwards.

Standardized Unexpected Earnings (SUE). As in Foster et al. (1984), we calculate unexpected earnings from quarterly Compustat data for each stock in each quarter q as the change in split-adjusted earnings per share from its value 4 quarters ago:

$$Unexpected\,earnings\,per\,share_q = \frac{EPSPXQ_q}{AJEXQ_q} - \frac{EPSPXQ_{q-4}}{AJEXQ_{q-4}},$$

where ESPSPXQ are quarterly earnings per share and AJEXQ denotes the number of shares outstanding in each quarter. Then, standardized unexpected earnings (SUE) are defined as unexpected earnings per share divided by the standard deviation of unexpected earnings per share over the previous 8 quarters. We require at least 6 quarterly observations for this rolling standard deviation. Moreover, the earnings announcement date has to be before the fiscal quarter end date. Data starts in January 1972 because the earnings announcement date RDQ is only available from 1972 onwards.

Tax Expense Surprise (TES). We follow Thomas and Zhang (2011), and calculate tax expense surprises (TES) for each stock in each quarter q as the change in tax expenses per share over the last 4 quarters scaled by assets per share from 4 quarters ago ($q-4$):

$$TES_q = \frac{\frac{TXTQ_q}{CSHPRQ_q \cdot AJEXQ_q} - \frac{TXTQ_{q-4}}{CSHPRQ_{q-4} \cdot AJEXQ_{q-4}}}{\frac{ATQ_{q-4}}{CSHPRQ_{q-4} \cdot AJEXQ_{q-4}}},$$

where TXTQ represents quarterly tax expenses, ATQ quarterly total assets, AJEXQ quarterly shares outstanding, and CSHPRQ the factor to adjust quarterly shares outstanding. We exclude firms that do not pay taxes from our sample and require the earnings announcement date to be after the fiscal quarter end date. We follow Hou et al. (2020) and start our calculation in January 1976 to ensure sufficient observations for this sorting variable

52-Week High (52W). We define the 52-week high (52W), similar to George and Hwang (2004), for each stock in each month t as the daily split-adjusted stock price at the end of each month scaled by the highest daily split-adjusted stock price over the previous 12 months.

E. Profitability

Asset Turnover (ATO). We follow Soliman (2008) and compute asset turnover (ATO) from Compustat data as sales (SALE) divided by net operating assets from the previous fiscal year:

$$ATO_t = \frac{SALE_t}{(AT_{t-1} - CHE_{t-1} - IVAO_{t-1}) - (AT_{t-1} - DLC_{t-1} - DLTT_{t-1} - MIB_{t-1} - PSTK_{t-1} - CEQ_{t-1})},$$

where Compustat item AT are total assets, item CHE are cash and short-term investments, and IVAO are other investments and advances. Moreover, item DLC represents debt in current liabilities, DLTT long-term debt, MIB minority interests, PSTK preferred stocks, and CEQ common equity. We follow Hou et al. (2020) and replace missing values of IVAO, DLC, DLTT, MIB, and PSTK with 0. Similar to Hou et al. (2020), we exclude observations with negative net operating assets.

Book Leverage (BL). Similar to Fama and French (1992), we compute the book leverage (BL) of each firm in each fiscal year by the ratio of total assets (Compustat item AT) and book equity. The definition of book equity follows from Davis et al. (2000) and is disclosed below when describing the book-to-market ratio.

Cash-Based Operating Profitability (CBOP). The definition of cash-based operating profitability (CBOP) closely follows Ball et al. (2016) and is based on Compustat data:

$$CBOP_t = \frac{REVT_t - COGS_t - XSGA_t + XRD_t - \Delta RECT_t - \Delta INVT_t - \Delta XPP_t + \Delta DRC_t + \Delta DRLT_t + \Delta AP_t + \Delta XACC_t}{AT_t},$$

where REVT is total revenue, COGS are cost of goods sold, XSGA are selling, general and administrative expenses, and XRD are R&D expenses. Moreover, $\Delta RECT_t$ is the change in accounts receivable, $\Delta INVT$ the change in inventory, ΔXPP is the change in prepaid expenses, $\Delta DRC_t + \Delta DRLT$ the change in deferred revenues, ΔAP the change in trade accounts payable and $\Delta XACC$ is the change in accrued expenses. We follow Hou et al. (2020) and set missing values of XRD and all missing changes to 0.

