The Corporate Bond Factor Replication Crisis*

Alexander Dickerson† Cesare Robotti‡ Giulio Rossetti‡

April 2026

Abstract

Corporate bond factor research faces a replication crisis. The crisis stems from two sources that inflate reported factor premia: transaction prices whose measurement error enters both sorting signals and return denominators, creating a correlated errors-in-variables bias, and asymmetric ex-post return filtering that embeds future information into factor construction. Applying our framework to a 'factor zoo' of 108 signals across nine thematic clusters, we show that the majority of previously documented factors do not produce statistically significant bond CAPM alphas after correction. We provide an open source framework via Open Bond Asset Pricing, including error-corrected TRACE data, bias-corrected factors, and software for reproducible research.

Keywords: Corporate bond factors; Open source; Factor zoo; Replication crisis; Measurement error; Look-ahead bias; Non-standard errors.

JEL Classification Codes: C12; C13; C58; G11; G12.

---

*An earlier version of this paper circulated under the title "Common pitfalls in the evaluation of corporate bond strategies." We gratefully acknowledge comments and suggestions from Patrick Augustin, Hendrik Bessembinder, Mike Chernov, Nicholas Chen, Mathieu Fournier, Alexey Ivashchenko, Soohun Kim, Jong-Myun Moon (Insight Investment), Yoshio Nozawa, Piotr Orłowski, Yancheng Qiu, Lukas Schmid, Botao Wu, and seminar and conference participants at Aarhus University, Louvain Finance, UNSW Business School, USI Lugano, University of Melbourne, University of Sydney, University of Technology Sydney, University of Toronto, Warwick Business School, and the 2023 ESSEC-CYU-Warwick Econometrics Workshop. The companion website to this paper, openbondassetpricing.com, provides error-corrected daily and monthly TRACE data, bias-corrected factors, and replication code. The data are also available on WRDS Contributed Data (daily) and WRDS Contributed Data (monthly). The companion factor construction software, PyBondLab, is available at pypi.org/project/PyBondLab/.

†School of Banking & Finance, The University of New South Wales; alexander.dickerson1@unsw.edu.au

‡Warwick Business School, The University of Warwick; cesare.robotti@wbs.ac.uk; giulio.rossetti.1@wbs.ac.uk

"... if you torture the data enough, nature will always confess"

(Coase, 1982)

1 Introduction

Corporate bond factor research faces a replication crisis. We construct a 'factor zoo' of 108 corporate bond factors and show that, after correcting for measurement error and look-ahead bias, most are spanned by the bond market factor. A small subset of factors retains significant bond CAPM alphas, primarily those formed on credit-spread-based value signals. The apparent risk-adjusted performance of many published factors is an artifact of two biases whose magnitudes have not been systematically quantified. The first, Latent Implementation Bias (LIB), arises because the same noisy transaction price enters both sorting signals and return denominators, creating a correlated errors-in-variables (CEIV) problem, and because the signal price is a historical transaction, not an actionable quote. The second, Look-Ahead Bias (LAB), arises from asymmetric ex-post return filtering that embeds future information into factor construction. Both systematically inflate measured factor premia and alphas. A third problem compounds the first two because corporate bonds lack a standardized data source, and researchers make idiosyncratic filtering and portfolio-construction choices whose collective variation rivals sampling uncertainty.

Among factors formed on yield, credit spread, value, and reversal signals, only two retain statistically significant bond CAPM alphas after adjustment for measurement error. Both are credit-spread-based value and spread-change factors. Credit-spread-based value factors retain roughly half their premium after adjustment and remain statistically significant. Short-term reversal illustrates the severity of LIB, with the premium dropping from -0.99% to -0.09% per month after correction (t-statistic: -4.46 to -0.51), and bias comprising over 90% of the documented effect. A second bias independently inflates reported premia. Many factors replicate only when researchers apply ex-post return filters, such as winsorizing or trimming at thresholds computed using full sample data, after the portfolio formation month. For the six-month momentum factor, the 0.30% monthly premium is entirely attributable to asymmetric ex-post winsorization. Without winsorization, the premium is zero. For twelve-month momentum, the base premium is negative (-0.13%) and ex-post filtering converts a losing factor into an apparent winner, with the bias concentrated during periods of financial distress. Idiosyncratic volatility factors and downside risk measures show similar patterns, with 57–78% of the measured alpha attributable to asymmetric ex-post return filtering.

Across 432 factor-specification combinations (108 signals × 2 weighting schemes × 2 sorting methods), only 26 (6.0%) bond CAPM alphas survive a Benjamini-Hochberg (BH) false discovery rate (FDR) correction, concentrated among credit-spread-based value factors. Results are consistent when FDR is applied to mean return p-values. Before correction, 119 of 432 specifications produce nominally significant premia ($t(\mu) > 1.96$), but only 22 survive a FDR correction.[1] Our point is not that the bond CAPM is the correct model, but rather that, once LIB and LAB are taken into account, most factors no longer earn economically meaningful premia or deliver superior risk-adjusted performance relative to the bond market factor.

Beyond LIB and LAB, a third source of variation affects measured premia. The same market characteristics that generate noisy prices and extreme returns (infrequent trading, wide bid-ask spreads, and no consolidated data source) also fragment the data infrastructure, since no two research teams process Trade Reporting and Compliance Engine (TRACE) data identically, and different choices, not all equally defensible, generate large dispersion in estimated premia. We distinguish between data uncertainty, which arises from how the admissible investment universe is defined, and methodological uncertainty, which arises from how a given signal is constructed into a factor. Across 648 data-filtering configurations (69,984 factor paths across all 108 signals), the interquartile range of estimated premia (the non-standard error, NSE) averages 0.35% per month, exceeding the average premium of 0.33% per month, with an NSE/SE ratio of 1.15. Holding the data fixed and varying only portfolio construction across 168 economically distinct specifications per signal (18,128 factor paths) produces comparable magnitudes, with the NSE/SE ratio rising to 1.45 and exceeding unity for all nine factor clusters. Researcher degrees of freedom rival sampling uncertainty as a source of variation in measured factor performance, yet fewer than 6% of specifications improve bond market risk-adjusted performance. Fig. 1 maps these three problems to their causes, the biases they produce, and our proposed solutions. The first bias, LIB, arises from two sources: the same noisy transaction price enters both the sorting signal and the return denominator, creating a CEIV problem (Blume and Stambaugh, 1983; Stambaugh, 1988; Duarte, Jones, and Wang, 2024), and the observed transaction price is not executable in over-the-counter (OTC) markets. Both are severe

[1] We report the BH adjustment as an illustrative benchmark to show that the already limited number of statistically significant specifications declines further once one accounts for multiple testing. More conservative procedures that allow for general dependence across strategies would only strengthen this conclusion.

in corporate bonds, where bid-ask spreads are an order of magnitude larger than in equities and trading is sporadic.[2] The second bias, LAB, arises from ex-post return filtering, where winsorization thresholds computed from the full sample embed future information into factor construction, mechanically inflating reported premia. Section 3 develops the formal framework for LIB, and Section 4 for LAB. We propose a protocol for credible corporate bond factor research. To address LIB, we develop gap procedures that ensure the signal and the return denominator use different prices, and that measure returns from the earliest feasible execution price. Breaking this link removes the CEIV component by construction while preserving genuine premia, so that the difference between standard and adjusted approaches identifies the bias without requiring observation of the true price. To address LAB, we implement ex-ante filtering, where winsorization thresholds use only historical information available at portfolio formation. The third element quantifies data and methodological uncertainty, documenting how standard filtering and portfolio-construction choices affect magnitude and inference even after bias corrections. All data and code are open source. The full replication code is available on GitHub. The companion software `PyBondLab` enables reproducible factor construction from any signal. The 'factor zoo', with bond-month data for 108 signals and pre-formed factors, is publicly available at Open Bond Asset Pricing and on WRDS Contributed Data (daily) and WRDS Contributed Data (monthly). Our work relates to several strands of literature. Harvey, Liu, and Zhu (2016) and Hou, Xue, and Zhang (2020) document concerns about the credibility of equity factor research, though Chen and Zimmermann (2022) and Jensen, Kelly, and Pedersen (2023) find high reproduction rates when methodologies are applied consistently. In equity options, Duarte, Jones, Khorram, Mo, and Wang (2025) document replication failures from ex-post sample filters that generate infeasible factors, inflating reported Sharpe ratios by an order of magnitude (over 1000%). We build on earlier work documenting pricing challenges in corporate bonds. Dickerson, Mueller, and Robotti (2023) show that most proposed risk factors, with liquidity as a marginal exception, have no incremental pricing power beyond the bond market factor. Dick-Nielsen, Feldhütter, Pedersen, and Stolborg (2023) attribute replication failures to data errors in TRACE. We find that measurement error and look-ahead bias are the primary drivers. Ghaderi, Plante, Roussanov, and Seo (2024) show that extending the cross-section to nearly a century reveals pricing power for several factors, emphasizing the value of

[2] Dickerson, Robotti, and Nozawa (2025) documents that in a representative year, 70% of bonds trade on 10 business days or fewer, and less than 0.50% of bonds trade every day.

longer samples. Our analysis differs from prior work in two respects. First, we provide a unifying framework for the corporate bond replication crisis, formalizing and quantifying LIB, LAB, and non-standard errors across 108 corporate bond factors. Second, we offer practical solutions through error-corrected open source data and software that enable researchers to construct bias-free factors and reproducible research going forward.

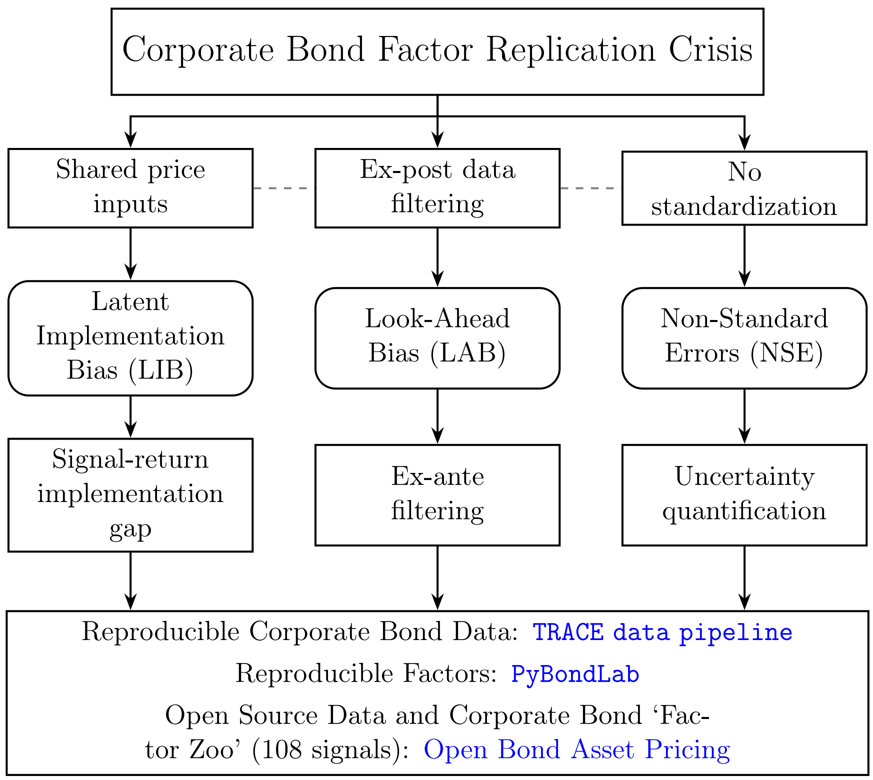

Figure 1: Framework for credible corporate bond factor research.

The figure maps the three sources of the corporate bond replication crisis to their causes, biases, and solutions. Shared price inputs and non-executable transaction prices give rise to Latent Implementation Bias (LIB). Ex-post data filtering introduces Look-Ahead Bias (LAB). The absence of standardized data and methods generates Non-Standard Errors (NSE) through researcher degrees of freedom. The dashed line indicates that infrequent trading and wide bid-ask spreads amplify all three problems. The solutions are a signal-return implementation gap for LIB, ex-ante filtering for LAB, and uncertainty quantification for NSE. The resulting open source protocol (data, software, and a corporate bond 'factor zoo' of 108 signals) is available at Open Bond Asset Pricing.

2 Open source corporate bond asset pricing data

We provide an open source end-to-end pipeline that transforms raw TRACE intraday transactions into a monthly corporate bond asset pricing data set. The pipeline first cleans transactions, then computes daily prices, yields, and credit spreads, and finally aggregates to monthly frequency. It differs from existing approaches in four ways. First, we include Rule 144A corporate bonds, which are excluded from most existing studies. Second, we track bonds through and after default rather than dropping them at the event boundary, avoiding censoring of the return distribution (Baumann, Kakhbod, Livdan, Nazemi, and Schürhoff, 2025; Baumann and Nazemi, 2025). Third, every filter decision is documented through companion data reports that are generated automatically when the pipeline is executed, plotting the complete price time series for each affected CUSIP so that researchers can inspect every correction and exclusion.[3] Fourth, all filter thresholds are configurable, and researchers can modify any parameter and re-run the pipeline from raw transactions up to the monthly data output. The sample covers August 2002 to December 2024.

2.1 From raw transactions to daily prices

The pipeline begins with raw TRACE enhanced and 144A transaction records, merged with Mergent Fixed Income Securities Database (FISD) for bond characteristics. We retain USD-denominated, fixed-rate, non-convertible, non-asset-backed bonds with $1,000 par value and original maturity of at least one year. Standard filters from Dick-Nielsen (2014) remove cancellations, corrections, reversals, and agency-side duplicates. Beyond these standard filters, we develop two transaction-level corrections for recording errors that survive the Dick-Nielsen procedures. The *decimal shift corrector* identifies prices recorded with incorrect decimal placement (for example, 10.5 entered as 105.0) and applies a multiplicative correction when a set of acceptance conditions is satisfied. The *bounce-back filter* detects transient price spikes that deviate from a backward-looking anchor and revert within a short window. Both filters are designed to correct errors where possible and flag anomalies otherwise, minimizing data loss. Appendix A provides complete filter specifications and parameter values.

After filtering, transactions are aggregated to daily frequency using volume-weighted aver-

[3]Detailed daily data error reports for TRACE enhanced are available on Open Bond Asset Pricing.

age prices. We then use `QuantLib` to compute duration, convexity, yield to maturity, and credit spreads, and merge credit ratings, equity identifiers, and Fama-French industry classifications. A final set of four daily-level filters targets anomalous prices in distressed bonds, addressing isolated ultra-low prices, transient spikes, stale plateaus, and intraday inconsistencies. Appendix A documents each filter. Detailed data reports with descriptive statistics, industry breakdowns, and additional information are available for download. Appendix IA.1 reports data availability and descriptive statistics for the daily data (Tables IA.I–IA.II). The code automatically generates companion data reports that plot the price time series of every bond CUSIP affected by a filter, providing a transparent record of each filtering decision.

2.2 Monthly returns and default handling

For each bond-month, we compute month-end returns (for comparability with prior literature), month-begin returns (computed from the first (last) available price within the first five (last five) business days of month $t+1$, measuring implementable performance after signal observation),[4] and duration-adjusted excess returns that remove interest rate exposure. The monthly price is the last available daily price within the last five business days of the month (New York Stock Exchange (NYSE) calendar). A valid return requires such a price in both months t and $t+1$ (bonds without a trade in either window are excluded for that month). Excess returns subtract the Fama-French one-month T-bill rate. The bond market factor (`MKTB`) is the value-weighted excess return on all corporate bonds in the sample. We use a single-factor bond CAPM (CAPMB) comprising `MKTB` to compute alphas throughout the paper.[5] Appendix A.4 provides the formal definitions, and Section 3 discusses the distinction between month-end and month-begin measurement.

Rule 144A bonds enter the sample from June 2014 onward and are processed identically to TRACE enhanced. These securities, issued under SEC Rule 144A to qualified institutional buyers, account for over 20% of new corporate bond issuance as of 2025. Excluding them omits a large and growing segment of the investable universe.[6]

[4] The median gap between the month-end price and the first available price the following month is one business day.

[5] In unreported results, we use multi-factor models that include default, credit, liquidity, and equity factors and observe lower alphas on average.

[6] Rule 144A transaction-level data became available on Wharton Research Data Services (WRDS) from June 2014 onward. The June 2014 entry point is therefore a data constraint, not a sample design choice. In unreported robustness checks, excluding 144A bonds from the full sample does not materially change our main

Dropping defaulted bonds from the sample censors extreme outcomes and can bias measured return distributions. We track bonds through and after default. In the default month, we compute the standard return, a default-event return excluding the final coupon, and a trading-in-default return on a flat basis without accrued interest. Appendix A.5 gives the complete return formulas for each case. Tables IA.III–IA.IV in Appendix IA.1 report monthly data coverage and descriptive statistics, while Tables IA.V–IA.VII document the prevalence and time concentration of extreme returns.

For price-based signals, we use two complementary gap procedures to break the mechanical link between signal measurement error and the return denominator. The first gap procedure observes the sorting variable at least one business day before the month-end price used for return computation, while the second retains the standard month-end signal but computes returns from prices observed in the first (last) five business days of the following month. Section 3 provides the formal treatment.

2.3 The corporate bond 'factor zoo'

We construct 108 signals spanning nine clusters (Spreads, Yields, Size; Value; Momentum & Reversal; Illiquidity; Volatility & Risk; Market Risk; Credit & Default Betas; Volatility & Liquidity Betas; and Macro & Other Betas). Table IA.VIII provides definitions and original citations; Internet Appendix IA.2 documents construction details.

The sample spans August 2002 to December 2024 (269 months), with 52,656 unique bonds and an average of 6,790 bond–month observations per month. Variable names are consistent across the database and code, and match Table IA.VIII. For example, `cs` denotes credit spread, `str` denotes short-term reversal, and `mom6_1` denotes six-month momentum. To our knowledge, this is the first open source corporate bond 'factor zoo' containing the majority of signals proposed in the literature.

The data pipeline addresses one dimension of the replication crisis, the absence of a standardized, transparent data source for corporate bond research. The remaining sections investigate three sources of bias that affect factor premia independently of the data source. Section 3 examines measurement error in price-based signals. Section 4 addresses look-ahead bias from ex-post return filtering. Section 5 quantifies the sensitivity of estimated premia to methodolog-

results.

ical variation.

3 Measurement error and latent implementation bias

The month-end bid–ask averaged TRACE price, which underlies most corporate bond factor research, is a synthetic construct. It is neither a precise measure of value nor an executable quote.[7]

Let $P_{i,t}$ and $s_{i,t}$ denote the true price and true sorting signal, and let $\hat{P}_{i,t}$ and $\hat{s}_{i,t}$ denote their observed counterparts. Following Blume and Stambaugh (1983), observed prices satisfy $\hat{P}_{i,t} = (1 + \delta_{i,t})P_{i,t}$, where $\delta_{i,t}$ is mean-zero price noise, and observed signals satisfy $\hat{s}_{i,t} = s_{i,t} + \eta_{i,t}$, where $\eta_{i,t}$ denotes signal measurement error. In TRACE, the month-end price is proxied by the volume-weighted average of transactions on the last trading day within the final five business days of the month, which averages across customer buy and sell transactions.[8] When the same price disturbance $\delta_{i,t}$ enters both $\eta_{i,t}$ and the return measurement error $\epsilon_{i,t} \approx \delta_{i,t+1} - \delta_{i,t}$ (Definition B.3 in Appendix B), the resulting correlation between signal and return measurement errors does not cancel in long-short factors. The component $\delta_{i,t+1}$ is realized after portfolio formation and, under serial independence, has zero conditional expectation given portfolio assignment. Thus, the bias-relevant component of the return measurement error is approximately $-\delta_{i,t}$: the price error at time t mechanically distorts the return denominator, understating the return when $\delta_{i,t}$ is positive and overstating it when negative. This error is correlated with the signal error because both load on the same price disturbance.

For all price-based signals in our 'factor zoo', higher price maps to a lower sorting signal (Proposition B.1 in Appendix B): a bond with $\delta_{i,t} > 0$ is sorted toward the short leg and simultaneously has its return understated; a bond with $\delta_{i,t} < 0$ is sorted toward the long leg with its return overstated.[9] The long leg holds bonds with systematically positive return

[7] Quote-based data from sources such as Bank of America Merrill Lynch or Bloomberg Barclays are not necessarily a panacea: dealer quotes can be stale. However, existing evidence suggests that these quotes are broadly reliable; see Choi (2013) and Andreani, Palhares, and Richardson (2024).

[8] The measurement error δ is a reduced-form representation encompassing bid-ask bounce, dealer inventory effects, informed trading, and stale pricing. Our framework does not require distinguishing among these sources; any discrepancy that enters both the sorting signal and the return denominator generates CEIV bias.

[9] Consider two bonds with true price \$100, true return 0%, and noise $\delta_A = +0.50\%$, $\delta_B = -0.50\%$. Bond A is sorted toward the short leg and earns a measured return of $100/100.50 - 1 = -0.50\%$. Bond B is sorted toward the long leg and earns $100/99.50 - 1 = +0.50\%$. The measured long-short return is 1.00%; the true long-short return is zero. Neither misranking nor return distortion alone produces this bias: with correct returns, both legs earn 0% (misranking attenuates); without sorting, the return errors cancel across bonds (by $\mathbb{E}[\delta_{i,t}] = 0$).

errors, and the short leg holds bonds with systematically negative return errors, spuriously widening the measured premium. Misranking alone does not create this bias: if the signal were noisy but returns measured without error, sorting on the noisy signal would attenuate the true premium toward zero (Appendix B, Remark B.3). The bias arises because the same $\delta_{i,t}$ that misranks bonds also contaminates their measured returns; misranking and return distortion are mechanically correlated, and this correlated error structure creates a spurious additive premium.

This is the CEIV mechanism, named by Duarte, Jones, and Wang (2024), building on the measurement-error literature of Blume and Stambaugh (1983), Fama (1984), and Stambaugh (1988). The CEIV mechanism builds on Blume and Stambaugh (1983), who showed that bid-ask bounce biases equally-weighted portfolio returns, and Stambaugh (1988), who addressed an analogous correlated measurement error in Treasury bill forward premium regressions by using prices from different maturities. Our gap procedure applies the same logic using temporal rather than cross-maturity independence. Duarte, Jones, and Wang (2024) applied the mechanism to option pricing and showed that their adjustment reverses the sign of the estimated volatility risk premium in individual stock options from positive to negative, the theoretically correct sign. In equities, Jegadeesh (1990) and Conrad, Gultekin, and Kaul (1997) recognized the same contamination in short-term reversal profits and addressed it by excluding the shared boundary price. In corporate bonds, Lair and Blonk (2024) show that month-end reversal profits vanish when implemented at month-begin, consistent with an implementation shortfall, though pricing distortions relative to independent index valuations retain predictive power even after controlling for bid-ask bounce.

CEIV bias is of first-order importance in the price noise, unlike the classic Blume and Stambaugh (1983) bias (σ_δ^2), which is of second-order importance and negligible in long-short factors. It also differs from standard errors-in-variables (EIV) attenuation (e.g., Fama and MacBeth, 1973), which biases regression coefficients toward zero. Appendix B derives the closed-form expression under distributional assumptions.

There is also an *implementation problem*. The month-end TRACE price records past transactions, not standing offers. A transaction price observed in a backtest is not necessarily available to trade at, because corporate bonds trade OTC and trading is not continuous. Standard backtests that use the signal price as the return denominator measure returns that no investor could actually capture. This gap between the observed signal and the earliest feasible trade

The bias arises because the same δ that determines portfolio assignment also determines the return error.

creates what we call LIB.[10] LIB encompasses both the CEIV bias from measurement error and the non-executability of observed prices. Both widen the gap between reported and achievable returns.

Breaking the link. Both problems share a common source: using the same month-end price for signal computation and return measurement.[11] The measurement problem arises because the same noise $\delta_{i,t}$ enters both; the implementation problem arises because the signal price is not executable. A time gap between signal and return can partially address both: the price used to compute the sorting signal must come from a different date than the price in the return denominator. Duarte, Jones, and Wang (2024) validate a conceptually analogous lag adjustment via simulation in their option-pricing context, confirming that the lag approach reliably removes the CEIV component when measurement errors are serially uncorrelated.[12] A *signal gap* computes the signal from an earlier price $\hat{P}_{i,t-\Delta}$, which contains noise $\delta_{i,t-\Delta}$. The return denominator still uses $\hat{P}_{i,t}$, containing $\delta_{i,t}$. Under serial independence (Assumption B.1 in Appendix B), $\delta_{i,t-\Delta}$ and $\delta_{i,t}$ are independent, and the CEIV-induced mechanical correlation is eliminated. A *return gap* measures the return from month-begin, where the investor observes the signal at month-end t and buys into the position at the first available price in month $t+1$. The return denominator is now $\hat{P}^{\text{bgn}}_{i,t+1}$, which contains $\delta^{\text{bgn}}_{i,t+1}$, independent of the $\delta_{i,t}$ that drove sorting under Assumption B.1.

Quantifying the bias. To proxy for the magnitude of LIB, we compute two types of monthly returns. The standard *month-end return* measures performance from the end of month t to the end of month $t+1$ (see Panel A of Fig. 2):

$$r^{\text{End}}_{i,t+1} = \frac{P^{\text{end}}_{i,t+1} + AI^{\text{end}}_{i,t+1} + C_{i,t+1}}{P^{\text{end}}_{i,t} + AI^{\text{end}}_{i,t}} - 1, \tag{1}$$

[10] We use the term "latent" because the bias is invisible to researchers who follow the standard approach: it does not appear as an explicit adjustment or error term but is embedded in the correlation structure between signal noise and return noise. The bias becomes apparent only when the signal-return price overlap is broken.

[11] Chen and Choi (2024) impose a one-month gap between signal observation and return computation, and Bartram, Grinblatt, and Nozawa (2025) vary the return denominator price to compute within-month returns; both recognize the nonsynchronous trading problem but neither provides a formal CEIV framework or quantifies the bias systematically across signals.

[12] This approach assumes measurement errors are serially uncorrelated, i.e., $\text{cov}(\delta_{i,t-\Delta}, \delta_{i,t}) = 0$ for $\Delta > 0$. If errors exhibit positive autocorrelation, the gap reduces but does not eliminate the mechanical correlation between portfolio weights $\omega_{i,t}$ (formed from price-based signals) and measured returns $r_{i,t+1}$. For illiquid bonds with stale or persistent pricing, some residual correlation may remain.

where P^{end} is the clean price observed within the last 5 business days of the month, AI is accrued interest, and C is any coupon payment. The *month-begin return* measures performance from the beginning to the end of month $t+1$ (see Panel C of Fig. 2):

$$r^{\text{Bgn}}_{i,t+1} = \frac{P^{\text{end}}_{i,t+1} + AI^{\text{end}}_{i,t+1} + C_{i,t+1}}{P^{\text{bgn}}_{i,t+1} + AI^{\text{bgn}}_{i,t+1}} - 1, \tag{2}$$

where P^{bgn} is observed within the first 5 business days of the month. The month-begin return captures what a trader could actually earn after observing a signal at month-end: the earliest feasible execution occurs at month-begin, not at the signal observation price. The month-end return approximately decomposes into the implementable return plus the latent implementation bias:

$$r^{\text{End}}_{i,t+1} \approx \underbrace{\frac{P^{\text{bgn}}_{i,t+1}}{P^{\text{end}}_{i,t}} - 1}_{\text{LIB}_{i,t+1}} + r^{\text{Bgn}}_{i,t+1}. \tag{3}$$

The LIB term measures the price change between signal observation at month-end t and the earliest feasible trade at month-begin $t+1$. Because this price change occurs before any investor could act on the signal, the LIB component of the month-end return is not achievable in practice. It captures both the statistical bias from shared price inputs and the economic wedge between what a backtest reports and what a trader could earn. Lair and Blonk (2024) propose a similar decomposition for the short-term reversal factor, which they term the implementation shortfall. Panel D of Fig. 2 illustrates this decomposition.

We compare factor performance under three approaches. *Approach 1 (Unadjusted)* uses the month-end price $P^{\text{end}}_{i,t}$ for both signal computation and return measurement, which is the standard practice in the literature. Portfolio weights $\omega_{i,t}$ are computed from signals observed at $P^{\text{end}}_{i,t}$, and returns are the month-end returns $r^{\text{End}}_{i,t+1}$. *Approach 2 (Adjusted Signal)* breaks the correlation by computing signals from prices observed at least 1 business day before month-end (up to 10 business days).[13] The signal noise now depends on $\delta_{i,t-\Delta}$, which is independent of $\delta_{i,t}$ in the return denominator. Portfolio weights $\omega^a_{i,t}$ are computed from these gapped signals; returns remain month-end returns. Panel B of Fig. 2 illustrates the signal gap. *Approach 3 (Adjusted Return)* breaks the correlation by measuring the return from month-begin: weights $\omega_{i,t}$ are computed from month-end signals, but returns are $r^{\text{Bgn}}_{i,t+1}$, which uses $P^{\text{bgn}}_{i,t+1}$ in the denominator.

[13] The median signal gap in the data is 1 business day. Capping the maximum gap at a lower number makes an immaterial difference to the results.

The return noise now depends on $\delta^{\text{bgn}}_{i,t+1}$, independent of the $\delta_{i,t}$ that drove sorting. This approach also measures the implementable return, that is, the return a trader could actually capture after observing the signal.

Approaches 2 and 3 both break the shared-price link between signal and return, whereas Approach 1 retains it. The difference between the standard and adjusted approaches identifies the bias empirically, without requiring observation of the true price. We quantify the bias as

$$\text{Bias}_{1-2} = (\omega_t - \omega^a_t) \times r^{\text{End}}_{t+1}, \tag{4}$$

$$\text{Bias}_{1-3} = \omega_t \times (r^{\text{End}}_{t+1} - r^{\text{Bgn}}_{t+1}). \tag{5}$$

Bias_{1-2} isolates the effect of using standard versus gapped signals while holding returns constant. Bias_{1-3} isolates the effect of using infeasible versus implementable returns while holding weights constant. Positive bias indicates that the standard approach overstates factor performance.

Table 1 presents bias estimates for seven price-based factors constructed from commonly used signals: yield to maturity (`ytm`), credit spread (`cs`), bond book-to-market (`bbtm`), 6-month change in log spreads (`dcs6`), IPR value (`val_ipr`), HZ value (`val_hz`), and short-term reversal (`str`). In Panel A, we sort bonds into deciles each month and form value-weighted portfolios (using bond market capitalization) that are long the top decile and short the bottom decile. In Panel B, we construct within-firm factors following Dick-Nielsen, Feldhütter, Pedersen, and Stolborg (2023): for each firm with at least two bonds, we form long-short portfolios based on within-firm signal rankings, then aggregate across firms using market-value weights.

Short-term reversal exhibits the largest bias. In Panel A (single-sort), the unadjusted factor earns −0.99% per month ($t = -4.46$), but the signal-adjusted factor earns only −0.09% ($t = -0.51$). The implied bias is 90 basis points, or 91% of the measured premium. The unadjusted CAPMB alpha of −0.77% becomes 0.12% after adjustment, with a bias of 89 basis points ($t = -8.20$). The return-adjusted approach yields comparable results: the premium falls to −0.17% with a bias of 82 basis points. The value factors show biases of 36–45 basis points. For `val_ipr`, the unadjusted premium of 0.96% per month falls to 0.51% after signal adjustment. The bias of 45 basis points ($t = 6.93$) accounts for 47% of the measured premium. For `val_hz`, the premium falls from 0.81% to 0.45%, a bias of 36 basis points. Spread-based factors (`ytm`, `cs`, `bbtm`) display more modest biases of 15–18 basis points per month, though all are highly statistically distinguishable from zero ($t > 5$). The `dcs6` factor, which sorts on spread changes from Kelly, Palhares, and Pruitt (2023), has a negative premium (−1.02% unadjusted, −0.50%

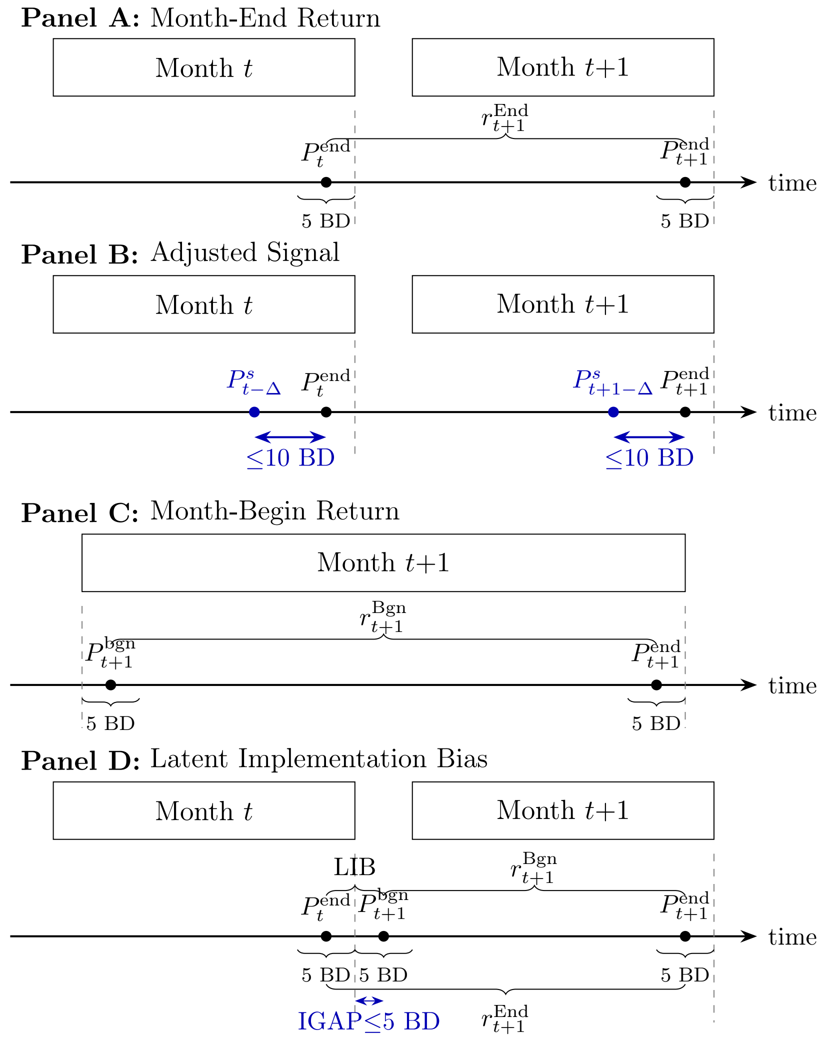

Figure 2: Return measurement windows and bias decomposition.

The figure illustrates the four return measurement windows used in the paper. Panel A plots the standard month-end return, measured from the last 5 business days (BD) of month t to the last 5 business days of month $t+1$. Panel B depicts the adjusted signal timing, where the investor observes signals at $P^s_{t-\Delta}$ and $P^s_{t+1-\Delta}$, between 1 and 10 business days before the month-end prices. Panel C plots the month-begin return, measured within month $t+1$ from the first 5 to the last 5 business days. Panel D decomposes the month-end return into LIB (from P^{end}_t to P^{bgn}_{t+1}) plus the month-begin return (from P^{bgn}_{t+1} to P^{end}_{t+1}). The implementation gap (IGAP) is the minimum one-day gap between the month-end and month-begin prices. All business days follow the NYSE trading calendar.

adjusted) and correspondingly negative bias (−52 basis points).

Panel B examines within-firm factors, which control for issuer-level heterogeneity by forming long-short portfolios within each firm. Dick-Nielsen, Feldhütter, Pedersen, and Stolborg (2023) report that within-firm factors based on price-based signals, specifically ytm, cs, bbtm, val_ipr, and str, exhibit the largest premia in their framework. After adjustment, only two factors retain statistically meaningful alphas: val_ipr ($\alpha = 0.18\%$, $t = 4.25$ with adjusted signal; $\alpha = 0.20\%$, $t = 4.30$ with adjusted return) and dcs6 ($\alpha = -0.22\%$, $t = -4.98$). The remaining factors (ytm, cs, bbtm, val_hz, and str) all have adjusted alphas with $|t| < 2$. Within-firm premia and alphas are lower on average than their single-sort counterparts, and the gap widens after adjustment.

Cumulative returns in Fig. 3 contrast standard and adjusted approaches. For short-term reversal (Panels A–B), \$1 invested in the unadjusted single-sort factor grows to \$12 by the end of 2024, but only to \$1.1 (adjusted signal) and \$1.4 (adjusted return); the within-firm factor falls from \$5.6 unadjusted to \$1.3 under both adjustments. Credit spread (Panels C–D) retains more of its single-sort premium (\$9.3 adjusted signal, \$6.5 adjusted return, vs. \$14 unadjusted), but the within-firm factor drops from \$14 unadjusted to \$1.8 (adjusted signal) and \$2.0 (adjusted return).

Panels A–B of Fig. 4 report average monthly biases with 95% confidence intervals; all biases are statistically different from zero. The largest biases appear for str and dcs6, followed by the value factors; spread-based factors exhibit smaller but still substantial biases of 15–30 basis points. Panels C–D decompose unadjusted returns into the implementable component and LIB. For reversal, LIB constitutes over 80% of the unadjusted premium; value and spread-change factors show LIB shares of 45–55%; yield and credit spread factors have LIB shares of 20–50%. Within-firm factors exhibit larger LIB shares than their single-sort counterparts.

To validate the LIB decomposition in Eq. (3), Table 2 compares the month-end return, month-begin return, their difference, and the directly estimated LIB. If the decomposition holds, the difference $r^{\text{End}}_{i,t+1} - r^{\text{Bgn}}_{i,t+1}$ should approximately equal the directly estimated LIB term $\hat{\eta}_{\text{LIB}}$, and the two series should be nearly perfectly correlated. The decomposition holds tightly for single-sort factors: all seven factors show correlations of 0.99 between $(r^{\text{End}} - r^{\text{Bgn}})$ and LIB, with residuals of at most 0.02 percentage points. Within-firm sorts show slightly lower correlations (0.93–0.98) and larger residuals (up to 0.09 percentage points), but the decomposition remains economically tight.

One concern is whether the gap procedure drives differences in premia through an altered holding period or economic events during the gap window, rather than by removing LIB. Table B.1 in Appendix B.5 provides a discriminating test. We compute the month-end minus

Table 1: Latent implementation bias in price-based factors.

The table reports premia and CAPMB alphas for seven price-based factors under standard and bias-adjusted approaches. Portfolios are value-weighted (using bond market capitalization) with excess returns over the one-month T-bill rate. Panel A sorts bonds into deciles each month, going long P10 and short P1. Panel B uses within-firm sorts: for each firm with at least two bonds, we rank bonds by signal, form a long-short portfolio, and aggregate across firms using market-value weights. μ is the monthly average return (%), α is the CAPMB alpha (%). $\Delta\mu$ and $\Delta\alpha$ measure the bias as the difference in premia and alphas between standard and adjusted approaches. Bias (1)−(2) compares standard versus signal-adjusted. Bias (1)−(3) compares standard versus return-adjusted. t-statistics (Newey-West, lags = $\lfloor T^{0.25} \rfloor$) in parentheses. Sample: 2002-09 to 2024-12, T=268.

	Unadjusted (1)		Adj. Signal (2)		Adj. Return (3)		Bias (1)−(2)		Bias (1)−(3)	
Factor	μ	α	μ	α	μ	α	$\Delta\mu$	$\Delta\alpha$	$\Delta\mu$	$\Delta\alpha$
Panel A: Single-Sort										
ytm	1.20	0.69	1.05	0.54	0.92	0.44	0.15	0.15	0.28	0.25
	(2.97)	(2.50)	(2.63)	(1.96)	(2.64)	(1.80)	(6.54)	(6.06)	(3.90)	(4.33)
cs	1.13	0.66	0.96	0.50	0.82	0.38	0.17	0.17	0.31	0.28
	(2.89)	(2.39)	(2.52)	(1.82)	(2.45)	(1.55)	(6.23)	(6.17)	(4.14)	(4.54)
bbtm	0.93	0.54	0.75	0.35	0.66	0.29	0.18	0.18	0.27	0.25
	(2.19)	(1.69)	(1.80)	(1.12)	(1.80)	(1.04)	(6.20)	(5.24)	(3.39)	(3.56)
dcs6	−1.02	−0.79	−0.50	−0.29	−0.46	−0.25	−0.52	−0.51	−0.56	−0.54
	(−4.58)	(−4.55)	(−2.52)	(−1.79)	(−2.50)	(−1.71)	(−8.21)	(−8.12)	(−7.83)	(−7.73)
val_ipr	0.96	0.76	0.51	0.31	0.51	0.33	0.45	0.45	0.45	0.43
	(4.82)	(5.24)	(2.68)	(2.12)	(3.05)	(2.61)	(6.93)	(6.42)	(7.12)	(7.52)
val_hz	0.81	0.47	0.45	0.12	0.43	0.12	0.36	0.34	0.38	0.35
	(3.80)	(3.71)	(2.29)	(1.13)	(2.32)	(1.08)	(6.60)	(6.43)	(6.21)	(6.01)
str	−0.99	−0.77	−0.09	0.12	−0.17	0.03	−0.90	−0.89	−0.82	−0.80
	(−4.46)	(−3.58)	(−0.51)	(0.68)	(−0.98)	(0.18)	(−8.70)	(−8.20)	(−7.34)	(−7.30)
Panel B: Within-Firm Sort										
ytm	0.56	0.31	0.30	0.06	0.26	0.04	0.26	0.25	0.30	0.27
	(4.42)	(4.08)	(2.85)	(1.44)	(2.68)	(0.94)	(5.42)	(5.54)	(5.60)	(5.58)
cs	0.63	0.47	0.26	0.11	0.24	0.10	0.38	0.36	0.40	0.37
	(5.57)	(4.43)	(3.10)	(1.72)	(3.11)	(1.62)	(6.34)	(6.36)	(6.42)	(6.11)
bbtm	0.38	0.28	0.19	0.10	0.17	0.09	0.19	0.18	0.21	0.19
	(4.02)	(4.11)	(2.50)	(1.86)	(2.32)	(1.64)	(6.05)	(6.58)	(5.64)	(5.56)
dcs6	−0.70	−0.65	−0.26	−0.22	−0.25	−0.22	−0.44	−0.43	−0.45	−0.43
	(−7.70)	(−8.08)	(−4.63)	(−4.98)	(−4.59)	(−4.92)	(−7.56)	(−7.44)	(−8.21)	(−8.25)
val_ipr	0.62	0.57	0.22	0.18	0.22	0.20	0.40	0.39	0.39	0.37
	(6.02)	(7.27)	(3.50)	(4.25)	(3.58)	(4.30)	(7.41)	(7.47)	(7.38)	(7.85)
val_hz	0.50	0.35	0.21	0.07	0.19	0.06	0.29	0.28	0.31	0.29
	(4.80)	(3.68)	(2.67)	(1.14)	(2.68)	(1.02)	(5.78)	(5.91)	(5.68)	(5.49)
str	−0.65	−0.61	−0.10	−0.07	−0.11	−0.10	−0.55	−0.54	−0.54	−0.52
	(−7.68)	(−7.12)	(−1.89)	(−1.36)	(−1.91)	(−1.69)	(−8.37)	(−8.51)	(−8.48)	(−8.77)

month-begin return difference for all 108 factors using unadjusted portfolio weights, comparing Approach 1 (month-end returns) with Approach 3 (month-begin returns). Among the 78 non-

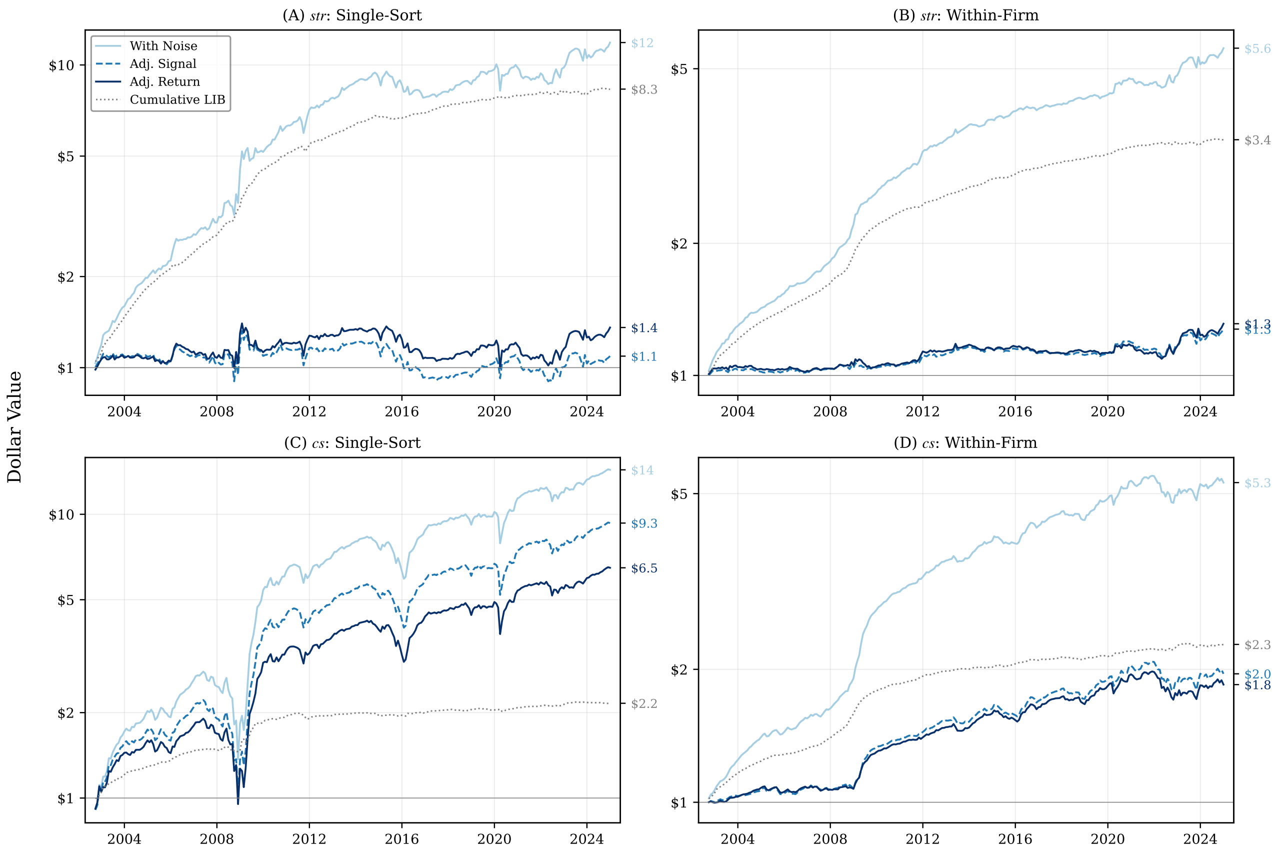

Figure 3: Cumulative factor returns under standard and adjusted approaches.

The figure shows the growth of $1 invested in long-short factors under standard and adjusted approaches. Panels A–B plot short-term reversal (`str`). Panels C–D plot credit spread (`cs`). The left column reports single-sort factors, while the right column within-firm sorts. The dotted gray line tracks cumulative Latent Implementation Bias (LIB). The y-axis uses log scale. The **`str`** factor is sign-corrected to have a positive premium. Value-weighted portfolios. Sample: 2002-09 to 2024-12, T=268.

price-based factors, where Corollary B.1 predicts no CEIV bias, the average absolute return difference is 0.03–0.10% per month with average $|t|$-statistics near unity. Only 10–18% of factors show a statistically significant gap, depending on the sort. Among the 30 price-based factors, 43–67% show a significant gap depending on the sort, with average absolute differences of 0.13–0.30% per month. For illiquidity factors, which are constructed from within-month averages of daily observations, the shared-price link between signal noise and return noise is dampened. Table IA.XIV in Internet Appendix IA.4 confirms that LIB estimates for five illiquidity characteristics are economically small and mostly statistically indistinguishable from zero. CAPMB alphas for all five illiquidity factors range from 0.10% to 0.20% per month with $|t| < 1.65$, and within-firm alphas are indistinguishable from zero: the illiquidity factor premium is at best marginal.

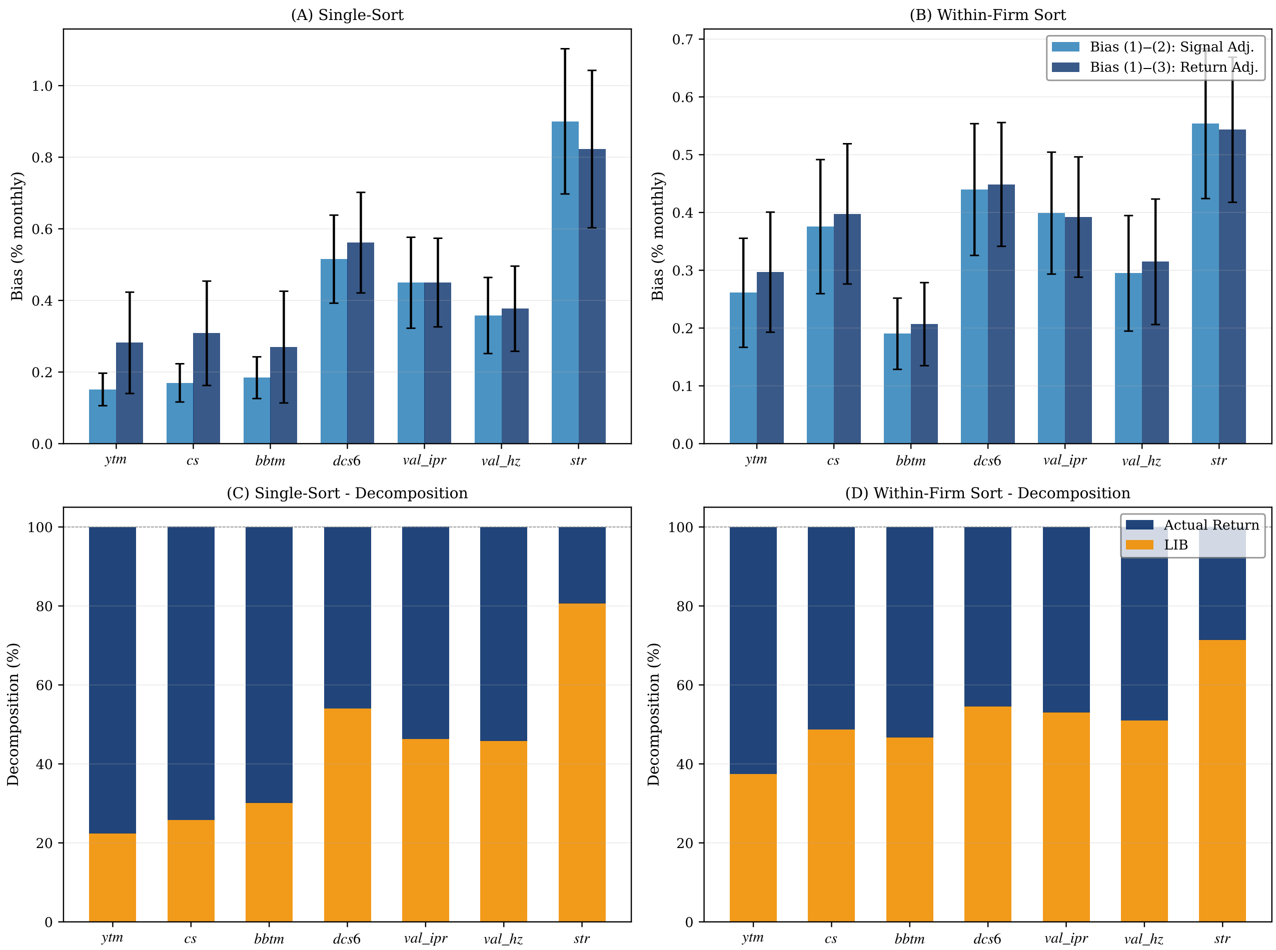

Figure 4: Latent implementation bias magnitude and return decomposition for price-based factors.

The figure displays the magnitude of Latent Implementation Bias (LIB) and its share of unadjusted returns for seven price-based factors. Panels A–B plot the magnitude of the bias for single-sort and within-firm factors. Bias (1)−(2) measures the difference between standard and signal-adjusted approaches. Bias (1)−(3) measures the difference between standard and return-adjusted approaches. Bias is sign-corrected to be positive. Error bars show 95% confidence intervals using Newey-West standard errors with lags = $\lfloor T^{0.25} \rfloor$. Panels C–D decompose unadjusted returns into the LIB portion (orange, percentage of the unadjusted mean attributable to LIB) and the implementable component (dark blue). Stacked bars sum to 100%. Sample: 2002-09 to 2024-12, T=268.

Measurement error in transaction prices inflates price-based factor premia by 15 to 90 basis points per month. Non-price-based factors and within-month illiquidity averages are largely unaffected. Infrequent trading and wide bid-ask spreads also produce volatile bond-level returns: monthly returns below −50% or above 50% are not uncommon, particularly during financial distress. Such returns, though economically important, invite ad hoc filtering, such as winsorizing or trimming at sample-wide thresholds, which introduces a second form of bias that we examine next.

Table 2: Latent implementation bias validation for price-based factors.

The table tests the decomposition $r^{\text{End}} \approx \eta_{\text{LIB}} + r^{\text{Bgn}}$ for seven price-based factors. Panel A reports single-sort factors. Panel B reports within-firm factors. μ_{End} is the month-end average return (%), μ_{Bgn} is the month-begin average return (%), $\Delta\mu = \mu_{\text{End}} - \mu_{\text{Bgn}}$, μ_{LIB} is the average LIB (%). $\hat{\rho}$ is the time series correlation between $(r^{\text{End}} - r^{\text{Bgn}})$ and LIB. Residual $= \Delta\mu - \mu_{\text{LIB}}$. t-statistics use Newey-West (lags $= \lfloor T^{0.25} \rfloor$). Sample: 2002-09 to 2024-12, T=268.

Factor	μ_{End}	μ_{Bgn}	$\Delta\mu$	$t_{\Delta\mu}$	μ_{LIB}	$t_{\mu_{\text{LIB}}}$	$\hat{\rho}$	Residual
Panel A: Single-Sort								
ytm	1.20	0.92	0.28	(3.90)	0.27	(3.67)	0.99	0.01
cs	1.13	0.82	0.31	(4.14)	0.29	(3.84)	0.99	0.02
bbtm	0.93	0.66	0.27	(3.39)	0.28	(3.47)	0.99	−0.01
dcs6	−1.02	−0.46	−0.56	(−7.83)	−0.55	(−7.63)	0.99	−0.01
val_ipr	0.96	0.51	0.45	(7.12)	0.44	(6.81)	0.99	0.01
val_hz	0.81	0.43	0.38	(6.21)	0.37	(5.87)	0.99	0.01
str	−0.99	−0.17	−0.82	(−7.34)	−0.80	(−7.18)	0.99	−0.02
Panel B: Within-Firm Sort								
ytm	0.56	0.26	0.30	(5.60)	0.21	(5.31)	0.93	0.09
cs	0.63	0.24	0.40	(6.42)	0.31	(6.21)	0.96	0.09
bbtm	0.38	0.17	0.21	(5.64)	0.18	(5.35)	0.96	0.03
dcs6	−0.70	−0.25	−0.45	(−8.21)	−0.38	(−7.95)	0.98	−0.06
val_ipr	0.62	0.22	0.39	(7.38)	0.33	(7.29)	0.95	0.06
val_hz	0.50	0.19	0.31	(5.68)	0.26	(5.76)	0.95	0.06
str	−0.65	−0.11	−0.54	(−8.48)	−0.46	(−8.19)	0.94	−0.08

4 Look-ahead bias

This section turns to ex-post return filtering, a second source of bias that affects all factors, including non-price-based ones. Historical cross-sectional backtests require some look ahead in sample definition. To compute a return for month $t+1$, a bond must trade near both month-end t and month-end $t+1$, so that bonds without valid future prices must be excluded at portfolio formation. This *required* look ahead is inherent to historical backtests and affects all researchers equally. The same issue arises in the Center for Research in Security Prices (CRSP) database, though less often because stocks trade daily. In TRACE, where many bonds trade infrequently, it is unavoidable. *Avoidable* look-ahead bias arises when return filtering thresholds are computed from the full sample rather than from data available at portfolio formation. We use "ex-ante" and "ex-post" to describe the information set used by the filtering rule rather than the return. Although all returns are realized, the distinction revolves around whether the threshold applied to those returns uses only information available at portfolio formation (ex-ante) or the sample including future months (ex-post). Ex-post filtering is widespread in corporate bond research despite being rarely stated explicitly. For many factors, reported factor premia and alphas are statistically significant only when ex-post filters are applied, commonly rationalized by outlier removal. At month t, the threshold $\tau^{\text{ex-post}} = \text{Percentile}_q(\{r_{i,s}\}_{s=1}^{T})$ incorporates information from months $t+1, \ldots, T$ that has not yet been realized. Filtering with such thresholds creates an infeasible trading strategy, one that no investor could have implemented in real time. We define the *Look-Ahead Bias* (LAB) as the difference between the infeasible (reported) factor return and the feasible (achievable) return. Duarte, Jones, Khorram, Mo, and Wang (2025) document the same mechanism in equity options, where ex-post sample filters selectively remove the worst-performing observations and inflate reported Sharpe ratios from 0.5 to above 5. The *base* (feasible) factor applies all transaction-level price filters (Dick-Nielsen corrections, decimal shift, bounce-back) and the signal gap procedure from Section 3, but does not apply any return-level winsorization or trimming. The *winsorized* (infeasible) factor additionally caps returns at percentile thresholds computed from the full sample. Transaction-level price filters already address data recording errors, whereas return-level winsorization applied to clean data selectively removes extreme but genuine returns. Tables IA.V and IA.VII document the prevalence and time concentration of extreme returns, with 15 bond-month observations having month-end returns below −95% and 385 exceeding 95%. Consider a long-short factor formed by sorting bonds on signal $\hat{s}_{i,t}$, where the long leg holds bonds in the top portfolio, the short leg holds bonds in the bottom portfolio, and both legs are equal-weighted. Let $r_{t+1}^{\text{LS}} = r_{t+1}^{L} - r_{t+1}^{S}$ denote the feasible long-short return using unmodified returns. Under ex-post winsorizing, each bond return $r_{i,t+1}$ is replaced by $\tilde{r}_{i,t+1} = \max(\tau_L, \min(r_{i,t+1}, \tau_U))$, where τ_L and τ_U are the lower and upper thresholds. The winsorizing adjustment $\Delta_{i,t+1} \equiv \tilde{r}_{i,t+1} - r_{i,t+1}$ is positive when returns are floored (left tail) and negative when capped (right tail). The look-ahead bias is

$$\text{LAB}_{t+1} \equiv \tilde{r}_{t+1}^{\text{LS}} - r_{t+1}^{\text{LS}} = \underbrace{\frac{1}{N_t^L} \sum_{i \in \text{Long}} \Delta_{i,t+1}}_{\text{LAB}_{t+1}^{L}} - \underbrace{\frac{1}{N_t^S} \sum_{i \in \text{Short}} \Delta_{i,t+1}}_{\text{LAB}_{t+1}^{S}}. \tag{6}$$

The bias decomposes into contributions from each leg, with LAB^L capturing adjustments to long positions and LAB^S capturing adjustments to short positions. A positive LAB indicates that the reported factor return overstates achievable performance.[14] The direction of LAB depends on where extreme returns concentrate. If the long leg contains more left-tail returns (e.g., high-risk bonds that crash during stress), left-tail winsorizing benefits the long leg disproportionately, inflating the factor return. If the short leg contains more right-tail returns, right-tail winsorizing protects the short leg from losses, inflating the factor return. The magnitude of LAB is largest during financial distress. In episodes like the 2008–2009 Great Recession or COVID-19 in early 2020, return distributions become highly asymmetric. One leg of the portfolio experiences a disproportionate share of extreme returns, and the winsorizing adjustment becomes unbalanced. Fig. 5 contrasts ex-post and ex-ante filtering. In Panel A, the threshold $\tau^{\text{ex-post}}$ is computed using returns from the full sample, including months $t+1, \ldots, T$ that lie in the investor's future. In Panel B, the ex-ante threshold $\tau_t^{\text{ex-ante}} = \text{Percentile}_q(\{r_{i,s}\}_{s=1}^t)$ uses only returns through month t, making it computable in real time. Ex-ante filtering uses only information available at portfolio formation and therefore yields a feasible trading strategy.