Capital Turnover (CTO). We measure capital turnover (CTO) from Compustat data as sales (SALE) divided by total assets (AT) from the previous fiscal year (Haugen and Baker (1996)).

Gross Profits to Assets (GPA). We follow Novy-Marx (2013) and obtain gross profits to assets (GPA) from Compustat data as total revenues (REVT) minus cost of goods sold (COGS), scaled by current total assets (AT).

Ohlson's O-Score (O). Ohlson (1980) suggests evaluating the financial stability of a firm with the following linear relation:

$$\begin{aligned} O_t = & -1.32 - 0.407 \cdot \log(AT_t) + 6.03 \cdot \frac{DLC_t + DLTT_t}{AT_t} - 1.43 \cdot \frac{ACT_t - LCT_t}{AT_t} + 0.076 \cdot \frac{LCT_t}{AT_t} - 1.72 \cdot \mathbb{1}_{LT_t > AT_t} - 2.37 \cdot \frac{NI_t}{AT_t} \\ & - 1.83 \cdot \frac{PI_t + DP_t}{LT_t} + 0.285 \cdot \mathbb{1}_{NI_t < 0 \text{ \& } NI_{t-1} < 0} - 0.521 \cdot \frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|}. \end{aligned}$$

All data items are obtained from Compustat: AT corresponds to total assets, DLC to short-term debt, DLTT to long-term debt, ACT to current assets, LCT to current liabilities, LT to total liabilities, PI to pretax income, DP to depreciation and amortization, and NI to net income. We follow Hou et al. (2020) and winsorize all variables except for dummy variables at the 1% and 99 % quantile of their respective distribution.

Operating Profits to Book Equity (OPE). We closely follow Fama and French (2015) and compute operating profits to book equity for each firm in each fiscal year t from Compustat annual data:

$$OPE_t = \frac{REVT_t - COGS_t - XSGA_t - XINT_t}{BE_t},$$

where REVT corresponds to total revenues, COGS to cost of goods sold, XSGA to selling, general and administrative expenses, XINT to interest expenses, and BE to book equity. The definition of book equity follows the disclosed definition for the variable book-to-market (below). Moreover, missing values in COGS, XSGA, and XINT are set to 0.

Return on Assets (ROA). We obtain data on return on assets (ROA) for each stock in each fiscal quarter q from Compustat quarterly data and closely follow Balakrishnan et al. (2010):

$$ROA_q = \frac{IBQ_q}{ATQ_{q-1}},$$

where IBQ corresponds to quarterly income before extraordinary items, and ATQ represents quarterly total assets. Moreover, the earnings announcement date of each record has to be after the fiscal quarter end date to ensure consistent recording. Data starts in January 1972 because the earnings announcement date RDQ is only available from 1972 onwards.

Return on Equity (ROE). Hou et al. (2015) define return on equity (ROE) for each firm in each fiscal quarter as:

$$ROE_q = \frac{IBQ_q}{BEQ_{q-1}},$$

where IBQ corresponds to quarterly income before extraordinary items and BEQ to quarterly book equity. Quarterly book equity (BEQ) is computed as the book equity of shareholders plus balance sheet deferred taxes and investment tax credit minus the book value of preferred stock. Depending on data availability, we measure shareholders' equity by SEQQ, or the sum of common equity (CEQQ) and the par value of preferred stock (PSTKQ), or if all previous items are unavailable by total assets (ATQ) minus total liabilities (LTQ). The book value of preferred stocks corresponds to PSTKRQ, to PSTKQ if PSTKRQ is unavailable or to 0 if both are unavailable. Balance sheet deferred taxes and investment tax credit is TXDITCQ, TXDBQ if TXDITCQ is missing or 0 if both are missing. Data starts in January 1972 because the earnings announcement date RDQ is only available from 1972 onwards.

Taxable Income to Book Income (TBI). We closely follow Green et al. (2013) and compute taxable income to book income (TBI) for each firm in each fiscal year t from Compustat annual data as:

$$TBI_t = \frac{PI_t}{NI_t},$$

where PI is pretax income and NI is net income. Moreover, we require positive pretax and net income.

Altman's Z-Score (Z). We measure the Altman (1968) Z-score for each firm in each fiscal year from Compustat annual data by the following definition:

$$Z_t = 1.2 \cdot \frac{ACT_t - LCT_t}{AT_t} + 1.4 \cdot \frac{RE_t}{AT_t} + 3.3 \cdot \frac{OIADP_t}{AT_t} + 0.6 \cdot \frac{ME_t}{LT_t} + \frac{SALE_t}{AT_t},$$

where ACT is current assets, LCT current liabilities, AT total assets, RE retained earnings, OIADP earnings before interest and taxes, ME market equity from the end of the fiscal year (from CRSP), LT total liabilities, and SALE corresponds to sales. Lastly, we winsorize the distributions of all 5 sub-variables of the Z-score at the 1% and 99% quantile in each fiscal year.

F. Size

Logarithm of Market Equity (ME). We follow Fama and French (1992) and compute the size of each stock (ME) in each month as the natural logarithm of the market equity. We obtain market equity data from CRSP by multiplying the shares outstanding (SHROUT) with the corresponding share price (PRC).

G. Trading Frictions

Amihud Illiquidity Measure (AMI). Amihud (2002) proposes to measure the illiquidity of each firm on each day d from daily CRSP data as the absolute daily return scaled by the daily dollar trading volume:

$$return\,to\,volume_d = \frac{|RET_d|}{PRC_d * VOL_d},$$

where RET is the daily return, PRC is the daily price, and VOL is the daily volume of stocks traded. The Amihud illiquidity measure (AMI) for each firm in each month t corresponds to the average return to volume estimate over the last 6 months. We require at least 50 observations for this average and adjust the trading volume of NASDAQ stocks according to Gao and Ritter (2010).

Beta Relative to the Market (BETA). We compute the market beta (BETA) for each stock in each month t from monthly CRSP data and similar as in Fama and MacBeth (1973). Specifically, we run the following time-series regression over the previous 5 years:

$$r_t^e = \alpha + \beta_1 \cdot (MKT_t - R_t^f) + u_t.$$

Moreover, we require at least 24 monthly observations for the regression above. The market beta for each firm in each month t corresponds to the regression coefficient β_1. Data on the market factor MKT_t is obtained from Kenneth French's website.

Frazzini and Pedersen Beta (BFP). Frazzini and Pedersen (2014) suggest measuring the beta (BFP) for each stock i and each month t from daily CRSP data as:

$$BFP_{i,t} = \frac{\hat{\rho} \cdot \hat{\sigma_i}}{\hat{\sigma_m}},$$

where $\hat{\sigma_i}$ corresponds to the standard deviation of each stock i measured as the standard deviation of daily logarithmic returns over the previous 750 days. Moreover, $\hat{\sigma_m}$ is the standard deviation of the market and is obtained as the standard deviation of daily logarithmic returns over the previous 750 days. Throughout the calculations of these standard deviations for each month t, we require at least 120 daily observations. Lastly, $\hat{\rho}$ is the return correlation between the market m and stock i. We estimate this return correlation for each month t over the last 750 daily returns. When estimating the return correlation, we use overlapping 3-day logarithmic returns for each stock i on each day d: $r_{i,d} = \sum_{k=-2}^{0} \ln(1 + r_{i,d+k})$ instead of 1-day raw returns.

Dollar Trading Volume (DTV). We follow Brennan et al. (1998) and compute dollar trading volume (DTV) from daily CRSP data as the average dollar trading volume from month $t-1$ to month $t-6$. We require at least 50 days of observations when computing this average. Dollar trading volume is defined as share price (PRC) multiplied by the number of shares traded on each day (VOL). Moreover, we adjust dollar trading volume from NASDAQ according to Gao and Ritter (2010).