4.1 Factors sensitive to ex-post return filtering

Table 3 reports factors from influential corporate bond studies (including momentum, idiosyncratic volatility, downside risk, and macroeconomic exposures) whose statistical significance depends on the choice of asymmetric return filtering. Table 3 shows that factors based on characteristics – volatility, Value at Risk (VaR), Expected Shortfall (ES), macro uncertainty – require left-tail winsorization to produce significant alphas, while momentum factors require right-tail winsorization. This pattern is consistent with the LAB decomposition in Eq. (6), whereby factors that sort on volatility, VaR, or macro uncertainty place bonds with large tail exposures in the long leg where left-tail winsorizing provides protection, while momentum factors place past losers in the short leg where right-tail winsorizing caps rebounds for bonds sold short. The winsorization threshold that "works" is determined by the sorting characteristic, specifically by whether the tail in which extreme returns concentrate corresponds to the long leg holding bonds with large downside exposure (left tail) or the short leg holding bonds prone to rebounds (right tail). Table 4 quantifies LAB for each factor. We sort bonds into value-weighted decile portfolios, going long P10 and short P1. Panel A reports factors sensitive to

[14] LAB can be decomposed into a weight component (different portfolio composition under filtering) and a return-evaluation component (different return magnitudes for the same bonds). The return-evaluation component dominates: winsorization directly alters measured returns while portfolio composition changes are negligible in value-weighted portfolios with many bonds.

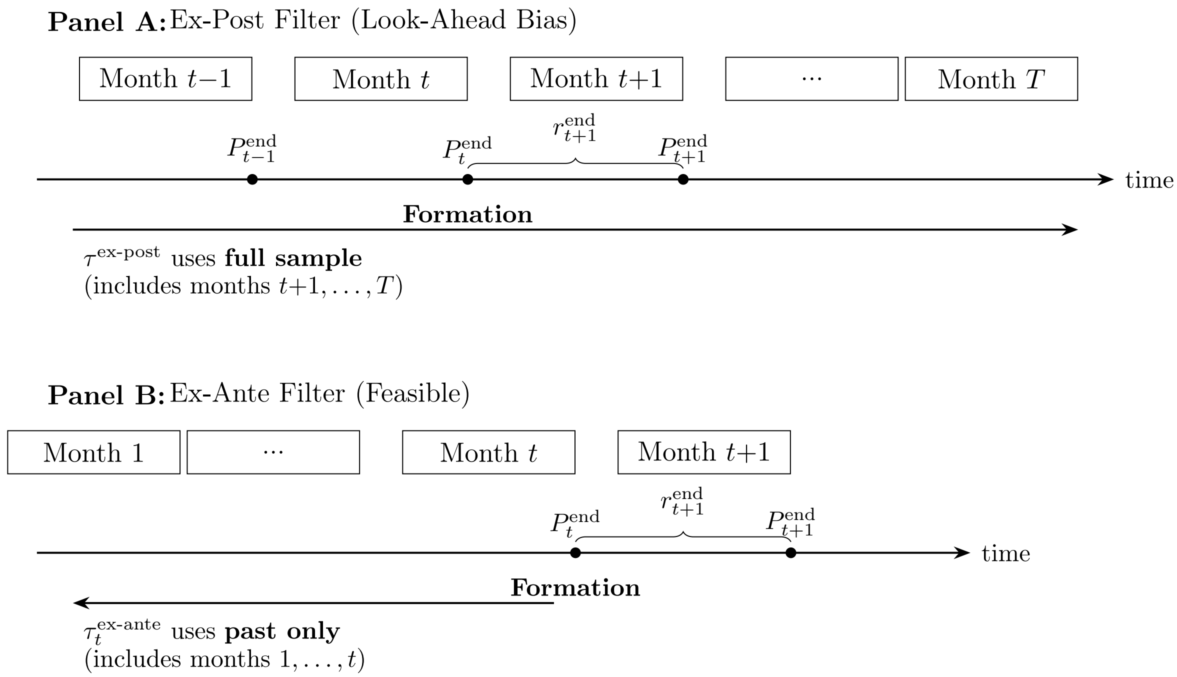

Figure 5: Comparison of ex-post (full-sample) and ex-ante (rolling-window) return filtering.

The figure contrasts ex-post and ex-ante return filtering. Panel A depicts ex-post filtering, where thresholds $\tau^{\text{ex-post}}$ are computed from the full sample including future months $t+1, \ldots, T$, embedding future information into portfolio construction. Panel B depicts the ex-ante alternative, where thresholds $\tau_t^{\text{ex-ante}}$ use only current data available at portfolio formation.

left-tail winsorization and Panel B reports factors sensitive to right-tail winsorization. For each factor, we compare the winsorized (infeasible) and base (feasible) premia and CAPMB alphas.

The idiosyncratic volatility factors exhibit the largest biases among factors sensitive to left-tail winsorization. For `ivol_bbw` (Bai, Bali, and Wen, 2021), the winsorized alpha of 0.53% per month ($t = 2.58$) drops to 0.18% ($t = 0.72$) without winsorization. The implied bias of 35 basis points accounts for 66% of the measured alpha. `ivol_vp` (Chung, Wang, and Wu, 2019) drops from 0.40% ($t = 2.48$) to 0.08% ($t = 0.44$), a 31-basis-point bias. Both factors sort high-volatility bonds into the long leg, where left-tail winsorization provides disproportionate protection during market stress. The macroeconomic uncertainty factor `b_dunc3` (Bali, Subrahmanyam, and Wen, 2023), which sorts on exposure to 3-month changes in the macroeconomic uncertainty index, has a winsorized alpha of −0.32% ($t = -1.87$) that shrinks to −0.11% ($t = -0.57$) in the feasible specification (a bias of 21 basis points). Without winsorization, the

Table 3: Corporate bond factors sensitive to ex-post return filtering.

The table lists factors from published corporate bond studies whose statistical significance depends on asymmetric return filtering. The "Wins. Threshold" column reports the representative percentile threshold at which significance changes, where 0.50^{th} indicates left-tail winsorization at the 0.5th percentile and 99.50^{th} indicates right-tail winsorization. Jostova, Nikolova, Philipov, and Stahel (2013) use the 99.50^{th} percentile (Footnote 16 of their paper). The remaining papers do not state their thresholds. Sample: 2002-08 to 2024-12, T=269.

Mnemonic	Factor Description	Journal	Wins. Threshold
`b_dunc`	Macro uncertainty (1-month Δ) beta	JFQA, Bali, Subrahmanyam, and Wen (2021b)	0.50^{th}
`b_dunc3`	Macro uncertainty (3-month Δ) beta	JFQA, Bali, Subrahmanyam, and Wen (2023)	0.50^{th}
`b_dunc6`	Macro uncertainty (6-month Δ) beta	JFQA, Bali, Subrahmanyam, and Wen (2023)	0.50^{th}
`b_unc`	Macro uncertainty level beta	JFQA, Bali, Subrahmanyam, and Wen (2023)	0.50^{th}
`mom3_1`	3-month momentum	RFS, Jostova, Nikolova, Philipov, and Stahel (2013)	99.50^{th}
`mom6_1`	6-month momentum	RFS, Jostova, Nikolova, Philipov, and Stahel (2013)	99.50^{th}
`mom12_1`	12-month momentum	RFS, Jostova, Nikolova, Philipov, and Stahel (2013)	99.50^{th}
`ltr48_12`	Long-term reversal (48-12)	JFE, Bali, Subrahmanyam, and Wen (2021a)	0.50^{th}
`ltr30_6`	Long-term reversal (30-6)	WP, Subrahmanyam (2023)	0.50^{th}
`ivol_bbw`	Idiosyncratic volatility (BBW)	JFE, Bai, Bali, and Wen (2021)	0.50^{th}
`ivol_vp`	Idiosyncratic volatility (VIX-PSB)	JFE, Chung, Wang, and Wu (2019)	0.50^{th}
`var_95`	Historical 95% value-at-risk	JFE, Bai, Bali, and Wen (2019)	0.50^{th}
`es_90`	Historical 90% expected shortfall	JFE, Bai, Bali, and Wen (2019)	0.50^{th}
`b_dvix_vp`	VIX innovation beta (PSB)	JFE, Chung, Wang, and Wu (2019)	0.50^{th}
`b_psb_m`	Pastor-Stambaugh beta (multi-factor)	JFE, Lin, Wang, and Wu (2011)	0.50^{th}
`b_amd_m`	Amihud beta (multi-factor)	JFE, Lin, Wang, and Wu (2011)	0.50^{th}

`ltr48_12` (Bali, Subrahmanyam, and Wen, 2021a) alpha shrinks from −0.33% ($t = -2.57$) to −0.23% ($t = -1.56$). In both cases, statistical significance depends on the infeasible filter.[15] Momentum factors in Panel B require right-tail winsorization. The 6-month momentum factor `mom6_1` (Jostova, Nikolova, Philipov, and Stahel, 2013) has a winsorized premium of 0.30% but a base premium of 0.00%. The alpha drops from 0.46% ($t = 2.39$) to 0.22% ($t = 1.08$). The 12-month momentum factor is more extreme, with the winsorized premium reaching 0.26% compared to a base premium of −0.13%, such that the 40-basis-point bias converts a negative premium into a positive one. Right-tail winsorization caps the rebounds of past losers in the short leg, artificially inflating momentum returns. Section IA.5 of the Internet Appendix decomposes these biases into contributions from the long and short legs. Even with winsorization, only `mom3_1` achieves a marginally significant premium ($t = 1.78$); `mom6_1` and `mom12_1` remain

[15] Several factors fail to produce detectable premia or alphas even with ex-post winsorization: `b_dunc`, `b_unc`, `b_dvix_vp`, and `b_psb_m` have $|t| < 1.96$ for both winsorized premia and alphas.

Table 4: Look-ahead bias by factor.

The table compares premia and CAPMB alphas with ($\tilde{\mu}$ and $\tilde{\alpha}$) and without (μ and α) asymmetric ex-post return winsorization for the factors in Table 3. Factors are sorted into value-weighted decile portfolios, long P10 and short P1. Bias is the difference between the winsorized (infeasible) and base (feasible) factor statistics. Panel A if for factors sensitive to left-tail winsorization (0.50^{th} percentile). Panel B is for factors sensitive to right-tail winsorization (99.50^{th} percentile). Factors have not been sign-corrected to have positive means. t-statistics use Newey-West standard errors with lags = $\lfloor T^{0.25} \rfloor$. Sample: 2002-08 to 2024-12, T=269.

	Mean Returns (%)			CAPMB Alpha (%)		
	$\tilde{\mu}_{\text{LS}}$	μ_{LS}	Bias	$\tilde{\alpha}_{\text{LS}}$	α_{LS}	Bias
Panel A: Left-tail winsorized factors (0.50^{th} percentile)						
`b_dunc`	−0.17	−0.04	−0.13	−0.22	−0.04	−0.18
	(−1.03)	(−0.22)	(−2.14)	(−1.38)	(−0.22)	(−2.20)
`b_dunc3`	−0.36	−0.20	−0.16	−0.32	−0.11	−0.21
	(−1.91)	(−1.04)	(−1.79)	(−1.87)	(−0.57)	(−1.88)
`b_unc`	−0.32	−0.18	−0.13	−0.20	−0.02	−0.18
	(−1.24)	(−0.78)	(−1.58)	(−0.91)	(−0.09)	(−1.68)
`ltr48_12`	−0.40	−0.32	−0.08	−0.33	−0.23	−0.10
	(−2.93)	(−2.21)	(−1.90)	(−2.57)	(−1.56)	(−1.65)
`ltr30_6`	−0.58	−0.45	−0.14	−0.43	−0.26	−0.17
	(−2.76)	(−1.98)	(−2.30)	(−2.37)	(−1.26)	(−2.32)
`ivol_bbw`	0.87	0.57	0.29	0.53	0.18	0.35
	(2.84)	(1.68)	(2.75)	(2.58)	(0.72)	(2.83)
`ivol_vp`	0.78	0.52	0.26	0.40	0.08	0.31
	(2.99)	(1.76)	(2.39)	(2.48)	(0.44)	(2.49)
`b_dvix_vp`	−0.02	0.02	−0.04	−0.01	0.04	−0.05
	(−0.19)	(0.14)	(−1.36)	(−0.09)	(0.38)	(−1.60)
`b_psb_m`	0.07	0.15	−0.08	0.12	0.23	−0.11
	(0.43)	(0.80)	(−2.36)	(0.85)	(1.35)	(−2.47)
`b_amd_m`	−0.29	−0.24	−0.05	−0.32	−0.26	−0.06
	(−1.99)	(−1.68)	(−2.93)	(−2.07)	(−1.74)	(−2.74)
`var_95`	1.05	0.78	0.26	0.56	0.24	0.32
	(3.50)	(2.65)	(2.69)	(3.01)	(1.42)	(2.71)
`es_90`	1.26	0.91	0.35	0.74	0.31	0.43
	(3.53)	(2.59)	(2.38)	(3.10)	(1.42)	(2.45)
Panel B: Right-tail winsorized factors (99.50^{th} percentile)						
`mom3_1`	0.33	0.17	0.16	0.45	0.31	0.13
	(1.78)	(0.90)	(2.22)	(2.26)	(1.65)	(2.10)
`mom6_1`	0.30	0.00	0.30	0.46	0.22	0.24
	(1.45)	(0.01)	(2.03)	(2.39)	(1.08)	(2.05)
`mom12_1`	0.26	−0.13	0.40	0.46	0.14	0.31
	(0.92)	(−0.39)	(2.21)	(1.60)	(0.51)	(2.33)

insignificant. The t-statistics on the winsorized CAPMB alphas for `mom3_1` ($t = 2.26$) and `mom6_1` ($t = 2.39$) exceed conventional thresholds, but are an artifact of the infeasible specifications and turn statistically insignificant without ex-post filtering. Fig. 6 illustrates the danger of asymmetric ex-post trimming using the 6-month momentum factor of Jostova, Nikolova, Philipov, and Stahel (2013) as a case study. Panels A and C apply ex-ante trimming: bonds with extreme returns are excluded from signal computation and portfolio formation using only historical information. Under ex-ante trimming, the CAPMB alpha is statistically indistinguishable from zero across all trimming thresholds in both tails. No feasible filter generates a detectable momentum alpha.

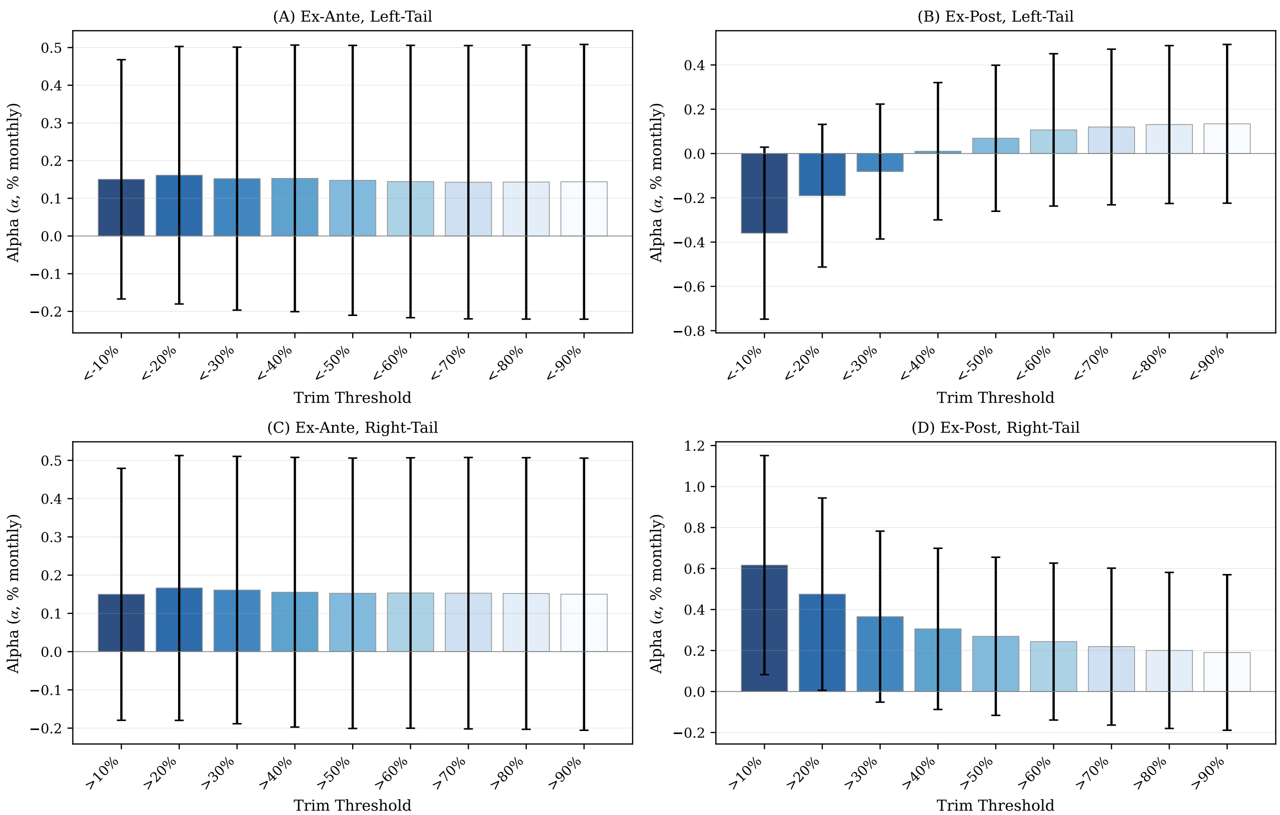

Figure 6: Sensitivity of momentum alpha to return filtering thresholds.

The figure plots CAPMB alphas for the six-month momentum factor (`mom6_1`), formed with a staggered six-month holding period following Jostova, Nikolova, Philipov, and Stahel (2013), across a range of return filtering thresholds. Panels A and C apply ex-ante filtering, excluding returns below the threshold only up to portfolio formation and excluding bonds with returns below the threshold over the prior month. Panels B and D apply ex-post filtering, adding full-sample return exclusions that embed look-ahead bias. Bars show α (in % monthly) with 95% confidence intervals using Newey-West standard errors with lags = $\lfloor T^{0.25} \rfloor$. The x-axis shows the trim threshold ($<-\tau\%$ for left-tail, $>\tau\%$ for right-tail). Value-weighted portfolios. Sample: 2003-03 to 2024-12, T=262.

In Panel D, right-tail ex-post trimming at a threshold of 10% per month generates a large alpha of approximately 0.60% that is statistically distinguishable from zero. The alpha declines monotonically as the threshold increases, requiring aggressive trimming of moderate positive returns to manufacture the premium. Panel B shows the mirror image for left-tail trimming, where at aggressive thresholds ($<-10\%$), the alpha is large and *negative*, the opposite of a momentum factor with a positive alpha. The alpha rises monotonically and crosses zero only at thresholds around -40%. This pattern illustrates the arbitrariness of ex-post filtering: a researcher who chooses right-tail trimming finds that momentum appears to work, whereas a researcher who chooses left-tail trimming finds the opposite. Neither result reflects an implementable trading strategy. Fig. 7 documents the time variation in LAB. Panels A and B plot monthly LAB for factors sensitive to left-tail and right-tail winsorization. The bias exhibits strong positive spikes during periods of market distress, specifically the Great Recession (2008–2009), the European debt crisis (2011–2012), and the COVID-19 shock (March 2020). For left-tail factors, these spikes arise because winsorization dampens large negative returns concentrated in the long leg during stress. For right-tail momentum factors, winsorization truncates the rebounds of past losers in the short leg. Panels C and D confirm this pattern, with scatter plots of LAB against the VIX level showing a strong positive convex relationship. The magnitude of LAB rises sharply when market volatility is elevated.

Fig. 8 shows cumulative returns for selected factors whose measured premia are sensitive to ex-post winsorization. The solid line plots the infeasible (winsorized) factor; the dashed line plots the feasible (unwinsorized) factor; the dotted line plots the cumulative LAB. For long-term reversal (`ltr48_12`), idiosyncratic volatility (`ivol_bbw`), and macroeconomic uncertainty (`b_dunc3`), the cumulative LAB rises sharply during the Great Recession and COVID-19, precisely when the infeasible factor appears to outperform. The gap between infeasible and feasible cumulative returns widens during crises, creating the illusion that these factors provide hedging benefits. For momentum (`mom3_1`), the same pattern emerges: the infeasible factor diverges from the feasible factor during stress episodes. In all cases, the "apparent outperformance" of the winsorized factor is attributable to the cumulative bias.

Correcting for measurement error and look-ahead biases eliminates the two largest sources of inflated factor premia and alphas documented in the corporate bond literature. But estimated premia remain sensitive to data-filtering and portfolio-construction choices that are neither standardized nor typically reported. Section 5 quantifies this residual fragility.

5 Data and methodological uncertainty

Even after correcting for measurement error (LIB) and ex-post filtering (LAB), corporate bond factor premia remain fragile. Variation from data-processing and portfolio-construction choices

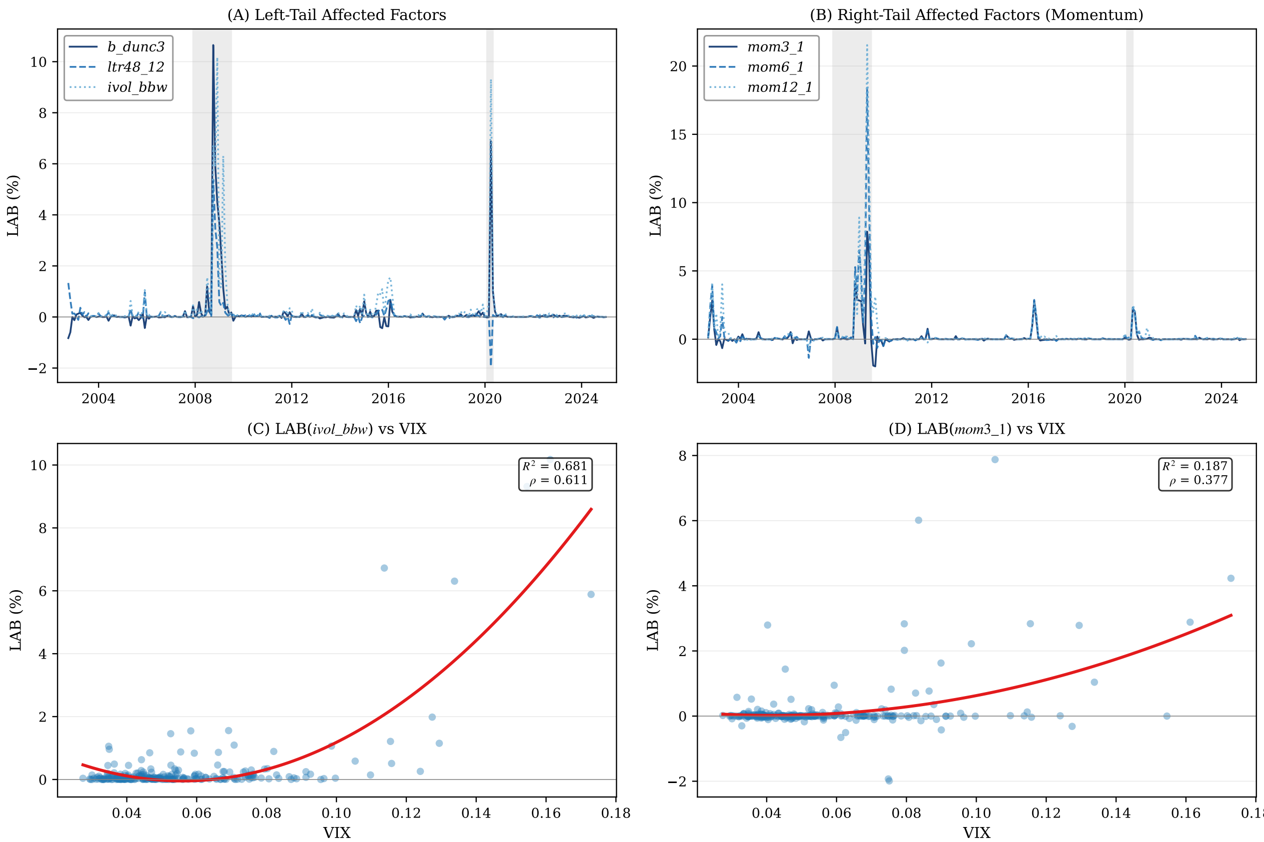

Figure 7: Time variation in look-ahead bias.

The figure plots the time series of look-ahead bias and its relationship with market volatility for selected factors. Panels A–B plot monthly LAB for left-tail factors (`b_dunc3`, `ltr48_12`, `ivol_bbw`) and right-tail momentum factors (`mom3_1`, `mom6_1`, `mom12_1`). The LAB for `b_dunc3` and `ltr48_12` is sign-corrected. Panels C–D display scatter plots of LAB versus the VIX level with quadratic fit. Shaded regions in Panels A–B indicate NBER recession periods (Great Recession: 2007:12–2009:06; COVID-19: 2020:02–2020:04). Sample: 2002-08 to 2024-12.

often exceeds both the point estimate and its sampling uncertainty. Sensitivity to specification choices is a well-documented concern in econometrics (Leamer, 1983), and recent work documents analogous fragility in equity factor research: Hou, Xue, and Zhang (2020) show that many equity factors fail to replicate under alternative data filters, and Jensen, Kelly, and Pedersen (2023) find that replicability depends on the specific methodological path. The corporate bond setting amplifies these concerns because wide bid-ask spreads, infrequent trading, and a short TRACE sample (2002–2024) magnify the influence of each filtering and aggregation decision. Uncertainty in estimated factor premia arises at multiple stages of the empirical asset pricing pipeline. We distinguish between two sources of this variation. *Data uncertainty* arises at portfolio formation and concerns which bond–month observations enter the admissible investment universe. Filtering choices such as return trimming thresholds, price range restrictions, and

Figure 8: Cumulative factor returns with and without ex-post winsorization.

The figure shows the growth of \$1 invested in four factors sensitive to ex-post return winsorization, under infeasible (with ex-post winsorization, solid light blue) and feasible (without winsorization, dashed dark blue) approaches. The dotted gray line tracks cumulative look-ahead bias, defined as the difference between the infeasible and feasible factor returns. The various panels plot `ltr48_12` (long-term reversal), `ivol_bbw` (idiosyncratic volatility), `b_dunc3` (macro uncertainty beta), and `mom3_1` (3-month momentum). Returns for `ltr48_12` and `b_dunc3` are sign-corrected. Shaded regions indicate NBER recession periods. The y-axis uses log scale. Sample: 2002-08 to 2024-12.

bid-ask bounce exclusions determine which bonds are available to trade at portfolio formation. *Methodological uncertainty* arises after the investment universe has been fixed and concerns how bonds are allocated into portfolios to form factors. Portfolio granularity, breakpoint definitions, weighting schemes, and subsample restrictions all affect measured premia even when the underlying data are identical. To quantify uncertainty arising from discretionary research choices, we follow Menkveld et al. (2024) and Walter, Weber, and Weiss (2024) and measure *non-standard errors* (NSE). Non-standard errors capture the variation in estimated factor premia attributable to alternative research and data-processing choices rather than sampling variation. While standard errors quantify uncertainty from finite samples, non-standard errors reflect uncertainty stemming from methodological decisions that are typically unreported. We define the non-standard error for factor v as the interquartile range of estimated premia across all estimation paths:

$$\text{NSE}^v = Q_{0.75}(\hat{\mu}^v) - Q_{0.25}(\hat{\mu}^v), \tag{7}$$

where $Q_\alpha(\hat{\mu}^v)$ denotes the α-quantile of the distribution of estimated premia $\hat{\mu}^v$ across paths. The ratio NSE/SE measures the relative magnitude of research-design uncertainty compared to sampling uncertainty. A ratio exceeding unity indicates that researcher degrees of freedom contribute more to result variation than statistical noise. Soebhag, Van Vliet, and Verwijmeren (2024) raise a similar concern for equity portfolio sorts, showing that seemingly innocuous aggregation choices alter inference about factor premia. In our setting, the average ratio reaches 1.15 for data uncertainty and 1.45 for methodological uncertainty, meaning two researchers applying different filters or portfolio constructions to the same corporate bond data can reach potentially opposite conclusions about whether a given factor premium is statistically significant. We apply this framework separately to data uncertainty and methodological uncertainty, reflecting their distinct stages in the research pipeline.