Idiosyncratic Skewness (ISKEW). We regress the daily excess returns of each stock on the Fama and French (1993) factor model:

$$r_t^e = \alpha + \beta_1 \cdot (MKT_t - R_t^f) + \beta_2 \cdot SMB_t + \beta_3 \cdot HML_t + u_t.$$

Throughout these regressions, we require at least 15 daily observations for each month. Idiosyncratic skewness (ISKEW) relative to the Fama and French (1993) model is then measured in each month as the skewness of residuals u_t (Bali et al. (2016)).

Idiosyncratic Volatility (IVOL). We follow Ang et al. (2006) and compute idiosyncratic volatility relative to the Fama and French (1993) factor model (IVOL) as the volatility of residuals from the following regression:

$$r_t^e = \alpha + \beta_1 \cdot (MKT_t - R_t^f) + \beta_2 \cdot SMB_t + \beta_3 \cdot HML_t + u_t.$$

In detail, we regress in each month the excess return of each stock on the Fama and French (1993) factor model using daily returns from CRSP and Kenneth French. Moreover, we require at least 15 daily observations for each month.

Maximum Daily Return (MDR). We compute the maximum daily return (MDR) for each stock in each month t similar to Bali et al. (2011) from daily CRSP data as the maximal daily return in each month t. Moreover, we require at least 15 return observations for each month t.

Short-Term Reversal (SREV). We follow Jegadeesh (1990) and measure the short-term reversal (SREV) for each firm in each month t from monthly CRSP data as the stock return during month t. We require a valid return on month t for all stocks. All sorting variables are subsequently lagged according to decision fork "sorting variable lag", i.e., this definition does not produce a look-ahead bias.

Share Turnover (TUR). Datar et al. (1998) propose to measure the daily share turnover (TUR) of each stock on each day as the number of shares traded (VOL) scaled by the number of shares outstanding (SHROUT) on the same day. The variable share turnover for each firm in each month t is then the average daily share turnover over the previous 6 months. Throughout this calculation, we require at least 50 daily observations. Lastly, we adjust the trading volume of NASDAQ stocks according to Gao and Ritter (2010).

H. Valuation

Assets to Market Equity (AM). Similar to Fama and French (1992) we measure assets to market equity (AM) for each stock in each fiscal year by total assets (AT) divided by market equity (CRSP) from the end of the fiscal year t. We exclude observations with negative total assets.

Book Equity to Market Equity (BM). This paper follows Davis et al. (2000) and computes the book-to-market ratio (BM) as book equity from Compustat divided by market equity from CRSP. Market equity is measured at the end of each fiscal year. Book equity corresponds to the book equity of shareholders plus balance sheet deferred taxes and investment tax credit (Compustat item TXDITC or TXDB + ITCB if TXDITC is unavailable) minus the book value of preferred stock. Depending on data availability, we measure shareholders' equity by SEQ, or the sum of common equity (CEQ) and the par value of preferred stock (PSTK), or if all previous items are unavailable by total assets (AT) minus total liabilities (LT). The book value of preferred stock corresponds in the following order either to the redemption value (PSTKRV), or the liquidation value (PSTKL), or if all previous items are unavailable to the par value (PSTK). We replace missing values of TXDITC or the book value of preferred stock with 0.

Cash-Flow to Market Equity (CFM). Lakonishok et al. (1994) suggest measuring the cash-flow-to-price ratio (CFM) from Compustat data as income before extraordinary items (IB) plus depreciation (DP), both divided by market equity (CRSP) from the end of the fiscal year. We exclude all stocks with negative cash flows.

Debt to Market Equity (DM). Following Bhandari (1988), the debt to market ratio (DM) is defined as short-term debt (Compustat item DLC) plus long-term debt (Compustat item DLTT) divided by market equity obtained from CRSP at the end of each fiscal year. We exclude stocks with missing DLC and DLTT observations.

Enterprise Book Equity to Market Equity (EBM). We follow Penman et al. (2007) and obtain enterprise book equity scaled by market equity (EBM) for each firm in each fiscal year t as net debt plus book equity scaled by net debt plus market equity:

$$EBM_t = \frac{(DLTT_t + DLC_t + PSTK_t + DVPA_t - TSTKP_t) - CHE_t + BE_t}{(DLTT_t + DLC_t + PSTK_t + DVPA_t - TSTKP_t) - CHE_t + ME_t},$$

where DLTT corresponds to long-term debt, DLC to current liabilities, PSTK to the value of preferred stock, DVPA to preferred dividends in arrears, TSTKP to preferred treasury stock, CHE to cash and short-term investments, and ME to the market equity from CRSP measured at the end of each fiscal year t. Lastly, book equity BE is computed as common equity (CEQ) plus TSTKP minus DVPA. Lastly, missing observations in DVPA and TSTKP are set to 0. We require that the sum of net debt and book equity as well as the sum of net debt plus market equity are positive.

Earnings to Market Equity (EM). We follow Basu (1983) and compute the earnings-to-price ratio (EM) as income before extraordinary items (Compustat item IB) divided by market equity from CRSP. Market equity corresponds to the end of each fiscal year. We exclude firms with negative earnings.

Net Debt to Market Equity (NDM). Similar to Penman et al. (2007), net debt to price (NDM) is measured from Compustat annual data for each stock in each fiscal year t in the following way:

$$NDM_t = \frac{(DLTT_t + DLC_t + PSTK_t + DVPA_t - TSTKP_t) - CHE_t}{ME_t},$$

where DLTT corresponds to long term-debt, DLC to current liabilities, PSTK to the value of preferred stock, DVPA to preferred dividends in arrears, TSTKP to preferred treasury stock, CHE to cash and short-term investments, and ME to the market equity from CRSP measured at the end of each fiscal year t. Lastly, missing observations in DVPA and TSTKP are set to 0.

Net Payout Yield (NPY). Boudoukh et al. (2007) suggest measuring the net payout yield (NPY) of each stock in the following way:

$$NPY_t = \frac{(DVC_t + PRSTKC_t + \Delta PSTKRV_t \cdot \mathbb{1}_{\Delta PSTKRV<0}) - (SSTK_t - \Delta PSTKRV_t \cdot \mathbb{1}_{\Delta PSTKRV>0})}{ME_t},$$

where DVC are dividends from common stock, PRSTKC is the purchase of common and preferred stock, PSTKRV is the value of the net number of preferred stocks outstanding, and SSTK reflects the sale of common and preferred stocks. $\mathbb{1}_{\Delta PSTKRV<0}$ is a dummy variable that has value of 1 if the annual change in PSTKRV is negative and 0 otherwise. Market equity (ME) is from CRSP and corresponds to the end of each fiscal year. Moreover, we exclude stocks with negative net payouts. Data starts in January 1972 because of sufficient data coverage for the sale of common and preferred stocks.

Operating Cash-Flow to Market Equity (OCM). We follow Desai et al. (2004) and compute the ratio of operating cash-flows to price (OCM) as operating cash flows from Compustat divided by the market equity at the end of each fiscal year from CRSP. Before 1988, we measure operating cash flows as funds from operations (FOPT) minus the change in working capital (item WCAP) and as net cash flows from operating activities (OANCF) thereafter. Moreover, we exclude firms with negative operating cash flows. Data starts in January 1972 because of sufficient data coverage for funds from operations.

Long-Term Reversal (REV). We measure the long-term reversal effect (REV) suggested by De Bondt and Thaler (1985) for each stock in each month t by the cumulative returns from month $t-60$ to month $t-13$.

Sales to Market Equity (SM). We compute the sales to price ratio (SM) as sales (Compustat item SALE) divided by the market equity at the end of each fiscal year (Barbee Jr et al., 1996). Stocks with negative sales are excluded.

IA.IX. Construction of Additional Characteristics

Age. We measure the age in terms of years for each stock in each month based on the number of previous months that we observe in the CRSP database.

Asset Growth. We estimate the asset growth of each stock from July in year t until June of year $t+1$ from Compustat data as the change in total assets (AT) from the fiscal year ending in $t-1$ to the fiscal year ending in $t-2$:

$$\text{Asset growth} = \frac{AT_{t-1} - AT_{t-2}}{AT_{t-2}}$$

Beta. In each month t we estimate the market beta for stock j as the slope coefficient from the following regression:

$$r_{j,t} - r_{f,t} = \alpha + \beta \cdot (r_{m,t} - r_{f,t}) + u_t$$

where r_j is the return of stock j, r_f the risk-free rate and r_m the market return. We run this regression in each month t using the observations from the previous 60 months. Moreover, we require a minimum of 24 monthly observations for each regression.