5.1 Data uncertainty

We study data uncertainty by varying ex-ante filtering rules that define the admissible corporate bond investment universe while holding portfolio construction fixed, with all factors using decile sorts, long the top decile (P10) and short the bottom decile (P1). All filters use only information available at portfolio formation month t, ensuring no look-ahead bias (see Fig. 2 for the timeline). Filtering is applied independently each month, so a bond excluded from portfolio formation at month t (because its characteristics violate the threshold computed from data through month t) is not held from t to $t+1$, but may re-enter the portfolio in any future month $t+k$ if it satisfies the threshold at that date. For each cross-sectional sorting variable, we evaluate 108 alternative filter configurations spanning three categories: *Return Trimming (48 configurations).* We exclude observations with monthly returns exceeding threshold τ, where $\tau \in \{0.20, 0.25, 0.30, \ldots, 0.95\}$ (16 breakpoints). Filters are applied to the left tail only (exclude $r < -\tau$), right tail only (exclude $r > \tau$), or both tails symmetrically, yielding $16 \times 3 = 48$ configurations (see, e.g., Jostova, Nikolova, Philipov, and Stahel, 2013). *Price Range Filters (30 configurations).* We restrict the sample to bonds trading within price bounds expressed as a percentage of par value ($1,000). The lower bound $P_L \in \{2, 4, 6, \ldots, 20\}\%$ (i.e., $20 to

\$200) excludes deep-discount bonds, while the upper bound $P_U \in \{150, 165, 180, \ldots, 285\}\%$ (i.e., \$1,500 to \$2,850) excludes extreme premium bonds. Each bound comprises 10 breakpoints. Filters are applied as lower bound only (10 configurations), upper bound only (10 configurations), or combined with matched pairs $[P_L, P_U] \in \{[2, 285], [4, 270], \ldots, [20, 150]\}$ (10 configurations), yielding 30 configurations total (see, e.g., Bai, Bali, and Wen, 2019, 2021; Bali, Subrahmanyam, and Wen, 2021a, 2023; Chung, Wang, and Wu, 2019). *Bid-Ask Bounce Filters (30 configurations).* We exclude observations where the absolute price change from the previous trade exceeds ϕ, where $\phi \in \{0.01, 0.02, \ldots, 0.10\}$ (10 breakpoints). Filters are applied to negative changes only (left tail), positive changes only (right tail), or both directions symmetrically, yielding $10 \times 3 = 30$ configurations (see, e.g., Bessembinder, Kahle, Maxwell, and Xu, 2008; Chordia, Goyal, Nozawa, Subrahmanyam, and Tong, 2017). Combining the three filter categories yields $48 + 30 + 30 = 108$ filter paths per signal. With equal-weighted (EW) and value-weighted (VW) portfolios across the full sample, investment grade (IG), and non-investment grade (NIG) subsamples, we obtain $108 \times 2 \times 3 = 648$ alternative estimation paths per factor, or 69,984 paths across all 108 signals. For each path, we compute (i) the average long-short return (premium), (ii) its Newey-West t-statistic with lags $\lfloor T^{0.25} \rfloor$, and (iii) the CAPMB alpha from regressing factor returns on the bond market factor.

Table 5 summarizes non-standard errors and the NSE/SE ratio by factor cluster across all 69,984 filter paths. The average non-standard error for premia is 0.35% per month, compared to an average premium of 0.33% per month. The interquartile range of filter-induced variation exceeds the point estimate. For CAPMB alphas, the average NSE of 0.27% per month is nearly twice as large as the average alpha of 0.15% per month.

The NSE varies across clusters but exceeds the average premium for every cluster except Value, Credit & Default Risk, and Macro & Other Risk. Spreads, Yields, Size (NSE 0.48%, mean premium 0.45%) and Volatility & Risk (NSE 0.49%, mean premium 0.44%) show the widest filter-induced dispersion. The NSE/SE ratio exceeds unity for four of nine clusters, with Spreads, Yields, Size (1.43), Value (1.43), and Volatility & Risk (1.36) showing the largest ratios. In Table IA.XVII of the Internet Appendix, we analyze the filter types that generate significant alphas where the baseline does not. A filter path is classified as an improvement if the associated CAPMB alpha t-statistic exceeds 1.96, exceeds the baseline, and produces a larger alpha. Most clusters show improvement rates of 0–3%: Credit & Default Risk shows 0% across all configurations, Volatility & Risk 0–2%, Market Risk 0–3%, and Macro & Other

Table 5: Non-standard errors and data uncertainty by factor cluster.

The table reports non-standard errors from data-filtering variation across 69,984 filter paths (648 per factor). For each cluster, we report the mean ($\bar{\mu}$, $\bar{\alpha}$), median (Med), and interquartile range (NSE $= Q_{0.75} - Q_{0.25}$) of factor premia and alphas across all filter paths. Ratio measures methodological uncertainty relative to sampling uncertainty, defined as Ratio $= \text{std}(r_p^v)/\overline{\sigma_p^v}$, where $\text{std}(r_p^v)$ is the standard deviation of the estimated premium or alpha across paths and $\overline{\sigma_p^v}$ is the corresponding average Newey-West standard error. N_{paths} is the number of filter configurations analyzed. Values in percent per month. Sample: 2002-09 to 2024-12, T=268.

	Premium (μ)				Alpha (α)				
Cluster	$\bar{\mu}$	Med	NSE	Ratio	$\bar{\alpha}$	Med	NSE	Ratio	N_{paths}
Spreads, Yields, Size	0.45	0.32	0.48	1.43	0.16	0.14	0.31	1.34	5,832
Value	0.43	0.38	0.30	1.43	0.27	0.22	0.22	0.97	3,240
Momentum & Reversal	0.24	0.17	0.25	0.93	0.16	0.15	0.35	1.12	13,608
Illiquidity	0.34	0.25	0.34	1.34	0.12	0.10	0.16	0.91	8,424
Volatility & Risk	0.44	0.39	0.49	1.36	0.09	0.04	0.19	0.85	10,368
Market Risk	0.30	0.32	0.35	0.80	0.09	0.10	0.37	1.01	5,832
Credit & Default Risk	0.56	0.53	0.42	0.79	0.16	0.15	0.20	0.65	2,592
Volatility & Liquidity Risk	0.21	0.18	0.26	0.87	0.15	0.11	0.34	1.03	8,424
Macro & Other Risk	0.29	0.25	0.27	0.85	0.19	0.19	0.18	0.75	11,664
All	0.33	0.26	0.35	1.15	0.15	0.13	0.27	1.01	69,984

Risk 0–1%. The Value cluster is the exception, with 21% of left-tail trimming paths and 15% of symmetric trimming paths producing significant improvements. Fig. IA.4 of the Internet Appendix displays premium t-statistics for the top four factors within each cluster across all 648 filter paths. We do not take a stand on which filters are economically justified, nor do we test every combination, as applying filters jointly across categories would produce approximately 4.2 million feasible configurations per signal. Even when each filter is tested individually, the NSE exceeds the average premium (0.35% vs 0.33% per month) and the average alpha (0.27% vs 0.15%), and most filter paths reduce rather than increase measured premia. Applying ad hoc data filters to transaction-level-cleaned data is more likely to destroy factor premia and alphas than to reveal them. The few configurations that do produce significant alphas are concentrated in a single cluster (Value) and a single filter type (left-tail return exclusions at portfolio formation). Value is also the only cluster that produces significant premia under the baseline (unfiltered) specification, so the improvement reflects amplification of an already detectable signal.

5.2 Methodological uncertainty

While data uncertainty originates from filtering bonds that enter the investable set, methodological uncertainty stems from how those bonds are sorted into portfolios to form factors. Five dimensions span the design space: portfolio granularity (tercile, quintile, decile), breakpoint universe (all bonds, investment grade only, large bonds only), weighting scheme (EW, VW), rating subsample (all, IG, NIG), and maturity bucket (all, short, intermediate, long).[16] Researchers routinely treat these dimensions as secondary design choices (see also Walter, Weber, and Weiss, 2024). Yet as Linnainmaa and Roberts (2018) show for equity factors, the mapping from a signal to a portfolio payoff can be as consequential as the signal. Combining these dimensions yields $2 \times 3 \times 3 \times 3 \times 4 = 216$ candidate portfolio constructions per signal. Not all are admissible. Breakpoints computed on IG bonds cannot sort NIG bonds (the universes do not overlap). This eliminates 24 specifications per signal. When portfolios are formed, e.g., exclusively within IG bonds, IG-based breakpoints coincide with full-universe breakpoints. This eliminates a further 24 redundant specifications. The effective grid contains 168 economically distinct portfolio constructions per signal, yielding $168 \times 108 = 18{,}144$ factor return series. Sixteen specifications (0.088%) produce months with empty long or short legs, all in decile sorts within the high-yield, long-maturity segment.[17] Excluding these leaves 18,128 well-defined factor return series.[18] Table 6 reports non-standard errors across these 168 portfolio construction paths. The average NSE for premia is 0.31% per month, compared to an average premium of 0.22%, and the interquartile range (IQR) of methodology-induced variation exceeds the point estimate. For CAPMB alphas, the average NSE of 0.20% per month exceeds the average alpha of 0.11%. The premium-related NSE/SE ratio averages 1.45 and exceeds unity for all nine factor clusters, ranging from 1.02 (Market Risk) to 2.00 (Spreads, Yields, Size).

Fig. IA.5 of the Internet Appendix displays CAPMB alpha t-statistics for the top four factors within each cluster across all 168 portfolio constructions. For many signals, the median t-statistic lies below 1.96, and the interquartile range spans both significant and insignificant

[16] Breakpoint variation changes a bond's relative rank given a reference distribution, while rating subsampling changes the set of bonds included in the sort. For example, "investment grade only" breakpoints compute quantile boundaries from the cross-section of a signal within the IG universe, then allocate all bonds (including high-yield) to portfolios based on those boundaries. The equity analogue is Fama and French's use of NYSE-only size breakpoints to sort all stocks.

[17] Table IA.XVIII confirms that even the most restrictive configurations contain a sufficient number of bonds in the long and short portfolios.

[18] All the main results are robust to excluding decile sorts from the specification grid.

Table 6: Non-standard errors and methodological uncertainty by factor cluster.

The table reports non-standard errors from portfolio-construction variation across 18,128 non-degenerate methodology paths (up to 168 per factor). For each cluster, we report the mean ($\bar{\mu}$, $\bar{\alpha}$), median (Med), and interquartile range (NSE $= Q_{0.75} - Q_{0.25}$) of factor premia and alphas across all methodology paths. Ratio measures methodological uncertainty relative to sampling uncertainty, defined as Ratio $= \text{std}(r_p^v)/\overline{\sigma_p^v}$, where $\text{std}(r_p^v)$ is the standard deviation of the estimated premium or alpha across paths and $\overline{\sigma_p^v}$ is the corresponding average Newey-West standard error. N_{paths} is the number of methodological configurations analyzed. Values in percent per month. Sample: 2002-09 to 2024-12, T=268.

	Premium (μ)				Alpha (α)				
Cluster	$\bar{\mu}$	Med	NSE	Ratio	$\bar{\alpha}$	Med	NSE	Ratio	N_{paths}
Spreads, Yields, Size	0.25	0.22	0.46	2.00	0.09	0.09	0.34	1.57	1,508
Value	0.29	0.25	0.27	1.62	0.19	0.17	0.20	1.13	838
Momentum & Reversal	0.13	0.09	0.24	1.28	0.08	0.06	0.23	1.21	3,520
Illiquidity	0.22	0.16	0.27	1.53	0.10	0.08	0.14	1.02	2,184
Volatility & Risk	0.32	0.26	0.43	1.66	0.10	0.07	0.19	1.07	2,688
Market Risk	0.21	0.20	0.27	1.02	0.09	0.08	0.24	1.15	1,512
Credit & Default Risk	0.43	0.38	0.35	1.11	0.16	0.13	0.20	0.79	672
Volatility & Liquidity Risk	0.15	0.12	0.20	1.10	0.11	0.09	0.23	1.10	2,184
Macro & Other Risk	0.21	0.19	0.26	1.12	0.14	0.13	0.16	0.89	3,022
All	0.22	0.17	0.31	1.45	0.11	0.09	0.20	1.14	18,128

values. Across these 36 factors (four top factors × nine clusters), 12 produce both positive and negative long–short premia under alternative portfolio constructions, and 29 produce both positive and negative CAPMB alphas. Table IA.XIX of the Internet Appendix decomposes improvement rates across portfolio construction dimensions. Out of all methodological variations, 94% of them do not improve the CAPMB alpha relative to the baseline. Value is the largest improver (16% overall, 36% in long-maturity constructions), and long maturity is the primary driver across clusters (9.5% vs 1.9% for intermediate maturity). As with data uncertainty, Value is the only cluster with a detectable baseline premium. Fig. IA.6 of the Internet Appendix displays premium t-statistics across all constructions for the top four factors per cluster. Given that 94% of the specifications fail to improve the CAPMB alpha relative to the baseline, factor construction choices are better guided by economic reasoning rather than by searching across the specification space.

6 Conclusion

Corporate bond factor research faces a replication crisis. Across a 'factor zoo' of 108 signals spanning nine thematic clusters, most previously documented factor alphas reflect two biases rather than genuine risk-adjusted performance. Latent Implementation Bias (LIB) inflates measured factor premia for price-based factors (yields, credit spreads, value, and short-term reversal) by up to 91%. LIB arises from two sources: the same noisy transaction price enters both the sorting signal and the return denominator, creating a correlated errors-in-variables (CEIV) problem, and the observed transaction price is not executable in over-the-counter (OTC) markets. Look-Ahead Bias (LAB) inflates measured factor premia for momentum, long-term reversal, idiosyncratic volatility, downside risk, and macroeconomic exposure factors by 57–100% through ex-post return filtering that embeds future information into portfolio construction. A third source of fragility persists even after both biases are corrected. Across 648 data-filtering configurations using only ex-ante information, non-standard errors average 0.35% per month, exceeding the average premium of 0.33%. Across 168 methodological configurations holding the investment universe fixed, the NSE/SE ratio averages 1.45 and exceeds unity for all nine factor clusters. Twelve of 36 top factors flip sign depending on the portfolio construction chosen. As emphasized in the Internet Appendix, across 432 factor-specification combinations, only 26 (6.0%) bond CAPM alphas survive after false discovery rate (FDR) correction. The survivors concentrate among credit-spread-based value factors. Applying FDR to mean return p-values produces consistent results, with 119 of 432 nominally significant but only 22 surviving correction. Across five illiquidity factors, no CAPMB alpha is distinguishable from zero.

Our contribution extends beyond identifying these replication failures. We provide open source data infrastructure, software, and a reproducible framework for corporate bond asset pricing research. The companion website Open Bond Asset Pricing hosts daily bond data constructed through a two-stage TRACE cleaning pipeline, pre-formed factors for all 108 signals, and data reports documenting every cleaning step. The data are also available through WRDS Contributed Data. The companion software `PyBondLab` provides the infrastructure for reproducible factor construction, with built-in signal gap procedures, ex-ante filtering, and flexible portfolio configurations. The replication code is publicly available on GitHub.

These resources define a path forward. Credible corporate bond factor research requires three elements: breaking the shared-price link through gap procedures that ensure the signal and the return denominator use different prices, filtering only with information available at portfolio formation, and reporting economically motivated factor specifications that limit the effects of data mining. Finally, given the nontrivial trade execution costs in OTC corporate bond markets, a natural extension is to evaluate whether the surviving factors remain profitable after accounting for corporate bond trade execution costs.

References

Abdi, F. and A. Ranaldo (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *Review of Financial Studies 30*, 4437–4480.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets 5*, 31–56.

Andreani, M., D. Palhares, and S. Richardson (2024). Computing corporate bond returns: A word (or two) of caution. *Review of Accounting Studies 29*, 3887–3906.

Ang, A., J. Chen, and Y. Xing (2006). Downside risk. *Review of Financial Studies 19*, 1191–1239.

Bai, J., T. G. Bali, and Q. Wen (2019). RETRACTED: Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics 131*, 619–642.

Bai, J., T. G. Bali, and Q. Wen (2021). Is there a risk-return tradeoff in the corporate bond market? Time-series and cross-sectional evidence. *Journal of Financial Economics 142*, 1017–1037.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics 131*, 1593–1636.

Bali, T. G., A. Subrahmanyam, and Q. Wen (2021a). Long-term reversals in the corporate bond market. *Journal of Financial Economics 139*, 656–677.

Bali, T. G., A. Subrahmanyam, and Q. Wen (2021b). The macroeconomic uncertainty premium in the corporate bond market. *Journal of Financial and Quantitative Analysis 56*, 1653–1678.

Bali, T. G., A. Subrahmanyam, and Q. Wen (2023). The macroeconomic uncertainty premium in the corporate bond market—Corrigendum. *Journal of Financial and Quantitative Analysis*.

Bao, J., J. Pan, and J. Wang (2011). The illiquidity of corporate bonds. *Journal of Finance 66*, 911–946.

Bartram, S. M., M. Grinblatt, and Y. Nozawa (2025). Book-to-market, mispricing, and the cross-section of corporate bond returns. *Journal of Financial and Quantitative Analysis 60*, 1185–1233.

Baumann, F., A. Kakhbod, D. Livdan, A. Nazemi, and N. Schürhoff (2025). Life after default: How dealer intermediation improves default recovery. Working Paper.

Baumann, F. and A. Nazemi (2025). Defaulted bonds: A hybrid asset priced by bond and equity markets. Working Paper.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological) 57*, 289–300.

Bessembinder, H., K. M. Kahle, W. F. Maxwell, and D. Xu (2008). Measuring abnormal bond performance. *Review of Financial Studies 22*, 4219–4258.

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

Blume, M. E. and R. F. Stambaugh (1983). Biases in computed returns: An application to the size effect. *Journal of Financial Economics 12*, 387–404.

Bollerslev, T., S. Z. Li, and B. Zhao (2020). Realized semicovariances. *Econometrica 88*, 1515–1551.

Ceballos, L. (2022). Inflation volatility risk and the cross-section of corporate bond returns. Working Paper.

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

Chen, Q. and J. Choi (2024). Reaching for yield and the cross section of bond returns. *Management Science 70*, 5226–5245.

Choi, J. (2013). What drives the value premium?: The role of asset risk and leverage. *Review of Financial Studies 26*, 2845–2875.

Choi, J., J. Han, S. S. Shin, and J. H. Yoon (2026). The more illiquid, the more expensive: A search-based explanation of the illiquidity premium. Working paper.

Chordia, T., A. Goyal, Y. Nozawa, A. Subrahmanyam, and Q. Tong (2017). Are capital market anomalies common to equity and corporate bond markets? An empirical investigation. *Journal of Financial and Quantitative Analysis 52*, 1301–1342.

Chung, K. H., J. Wang, and C. Wu (2019). Volatility and the cross-section of corporate bond returns. *Journal of Financial Economics 133*, 397–417.

Coase, R. H. (1982). *How Should Economists Choose?* G. Warren Nutter Lectures in Political Economy.

Washington, D.C.: American Enterprise Institute.

Conrad, J., M. N. Gultekin, and G. Kaul (1997). Profitability of short-term contrarian strategies: Implications for market efficiency. *Journal of Business & Economic Statistics 15*, 379–386.

Corwin, S. A. and P. Schultz (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance 67*, 719–760.

Danyliv, O., B. Bland, and D. Nicholass (2014). A convenient liquidity measure. *Journal of Trading 9*, 38–49.

Dick-Nielsen, J. (2014). How to clean Enhanced TRACE data. Working Paper.

Dick-Nielsen, J., P. Feldhütter, and D. Lando (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics 103*, 471–492.

Dick-Nielsen, J., P. Feldhütter, L. H. Pedersen, and C. Stolborg (2023). Corporate bond factors: Replication failures and a new framework. Working Paper.

Dickerson, A., P. Mueller, and C. Robotti (2023). Priced risk in corporate bonds. *Journal of Financial Economics 150, article 103707.*

Dickerson, A., C. Robotti, and Y. Nozawa (2025). Factor investing with delays. Working Paper.

Duarte, J., C. S. Jones, M. Khorram, H. Mo, and J. L. Wang (2025). Too good to be true: Look-ahead bias in empirical options research. *Review of Financial Studies*. Forthcoming.

Duarte, J., C. S. Jones, and J. L. Wang (2024). Very noisy option prices and inference regarding the volatility risk premium. *Journal of Finance 79*, 3581–3621.

Elkamhi, R., C. Jo, and Y. Nozawa (2024). A one-factor model of corporate bond premia. *Management Science 70*, 1875–1900.

Fama, E. F. (1984). The information in the term structure. *Journal of Financial Economics 13*, 509–528.

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

Fong, K. Y., C. W. Holden, and C. A. Trzcinka (2017). What are the best liquidity proxies for global research? *Review of Finance 21*, 1355–1401.

Gebhardt, W. R., S. Hvidkjaer, and B. Swaminathan (2005). The cross-section of expected corporate bond returns: Betas or characteristics? *Journal of Financial Economics 75*, 85–114.

Ghaderi, M., S. Plante, N. L. Roussanov, and S. B. Seo (2024). Pricing of corporate bonds: Evidence from a century-long cross-section. Working Paper.

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

Harvey, C. R. and A. Siddique (2000). Conditional skewness in asset pricing tests. *Journal of Finance 55*, 1263–1295.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*, 1–35.

Hong, G. and A. Warga (2000). An empirical study of bond market transactions. *Financial Analysts Journal 56*, 32–46.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*, 2019–2133.

Houweling, P. and J. Van Zundert (2017). Factor investing in the corporate bond market. *Financial Analysts Journal 73*, 100–115.

Israel, R., D. Palhares, and S. Richardson (2018). Common factors in corporate bond returns. *Journal of Investment Management 16*, 17–46.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*, 881–898.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). Is there a replication crisis in finance? *Journal of Finance 78*, 2465–2518.

Jostova, G., S. Nikolova, A. Philipov, and C. W. Stahel (2013). Momentum in corporate bond returns. *Review of Financial Studies 26*, 1649–1693.

Kelly, B., D. Palhares, and S. Pruitt (2023). Modeling corporate bond returns. *Journal of Finance 78*, 1967–2008.

Koijen, R. S., H. Lustig, and S. Van Nieuwerburgh (2017). The cross-section of managerial ability, incentives, and risk preferences. *Journal of Monetary Economics 91*, 1–17.

Lair, T. and J. Blonk (2024). Valuations in the dark: When independent valuators influence corporate bond returns. Working Paper.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review 73*, 31–43.

Lin, H., J. Wang, and C. Wu (2011). Liquidity risk and expected corporate bond returns. *Journal of Financial Economics 99*, 628–650.

Linnainmaa, J. T. and M. R. Roberts (2018). The history of the cross-section of stock returns. *Review of Financial Studies 31*, 2606–2649.

Liu, Y. and J. C. Wu (2021). Reconstructing the yield curve. *Journal of Financial Economics 142*, 1395–1425.

Menkveld, A. J., A. Dreber, F. Holzmeister, J. Huber, M. Johannesson, M. Kirchler, S. Neususs, M. Razen, U. Weitzel, and et al. (2024). Non-standard errors. *Journal of Finance 79*, 2339–2390.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics 103*, 429–453.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*, 642–685.

Richardson, S. and D. Palhares (2018). (Il)liquidity premium in credit markets: A myth? *Journal of Fixed Income 28*, 3–31.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance 39*, 1127–1139.

Soebhag, A., B. Van Vliet, and P. Verwijmeren (2024). Non-standard errors in asset pricing: Mind your sorts. *Journal of Empirical Finance 78*, 101517.

Stambaugh, R. F. (1988). The information in forward rates: Implications for models of the term structure. *Journal of Financial Economics 21*, 41–70.

Subrahmanyam, A. (2023). Corporate bond data projects: Some clarifications. Working Paper.

Tobek, O. (2016). Liquidity proxies using daily trading volume. Working Paper.

van Binsbergen, J. H., Y. Nozawa, and M. Schwert (2025). Duration-based valuation of corporate bonds. *Review of Financial Studies 38*, 158–191.

Walter, D., R. Weber, and P. Weiss (2024). Methodological uncertainty in portfolio sorts. Working Paper.

Wang, J., D. Wu, and L. Yang (2024). Cross-bond momentum spillovers. Working Paper.

Appendix

A Data filters and monthly returns

This appendix provides detailed specifications for the transaction-level and daily-level filters described in Section 2, followed by the formal definitions of monthly returns and the treatment of defaulted bonds. All thresholds are expressed in percentage-of-par units ($1,000 face value), so a threshold of 35 corresponds to $350. Section 5 examines the sensitivity of our results to these parameter choices across 648 filter configurations. All filters are implemented in Python and available at `github.com/Alexander-M-Dickerson/trace-data-pipeline`.

A.1 Decimal shift corrector

Corporate bond prices in TRACE are expressed as a percentage of par value, where par equals $1,000 for all bonds in our sample. A price of 100 represents $1,000 (100% of par); a price of 98.5 represents $985. Prices occasionally exhibit multiplicative errors arising from incorrect decimal placement during data entry (for example, a price of 98.50 recorded as 985.0 or 9.85). The decimal shift corrector identifies and corrects such errors while preserving legitimate price movements. Further details and worked examples are available at `github.com/.../README_decimal_shift_corrector.md`.

Anchor construction. For each transaction i with observed price P_i, we construct a *rolling unique-median anchor* A_i: the median of distinct prices within a centered window of $2w + 1$ observations (default $w = 5$, yielding 11 transactions). When centered windows are unavailable near sample boundaries, the filter falls back to forward-looking or backward-looking windows, and if both are unavailable, to the median of the bond's full price series.

Error measurement. The *raw relative error* measures how far the observed price deviates from the anchor:

$$\varepsilon_i^{\text{raw}} = \frac{|P_i - A_i|}{A_i}. \tag{A.8}$$

A transaction qualifies for correction testing only if this raw error exceeds the gate threshold $\tau_{\text{bad}} = 0.05$ (5%), establishing that a meaningful discrepancy exists.

Candidate correction. We test four multiplicative correction factors: $\mathcal{F} = \{0.1, 0.01, 10, 100\}$, corresponding to common decimal-shift scenarios. For each candidate factor $f \in \mathcal{F}$, the corrected price is $P_i^{(f)} = f \cdot P_i$, and the *corrected relative error* is

$$\varepsilon_i^{(f)} = \frac{|P_i^{(f)} - A_i|}{A_i}. \tag{A.9}$$

Acceptance gates. A correction is accepted only if all four conditions hold:

1. **Raw error gate:** $\varepsilon_i^{\text{raw}} > \tau_{\text{bad}} = 0.05$ (5%), confirming a meaningful discrepancy.
2. **Alignment gate:** At least one of: (a) $\varepsilon_i^{(f)} \le \tau_{\text{pct}} = 0.02$ (2%); (b) the absolute error $|P_i^{(f)} - A_i| \le \tau_{\text{abs}} = 8.0$ price points; or (c) both A_i and $P_i^{(f)}$ fall within ± 15 points of par (par-proximity rule).
3. **Improvement requirement:** $\varepsilon_i^{(f)} \le \gamma \cdot \varepsilon_i^{\text{raw}}$, where $\gamma = 0.20$; the corrected error must be at most 20% of the raw error.
4. **Plausibility check:** The corrected price lies within $[5, 300]$ (5% to 300% of par).

When multiple factors satisfy all gates, we select the factor minimizing $\varepsilon_i^{(f)}$.

Table A.1 summarizes the default parameter values.

Table A.1: Decimal shift corrector parameters.

Parameter	Default	Description
w	5	Half-window size (effective window = $2w + 1 = 11$)
τ_{pct}	0.02	Relative error acceptance threshold
τ_{abs}	8.0	Absolute error acceptance threshold (price points)
τ_{bad}	0.05	Minimum raw error to consider correction
γ	0.20	Required proportional improvement
$[\underline{P}, \bar{P}]$	$[5, 300]$	Plausible corrected price range
par-proximity band	15.0	Par-proximity tolerance (points from par)

A.2 Bounce-back filter

The bounce-back filter detects transient price spikes (large deviations from baseline that quickly revert) which typically indicate data entry errors rather than genuine market movements. Further details and worked examples are available at github.com/.../README_bounce_back_filter.md. The detection threshold is $\tau = 35$ price points (35% of par, or \$350 on a \$1,000 face-value bond).

Anchor construction. For each transaction i with observed price P_i, we compute a trailing baseline B_i as the median of unique prices from the prior w trades (default $w = 5$):

$$B_i = \text{median}\big(\text{unique}(\{P_{i-w}, \ldots, P_{i-1}\})\big). \tag{A.10}$$

Taking unique prices removes repeated-print bias from consecutive identical trades.

Candidate detection. An observation becomes a bounce-back candidate if any of the following three conditions holds:

1. **Large one-step jump:** $|\Delta P_i| \geq \tau - \delta$, where $\Delta P_i = P_i - P_{i-1}$ and $\delta = 1$ is a small slack tolerance.

2. **Large displacement from baseline:** $|P_i - B_i| \geq \tau - \delta$.

3. **Par-spike heuristic:** $P_i = 100$ (par) and $|P_i - B_i| \geq \alpha \cdot \tau$, where $\alpha = 0.25$.

Reversion detection. Once a candidate is identified, the filter scans forward up to $L = 5$ rows seeking evidence of reversion via either:

- **Path A (opposite move):** A subsequent price change of opposite sign with magnitude $\geq \tau - \delta$.

- **Path B (return to anchor):** A subsequent price within $\alpha \cdot \tau$ of the original baseline B_i.

Flagging logic. Upon detecting reversion, the filter: (i) reassigns the flag to the preceding row if it shows greater displacement from baseline ($\geq$ 5 points); (ii) extends flags forward within a five-row span while prices remain displaced; and (iii) flags persistent par blocks (runs of $\ell_{\min} \geq 3$ consecutive trades at par) with a two-row cooldown to suppress cascading flags.

Table A.2 summarizes the default parameter values.

Table A.2: Bounce-back filter parameters.

Parameter	Default	Description
τ	35.0	Minimum price change to trigger detection (\$350)
L	5	Forward rows scanned for reversion
w	5	Trailing window for anchor calculation
α	0.25	Tolerance fraction for return-to-anchor criterion
$\ell_{\min}$	3	Minimum consecutive par trades to flag
cooldown	2	Rows skipped after flagging par blocks

A.3 Distressed bond filters

After aggregating transactions to daily frequency, a second filtering stage targets anomalies in distressed bond prices. Four sub-filters address distinct error patterns, each using lookback and lookforward windows of $L = 5$ days and ratio-based thresholds that adapt to varying price scales. All prices are expressed as a percentage of par (\$1,000 face value). Further details and worked examples are available at `github.com/.../README_distressed_filter.md`.

Filter 1: Anomaly detection (downward outliers). This filter identifies isolated ultra-low prices far below surrounding observations. A price P_t on day t becomes a candidate if either: (i) $P_t < \tau_{\text{low}} = 0.10$ (i.e., below 0.10% of par, or \$1), or (ii) P_t matches a suspicious round number in $\{0.001, 0.01, 0.05, 0.10, 0.25, 0.50, 1.00\}$. To assess whether the candidate is anomalous, we collect all prices from the surrounding $\pm L$-day window that exceed P_t, and compute their median M_{surr}. If the surrounding prices are much higher than the candidate (specifically, if $M_{\text{surr}}/P_t \geq \rho_{\text{anomaly}} = 3.0$), then the candidate is flagged as a likely data error.

Filter 2: Spike detection (upward outliers). This filter detects temporary upward price spikes that quickly revert. A price P_t becomes a candidate if either: (i) $P_t > \tau_{\text{high}} = 5.0$ (i.e., above 5% of par, or \$50), or (ii) P_t is a round number exceeding 0.50 (\$5). To confirm a spike, we compute the pre-spike median M_{pre} from the L days preceding day t. The spike is flagged only if two conditions hold. First, the spike magnitude satisfies $P_t/M_{\text{pre}} \geq \rho_{\text{spike}} = 3.0$ (the price is at least 3× the pre-spike level). Second, recovery occurs within the subsequent L days: at least one price falls to $P_j \leq \rho_{\text{recovery}} \cdot M_{\text{pre}}$, where $\rho_{\text{recovery}} = 2.0$.

Filter 3: Plateau detection. This filter identifies consecutive days with identical ultra-low or round prices, often indicating stale or placeholder values. A price P_t opens a candidate if either: (i) $P_t < \tau_{\text{plateau}} = 0.15$ (below 0.15% of par, or \$1.50), or (ii) P_t is a round number. The algorithm identifies runs of $\ell \geq \ell_{\min} = 2$ consecutive days with exactly the same price. Let P_{pre} denote the price on the day before the plateau begins, and P_{post} denote the price on the day after the plateau ends. The plateau is flagged if any of three conditions holds: (i) $P_{\text{pre}}/P_t \geq \rho_{\text{plateau}} = 3.0$; (ii) $P_{\text{post}}/P_t \geq \rho_{\text{plateau}}$; or (iii) P_t is a round number. The intuition is that genuine low prices should persist, whereas placeholder values typically appear briefly before reverting.

Filter 4: Intraday inconsistency. This filter flags large discrepancies between different intraday price measures. For each day t, let $\mathcal{I}_t$ denote the set of available intraday prices (high, low) and let $\bar{P}_t = |\mathcal{I}_t|^{-1} \sum_{p \in \mathcal{I}_t} p$ be their mean. An observation is flagged if both conditions hold: (i) at least one price in $\mathcal{I}_t$ falls below $\tau_{\text{intraday}} = 20.0$ (20% of par, or \$200), and (ii) the intraday range exceeds $\gamma_{\text{range}} = 0.75$ (75%) of the mean:

$$\frac{\max(\mathcal{I}_t) - \min(\mathcal{I}_t)}{\bar{P}_t} > 0.75.$$

The first condition ensures the filter activates only for days with at least one low price, avoiding false positives on normally-priced bonds with typical intraday volatility. Table A.3 summarizes the default parameter values.

A.4 Monthly return computation

We compute three return series for each bond-month, as referenced in Section 2.2.

Month-end return. The month-end return measures total performance from the end of month t to the end of month $t+1$:

$$r_{i,t+1}^{\text{End}} = \frac{P_{i,t+1}^{\text{end}} + AI_{i,t+1}^{\text{end}} + C_{i,t+1}}{P_{i,t}^{\text{end}} + AI_{i,t}^{\text{end}}} - 1, \tag{A.11}$$

Table A.3: Distressed bond filter parameters.

Parameter	Default	Description
τ_{low}	0.10%	Ultra-low threshold (Filter 1)
τ_{high}	5.0%	High spike threshold (Filter 2)
τ_{plateau}	0.15%	Plateau opening threshold (Filter 3)
ρ_{anomaly}	3.0	Ratio for downward outliers (Filter 1)
ρ_{spike}	3.0	Spike magnitude ratio (Filter 2)
ρ_{recovery}	2.0	Recovery ratio (Filter 2)
ρ_{plateau}	3.0	Pre/post displacement ratio (Filter 3)
γ_{range}	0.75	Intraday range threshold (Filter 4)
τ_{intraday}	20.0%	Price threshold for Filter 4 activation
L	5	Lookback/lookforward window (days)
ℓ_{min}	2	Minimum plateau length (days)

where $P_{i,t}$ is the volume-weighted clean price, $AI_{i,t}$ is accrued interest, and $C_{i,t+1}$ is the coupon payment (if any) during month $t+1$. A return is valid only if the bond trades within the last five business days of both month t and month $t+1$ (NYSE calendar). This measure is the standard return used for comparability with prior literature.

Month-begin return. The month-begin return measures performance within a single month, from the first available trade to the last:

$$r_{i,t+1}^{\text{Bgn}} = \frac{P_{i,t+1}^{\text{end}} + AI_{i,t+1}^{\text{end}} + C_{i,t+1}}{P_{i,t+1}^{\text{bgn}} + AI_{i,t+1}^{\text{bgn}}} - 1, \tag{A.12}$$

where superscripts denote prices from the last and first five business days of month $t+1$, respectively. This return is valid only if the bond trades in both windows. Because the entry price post-dates signal observation, the month-begin return measures *implementable* performance: it is the return a trader could earn after observing a signal at month-end t and executing at the earliest available price in month $t+1$.

Return decomposition. The month-end return decomposes approximately as

$$r_{i,t+1}^{\text{End}} \approx \underbrace{\frac{P_{i,t+1}^{\text{bgn}}}{P_{i,t}} - 1}_{\text{LIB}_{i,t+1}} + r_{i,t+1}^{\text{Bgn}}, \tag{A.13}$$

where $\text{LIB}_{i,t+1}$ is the clean-price return between the signal observation price and the execution price. Section 3 discusses the economic interpretation of this decomposition.

Excess and duration-adjusted returns. We form excess returns by subtracting the Fama-French risk-free rate: $r^x_{i,t+1} = r_{i,t+1} - r^f_{t+1}$. Duration-adjusted excess returns instead subtract a duration-matched U.S. Treasury return:

$$r^x_{i,t+1} = r_{i,t+1} - r^{\text{Tsy}}_{i,t+1}, \tag{A.14}$$

where $r^{\text{Tsy}}_{i,t+1}$ is obtained by linearly interpolating key-rate U.S. Treasury bond returns from WRDS using each bond's modified duration, following Andreani, Palhares, and Richardson (2024).

A.5 Default handling

When a bond enters or trades under default, we assume coupon payments cease and adjust the return formula accordingly. We identify a bond as in default if its S&P rating equals 22 (D) or its Moody's rating equals 21 (D).

Default-event return. When a bond transitions *into* default at time $t+1$ (rated non-default at t, default at $t+1$), we set the coupon to zero. The return uses the clean price at default in the numerator and the dirty price before default in the denominator:

$$r^{\text{def}}_{i,t+1} = \frac{P^{\text{clean}}_{i,t+1}}{P_{i,t} + AI_{i,t}} - 1. \tag{A.15}$$

Trading-in-default return. When a bond remains in default at both t and $t+1$, no coupon accrues. We compute the return on a flat basis using clean prices only:

$$r^{\text{flat}}_{i,t+1} = \frac{P^{\text{clean}}_{i,t+1}}{P^{\text{clean}}_{i,t}} - 1. \tag{A.16}$$

This return is capped at the standard return (A.11) to prevent cases where ignoring accrued interest produces an artificially higher return.

B Sorting-induced bias: Formal treatment

A correlated errors-in-variables (CEIV) bias arises in long–short factor returns constructed from price-based signals whenever the observed sorting signal and the observed return load on a common disturbance. In this case, portfolio assignment becomes informative about the return measurement error, and the bias term in the long–short decomposition does not cancel out.

Our general result only requires that price measurement error exhibit cross-sectional variation, that this disturbance enter the observed signal in a non-degenerate way, and that the return measurement error contain a component driven by the same disturbance. In particular, when the signal measurement error is a non-constant (locally monotone) function of a cross-sectionally heterogeneous price disturbance, portfolio sorting induces selection on the disturbance directly. If the return measurement error also depends non-trivially on the same component, assets assigned to the long and short portfolios systematically differ in their expected return measurement error. As a result, even when the true signal has no predictive content for future returns, long–short portfolio returns are biased.

This bias is not driven by misranking per se. While measurement error can lead to misallocation of assets across portfolios, misranking alone does not generate a mechanical premium. The bias arises specifically from the shared component between the signal and the return denominator. Consistent with this distinction, we show that for non-price signals, such as characteristics whose measurement error does not enter the return calculation, measurement error may attenuate true premia or alter portfolio composition, but it does not generate the mechanical CEIV bias identified here.

Having established the existence and sign of the bias under general conditions, we then impose stronger distributional assumptions to characterize its magnitude. Under linear signal contamination and joint normality, we derive a closed-form expression for the bias as a function of portfolio granularity (finer sorts amplify the bias), the share of signal variance attributable to noise (ρ), and the volatility of price measurement error (σ_δ).

B.1 Measurement structure

Definition B.1 (Price noise)**.** *Following Blume and Stambaugh (1983), the observed price* $\hat{P}_{i,t}$ *relates to the true price* $P_{i,t}$ *by*

$$\hat{P}_{i,t} = (1 + \delta_{i,t}) P_{i,t}, \tag{A.17}$$

where $\delta_{i,t}$ *is a mean-zero measurement error with finite variance* $\sigma_\delta^2 \equiv \mathrm{Var}(\delta_{i,t})$.

Definition B.2 (Signal decomposition)**.** *The observed sorting signal* $\hat{s}_{i,t}$ *decomposes as*

$$\hat{s}_{i,t} = s_{i,t} + \eta_{i,t}, \tag{A.18}$$

where $s_{i,t}$ *is the true signal and* $\eta_{i,t}$ *is the signal's measurement error.*

Definition B.3 (Return decomposition)**.** *Define the observed return by* $\hat{r}_{i,t+1} \equiv \hat{P}_{i,t+1}/\hat{P}_{i,t} - 1$*. Then,*

$$\hat{r}_{i,t+1} = r_{i,t+1} + \epsilon_{i,t}, \tag{A.19}$$

where $r_{i,t+1} \equiv P_{i,t+1}/P_{i,t} - 1$ *is the true return, and*

$$\epsilon_{i,t} \equiv \frac{1 + \delta_{i,t+1}}{1 + \delta_{i,t}} - 1. \tag{A.20}$$

A first-order approximation yields $\epsilon_{i,t} \approx \delta_{i,t+1} - \delta_{i,t}$*.*

Assumption B.1 (Serial independence)**.** *For each* i*,* $\delta_{i,t}$ *is independent across* t *with* $\mathbb{E}[\delta_{i,t}] = 0$ *and* $\mathrm{Var}(\delta_{i,t}) = \sigma_\delta^2 < \infty$.

B.2 Portfolio formation

Each month, we sort bonds by their observed signal $\hat{s}_{i,t}$ and form long-short factors from the extreme tails. Let α denote the tail fraction, that is, the share of bonds assigned to each extreme portfolio (e.g., $\alpha = 0.10$ for decile sorts, $\alpha = 0.20$ for quintile sorts). The long portfolio L contains bonds in the top α fraction of the signal distribution; the short portfolio S contains bonds in the bottom α fraction:

$$L \equiv \{\hat{s}_{i,t} \geq c_L\}, \qquad S \equiv \{\hat{s}_{i,t} \leq c_S\},$$

where c_L and c_S are the corresponding quantile cutoffs. The population long-short return is

$$\hat{r}_{t+1}^{LS} \equiv \mathbb{E}[\hat{r}_{i,t+1} \mid L] - \mathbb{E}[\hat{r}_{i,t+1} \mid S]. \tag{A.21}$$

Throughout, expectations are with respect to the cross-section at time t.

B.3 General mechanism: Why the bias does not cancel

B.3.1 Long-short decomposition

Lemma B.1 (Long–short decomposition)**.** *For any sorting rule and any return decomposition* $\hat{r} = r + \epsilon$*,*

$$\hat{r}_{t+1}^{LS} = \underbrace{\mathbb{E}[r_{i,t+1} \mid L] - \mathbb{E}[r_{i,t+1} \mid S]}_{r_{t+1}^{LS}} + \underbrace{\mathbb{E}[\epsilon_{i,t} \mid L] - \mathbb{E}[\epsilon_{i,t} \mid S]}_{\mathrm{Bias}_{t+1}}. \tag{A.22}$$

Proof. Substitute $\hat{r}_{i,t+1} = r_{i,t+1} + \epsilon_{i,t}$ into Eq. (A.21) and rearrange terms. □

Remark B.1 (When does the bias cancel out?)**.** *By Eq.* (A.22)*, the bias cancels out if and only if* $\mathbb{E}[\epsilon_{i,t} \mid L] = \mathbb{E}[\epsilon_{i,t} \mid S]$*. A sufficient (but not necessary) condition is* conditional mean independence*:*

$$\mathbb{E}[\epsilon_{i,t} \mid \hat{s}_{i,t}] = 0 \quad \textit{for all values of } \hat{s}_{i,t}.$$

The bias is zero only if portfolio assignment is mean-independent of the return error; price-based portfolio sorting violates this condition because it conditions on the very disturbance that enters the return denominator, so that long–short differencing cannot eliminate the bias.

B.3.2 Sorting on the bias

The next result formalizes the key intuition: for price-based signals, portfolio assignment depends on the same disturbance that enters the return measurement error, hence long and short portfolios have different conditional mean return errors.

Assumption B.2 (Overlap between signal noise and return error)**.** *There exists a scalar disturbance* $\delta_{i,t}$ *such that the signal error and the return measurement error can be written as*

$$\eta_{i,t} = g(\delta_{i,t}), \qquad \epsilon_{i,t} = h(\delta_{i,t}) + u_{i,t},$$

where g *and* h *are non-constant functions, and* $\mathbb{E}[u_{i,t} \mid \delta_{i,t}] = 0$.

Assumption B.3 (Monotone contamination)**.** *The function* g *is (weakly) monotone on the support of* $\delta_{i,t}$*.*

For the price-based signals in our 'factor zoo' (yields, spreads, value ratios, past returns), the signal error is a monotone function of the price disturbance; see Section 3 for discussion.

Assumption B.4 (Signal-noise independence)**.** *The true signal* $s_{i,t}$ *is independent of the price measurement error* $\delta_{i,t}$*.*

This assumption holds when the true characteristic is determined independently of the trading process that generates measurement error. It may be violated if bonds with extreme true signals (e.g., very high yields) are also more illiquid and therefore have larger $|\delta_{i,t}|$; such dependence through the variance channel could amplify or dampen the bias beyond what Proposition B.1 predicts.

Proposition B.1 (Sorting on the bias: sign result)**.** *Under Assumptions B.1–B.4, if $h(\delta)$ is (weakly) decreasing and $g(\delta)$ is (weakly) decreasing, then*

$$\mathbb{E}[\epsilon_{i,t} \mid L] \;\geq\; \mathbb{E}[\epsilon_{i,t}] \;\geq\; \mathbb{E}[\epsilon_{i,t} \mid S],$$

with strict inequalities whenever the sorting is non-trivial. In particular, the long–short bias is nonnegative:

$$\text{Bias}_{t+1} = \mathbb{E}[\epsilon_{i,t} \mid L] - \mathbb{E}[\epsilon_{i,t} \mid S] \;\geq\; 0.$$

Proof. By Definition B.2, $\hat{s}_{i,t} = s_{i,t} + g(\delta_{i,t})$. Under Assumption B.4, $s_{i,t} \perp \delta_{i,t}$. Holding $s_{i,t}$ fixed, since g is decreasing, higher values of $\hat{s}_{i,t}$ correspond to (stochastically) lower values of $\delta_{i,t}$, and lower values correspond to (stochastically) higher values of $\delta_{i,t}$. Thus, relative to the unconditional distribution, the event L (upper tail) tilts the distribution of $\delta_{i,t}$ downward, while S (lower tail) tilts it upward.

Under Assumption B.2, $\mathbb{E}[\epsilon_{i,t} \mid \delta_{i,t}] = h(\delta_{i,t})$. Since h is weakly decreasing, downward shifts in $\delta_{i,t}$ weakly increase $\mathbb{E}[\epsilon_{i,t} \mid \cdot]$ and upward shifts weakly decrease it. Therefore,

$$\mathbb{E}[\epsilon_{i,t} \mid L] \geq \mathbb{E}[\epsilon_{i,t}] \geq \mathbb{E}[\epsilon_{i,t} \mid S],$$

and the bias is nonnegative. Serial independence (Assumption B.1) justifies interpreting the bias-relevant component of $\epsilon_{i,t}$ as driven by the time-t disturbance (e.g., in the first-order approximation $\epsilon_{i,t} \approx \delta_{i,t+1} - \delta_{i,t}$, the $\delta_{i,t+1}$ term has mean zero conditional on time-t sorting). □

Remark B.2 (No need for true predictability)**.** *Proposition B.1 does not require the true signal $s_{i,t}$ to predict true returns. Even in a pure-noise world where $r_{i,t+1} \equiv 0$ (so $r^{LS}_{t+1} = 0$), the bias term in Lemma B.1 can be nonzero because sorting is conditioning on the disturbance that enters $\epsilon_{i,t}$.*

If a disturbance $\delta_{i,t}$ enters both the signal error $\eta_{i,t}$ and the return error $\epsilon_{i,t}$, then portfolio assignment based on the noisy signal $\hat{s}_{i,t}$ conditions on a component of $\epsilon_{i,t}$. As a result, $\mathbb{E}[\epsilon_{i,t} \mid L] \neq \mathbb{E}[\epsilon_{i,t} \mid S]$ and the bias does not cancel out. This holds even when the true signal $s_{i,t}$ has no predictive power.

For non-price signals (e.g., rating), the measurement error does not enter the return measurement error, so sorting on a noisy non-price signal does not create CEIV bias.

B.3.3 Non-price signals do not generate CEIV bias

Let a *non-price* characteristic (e.g., rating) be observed with error:

$$\hat{s}_{i,t} = s_{i,t} + \xi_{i,t},$$

where $\xi_{i,t}$ is a measurement error unrelated to return measurement.

Corollary B.1 (No CEIV bias for non-price signal noise)**.** *Suppose portfolio assignment depends on* $(s_{i,t}, \xi_{i,t})$ *only through* $\hat{s}_{i,t} = s_{i,t} + \xi_{i,t}$*. If*

$$\mathbb{E}[\epsilon_{i,t} \mid s_{i,t}, \xi_{i,t}] = 0, \tag{A.23}$$

then $\mathbb{E}[\epsilon_{i,t} \mid L] = \mathbb{E}[\epsilon_{i,t} \mid S] = 0$*, hence* $\text{Bias}_{t+1} = 0$ *and* $\hat{r}^{LS}_{t+1} = r^{LS}_{t+1}$*.*

Proof. Using the law of iterated expectations and (A.23),

$$\mathbb{E}[\epsilon_{i,t} \mid L] = \mathbb{E}\big[\mathbb{E}[\epsilon_{i,t} \mid s_{i,t}, \xi_{i,t}] \mid L\big] = \mathbb{E}[0 \mid L] = 0,$$

and similarly for S. □

Remark B.3 ("no CEIV bias" does not mean "measurement error is harmless")**.** *Corollary B.1 rules out the* mechanical *long–short bias that arises when the same disturbance contaminates both the sorting variable and the return denominator. It does* not *imply that non-price characteristics measured with error are innocuous: misclassification can attenuate* r^{LS}*, change portfolio composition and risk exposures, and can affect inference. Noisy non-price signals can lead to misranking but they do not create a spurious long-short premium unless the signal noise also contaminates the return measurement error.*

B.4 Closed-form characterization under stricter assumptions

Proposition B.1 establishes the sign of the bias under general conditions, and Corollary B.1 shows that non-price signals are immune. We now impose distributional assumptions to derive the bias magnitude in closed form.

Assumption B.5 (Linear price contamination)**.** *For price-based signals,* $\eta_{i,t} = a\,\delta_{i,t}$ *for some constant* $a \neq 0$*.*

Assumption B.6 (Joint normality and orthogonality)**.** *The vector* $(s_{i,t}, \eta_{i,t}, \epsilon_{i,t})$ *is jointly normal and* $s_{i,t} \perp \eta_{i,t}$*. Since* $\eta_{i,t} = a\,\delta_{i,t}$ *under Assumption B.5, normality of* $\eta_{i,t}$ *implies normality of* $\delta_{i,t}$*.*

Proposition B.2 (Correlation structure). *Under Assumptions B.1 and B.5, and using the first-order approximation $\epsilon_{i,t} \approx \delta_{i,t+1} - \delta_{i,t}$, we have*

$$\mathrm{Cov}(\eta_{i,t}, \epsilon_{i,t}) = -a\sigma_\delta^2 \neq 0. \tag{A.24}$$

Proof. Under Assumption B.1, $\delta_{i,t+1}$ is independent of $\delta_{i,t}$ and has mean zero conditional on time-t sorting. Thus, the bias-relevant component of $\epsilon_{i,t}$ is $-\delta_{i,t}$, so

$$\mathrm{Cov}(\eta_{i,t}, \epsilon_{i,t}) \approx \mathrm{Cov}(a\delta_{i,t}, -\delta_{i,t}) = -a\ \mathrm{Var}(\delta_{i,t}) = -a\sigma_\delta^2.$$

□

Define $\sigma_s^2 \equiv \mathrm{Var}(s_{i,t})$, $\sigma_\eta^2 \equiv \mathrm{Var}(\eta_{i,t})$, and $\sigma_{\hat{s}}^2 \equiv \mathrm{Var}(\hat{s}_{i,t}) = \sigma_s^2 + \sigma_\eta^2$ (by $s \perp \eta$ in Assumption B.6). Let the *noise share* of the observed signal be

$$\rho \equiv \frac{\sigma_\eta}{\sigma_{\hat{s}}} \in [0, 1].$$

Note that ρ denotes the noise share (a ratio of standard deviations), not a correlation coefficient. Under Assumption B.5 with $\sigma_\eta = |a|\sigma_\delta$, the bias in Eq. (A.26) can equivalently be written as $2\kappa|a|\sigma_\delta^2/\sigma_{\hat{s}}$. We describe the bias as first-order in σ_δ in the sense that it scales linearly with price noise for a fixed noise share ρ. Finally, define a positive tail constant

$$\kappa \equiv -\mathbb{E}\big[Z \mid Z < \Phi^{-1}(\alpha)\big] > 0, \qquad Z \sim \mathcal{N}(0, 1). \tag{A.25}$$

Theorem B.1 (Sorting-induced bias: closed form). *Under Assumptions B.1, B.4, B.5, and B.6, the bias in a long–short factor sorted on a price-based signal is*

$$\mathrm{Bias}_{t+1} = 2\kappa\, \rho\, \sigma_\delta, \tag{A.26}$$

where κ is defined in Eq. (A.25) and σ_δ is the previously-defined standard deviation of the price noise.

Proof. Under joint normality, the conditional expectation of η given $\hat{s}$ is linear:

$$\mathbb{E}[\eta_{i,t} \mid \hat{s}_{i,t}] = \frac{\mathrm{Cov}(\eta_{i,t}, \hat{s}_{i,t})}{\mathrm{Var}(\hat{s}_{i,t})}\, \hat{s}_{i,t} = \frac{\sigma_\eta^2}{\sigma_{\hat{s}}^2}\, \hat{s}_{i,t},$$

since $\hat{s} = s + \eta$ and $s \perp \eta$.

For the long leg $L = \{\hat{s} \geq c_L\}$, where c_L is the upper α-quantile, standard normal truncation implies

$$\mathbb{E}[\hat{s}_{i,t} \mid L] = +\kappa\, \sigma_{\hat{s}}.$$

Hence,

$$\mathbb{E}[\eta_{i,t} \mid L] = \frac{\sigma_\eta^2}{\sigma_{\hat{s}}^2}\,\mathbb{E}[\hat{s}_{i,t} \mid L] = +\kappa\,\frac{\sigma_\eta^2}{\sigma_{\hat{s}}} = +\kappa\,\rho\,\sigma_\eta.$$

Under Assumption B.5, $\eta = a\delta$ and $\sigma_\eta = |a|\sigma_\delta$. Without loss of generality, assume $a < 0$ (as for our price-based signals); if $a > 0$, redefine the signal as $-\hat{s}_{i,t}$, which swaps L and S and reverses the sign of a without changing the bias magnitude. Then

$$\mathbb{E}[\delta_{i,t} \mid L] = \frac{1}{a}\mathbb{E}[\eta_{i,t} \mid L] = -\kappa\,\rho\,\sigma_\delta,$$

and by symmetry $\mathbb{E}[\delta_{i,t} \mid S] = +\kappa\,\rho\,\sigma_\delta$.

Finally, the first-order approximation (Definition B.3) gives $\epsilon_{i,t} \approx \delta_{i,t+1} - \delta_{i,t}$. Under Assumption B.1, $\delta_{i,t+1}$ is independent of time-t information, so $\mathbb{E}[\delta_{i,t+1} \mid L] = \mathbb{E}[\delta_{i,t+1}] = 0$. Hence $\mathbb{E}[\epsilon_{i,t} \mid L] \approx -\mathbb{E}[\delta_{i,t} \mid L]$, and similarly for S. Therefore,

$$\mathbb{E}[\epsilon_{i,t} \mid L] - \mathbb{E}[\epsilon_{i,t} \mid S] \approx -\mathbb{E}[\delta_{i,t} \mid L] + \mathbb{E}[\delta_{i,t} \mid S] = 2\kappa\,\rho\,\sigma_\delta,$$

which yields the expression in Eq. (A.26). The second-order remainder, approximately $\delta_{i,t}^2 - \delta_{i,t}\delta_{i,t+1}$, has positive conditional expectation in both tails but cancels exactly in the long–short difference under normality, because the symmetric truncation of $\delta_{i,t}$ induced by the long and short portfolios yields identical conditional second moments; under asymmetric noise distributions, a second-order residual may remain. □

Corollary B.2 (Pure-noise limit)**.** *If the true signal has no cross-sectional variation, $\sigma_s = 0$ (equivalently, $\rho = 1$), then*

$$\text{Bias}_{t+1} = 2\kappa\,\sigma_\delta.$$

Corollary B.3 (Reversal signals)**.** *Under Assumptions B.1, B.4, and B.6, for reversal signals where $\eta_{i,t} = \delta_{i,t} - \delta_{i,t-1}$, the bias is $|\text{Bias}| = \sqrt{2}\kappa\rho\sigma_\delta$.*

Proof. The reversal structure $\eta_{i,t} = \delta_{i,t} - \delta_{i,t-1}$ does not satisfy the linear form $\eta = a\delta_t$ in Assumption B.5; we therefore compute $\text{Cov}(\delta_{i,t}, \hat{s}_{i,t})$ directly.