Book Equity. We follow Davis et al. (2000) and define book equity (BE) as shareholders' equity plus deferred taxes and investment tax credit (COMPUSTAT item TXDITC) minus book value of preferred stocks. Missing TXDITC observations are set to 0. Particularly, shareholders' equity is shareholders' equity (SEQ) or common equity (CEQ) plus the carrying value of preferred stocks (PSTK). If the aforementioned data is not available shareholders' equity is computed as total assets (AT) minus total liabilities (LT). The book value of preferred stocks reflects either the redemption value (PSTKRV), the liquidating value (PSTKL) or the carrying value of preferred stocks (PSTK). Following this precise order, we replace the book value of preferred stocks in case 1 of the aforementioned data items is not available. Lastly, we follow Davis et al. (2000) and add hand collected book equity data from Moody's manual.

Book-to-Market. We follow Fama and French (1992) and obtain the book-to-market ratio for each stock from July in year t until June in year $t+1$ by scaling the book equity from the fiscal year ending in year $t-1$ with the market equity from CRSP, which we measure at the end of December in year $t-1$. The market-to-book ratio is then the inverse of this ratio.

Dividend Ratio. We calculate the dividend ratio of each stock from July in year t until June in year $t+1$ by the ratio of common dividends (DVC) to income before extraordinary items (IB). Both accounting variables are from the fiscal year ending in $t-1$.

Issuance Ratio. We estimate the issuance ratio of each stock from July in year t until June in year $t+1$ as:

$$\text{Issuance Ratio} = \frac{SSTK_{t-1} - \mathbb{1}_{\Delta PSTKRV>0}(PSTKRV_{t-1} - PSTKRV_{t-2})}{BE_{t-1}}$$

SSTK corresponds to the sale of common and preferred stock, PSTKRV is the value of preferred stocks outstanding, $\mathbb{1}_{\Delta PSTKRV>0}$ is an indicator being 1 if the change in PSTKRV is positive and 0 otherwise. The time subscripts correspond to the fiscal year ending in the denoted year.

Momentum. Momentum corresponds to the compounded holding period return over the preceding 12 months skipping the most recent month, as in Fama and French (1996).

Operating Profitability. We follow Fama and French (2015) and obtain operating profitability for each stock from July in year t until June in year $t+1$ as:

$$\text{Operating Profitability} = \frac{REVT_{t-1} - COGS_{t-1} - XSGA_{t-1} - XINT_{t-1}}{BE_{t-1}}$$

REVT are revenues, COGS costs of goods sold, XSGA selling and administrative expenses, XINT interest expenses, and BE is book equity. All accounting variables are from the fiscal year ending in $t-1$. We replace missing values of COGS, XSGA and XINT with 0 as long as at least 1 of these 3 accounting variables is available.

Real-Estate Ratio. We define the real-estate ratio (RER) similar to Tuzel (2010) with Compustat data. Prior to 1983, it corresponds to the sum of buildings (PPENB) and capital leases (PPENLS) scaled by net property, plant and equipment (PPENT). After the end of 1983, it is measured as the sum of buildings at cost (FATB) and leases at cost (FATL), both divided by gross property, plant and equipment (PPEGT). Subsequently, we winsorize the real estate ratios in each fiscal year at the 1% and 99% percentile. The industry-adjusted real-estate ratio is obtained by subtracting the industry average real-estate ratio from each stock-specific real-estate ratio. We use 2-digit SIC codes to assign stocks to industries. We always require at least 5 observations to calculate the industry average each year. Note that real estate data starts in 1969, limiting the observation period for this specific sorting variable. Data for the real estate ratio starts in January 1970 due to data availability.

Size. We calculate the size of each stock as the natural logarithm of the market capitalization denoted in U.S. Dollar from CRSP.

Total Payout Ratio. We estimate the total payout ratio of each stock from July in year t until June in year $t+1$ by dividing net payouts for the fiscal year ending in $t-1$ with the book equity from the fiscal year ending in $t-1$.