Under Assumption B.1, the bias-relevant component of the return measurement error is $\epsilon_{i,t} \approx -\delta_{i,t}$. For the reversal signal $\eta_{i,t} = \delta_{i,t} - \delta_{i,t-1}$, serial independence implies $\text{Cov}(\delta_{i,t}, \delta_{i,t-1}) = 0$, so that

$$\text{Var}(\eta_{i,t}) = \text{Var}(\delta_{i,t}) + \text{Var}(\delta_{i,t-1}) = 2\sigma_\delta^2, \qquad \text{Cov}(\delta_{i,t}, \eta_{i,t}) = \sigma_\delta^2.$$

Under joint normality (Assumption B.6),

$$\mathbb{E}[\delta_{i,t} \mid \hat{s}_{i,t}] = \frac{\mathrm{Cov}(\delta_{i,t}, \hat{s}_{i,t})}{\mathrm{Var}(\hat{s}_{i,t})}\, \hat{s}_{i,t} = \frac{\sigma_\delta^2}{\sigma_{\hat{s}}^2}\, \hat{s}_{i,t},$$

where $\mathrm{Cov}(\delta_{i,t}, \hat{s}_{i,t}) = \mathrm{Cov}(\delta_{i,t}, \eta_{i,t}) = \sigma_\delta^2$ (using $s_{i,t} \perp \delta_{i,t}$ and Assumption B.1) and $\mathrm{Var}(\hat{s}_{i,t}) = \sigma_s^2 + \sigma_\eta^2$ (using $s_{i,t} \perp \eta_{i,t}$ from Assumption B.6). For the long leg, $\mathbb{E}[\hat{s}_{i,t} \mid L] = +\kappa\sigma_{\hat{s}}$, hence

$$\mathbb{E}[\delta_{i,t} \mid L] = +\kappa\frac{\sigma_\delta^2}{\sigma_{\hat{s}}} = +\kappa\frac{\rho}{\sqrt{2}}\sigma_\delta,$$

where $\rho = \sigma_\eta/\sigma_{\hat{s}}$ and $\sigma_\eta = \sqrt{2}\sigma_\delta$. By symmetry, $\mathbb{E}[\delta_{i,t} \mid S] = -\kappa\frac{\rho}{\sqrt{2}}\sigma_\delta$. The bias is

$$\mathrm{Bias} = \mathbb{E}[\epsilon \mid L] - \mathbb{E}[\epsilon \mid S] \approx -\mathbb{E}[\delta \mid L] + \mathbb{E}[\delta \mid S] = -\sqrt{2}\kappa\rho\sigma_\delta,$$

so $|\mathrm{Bias}| = \sqrt{2}\kappa\rho\sigma_\delta$. □

B.5 Empirical magnitude of the implementation gap

Table B.1 quantifies the latent implementation bias across all 108 factors. For each factor, we compute the difference $\Delta_{f,t} = r_{f,t}^{\mathrm{End}} - r_{f,t}^{\mathrm{Bgn}}$ between month-end and month-begin long–short factor returns, holding the signal fixed at its month-end value. A large and statistically significant Δ indicates that the one-day implementation gap materially affects the measured premium.

The results are split sharply by signal type. For non-price signals (credit ratings, momentum, factor betas), the signal noise $\xi_{i,t}$ is independent of the price disturbance $\delta_{i,t}$, so $\mathrm{Cov}(\xi_{i,t}, \epsilon_{i,t}) = 0$ and Corollary B.1 implies zero CEIV bias. Among the 30 price-based factors, 43–67% exhibit a statistically significant gap across the four sort specifications (single-sort and within-firm, value- and equal-weighted), with average absolute differences of 0.13–0.30% per month and average absolute t-statistics of 3.08–4.05. Among the 78 non-price factors, only 10–18% show a significant gap, with average absolute differences of 0.03–0.10% per month and average absolute t-statistics near one. The one-day implementation gap has no material effect on average premia for non-price signals, consistent with Corollary B.1.

Table B.1: Latent implementation bias across 108 factors.

The table tests whether the one-day implementation gap produces a significant return difference for each of the 108 factors. For each factor and portfolio sort, $\Delta_{f,t} = r_{f,t}^{\text{End}} - r_{f,t}^{\text{Bgn}}$ measures the difference between month-end and month-begin long–short returns when both series use signal-adjusted sorting variables. Sig./Total counts factors with $|t(\Delta)| > 1.96$ (Newey-West). $\overline{|\Delta|}$ is the average absolute monthly return difference (%). $\overline{|t|}$ is the average absolute t-statistic. Sample: 2002-09 to 2024-12.

	Price-Based				**Non-Price-Based**			
Sort	**Sig./Total**	**%**	$\overline{\lvert\Delta\rvert}$	$\overline{\lvert t\rvert}$	**Sig./Total**	**%**	$\overline{\lvert\Delta\rvert}$	$\overline{\lvert t\rvert}$
Single-Sorted (VW)	13/30	43%	0.18	3.08	11/78	14%	0.05	1.10
Single-Sorted (EW)	15/30	50%	0.30	3.18	14/78	18%	0.10	1.20
Within-Firm (VW)	14/30	47%	0.13	3.32	8/78	10%	0.03	1.07
Within-Firm (EW)	20/30	67%	0.19	4.05	11/78	14%	0.04	1.26

Internet Appendix for

The Corporate Bond Factor Replication Crisis

Abstract

This Internet Appendix provides additional information, tables, figures, and empirical results supporting the main text.

IA.1 Descriptive statistics

Tables IA.I–IA.II report data availability and descriptive statistics for the daily TRACE data. Table IA.I documents the coverage of key variables across rating categories; Table IA.II presents pooled and cross-sectional summary statistics for prices, yields, spreads, duration, volume, and ratings.

Tables IA.III–IA.IV present the corresponding statistics for the monthly data. Table IA.III reports variable availability after resampling to monthly frequency; Table IA.IV reports return and characteristic distributions.

Tables IA.V–IA.VII document the prevalence and time concentration of extreme returns. Table IA.V reports tail percentiles, higher moments, and the frequency of returns exceeding various thresholds. Table IA.VI shows that extreme returns cluster during financial distress (2008–2009, 2020). Table IA.VII provides annual return summary statistics together with the within-year correlation between month-end and month-begin returns. More extensive data reports are publicly available at openbondassetpricing.com.

Table IA.I: Data Availability by Rating Category.

	Panel A: All		**Panel B: Inv. Grade**		**Panel C: Non-Inv. Grade**		**Panel D: Defaulted**	
Variable	Obs.	% Missing	Obs.	% Missing	Obs.	% Missing	Obs.	% Missing
Price (VW)	29,776,137	0.00	22,890,794	0.00	6,807,429	0.00	77,914	0.00
Price (Bid)	20,659,079	30.62	15,576,988	31.95	5,019,498	26.26	62,593	19.66
Price (Ask)	21,539,587	27.66	16,599,053	27.49	4,894,485	28.10	46,049	40.90
Spread	29,211,584	1.90	22,423,512	2.04	6,710,255	1.43	77,817	0.12
Rating (SP)	28,916,834	2.89	22,288,333	2.63	6,554,745	3.71	73,756	5.34
Rating (MD)	28,720,454	3.55	22,154,977	3.21	6,518,473	4.24	47,004	39.67
PERMNO	26,805,234	9.98	21,368,128	6.65	5,370,695	21.11	66,411	14.76

The table reports data availability for key variables across rating categories. For each panel, we report the number of non-missing observations and the percentage of missing values. Panel A includes all bonds in the sample. Panel B includes investment grade bonds (S&P ratings 1–10, AAA to BBB−). Panel C includes non-investment grade bonds (S&P ratings 11–21, BB+ to CCC−). Panel D includes defaulted bonds (S&P rating 22, D). The sample spans the period 2002-07-01 to 2025-03-31. All other variables in the dataset (not shown) have zero missing observations. Accrued interest is computed assuming a 2-day settlement period based on the modified following rule using the `QuantLib` Python package. Spreads (credit spreads) are computed by interpolating over `bond_maturity` using the constant maturity zero-coupon yield curve data at key rates, made available from Liu and Wu (2021)'s `website`.

Table IA.II: TRACE daily descriptive statistics.

Variable	Mean	Median	SD	P1	P5	P95	P99
Panel A: Pooled							
Price (VW)	100.89	101.24	13.45	55.89	79.44	120.41	135.85
Price (EW)	100.90	101.26	13.44	55.82	79.46	120.37	135.80
Price (ParW)	100.89	101.24	13.45	55.88	79.44	120.41	135.85
Price (Bid)	100.31	100.94	13.60	53.04	78.41	119.49	134.68
Price (Ask)	101.01	101.28	12.70	59.08	80.84	119.12	134.70
Price (Full)	102.10	102.34	13.58	57.33	80.51	122.06	137.61
YTM	5.20	4.80	4.83	0.63	1.37	9.71	20.04
Spread	2.53	1.46	4.76	0.18	0.39	7.03	17.85
Duration (Macaulay)	6.40	5.26	4.24	1.07	1.47	15.26	17.94
Duration (Modified)	6.25	5.12	4.15	1.04	1.44	14.92	17.61
Bond Maturity	9.27	6.23	8.72	1.11	1.55	27.70	30.14
Bond Age	4.49	3.21	4.60	0.05	0.27	13.80	23.15
Convexity	76.06	31.68	105.57	1.61	2.84	317.64	433.57
Volume (Dollar)	3.07	0.45	10.64	0.00	0.01	13.64	39.22
Volume (Par)	3.08	0.45	10.67	0.00	0.01	13.69	39.43
Bid Count	2.47	1.00	7.66	1.00	1.00	7.00	14.00
Ask Count	2.54	1.00	13.66	0.00	0.00	8.00	22.00
Rating (SP)	8.66	8.00	3.60	1.00	3.00	16.00	18.00
Rating (MD)	8.73	8.00	3.76	1.00	3.00	16.00	18.00
Panel B: Cross-sectional							
Price (VW)	100.97	101.50	11.79	62.90	82.85	117.79	128.65
Price (EW)	100.98	101.52	11.78	62.90	82.88	117.75	128.60
Price (ParW)	100.96	101.50	11.79	62.89	82.84	117.79	128.65
Price (Bid)	100.29	101.11	11.99	60.52	81.41	116.68	127.48
Price (Ask)	101.33	101.78	11.18	64.66	84.39	117.07	128.30
Price (Full)	102.24	102.67	11.95	64.39	84.06	119.63	130.67
YTM	5.51	4.72	3.81	2.17	2.58	10.58	22.80
Spread	2.77	1.65	3.78	0.26	0.51	7.99	20.51
Duration (Macaulay)	6.27	5.22	4.02	1.07	1.48	14.55	16.25
Duration (Modified)	6.11	5.08	3.92	1.04	1.44	14.21	15.88
Bond Maturity	9.18	6.25	8.51	1.11	1.55	27.24	31.39
Bond Age	4.37	3.17	4.32	0.06	0.28	13.34	20.08
Convexity	72.43	31.40	96.09	1.61	2.83	298.52	373.45
Volume (Dollar)	3.06	0.41	10.02	0.00	0.01	14.01	39.48
Volume (Par)	3.07	0.41	10.09	0.00	0.01	14.04	39.59
Bid Count	2.38	1.27	4.04	1.00	1.00	6.58	13.73
Ask Count	2.44	1.07	7.28	0.00	0.00	8.18	21.37
Rating (SP)	8.59	8.04	3.71	1.57	2.88	15.72	18.45
Rating (MD)	8.65	8.16	3.85	1.26	3.16	16.06	18.49

The table presents descriptive statistics for the cleaned TRACE daily dataset. Panel A shows statistics pooled across all cusip-date observations. Panel B shows time-series averages of daily cross-sectional statistics. The sample spans the period 2002-07-01 to 2025-03-31. All prices are in percentage of par, 100% implies a dollar value of $1000. All par values are exactly $1000. Yield to maturity (`ytm`) and Spread (`credit_spread`) are in percentage points. Duration, Bond Maturity, and Age are in years. Volumes (`dvolume` and `qvolume`) are in millions of U.S. dollars. Ratings are in numeric format (AAA = 1, ..., D = 22). Variables `ytm`, `credit_spread`, `dvolume`, and `qvolume` are winsorized at the 0.5th and 99.5th percentiles within each date. Accrued interest is computed assuming a 2-day settlement period based on the modified following rule using the `QuantLib` Python package. Spreads (credit spreads) are computed by interpolating over `bond_maturity` using the constant maturity zero-coupon yield curve data at key rates, made available from Liu and Wu (2021)'s `website`.

Table IA.III: Monthly Panel Data Availability by Rating Category.

	Panel A: All		Panel B: Inv. Grade		Panel C: Non-Inv. Grade		Panel D: Defaulted	
Variable	Obs.	% Missing	Obs.	% Missing	Obs.	% Missing	Obs.	% Missing
Total	2,662,981	–	1,436,167	–	418,304	–	5,075	–
Price (VW)	1,794,196	32.62	1,381,521	3.80	407,798	2.51	4,877	3.90
Month-End Return	1,859,498	30.17	1,436,119	0.00	418,304	0.00	5,075	0.00
Month-Begin Return	1,715,302	35.59	1,315,666	8.39	394,918	5.59	4,718	7.03
YTM	1,794,165	32.63	1,381,490	3.81	407,798	2.51	4,877	3.90
Spread	1,761,829	33.84	1,354,456	5.69	402,502	3.78	4,871	4.02
Composite Rating (SP)	1,859,546	30.17	1,436,167	0.00	418,304	0.00	5,075	0.00
Composite Rating (MD)	1,859,546	30.17	1,436,167	0.00	418,304	0.00	5,075	0.00
PERMNO	1,657,912	37.74	1,331,050	7.32	322,548	22.89	4,314	15.00

The table reports data availability for key variables in the monthly panel across rating categories. For each panel, we report the number of non-missing observations and the percentage of missing values after resampling each bond to a contiguous monthly time series. Panel A includes all bonds in the sample. Panel B includes investment grade bonds (S&P ratings 1–10, AAA to BBB−). Panel C includes non-investment grade bonds (S&P ratings 11–21, BB+ to CCC−). Panel D includes defaulted bonds (S&P rating 22, D). The sample spans the period 2002-08-31 to 2025-03-31.

Table IA.IV: Monthly Panel Descriptive Statistics — All Corporate Bonds.

Variable	Mean	Median	SD	P1	P5	P95	P99
Panel A: Pooled							
Total End Return (%)	0.47	0.37	4.98	−10.38	−4.18	5.09	11.59
Total Begin Return (%)	0.40	0.34	4.51	−9.77	−4.02	4.77	10.61
Dur. Adj. End Return (%)	0.29	0.17	4.90	−10.97	−3.56	4.33	11.09
Dur. Adj. Begin Return (%)	0.23	0.14	4.45	−10.46	−3.46	4.05	10.30
Latent Imp. Bias	0.00	0.00	0.02	−0.04	−0.02	0.02	0.04
End Holding Period	21.83	22.00	1.46	19.00	20.00	25.00	26.00
Begin Holding Period	18.95	19.00	1.92	13.00	15.00	22.00	22.00
Implementation Gap	1.50	1.00	0.95	1.00	1.00	4.00	5.00
Signal Gap	1.68	1.00	1.39	1.00	1.00	5.00	8.00
YTM (%)	5.21	4.80	4.82	0.64	1.38	9.74	20.35
Spread (%)	2.54	1.46	4.75	0.18	0.40	7.04	18.09
Bond Maturity	9.60	6.38	8.94	1.11	1.54	27.71	30.13
Bond Age	4.76	3.41	4.75	0.15	0.39	14.58	23.66
Market Cap.	677.21	507.99	685.92	4.52	15.06	1961.85	3321.84
Composite Rating (SP)	8.66	8.00	3.62	1.00	3.00	16.00	18.00
Composite Rating (MD)	8.76	8.00	3.77	1.00	3.00	16.00	19.00
Bid-Ask Spread (%)	0.85	0.39	1.77	−1.15	−0.13	3.34	6.10
Panel B: Cross-sectional							
Total End Return (%)	0.54	0.42	3.78	−7.82	−3.59	5.03	10.41
Total Begin Return (%)	0.48	0.38	3.44	−7.53	−3.43	4.64	9.65
Dur. Adj. End Return (%)	0.31	0.19	3.72	−7.91	−3.66	4.63	10.08
Dur. Adj. Begin Return (%)	0.24	0.14	3.39	−7.64	−3.52	4.28	9.36
Latent Imp. Bias	0.00	0.00	0.02	−0.04	−0.02	0.02	0.05
End Holding Period	21.89	21.36	1.03	21.26	21.26	24.24	25.24
Begin Holding Period	18.86	19.53	1.52	13.82	15.66	19.96	19.97
Implementation Gap	1.54	1.03	0.96	1.00	1.00	3.76	4.98
Signal Gap	1.73	1.01	1.41	1.00	1.00	4.81	7.49
YTM (%)	5.54	4.77	3.79	2.18	2.60	10.54	22.98
Spread (%)	2.79	1.67	3.76	0.27	0.53	7.96	20.67
Bond Maturity	9.49	6.41	8.73	1.11	1.54	27.30	31.40
Bond Age	4.64	3.37	4.48	0.16	0.40	14.09	20.56
Market Cap.	636.30	456.70	651.74	5.28	29.41	1836.22	3204.10
Composite Rating (SP)	8.64	8.00	3.72	1.60	2.90	15.81	18.49
Composite Rating (MD)	8.73	8.24	3.85	1.34	3.22	16.16	18.63
Bid-Ask Spread (%)	0.99	0.55	1.55	−1.31	−0.16	3.56	6.14

The table presents descriptive statistics for the monthly corporate bond panel. The sample includes all corporate bonds (All ratings). The sample contains 1,859,498 bond-month observations with non-missing returns (100.00% of total). Panel A shows statistics pooled across all cusip-month observations. Panel B shows time-series averages of monthly cross-sectional statistics. The sample spans the period 2002-08-31 to 2025-03-31. All prices are in percentage of par. Yield to maturity, spread, and returns are in percentage points. Duration, bond maturity, and age are in years. Ratings are in numeric format (AAA = 1, ..., D = 22).

Table IA.V: Extreme Returns Analysis.

	End Return (%)				Begin Return (%)			
	All	IG	NIG	Def	All	IG	NIG	Def
Panel A: Tail Percentiles								
P0.01	−73.03	−40.55	−84.23	−99.81	−71.13	−38.42	−84.15	−99.82
P0.05	−48.06	−27.15	−65.36	−99.68	−45.98	−25.13	−61.97	−98.57
P0.10	−36.25	−20.49	−55.32	−98.45	−34.41	−18.86	−53.12	−96.20
P99.90	42.13	21.51	80.65	204.04	37.26	19.21	67.02	139.03
P99.95	61.05	27.80	118.77	413.99	54.15	25.30	95.29	157.63
P99.99	137.49	48.76	204.77	743.90	116.06	43.70	198.72	462.42
Panel B: Higher Moments								
Skewness	14.74	11.35	8.03	6.94	9.12	1.85	7.62	2.67
Excess Kurtosis	1369.69	1759.45	259.29	160.86	562.36	148.98	253.43	47.23
Panel C: Extreme Return Counts								
<−20%	6,129	1,527	4,007	590	5,006	1,141	3,367	496
>+20%	6,673	1,710	4,272	690	5,208	1,191	3,442	575
<−50%	845	71	578	196	679	54	469	156
>+50%	1,357	138	1,017	202	988	109	721	158
<−100%	0	0	0	0	0	0	0	0
>+100%	353	22	289	42	220	12	181	27
N(>+500%)	4	1	1	2	1	0	0	1

The table presents extreme returns analysis for the monthly corporate bond panel. Values are pooled over all bonds and months within each rating category. Panel A reports tail percentiles of the return distribution (in %). Panel B reports skewness and excess kurtosis. Panel C reports the count of observations exceeding the specified thresholds by direction, including returns exceeding +500%. Observations: All: 1,859,546; IG: 1,436,167; NIG: 418,304; Def: 5,075. Sample period: 2002-08-31 to 2025-03-31.

Table IA.VI: Time Concentration of Extreme Returns.

	End Return					Begin Return				
Year	N	<−20%	>+20%	<−95%	>+95%	N	<−20%	>+20%	<−95%	>+95%
2002	20,964	229	432	1	7	18,592	184	324	1	6
2003	57,134	120	432	1	16	50,403	84	311	0	8
2004	57,853	56	90	0	2	50,537	39	61	0	1
2005	57,402	81	49	0	0	49,764	70	37	0	0
2006	58,115	23	94	0	0	50,200	18	64	0	0
2007	52,255	64	27	0	0	44,435	49	19	0	0
2008	49,163	1,876	1,030	10	180	42,178	1,408	787	10	144
2009	59,089	963	2,269	0	115	52,597	717	1,669	0	45
2010	66,827	42	97	0	3	60,941	35	68	0	2
2011	67,390	91	52	0	1	61,281	65	61	0	0
2012	69,488	41	55	0	0	63,705	33	48	0	0
2013	73,049	29	47	0	0	67,957	25	34	0	0
2014	78,128	89	22	0	0	72,334	72	17	0	0
2015	86,659	417	93	0	3	80,149	383	90	0	2
2016	90,477	255	657	1	21	83,988	225	532	0	15
2017	94,378	44	39	0	0	88,076	35	28	0	0
2018	99,230	66	27	0	1	92,777	61	27	0	0
2019	101,889	136	69	1	0	96,124	117	63	1	0
2020	107,730	1,266	803	1	26	101,557	1,179	720	1	18
2021	113,401	20	52	0	5	107,116	19	40	0	4
2022	116,184	101	58	0	0	110,350	86	53	0	2
2023	120,518	79	91	0	3	114,628	71	71	0	3
2024	129,264	29	82	0	2	124,097	25	77	0	0
2025	32,911	7	5	0	0	31,516	4	7	0	0
Total	1,859,498	6,124	6,672	15	385	1,715,302	5,004	5,208	13	250

The table reports the annual frequency of extreme monthly returns (absolute returns exceeding 20%, 95%) for the monthly corporate bond panel. For each year, we report the number of non-missing observations (N) for each return type and the count of returns falling below or above the specified thresholds. Sample period: 2002-08-31 to 2025-03-31.

Table IA.VII: Annual Return Summary Statistics.

	End Return (%)						Begin Return (%)						
Year	N	Mean	SD	Med.	Min	Max	N	Mean	SD	Med.	Min	Max	ρ
2002	20,964	1.56	8.19	1.11	−96.76	163.17	18,592	1.38	7.73	1.02	−96.30	153.49	0.78
2003	57,134	1.19	5.79	0.75	−98.57	290.90	50,403	1.20	5.28	0.89	−93.06	519.51	0.64
2004	57,853	0.57	4.78	0.54	−86.30	652.82	50,537	0.58	2.89	0.56	−86.46	102.51	0.51
2005	57,402	−0.03	3.27	0.21	−66.11	42.60	49,764	−0.01	3.20	0.18	−62.95	67.63	0.85
2006	58,115	0.82	2.87	0.56	−75.02	48.62	50,200	0.70	2.76	0.47	−74.98	49.01	0.84
2007	52,255	0.20	3.00	0.40	−72.62	94.48	44,435	0.15	2.86	0.35	−69.59	71.08	0.85
2008	49,163	−0.45	14.91	−0.01	−99.81	299.56	42,178	−0.37	14.63	−0.06	−99.83	339.15	0.91
2009	59,089	2.91	12.26	1.42	−92.68	900.08	52,597	2.48	9.71	1.32	−82.82	254.10	0.80
2010	66,827	0.91	3.78	0.67	−91.77	539.31	60,941	0.85	3.16	0.65	−90.30	222.54	0.75
2011	67,390	0.60	3.23	0.52	−93.85	101.88	61,281	0.45	3.15	0.42	−87.71	80.71	0.86
2012	69,488	0.89	2.61	0.59	−59.35	88.15	63,705	0.80	2.47	0.52	−57.97	69.30	0.85
2013	73,049	0.11	2.50	0.23	−54.55	82.88	67,957	0.15	2.36	0.24	−45.54	78.23	0.89
2014	78,128	0.45	2.24	0.37	−64.59	72.18	72,334	0.46	2.12	0.35	−64.67	64.52	0.89
2015	86,659	−0.19	3.79	0.07	−93.57	124.60	80,149	−0.20	3.67	0.07	−94.65	119.51	0.94
2016	90,477	0.84	5.25	0.39	−96.76	227.70	83,988	0.80	4.80	0.38	−91.70	178.95	0.91
2017	94,378	0.53	1.94	0.37	−78.29	87.47	88,076	0.47	1.79	0.35	−77.95	69.47	0.86
2018	99,230	−0.22	2.14	−0.02	−60.31	96.29	92,777	−0.14	2.03	0.03	−60.75	54.20	0.91
2019	101,889	1.12	2.78	0.74	−96.45	71.98	96,124	1.01	2.62	0.68	−96.02	61.23	0.92
2020	107,730	0.84	7.09	0.71	−96.99	367.66	101,557	0.76	6.91	0.67	−96.41	192.18	0.94
2021	113,401	0.09	2.44	0.04	−60.17	291.56	107,116	0.01	2.33	0.01	−60.12	281.40	0.95
2022	116,184	−1.16	4.06	−1.12	−82.74	75.52	110,350	−1.27	3.84	−1.28	−83.71	111.64	0.95
2023	120,518	0.80	4.04	0.41	−84.15	477.47	114,628	0.76	3.60	0.37	−78.55	398.48	0.95
2024	129,264	0.31	2.43	0.67	−63.66	121.45	124,097	0.23	2.37	0.53	−63.76	86.52	0.92
2025	32,911	0.64	1.60	0.59	−42.74	28.68	31,516	0.56	1.62	0.56	−28.44	39.29	0.91
Total	1,859,498	0.47	4.46	0.46	−99.81	900.08	1,715,302	0.40	4.08	0.40	−99.83	519.51	0.86

The table reports annual summary statistics for monthly returns (in %). For each year, we report the number of non-missing observations (N), mean, standard deviation (SD), median (Med.), minimum (Min), and maximum (Max) of returns. ρ is the within-year correlation between End and Begin returns using matched non-missing pairs. The Total row shows weighted averages for means, simple averages for SD and ρ, median of medians, and overall min/max. Sample period: 2002-08-31 to 2025-03-31.

IA.2 An open source corporate bond 'factor zoo'

We provide an open source data set of 108 corporate bond signals and their corresponding factors, publicly available at https://openbondassetpricing.com. The data set comprises two components. First, *bond-month-level data* contains all 108 signals along with bond identifiers, return variables, and control characteristics, enabling researchers to construct custom factors or apply alternative weighting schemes. Second, *pre-formed factors* for each signal are constructed with a 1-month holding period using decile sorts for the full sample (all bonds, All) and quintile sorts for the investment grade (IG) and non-investment grade (NIG) subsamples. We also form within-firm sorts, where for each firm with at least two bonds we split at the within-firm median signal, forming a long-short portfolio that is then aggregated across firms using market-value weights. Value-weighted and equal-weighted returns are provided for each sample (All, IG, NIG).

All price-based signals (yields, spreads, value measures, short-term reversal) are measurement-error adjusted by default, observed with a minimum 1-business-day gap before the month-end price used for return computation. This gap breaks the shared-price link documented in Section 3. Return-based signals (betas, momentum, reversals, VaR, Expected Shortfall) are computed in two versions: a standard version using excess returns ($r - r^f$) and a duration-adjusted version using Treasury-hedged returns ($r - r^{Tsy}$). Rolling window estimations use 36 months of data with a minimum of 12 observations. For long-term reversal signals requiring extended history, we use an expanding window that begins at 12 months and ramps up to the target horizon.

The 108 signals are organized into nine thematic clusters:

I. Spreads, Yields, Size (9 signals): Bond maturity, age, yield to maturity, credit spread, duration, convexity, size, spread changes, and average spreads.

II. Value (5 signals): Bond book-to-market ratio and fair-value deviations using the methodologies of Houweling and Van Zundert (2017) and Israel, Palhares, and Richardson (2018), with duration-times-spread adjustments.

III. Momentum & Reversal (22 signals): Standard momentum at 3-, 6-, 9-, and 12-month horizons; intermediate momentum; systematic and idiosyncratic momentum decompositions; industry momentum; long-term reversal at multiple horizons; industry long-term reversal; and short-term reversal.

IV. Illiquidity (14 signals): Price impact, Amihud illiquidity and its volatility, Roll spread, bid-ask spreads (relative, absolute, Corwin-Schultz, Abdi-Ranaldo), zero-return proportion, FHT spread, volatility of volume, and LIX liquidity.

V. Volatility & Risk (16 signals): Daily volatility, skewness, and kurtosis; realized volatility, signed jumps, skewness, and kurtosis; Value-at-Risk and Expected Shortfall; systematic and idiosyncratic volatility decompositions; idiosyncratic volatility from multiple factor models; and idiosyncratic skewness.

VI. Market Risk (9 signals): Equity and bond market betas from joint regressions; univariate bond market beta; duration-adjusted market and term betas; downside and upside market betas; term beta; and daily market beta.

VII. Credit & Default Betas (4 signals): Betas on downside risk, credit risk, liquidity risk, and default factors.

VIII. Volatility & Liquidity Betas (14 signals): VIX innovation betas from multiple specifications; downside and upside VIX betas; Pastor-Stambaugh liquidity betas; Amihud illiquidity betas; coskewness beta; VIX level beta; daily VIX beta; and aggregate illiquidity beta.

IX. Macro & Other Betas (15 signals): Betas on macro uncertainty (level and changes at multiple horizons, real and financial components); inflation and inflation volatility; credit spread changes and levels; intermediary capital; realized volatility and jump factors; yield curve level and spread; and economic policy uncertainty indices.

For all signals requiring a rolling window (betas, VaR, ES), we use a 36-month rolling window with a minimum of 12 observations, denoted 36(12). For long-term reversal signals requiring more than 12 months of history, we use an expanding window that starts at 12-3 and ramps up to the target horizon (e.g., 48-12 or 30-6) to preserve sample coverage. Nearly all signals have a start date of August 2002, coinciding with the launch of TRACE. For signals requiring a rolling window, we use monthly Intercontinental Exchange (ICE)/Bank of America Merrill Lynch (BAML) quote data prior to TRACE to construct the initial rolling estimates, enabling an August 2002 start date even for 36-month rolling betas. Table IA.VIII provides complete variable definitions and citations.

Table IA.VIII: Signal Definitions and Citations

Mnemonic	Name	Description	Citation
Bond Identifiers and Return Metrics			
`cusip`	Bond Identifier	9-digit CUSIP bond identifier.	–
`date`	Date	True month-end date (YYYY-MM-DD).	–
`issuer_cusip`	Issuer CUSIP	6-digit firm identifier (first 6 digits of CUSIP).	–
`permno`	CRSP PERMNO	CRSP permanent security identifier.	–
`permco`	CRSP PERMCO	CRSP permanent company identifier.	–
`gvkey`	Compustat GVKEY	Compustat global company key.	–
`hprd`	Holding Period	Month-end holding period in calendar days (see Fig. 2).	–
`lib`	Latent Implementation Bias	Clean price return from month-end to month-begin: LIB $= P_t^{\text{bgn}}/P_{t-1}^{\text{end}} - 1$ (see Fig. 2).	–
`libd`	LIB (Dirty)	LIB computed using dirty prices (includes accrued interest and coupon).	–
`ret_type`	Return Type	Return classification: `standard`, `trad_in_def`, or `default_evnt`.	–
`ff17num`	FF17 Industry	Fama-French 17-industry classification.	–
`ff30num`	FF30 Industry	Fama-French 30-industry classification.	–

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
mcap_s	Market Cap (Start)	Bond market capitalization at end of month $t-1$.	–
mcap_e	Market Cap (End)	Bond market capitalization at end of month t.	–
tret	Treasury Return	Duration-matched U.S. Treasury portfolio return. Used to compute duration-adjusted returns: $r^x = r - \texttt{tret}$.	–
ret_vw	Total Return (End)	Month-end to month-end total return (see Fig. 2).	–
ret_vw_bgn	Total Return (Begin)	Month-begin to month-end total return within the same month (see Fig. 2).	–
dt_s	Date Start	Trade date for month-end price in month $t-1$ (last 5 BD); refer to Panel A of Fig. 2.	–
dt_e	Date End	Trade date for month-end price in month t (last 5 BD); refer to Panel A of Fig. 2.	–
dt_s_bgn	Date Start (Begin)	Trade date for month-begin price in month t (first 5 BD); refer to Panel C of Fig. 2.	–
dt_e_bgn	Date End (Begin)	Trade date for month-end price in month t (last 5 BD) for begin returns; refer to Panel C of Fig. 2.	–
hprd_bgn	Holding Period (Begin)	Month-begin holding period in calendar days.	–
igap_bgn	Implementation Gap	Business days between month-end price ($t-1$) and month-begin price (t); capped at 5 BD. Refer to Panel D of Fig. 2.	–
sig_dt	Signal Date	Date when price-based signal was observed (minimum 1 BD before month-end price). Refer to Panel B of Fig. 2.	–
sig_gap	Signal Gap	Business days between signal observation and month-end price; ranges 1–10 BD. Refer to Panel B of Fig. 2.	–
rfret	Risk-Free Rate	Monthly risk-free rate from Fama-French. Used for excess returns: $r^x = r - r^f$.	–
Bond Characteristics			
spc_rat	S&P Composite Rating	Composite credit rating: S&P rating if available, otherwise Moody's rating. Scale: 1 (AAA) to 21 (CCC-), 22 = Default.	–

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
mdc_rat	Moody's Composite Rating	Composite credit rating: Moody's rating if available, otherwise S&P rating. Scale: 1 (AAA) to 20 (CCC-), 21 = Default.	–
call	Callable Indicator	Indicator for embedded call option (1 = callable, 0 = non-callable).	–
fce_val	Face Value	Bond amount outstanding (face value); units of the bond outstanding.	–
144a	Rule 144A Indicator	Dummy variable: 1 if bond is Rule 144A, 0 otherwise.	–
country	Country	Country of issuance (e.g., USA for U.S. bonds).	–
Cluster I: Spreads, Yields, Size			
tmat	Time to Maturity	Years remaining until bond maturity.	–
age	Bond Age	Years since bond issuance.	Kelly et al. (2023)
ytm	Yield to Maturity	Annualized yield to maturity. Computed with QuantLib.	Gebhardt et al. (2005)
cs	Credit Spread	Annualized credit spread: yield minus maturity-matched U.S. Treasury yield.	Kelly et al. (2023)
md_dur	Modified Duration	Modified duration measuring price sensitivity to yield changes. Computed with QuantLib.	Kelly et al. (2023)
convx	Convexity	Second-order price sensitivity to yield changes. Computed with QuantLib.	–
sze	Bond Size	Bond market capitalization: dirty price times amount outstanding ($ millions).	Houweling and Van Zundert (2017)
dcs6	6-Month Spread Change	Log change in credit spread over prior 6 months: $\log(cs_{t-6}) - \log(cs_t)$. If spread is missing exactly 6 months ago, the nearest adjacent spread is used within a ± 2 month band, with $+$ favored over $-$.	Kelly et al. (2023)
cs_mu12_1	12-Month Average Spread	Rolling 12-month average credit spread, skipping the prior month. Requires minimum 6 observations.	Elkamhi et al. (2024)
Cluster II: Value			
bbtm	Bond Book-to-Market	Par price divided by market price of the bond. All bonds have a par value of $1,000 (100%).	Bartram et al. (2025)

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
val_hz	Value (HZ)	Percentage deviation of observed credit spread from fitted "fair" spread: $(cs - \widehat{cs})/\widehat{cs}$. Controls: rating, industry, maturity, spread change, callable. Fitted spread is estimated each month via cross-sectional regression; because cs is used, this signal is also exposed to microstructure noise.	Houweling and Van Zundert (2017)
val_hz_dts	Value (HZ, DtS-adjusted)	HZ value signal demeaned within duration-times-spread quintiles to control for systematic spread-duration risk.	Houweling and Van Zundert (2017)
val_ipr	Value (IPR)	Log-spread residual from fair-value regression: $\log(cs) - X'\hat{\beta}$. Controls: rating, industry, log duration, volatility, callable. Fitted spread is estimated each month via cross-sectional regression; because cs is used, this signal is also exposed to microstructure noise.	Israel et al. (2018)
val_ipr_dts	Value (IPR, DtS-adjusted)	IPR value signal demeaned within duration-times-spread quintiles.	Israel et al. (2018)
Cluster III: Momentum & Reversal			
mom3_1	3-Month Momentum	Cumulative return months $t-2$ to $t-1$, skipping prior month.	Gebhardt et al. (2005)
mom6_1	6-Month Momentum	Cumulative return months $t-5$ to $t-1$.	Gebhardt et al. (2005)
mom12_1	12-Month Momentum	Cumulative return months $t-11$ to $t-1$.	Gebhardt et al. (2005)
mom12_7	Intermediate Momentum	Cumulative return months $t-11$ to $t-7$, excluding recent returns.	Novy-Marx (2012)
sysmom6_1	Systematic Momentum (6-1)	Sum of fitted values over months $t-5$ to $t-1$ from a 36(12) rolling CAPMB regression (bond market factor).	–
idimom6_1	Idiosyncratic Momentum (6-1)	Sum of residuals over months $t-5$ to $t-1$ from a 36(12) rolling CAPMB regression (bond market factor).	Blitz et al. (2011)
imom12_1	Industry Momentum (12-1)	Equal-weighted average 12-1 momentum of other bonds in the same FF17 industry.	Wang et al. (2024)
ltr48_12	Long-Term Reversal (48-12)	Cumulative return months $t-47$ to $t-12$. Uses expanding window starting at 12-3, ramping to 48-12.	Bali et al. (2021a)

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
ltr30_6	Long-Term Reversal (30-6)	Cumulative return months $t-29$ to $t-6$. Uses expanding window starting at 12-3, ramping to 30-6.	Subrahmanyam (2023)
iltr48_12	Industry LTR (48-12)	Equal-weighted average 48-12 LTR of other bonds in the same FF17 industry.	–
str	Short-Term Reversal	Prior month return r_{t-1}.	Bai et al. (2019)
mom9_1	9-Month Momentum	Cumulative return months $t-8$ to $t-1$.	Gebhardt et al. (2005)
sysmom3_1	Systematic Momentum (3-1)	Sum of fitted values over months $t-2$ to $t-1$ from a 36(12) rolling CAPMB regression.	–
sysmom12_1	Systematic Momentum (12-1)	Sum of fitted values over months $t-11$ to $t-1$ from a 36(12) rolling CAPMB regression.	–
idimom3_1	Idiosyncratic Momentum (3-1)	Sum of residuals over months $t-2$ to $t-1$ from a 36(12) rolling CAPMB regression.	Blitz et al. (2011)
idimom12_1	Idiosyncratic Momentum (12-1)	Sum of residuals over months $t-11$ to $t-1$ from a 36(12) rolling CAPMB regression.	Blitz et al. (2011)
imom1	Industry Momentum (1)	Equal-weighted average prior-month return of other bonds in the same FF17 industry.	Wang et al. (2024)
imom3_1	Industry Momentum (3-1)	Equal-weighted average 3-1 momentum of other bonds in the same FF17 industry.	Wang et al. (2024)
ltr24_3	Long-Term Reversal (24-3)	Cumulative return months $t-23$ to $t-3$. Uses expanding window starting at 12-3.	–
iltr30_6	Industry LTR (30-6)	Equal-weighted average 30-6 LTR of other bonds in the same FF17 industry.	–
iltr24_3	Industry LTR (24-3)	Equal-weighted average 24-3 LTR of other bonds in the same FF17 industry.	–
Cluster IV: Illiquidity			
pi	Price Impact	Pastor-Stambaugh liquidity: negated coefficient from return reversal regression on signed volume. Requires a minimum of 5 daily returns.	Pástor and Stambaugh (2003)
ami	Amihud Illiquidity	Within-month mean of daily $\|r_t\|/\text{dvol}_t$. Requires a minimum of 5 daily returns.	Amihud (2002)
ami_v	Amihud Volatility	Monthly standard deviation of daily Amihud ratios. Requires a minimum of 5 daily returns.	Dick-Nielsen et al. (2012)

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
roll	Roll Spread	Implicit bid-ask spread: $2\sqrt{\max(-\text{Cov}(r_t, r_{t-1}), 0)}$. Requires a minimum of 5 daily returns.	Roll (1984)
ilq	Roll Autocovariance	Negative autocovariance of log returns $\times 100$. Requires a minimum of 5 daily returns.	Bao et al. (2011)
spd_rel	Relative Bid-Ask Spread	Volume-weighted $(P^{ask} - P^{bid})/\text{mid}$. Requires a minimum of 5 prices.	Hong and Warga (2000)
cs_sprd	Corwin-Schultz Spread	High-low spread estimator using two-day price ranges. Requires a minimum of 5 prices.	Corwin and Schultz (2012)
ar_sprd	Abdi-Ranaldo Spread	Closing price spread estimator. Requires a minimum of 5 prices.	Abdi and Ranaldo (2017)
p_zro	Zero-Return Proportion	Fraction of business days with no valid price.	Fong et al. (2017)
p_fht	FHT Spread	Spread derived from zero-return proportion: $2\sigma\Phi^{-1}((1+p_{zro})/2)$. Requires a minimum of 5 daily returns.	Fong et al. (2017)
vov	Volatility of Volume	Liquidity proxy: $2.5\times\sigma^{0.6}/\bar{V}^{0.25}$. Requires a minimum of 5 daily returns.	Tobek (2016)
lix	LIX Liquidity	$\log_{10}[(V \times P_{\text{close}})/(P_{\text{high}} - P_{\text{low}})]$. Requires a minimum of 5 prices.	Danyliv et al. (2014)
spd_abs	Absolute Bid-Ask Spread	Volume-weighted $(P^{ask} - P^{bid})$ in dollars. Requires a minimum of 5 prices.	–
Cluster V: Volatility & Risk			
dvol	Daily Volatility	Standard deviation of daily returns within the month.	–
dskew	Daily Skewness	Central skewness of daily returns within the month.	–
dkurt	Daily Kurtosis	Excess kurtosis of daily returns within the month.	–
rvol	Realized Volatility	Within-month square root of sum of squared daily returns: $\sqrt{\sum r_t^2}$. Requires a minimum of 5 daily returns.	–
rsj	Realized Signed Jump	Within-month asymmetry in positive vs. negative squared returns: $(RV^+ - RV^-)/RV$. Requires a minimum of 5 daily returns.	Bollerslev et al. (2020)

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
rsk	Realized Skewness	Within-month third moment of daily returns scaled by realized volatility. Requires a minimum of 5 daily returns.	–
rkt	Realized Kurtosis	Within-month fourth moment of daily returns scaled by realized volatility. Requires a minimum of 5 daily returns.	–
var_95	95% Value-at-Risk	5th percentile loss from empirical daily return distribution over a 36(12) rolling window.	Bai et al. (2019)
es_90	90% Expected Shortfall	Mean of worst 10% of daily returns over a 36(12) rolling window.	Bai et al. (2019)
var_90	90% Value-at-Risk	10th percentile loss from empirical daily return distribution over a 36(12) rolling window.	Bai et al. (2019)
dvol_sys	Systematic Volatility	Standard deviation of systematic returns (CAPMB fitted values) within the month. Requires a minimum of 5 daily returns.	–
dvol_idio	Idiosyncratic Volatility	Standard deviation of idiosyncratic returns (CAPMB residuals) within the month. Requires a minimum of 5 daily returns.	–
ivol_mkt	Idiosyncratic Volatility (MKT)	Residual volatility from joint MKTRF+MKTB regression.	–
ivol_bbw	Idiosyncratic Volatility (BBW)	Residual volatility from BBW 4-factor regression.	Bai et al. (2019)
ivol_vp	Idiosyncratic Volatility (VP)	Residual volatility from VOLPSB regression.	Chung et al. (2019)
iskew	Idiosyncratic Skewness	Skewness of residuals from coskewness regression.	Harvey and Siddique (2000)
Cluster VI: Market Risk			
b_mktrf_mkt	Equity Market Beta	Beta on MKTRF from joint regression with MKTB (36-month rolling).	–
b_mktb_mkt	Bond Market Beta	Beta on MKTB from joint regression with MKTRF (36-month rolling).	–
b_mktb	Bond Market Beta (Univariate)	Beta from univariate regression on MKTB.	Dickerson et al. (2023)
b_mktbx_dcapm	Duration-Adj Market Beta	Beta on MKTBX from duration-adjusted CAPM.	van Binsbergen et al. (2025)
b_term_dcapm	Term Premium Beta	Beta on TERM = MKTB − MKTBX from duration-adjusted CAPM.	van Binsbergen et al. (2025)

Continued on next page

Mnemonic	Name	Description	Citation
b_mktb_dn	Downside Market Beta	Beta on $\min(\text{MKTB}, 0)$ from asymmetric market model.	Ang et al. (2006)
b_mktb_up	Upside Market Beta	Beta on $\max(\text{MKTB}, 0)$ from asymmetric market model.	Ang et al. (2006)
b_termb	Term Beta	Beta on TERMB from market regression.	Gebhardt et al. (2005)
db_mkt	Daily Market Beta	Beta from within-month daily regression on cross-sectional mean return.	–
Cluster VII: Credit & Default Betas			
b_drf	Downside Risk Beta (Univariate)	Beta from univariate regression on DRF.	Dickerson et al. (2023)
b_crf	Credit Risk Beta (Univariate)	Beta from univariate regression on CRF.	Dickerson et al. (2023)
b_lrf	Liquidity Risk Beta (Univariate)	Beta from univariate regression on LRF.	Dickerson et al. (2023)
b_defb	Default Beta	Beta on DEFB from duration-adjusted market regression.	Gebhardt et al. (2005)
Cluster VIII: Volatility & Liquidity Betas			
b_dvix	VIX Innovation Beta	Sum of contemporaneous and lagged ΔVIX betas from MKTB+MKTRF regression.	–
b_dvix_va	VIX Beta (Amihud)	ΔVIX beta from FF3+VIX+Amihud specification.	Chung et al. (2019)
b_dvix_vp	VIX Beta (PSB)	ΔVIX beta from FF3+VIX+PSB specification.	Chung et al. (2019)
b_dvix_dn	Downside VIX Beta	Beta on $\min(\Delta\text{VIX}, 0)$ from asymmetric VIX model.	–
b_dvix_up	Upside VIX Beta	Beta on $\max(\Delta\text{VIX}, 0)$ from asymmetric VIX model.	–
b_psb	Pastor-Stambaugh Beta	Beta on bond market liquidity factor PSB.	Lin et al. (2011)
b_psb_m	PSB Beta (Multi-Factor)	PSB beta controlling for FF3+MKTBX+TERM.	Lin et al. (2011)
b_amd_m	Amihud Beta (Multi-Factor)	Amihud beta controlling for FF3+MKTBX+TERM.	Lin et al. (2011)
b_amd	Amihud Beta	Beta on aggregate Amihud illiquidity factor.	Lin et al. (2011)
b_coskew	Coskewness Beta	Beta on MKTB^2 from coskewness regression.	Harvey and Siddique (2000)

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
b_vix	VIX Level Beta	Beta on VIX level from monthly regression.	Chung et al. (2019)
b_dvixd	Daily VIX Innovation Beta	Beta on daily ΔVIX within the month. Requires a minimum of 5 daily returns.	–
b_illiq	Illiquidity Beta	Beta on aggregate bond market illiquidity factor.	–
Cluster IX: Macro & Other Betas			
b_dunc	Macro Uncertainty Beta	Beta on changes in Jurado, Ludvigson, and Ng (2015) macro uncertainty index.	Bali et al. (2021b)
b_duncr	Real Uncertainty Beta	Beta on changes in Jurado, Ludvigson, and Ng (2015) real uncertainty component.	Bali et al. (2021b)
b_duncf	Financial Uncertainty Beta	Beta on changes in Jurado, Ludvigson, and Ng (2015) financial uncertainty component.	Bali et al. (2021b)
b_unc	Uncertainty Level Beta	Beta on Jurado, Ludvigson, and Ng (2015) macro uncertainty level.	Bali et al. (2021b)
b_dunc3	3-Month Uncertainty Change Beta	Beta on 3-month change in Jurado, Ludvigson, and Ng (2015) macro uncertainty.	Bali et al. (2021b)
b_dunc6	6-Month Uncertainty Change Beta	Beta on 6-month change in Jurado, Ludvigson, and Ng (2015) macro uncertainty.	Bali et al. (2021b)
b_dcpi	Inflation Beta	Beta on monthly CPI changes.	–
b_cpi_vol6	Inflation Volatility Beta	Beta on 6-month rolling CPI volatility.	Ceballos (2022)
b_dcredit	Credit Spread Change Beta	Beta on monthly changes in BAA-AAA spread.	–
b_credit	Credit Spread Level Beta	Beta on BAA-AAA credit spread level.	Dickerson et al. (2023)
b_cptlt	Intermediary Capital Beta	Beta on traded intermediary capital ratio.	He et al. (2017)
b_rvol	Realized Volatility Beta	Beta on aggregate realized volatility factor.	–
b_rsj	Realized Jump Beta	Beta on aggregate realized signed jump factor.	Bollerslev et al. (2020)
b_lvl	Level Factor Beta	Beta on the average over key-rate U.S. Treasury yields.	Koijen et al. (2017)
b_ysp	Yield Spread Beta	Beta on yield spread factor.	Koijen et al. (2017)
b_epu	Economic Policy Uncertainty Beta	Beta on EPU index level.	Baker et al. (2016)

Continued on next page

Table IA.VIII continued from previous page

Mnemonic	Name	Description	Citation
b_epum	Monetary Policy Uncertainty Beta	Beta on monetary policy uncertainty index level.	Baker et al. (2016)
b_eput	Trade Policy Uncertainty Beta	Beta on trade policy uncertainty index level.	Baker et al. (2016)

IA.3 Factors that survive adjustment for bond market risk

Tables IA.X and IA.XI report annualized long-short factor returns for all signals with CAPMB $t(\alpha) > 1.96$, using value-weighted and equal-weighted portfolios, respectively. Price-based signals use the adjusted signal as standard (observed at least 1 business day before the month-end price used for return computation); returns are month-end. Factors are ranked by alpha from highest to lowest. In each table, Panel A shows single-sort decile factors and Panel B shows within-firm factors. Shaded rows pass the Benjamini and Hochberg (1995, BH) false discovery rate (FDR) correction at 5%, applied independently within each panel (i.e., the 108 alpha p-values in each panel are corrected separately, not pooled across weighting schemes or sorting methods).

Across 108 signals $\times$ 2 weighting schemes $\times$ 2 sorting methods = 432 specifications, only 65 (15.0%) produce a CAPMB alpha with $t > 1.96$, and only 26 (6.0%) bond CAPM alphas survive the FDR correction. The breakdown by specification is

	$t(\alpha) > 1.96$		FDR survivors	
	Count	%	Count	%
VW, Single-Sort	9	8.3	0	0.0
VW, Within-Firm	21	19.4	9	8.3
EW, Single-Sort	8	7.4	1	0.9
EW, Within-Firm	27	25.0	16	14.8
Total (of 432)	65	15.0	26	6.0

No value-weighted single-sort bond CAPM alpha survives the FDR correction. Within-firm sorts retain more factors, but the majority still fail after the multiple testing adjustment. Applying the same FDR correction to mean return p-values produces consistent results. Without FDR, 119 of 432 specifications (27.5%) produce a statistically significant mean return ($t(\mu) > 1.96$), compared to 65 (15.0%) for bond CAPM alphas. After FDR correction, the gap narrows sharply:

	$t(\mu) > 1.96$		FDR survivors	
	Count	%	Count	%
VW, Single-Sort	26	24.1	0	0.0
VW, Within-Firm	24	22.2	5	4.6
EW, Single-Sort	34	31.5	0	0.0
EW, Within-Firm	35	32.4	17	15.7
Total (of 432)	119	27.5	22	5.1

No single-sort factor premium survives FDR in either weighting scheme. Only within-firm sorts produce robust mean premia (5 VW, 17 EW). Although roughly twice as many factors have nominally significant premia (27.5%) as have significant alphas (15.0%) before FDR, the correction reduces both to similar levels (5.1% vs 6.0%), reinforcing the message that most of the 'factor zoo' is spanned by the bond market factor.

The 26 FDR survivors split into two groups. The first consists of value and spread-change factors (`val_ipr`, `val_hz_dts`, `dcs6`) that retain significant premia after breaking the shared-price link. The second consists of systematic risk exposure factors (coskewness, illiquidity, uncertainty, Pastor-Stambaugh and realized volatility betas, realized signed jumps, idiosyncratic and daily skewness) whose measurement error does not enter the return denominator and are immune to LIB (Corollary B.1). Standard momentum is absent. None of the five illiquidity characteristic factors from Table IA.XIV appear. Bond age is the only single-sort FDR survivor (equal-weighted). Twenty-five of the 26 FDR survivors are within-firm sorts.

Table IA.IX decomposes the FDR survivors by factor cluster. Panel A reports counts; Panel B lists the surviving factors within each cluster.

IA.3.1 Value-Weighted Results

Table IA.X: Factors with significant alphas after bias correction (value-weighted).

The table reports annualized long-short factor returns for all factors with CAPMB $t(\alpha) > 1.96$, ranked by α. Price-based factors use the adjusted signal (observed ≥1 business day before the month-end return price). Panel A reports single-sort portfolios. Panel B reports within-firm sorted portfolios. μ is mean return (%), α is CAPMB alpha (%), IR = $\alpha/\sigma(\varepsilon)$. t-statistics use Newey-West standard errors (lags = $\lfloor T^{0.25} \rfloor$). Shaded rows pass the Benjamini and Hochberg FDR correction at 5%, applied independently within each panel. Asterisked factors are sign-corrected. Sample: 2002-09 to 2024-12.

Factor	**T**	**Date$_{\text{Start}}$**	**Date$_{\text{End}}$**	μ	**SD**	$t(\mu)$	**SR**	α	$t(\alpha)$	**IR**
Panel A: Single-Sorted Portfolios (9 of 108 factors)										
`b_coskew*`	268	2002-09	2024-12	4.41	8.90	1.74	0.50	5.41	2.03	0.62
`b_illiq*`	257	2003-08	2024-12	3.43	10.92	1.72	0.31	4.64	2.05	0.43
`b_dvix_dn`	268	2002-09	2024-12	4.81	6.61	2.51	0.73	4.53	2.37	0.69
`imom12_1`	268	2002-09	2024-12	2.24	8.00	1.19	0.28	4.46	2.44	0.64
`val_ipr`	268	2002-09	2024-12	6.08	8.60	2.68	0.71	3.71	2.12	0.49
`b_amd*`	257	2003-08	2024-12	3.22	6.94	1.99	0.46	3.56	2.10	0.51

Continued on next page

Table IA.X continued from previous page

Factor	**T**	**Date$_{\text{Start}}$**	**Date$_{\text{End}}$**	μ	**SD**	$t(\mu)$	**SR**	α	$t(\alpha)$	**IR**
imom1	268	2002-09	2024-12	3.05	8.33	1.94	0.37	3.44	2.11	0.41
val_ipr_dts	268	2002-09	2024-12	2.61	5.81	2.43	0.45	2.74	2.26	0.47
age	268	2002-09	2024-12	1.19	2.28	2.16	0.52	0.92	2.05	0.41
Panel B: Within-Firm Sorted Portfolios (21 of 108 factors)										
dcs6_wf*	268	2002-09	2024-12	3.16	2.59	4.63	1.22	2.66	4.98	1.09
val_ipr_dts_wf	268	2002-09	2024-12	1.52	2.69	2.39	0.57	2.47	4.50	1.17
val_ipr_wf	268	2002-09	2024-12	2.61	2.37	3.50	1.10	2.22	4.25	0.97
b_illiq_wf*	257	2003-08	2024-12	1.02	2.59	1.86	0.39	1.72	3.47	0.78
b_cptlt_wf	268	2002-09	2024-12	0.99	2.88	1.50	0.34	1.64	2.33	0.62
b_coskew_wf*	268	2002-09	2024-12	1.23	2.39	2.15	0.51	1.61	2.82	0.70
b_amd_m_wf*	257	2003-08	2024-12	1.20	2.24	1.69	0.53	1.47	2.16	0.67
b_rvol_wf*	257	2003-08	2024-12	0.62	2.65	1.24	0.23	1.39	2.71	0.63
b_dunc6_wf*	268	2002-09	2024-12	1.34	2.13	1.92	0.63	1.36	2.15	0.64
b_duncf_wf*	268	2002-09	2024-12	0.53	2.34	1.14	0.22	1.32	3.04	0.71
b_dunc_wf*	268	2002-09	2024-12	0.95	2.31	2.10	0.41	1.27	2.86	0.56
b_psb_wf	257	2003-08	2024-12	0.65	2.31	1.44	0.28	1.25	2.96	0.62
val_hz_dts_wf	268	2002-09	2024-12	0.77	2.12	2.09	0.36	1.21	2.85	0.61
b_amd_wf*	257	2003-08	2024-12	0.94	2.26	1.59	0.41	1.19	2.27	0.54
rsj_wf*	268	2002-09	2024-12	1.67	2.33	3.67	0.72	1.17	2.44	0.54
b_duncr_wf*	268	2002-09	2024-12	0.93	2.06	2.04	0.45	1.03	2.21	0.50
b_psb_m_wf	257	2003-08	2024-12	0.50	2.19	1.13	0.23	0.98	2.28	0.49
rsk_wf*	268	2002-09	2024-12	1.06	1.46	4.12	0.72	0.79	2.57	0.57
sze_wf*	268	2002-09	2024-12	0.66	1.93	1.77	0.34	0.67	2.00	0.35
age_wf	268	2002-09	2024-12	0.15	1.46	0.63	0.10	0.63	2.59	0.52
rkt_wf*	268	2002-09	2024-12	0.48	0.94	2.42	0.51	0.34	2.18	0.37

Table IA.IX: FDR survivors by factor cluster.

The table reports factors whose CAPMB α passes the Benjamini and Hochberg (1995) false discovery rate correction at 5%, applied independently within each specification. Panel A counts survivors by cluster. Panel B lists the surviving factors (asterisked factors are sign-corrected). SS = single-sort decile factors. WF = within-firm factors. Of 108 factors $\times$ 4 specifications = 432 tests, 26 (6.0%) bond CAPM alphas survive.

Panel A: Count of FDR Survivors

	VW, SS	VW, WF	EW, SS	EW, WF	Total
I. Spreads, Yields, Size	0	1	1	2	4
II. Value	0	3	0	3	6
III. Momentum & Reversal	0	0	0	2	2
IV. Illiquidity	0	0	0	0	0
V. Volatility & Risk	0	0	0	4	4
VI. Market Risk	0	0	0	1	1
VII. Credit & Default Betas	0	0	0	0	0
VIII. Vol. & Liq. Betas	0	3	0	3	6
IX. Macro & Other Betas	0	2	0	1	3
Total	0	9	1	16	26

Panel B: Surviving Factors by Cluster

Cluster	Factors
I. Spreads, Yields, Size	`dcs6*`, `age`, `ytm`
II. Value	`val_ipr`, `val_ipr_dts`, `val_hz_dts`
III. Momentum & Reversal	`sysmom12_1*`, `idimom12_1`
V. Volatility & Risk	`rsj*`, `rsk*`, `dskew*`, `iskew*`
VI. Market Risk	`b_mktb_up*`
VIII. Vol. & Liq. Betas	`b_coskew*`, `b_illiq*`, `b_psb`, `b_rvol*`
IX. Macro & Other Betas	`b_dunc*`, `b_duncf*`

IA.3.2 Equal-Weighted Results

Table IA.XI: Factors with significant alphas after bias correction (equal-weighted).

The table reports annualized long-short factor returns for all factors with CAPMB $t(\alpha) > 1.96$, ranked by α. Price-based factors use the adjusted signal (observed $\geq$1 business day before the month-end return price). Panel A reports single-sort portfolios. Panel B reports within-firm sorted portfolios. μ is mean return (%), α is CAPMB alpha (%), IR = $\alpha/\sigma(\varepsilon)$. t-statistics use Newey-West standard errors (lags = $\lfloor T^{0.25} \rfloor$). Shaded rows pass the Benjamini and Hochberg FDR correction at 5%, applied independently within each panel. Asterisked factors are sign-corrected. Sample: 2002-09 to 2024-12.

Factor	T	$\text{Date}_{\text{Start}}$	Date_{End}	μ	SD	$t(\mu)$	SR	α	$t(\alpha)$	IR
Panel A: Single-Sorted Portfolios (8 of 108 factors)										
ytm	268	2002-09	2024-12	16.30	25.28	2.68	0.64	8.30	2.08	0.39
cs	268	2002-09	2024-12	15.23	24.25	2.58	0.63	7.94	2.00	0.39
ltr30_6*	268	2002-09	2024-12	8.56	13.07	2.51	0.66	5.42	2.00	0.46
imom12_1	268	2002-09	2024-12	0.36	16.71	0.11	0.02	4.92	1.98	0.33
md_dur*	268	2002-09	2024-12	1.52	11.00	0.62	0.14	4.91	2.27	0.53
convx*	268	2002-09	2024-12	0.82	10.64	0.36	0.08	4.68	2.32	0.57
age	268	2002-09	2024-12	2.68	2.91	3.31	0.92	2.46	3.71	0.85
p_zro	268	2002-09	2024-12	2.12	4.02	2.23	0.53	1.44	2.24	0.38
Panel B: Within-Firm Sorted Portfolios (27 of 108 factors)										
dcs6_wf*	268	2002-09	2024-12	2.80	1.66	5.94	1.69	2.46	6.77	1.59
b_lvl_wf	268	2002-09	2024-12	2.00	3.72	1.68	0.54	2.42	2.43	0.66
val_ipr_wf	268	2002-09	2024-12	2.43	1.68	4.39	1.45	2.30	5.03	1.38
val_ipr_dts_wf	268	2002-09	2024-12	1.35	2.14	2.83	0.63	2.23	5.51	1.53
sysmom12_1_wf*	268	2002-09	2024-12	2.44	3.40	3.31	0.72	2.20	3.04	0.65
b_mktb_up_wf*	268	2002-09	2024-12	0.26	3.69	0.33	0.07	1.82	3.41	0.74
sysmom6_1_wf*	268	2002-09	2024-12	2.03	3.52	2.95	0.58	1.70	2.30	0.49
val_hz_dts_wf	268	2002-09	2024-12	1.20	1.60	3.66	0.75	1.65	4.63	1.18
b_coskew_wf*	268	2002-09	2024-12	1.34	2.09	2.70	0.64	1.55	3.05	0.76
idimom12_1_wf	268	2002-09	2024-12	1.57	1.77	3.39	0.88	1.55	2.94	0.87
cs_wf	268	2002-09	2024-12	2.91	3.28	3.89	0.89	1.54	2.52	0.69
b_illiq_wf*	257	2003-08	2024-12	1.03	2.05	2.11	0.50	1.44	3.45	0.76
ytm_wf	268	2002-09	2024-12	3.49	4.08	3.57	0.85	1.35	3.23	0.86
rsj_wf*	268	2002-09	2024-12	1.60	1.58	5.17	1.01	1.25	3.89	0.86
b_dunc_wf*	268	2002-09	2024-12	1.01	2.28	1.59	0.44	1.23	2.04	0.55
b_epum_wf*	268	2002-09	2024-12	1.43	2.14	2.28	0.67	1.20	2.23	0.57
b_termb_wf*	268	2002-09	2024-12	0.11	3.49	0.16	0.03	1.17	1.97	0.39
iskew_wf*	268	2002-09	2024-12	1.50	1.63	2.86	0.92	1.13	2.76	0.75
bbtm_wf	268	2002-09	2024-12	2.07	2.94	2.93	0.70	1.08	1.98	0.46
b_crf_wf	268	2002-09	2024-12	0.93	2.79	1.70	0.33	1.05	2.26	0.38
b_rvol_wf*	257	2003-08	2024-12	0.50	2.08	1.46	0.24	1.04	3.67	0.58
b_duncf_wf*	268	2002-09	2024-12	0.19	2.17	0.36	0.09	1.01	2.92	0.62
rsk_wf*	268	2002-09	2024-12	1.08	1.23	4.76	0.88	0.86	3.58	0.74
idimom6_1_wf	268	2002-09	2024-12	0.72	1.69	2.20	0.43	0.71	2.16	0.42

Continued on next page

Table IA.XI continued from previous page

Factor	T	$\text{Date}_{\text{Start}}$	Date_{End}	μ	SD	$t(\mu)$	SR	α	$t(\alpha)$	IR
b_psb_wf	257	2003-08	2024-12	0.30	1.95	0.81	0.15	0.70	2.00	0.39
dskew_wf*	268	2002-09	2024-12	0.59	0.78	3.52	0.77	0.54	3.32	0.70
sze_wf*	268	2002-09	2024-12	0.49	1.19	2.23	0.41	0.52	2.49	0.44

IA.4 Measurement error and latent implementation bias: Additional results

Table IA.XII complements Panel A of Table 1 by reporting single-sort results separately for IG and NIG bonds. After adjustment, no factor retains a statistically significant alpha at the 5% level for IG bonds. For NIG bonds, no adjusted alpha is statistically significant at the 5% level. val_hz comes closest ($\alpha = 0.37\%$, $t = 1.91$ based on the adjusted signal).

Table IA.XIII complements Panel B of Table 1 by reporting within-firm results by rating. In Panel A (All Bonds), only dcs6 and val_ipr retain statistically significant alphas after adjustment: -0.22% ($t = -4.98$) and 0.18% ($t = 4.25$) per month, respectively. Panel C (NIG) shows a different pattern: within-firm factors for ytm, cs, bbtm, dcs6, and str all exhibit significant adjusted alphas, while val_ipr and val_hz do not.

Fig. IA.1 displays the average LIB by rating category. Panel A shows single-sort factors. Panel B shows within-firm sorts. The bias magnitude is similar across IG and NIG bonds for most factors.

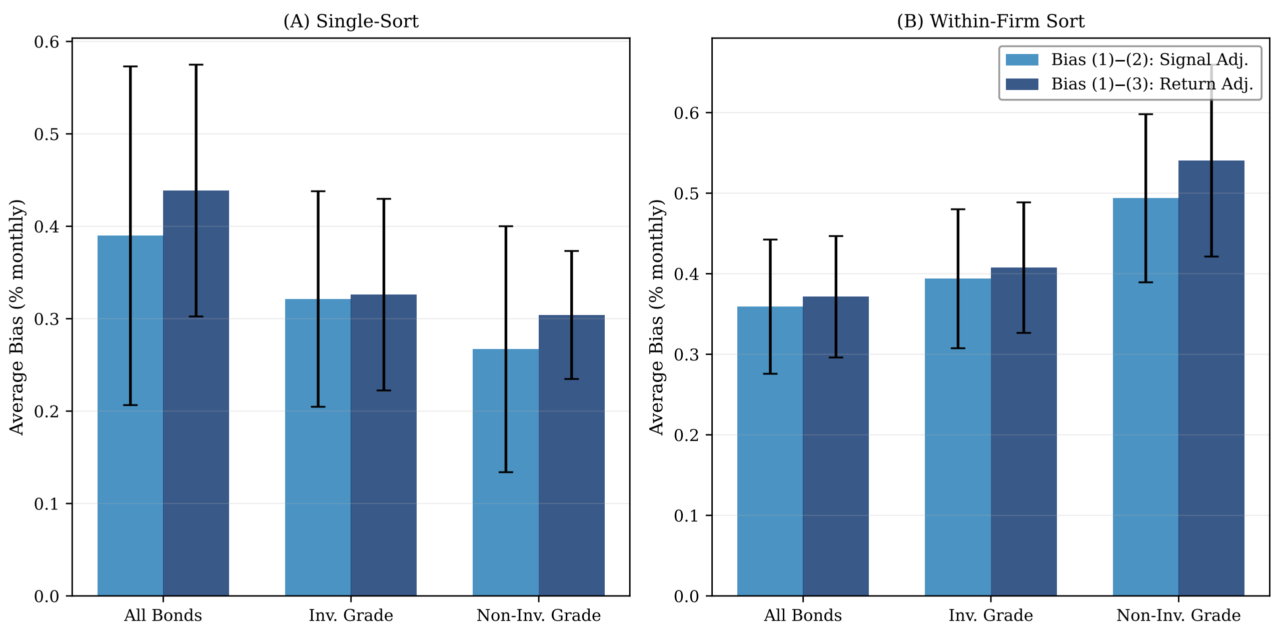

Figure IA.1: Average LIB by rating category for value-weighted portfolios.

Panel A shows single-sort factors. Panel B shows within-firm factors. Bars show average absolute bias across 7 signals. Bias is sign-corrected to be positive (absolute magnitude). Error bars show 95% confidence intervals based on cross-signal variation. Sample: 2002-09 to 2024-12, T=268.

IA.4.1 Illiquidity factors

Illiquidity characteristics are computed as within-month averages of daily observations (each requiring at least 5 daily values), so the signal does not share the month-end price that enters the return denominator. The shared-price link that generates LIB for point-in-time signals (credit spread, short-term reversal) is dampened by construction. Table IA.XIV examines five illiquidity factors: price impact (`ilq`) from Bao, Pan, and Wang (2011), Amihud illiquidity (`ami`) from Amihud (2002), Corwin-Schultz spread (`cs_sprd`) from Corwin and Schultz (2012), Abdi-Ranaldo spread (`ar_sprd`) from Abdi and Ranaldo (2017), and relative spread (`spd_rel`) from Hong and Warga (2000). All biases are economically small and mostly statistically indistinguishable from zero. In Panel A, premia range from 23 to 52 basis points per month, but CAPMB alphas are uniformly small (10–20 basis points with t-statistics between 0.99 and 1.64). None are statistically distinguishable from zero at conventional levels. The largest premium (`ar_sprd` at 52 basis points) has an alpha of just 0.20% ($t = 1.44$). Across all five characteristics, evidence for an illiquidity premium is weak. Within-firm sorts, which isolate variation in illiquidity across bonds of the same issuer, produce weaker results. Premia shrink to 6–11 basis points per month, and alphas are indistinguishable from zero (ranging from -0.03% to 0.03% with $|t| < 1.1$). These results confirm and extend practitioner findings by Richardson and Palhares (2018), who document that the illiquidity premium in credit markets is largely a 'myth' once maturity, credit quality, and other bond characteristics are controlled for. Our within-firm sorts provide a tighter test by comparing bonds of the same issuer. The findings are also consistent with Choi, Han, Shin, and Yoon (2026), who show that in OTC search-based markets, the sign of the illiquidity premium is state-dependent: illiquid bonds can trade at *higher* prices than liquid bonds with similar cash flows during episodes of selling pressure. This search-based mechanism suggests that any unconditional illiquidity premium is attenuated by episodes in which the premium reverses, helping explain the weak average alphas we document. In a nutshell, the illiquidity characteristic premium in corporate bond markets appears substantially smaller than commonly thought.

IA.5 Look-ahead bias: Additional results

Tables IA.XV and IA.XVI decompose LAB into contributions from the long and short legs. For factors sensitive to left-tail winsorization (Panel A), bias concentrates in the long leg for risk-based factors (`ivol_bbw`, `ivol_vp`, `var_95`, `es_90`) but in the short leg for beta-based factors (`b_dunc3`, `b_unc`). For momentum factors sensitive to right-tail winsorization (Panel B), bias concentrates in the short leg, where right-tail winsorization caps the rebounds of past losers.

Fig. IA.2 decomposes the average LAB into contributions from the long and short legs, confirming the asymmetric mechanism described in Section 4. Panel A shows factors sensitive to left-tail winsorization, where the bias concentrates almost entirely in the long leg, with the short leg contributing marginally negative bias. For `ivol_bbw`, `ltr48_12`, and `b_dunc3`, left-tail winsorization protects the long leg from large negative returns during stress because high-risk bonds populate the long leg and experience disproportionate left-tail losses that winsorization removes. Panel B shows the mirror image for momentum factors, where the bias concentrates in the short leg, with the long leg contributing marginally negative bias. Past losers in the short leg rebound sharply during recoveries, and right-tail winsorization caps these gains, artificially improving the short position. Panel C compares LAB across rating categories. The bias is two to three times larger for NIG bonds than for IG bonds across all factors, because NIG bonds exhibit more extreme returns in both tails.

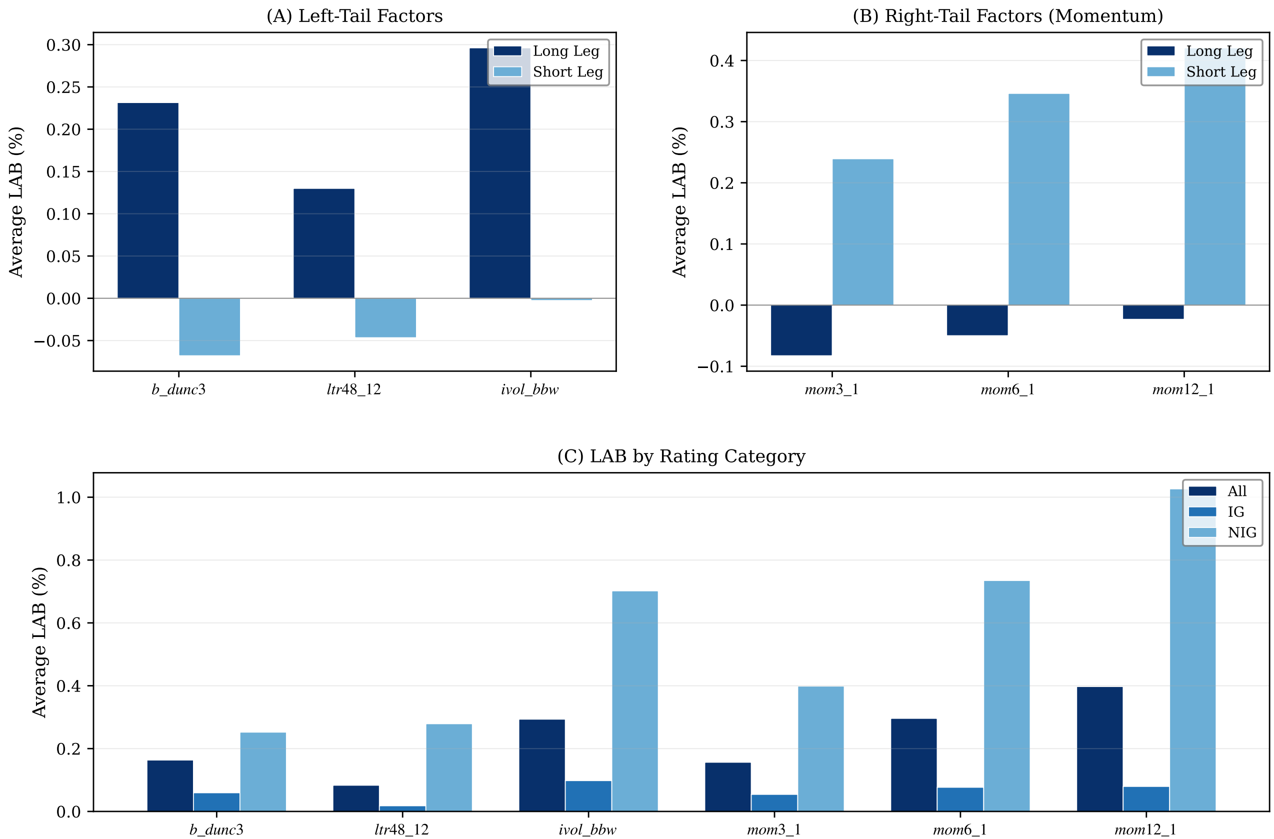

Figure IA.2: Look-ahead bias decomposition by leg and rating.

The figure decomposes the average look-ahead bias into long and short leg contributions. Panels A–B plot factors sensitive to left-tail and right-tail winsorization. For `b_dunc3` and `ltr48_12`, long and short legs are swapped due to sign correction. In both panels, the short leg is scaled by -1 to show its contribution to the long-short factor (negative because we are short the losers). Panel C compares LAB across all bonds, IG, and NIG samples, with sign-corrected values so that positive LAB indicates upward bias. Sample: 2002-08 to 2024-12.

IA.6 Data and methodological uncertainty: Additional results

This section provides additional figures and tables for the data and methodological uncertainty analysis in Section 5. Table IA.XVII decomposes the improvement in factor significance by filter type and tail location. For each factor cluster, we report the count and percentage of filter configurations that improve the factor's CAPMB alpha relative to the baseline (unfiltered) specification. A filter is classified as an improvement if $t_\alpha > 1.96$, $t_\alpha > t_{\alpha,\text{baseline}}$, and $\alpha >$

α_{baseline}. Value factors show the highest improvement rate for left-tail trimming (21%), where bonds with large negative prior returns are excluded from the investable universe at portfolio formation. Momentum & Reversal factors show improvement spread across filter types, with 6% of paths improving under symmetric bid-ask bounce filters. Volatility & Risk factors show improvement concentrated in right-tail trimming (2%), while left-tail and symmetric filters produce no improvement.

Table IA.XVIII reports the cross-sectional size of the top and bottom portfolios combined across design dimensions, pooled over all 108 factors and averaged across admissible portfolio constructions. Across all 18,128 non-degenerate strategies, the top and bottom portfolios combined contain, on average, 813 bonds (median 752), with a 5th-percentile month-level count of 423. The average within-strategy share of months with fewer than 20 bonds is 0.2%. Portfolio granularity is the primary determinant of portfolio size. Tercile sorts yield an average of 1,261 bonds in the top and bottom portfolios combined, quintiles 772, and deciles 404. Even under decile sorting, the 5th-percentile count remains 209 bonds and only 0.5% of months fall below 20 bonds. The all-bonds universe averages 1,147 combined bonds, investment grade 856, and high-yield 267. Although the high-yield segment is mechanically thinner, its 5th-percentile count remains 145 bonds and the share of months with fewer than 20 bonds is 0.5%. Using the full maturity spectrum yields 1,618 bonds in the top and bottom portfolios combined, while restricting to short, intermediate, and long maturities reduces this to 645, 545, and 441. Even in the long-maturity segment, the 5th-percentile count remains 213 bonds and the incidence of months with fewer than 20 bonds is 0.6%. Breakpoint construction affects portfolio size primarily through composition. Using investment grade breakpoints produces larger top-and-bottom portfolio counts (1,195) than the full universe (734) or large-bond breakpoints (763). This difference reflects the admissible pairing of breakpoint universes with rating subsamples, which alters the effective cross-sectional pool underlying the sort. The tail-size bins summarize the distribution of cross-sectional thickness across strategies. Approximately 2,258 strategies (12%) have average top-and-bottom portfolio counts below 200 bonds, averaging 89 bonds with 1.2% of months below 20 bonds. A further 7,138 strategies (39%) fall in the 200–600 range, averaging 375 bonds with no months below 20 bonds. The largest group, 8,732 strategies (48%), exceeds 600 average bonds, averaging 1,358 bonds with no months below 20 bonds. Even in the most restrictive configurations, long-short factors are formed from large underlying portfolios, so the dispersion documented in the main analysis cannot be attributed to small cross-sectional samples.

Table IA.XIX decomposes improvement rates across portfolio construction dimensions, using the same criteria as in the data uncertainty analysis. Only 6% of specifications improve relative to baseline, and improvements are highly uneven across factor clusters and design choices. Across the Rating dimension, improvement rates are modest and similar for IG (4%) and NIG (5%). Value improves in 16% of IG and 14% of NIG specifications, while Momentum & Reversal improves more frequently in NIG (6%) than IG (3%). Credit & Default Risk (4 factors) shows no improvement within IG or NIG subsamples. The strongest asymmetry arises in the Maturity dimension. Long-maturity portfolios account for 10% improvements overall, compared to only 2% in the mid segment. This pattern is pronounced for Value (36% improvement in long maturities) and is also visible for Illiquidity (13%) and Momentum & Reversal (9%). Variation across breakpoint universes yields moderate and relatively uniform improvement rates (around 5%), with no dominance of a specific breakpoint construction.

Fig. IA.4 shows premium t-statistics for the top four factors per cluster across all 648 filter paths. Wide distributions indicate that statistical significance depends heavily on data processing choices. Factors whose distributions span both sides of the 1.96 threshold are particularly fragile, with published significance status changing under minor methodological variations.

Table IA.XII: Latent implementation bias in price-based factors (value-weighted, single-sort).

The table complements Table 1 by reporting single-sort results separately for IG and NIG bonds. μ is the mean monthly return (%), α is the CAPMB alpha (%). Bias (1)−(2) compares standard versus signal-adjusted. Bias (1)−(3) compares standard versus return-adjusted. t-statistics (Newey-West, lags = $\lfloor T^{0.25} \rfloor$) in parentheses. Sample: 2002-09 to 2024-12, T=268.

	Unadjusted (1)		Adj. Signal (2)		Adj. Return (3)		Bias (1)−(2)		Bias (1)−(3)	
Factor	μ	α	μ	α	μ	α	$\Delta\mu$	$\Delta\alpha$	$\Delta\mu$	$\Delta\alpha$
Panel A: All Bonds										
ytm	1.20	0.69	1.05	0.54	0.92	0.44	0.15	0.15	0.28	0.25
	(2.97)	(2.50)	(2.63)	(1.96)	(2.64)	(1.80)	(6.54)	(6.06)	(3.90)	(4.33)
cs	1.13	0.66	0.96	0.50	0.82	0.38	0.17	0.17	0.31	0.28
	(2.89)	(2.39)	(2.52)	(1.82)	(2.45)	(1.55)	(6.23)	(6.17)	(4.14)	(4.54)
bbtm	0.93	0.54	0.75	0.35	0.66	0.29	0.18	0.18	0.27	0.25
	(2.19)	(1.69)	(1.80)	(1.12)	(1.80)	(1.04)	(6.20)	(5.24)	(3.39)	(3.56)
dcs6	−1.02	−0.79	−0.50	−0.29	−0.46	−0.25	−0.52	−0.51	−0.56	−0.54
	(−4.58)	(−4.55)	(−2.52)	(−1.79)	(−2.50)	(−1.71)	(−8.21)	(−8.12)	(−7.83)	(−7.73)
val_ipr	0.96	0.76	0.51	0.31	0.51	0.33	0.45	0.45	0.45	0.43
	(4.82)	(5.24)	(2.68)	(2.12)	(3.05)	(2.61)	(6.93)	(6.42)	(7.12)	(7.52)
val_hz	0.81	0.47	0.45	0.12	0.43	0.12	0.36	0.34	0.38	0.35
	(3.80)	(3.71)	(2.29)	(1.13)	(2.32)	(1.08)	(6.60)	(6.43)	(6.21)	(6.01)
str	−0.99	−0.77	−0.09	0.12	−0.17	0.03	−0.90	−0.89	−0.82	−0.80
	(−4.46)	(−3.58)	(−0.51)	(0.68)	(−0.98)	(0.18)	(−8.70)	(−8.20)	(−7.34)	(−7.30)
Panel B: Investment Grade										
ytm	0.60	0.22	0.44	0.07	0.40	0.04	0.16	0.15	0.20	0.18
	(3.13)	(2.31)	(2.35)	(0.77)	(2.21)	(0.42)	(7.94)	(7.92)	(6.00)	(5.54)
cs	0.52	0.22	0.31	0.02	0.29	0.00	0.21	0.20	0.23	0.22
	(2.97)	(1.83)	(1.79)	(0.14)	(1.80)	(0.04)	(7.18)	(7.06)	(6.50)	(6.03)
bbtm	0.46	0.29	0.29	0.13	0.29	0.14	0.17	0.17	0.18	0.16
	(3.12)	(2.44)	(2.12)	(1.06)	(2.24)	(1.26)	(5.80)	(5.30)	(4.93)	(4.76)
dcs6	−0.71	−0.59	−0.29	−0.18	−0.29	−0.17	−0.42	−0.41	−0.42	−0.42
	(−5.53)	(−5.07)	(−2.47)	(−1.55)	(−2.52)	(−1.57)	(−7.98)	(−7.47)	(−8.77)	(−8.24)
val_ipr	0.66	0.50	0.31	0.16	0.31	0.17	0.35	0.35	0.35	0.34
	(5.57)	(5.64)	(2.92)	(1.77)	(3.08)	(1.95)	(7.60)	(7.18)	(8.07)	(8.16)
val_hz	0.56	0.27	0.27	−0.01	0.27	−0.00	0.29	0.29	0.30	0.28
	(3.73)	(2.77)	(1.92)	(−0.17)	(2.02)	(−0.03)	(6.05)	(5.72)	(5.95)	(5.39)
str	−0.64	−0.56	0.00	0.07	−0.03	0.02	−0.64	−0.63	−0.61	−0.59
	(−5.23)	(−3.84)	(0.04)	(0.69)	(−0.28)	(0.21)	(−9.08)	(−8.33)	(−8.44)	(−8.01)
Panel C: Non-Investment Grade										
ytm	1.15	0.73	1.01	0.56	0.87	0.49	0.14	0.16	0.28	0.24
	(2.54)	(2.10)	(2.26)	(1.66)	(2.22)	(1.60)	(4.23)	(4.47)	(3.38)	(3.47)
cs	1.13	0.73	0.96	0.54	0.85	0.50	0.17	0.19	0.28	0.23
	(2.60)	(2.17)	(2.21)	(1.63)	(2.27)	(1.66)	(4.53)	(4.12)	(3.42)	(3.53)
bbtm	1.20	0.62	0.99	0.46	0.89	0.37	0.21	0.16	0.31	0.25
	(2.26)	(1.64)	(1.86)	(1.20)	(1.99)	(1.13)	(2.78)	(3.73)	(2.95)	(3.09)
dcs6	−0.73	−0.46	−0.39	−0.10	−0.38	−0.15	−0.34	−0.36	−0.35	−0.31
	(−2.85)	(−2.36)	(−1.64)	(−0.60)	(−1.78)	(−0.89)	(−6.20)	(−5.55)	(−5.25)	(−5.77)
val_ipr	0.43	0.37	0.24	0.17	0.20	0.15	0.19	0.19	0.24	0.21
	(2.01)	(2.05)	(1.10)	(0.97)	(0.99)	(0.94)	(5.02)	(5.13)	(4.89)	(5.33)
val_hz	0.68	0.49	0.54	0.37	0.51	0.34	0.14	0.12	0.17	0.15
	(2.52)	(2.39)	(2.18)	(1.91)	(2.12)	(1.82)	(3.61)	(3.25)	(3.75)	(4.08)
str	−0.64	−0.29	0.04	0.21	−0.14	0.18	−0.68	−0.50	−0.50	−0.46
	(−1.67)	(−1.12)	(0.20)	(1.14)	(−0.41)	(0.72)	(−3.10)	(−4.33)	(−5.62)	(−6.19)

Table IA.XIII: Latent implementation bias in price-based factors (value-weighted, within-firm).

The table complements Table 1 by reporting within-firm results separately for all bonds, IG, and NIG samples. μ is the mean monthly return (%), α is the CAPMB alpha (%). Bias (1)−(2) compares standard versus signal-adjusted. Bias (1)−(3) compares standard versus return-adjusted. t-statistics (Newey-West, lags = $\lfloor T^{0.25} \rfloor$) in parentheses. Sample: 2002-09 to 2024-12, T=268.

	Unadjusted (1)		Adj. Signal (2)		Adj. Return (3)		Bias (1)−(2)		Bias (1)−(3)	
Factor	μ	α	μ	α	μ	α	$\Delta\mu$	$\Delta\alpha$	$\Delta\mu$	$\Delta\alpha$
Panel A: All Bonds										
ytm	0.56	0.31	0.30	0.06	0.26	0.04	0.26	0.25	0.30	0.27
	(4.42)	(4.08)	(2.85)	(1.44)	(2.68)	(0.94)	(5.42)	(5.54)	(5.60)	(5.58)
cs	0.63	0.47	0.26	0.11	0.24	0.10	0.38	0.36	0.40	0.37
	(5.57)	(4.43)	(3.10)	(1.72)	(3.11)	(1.62)	(6.34)	(6.36)	(6.42)	(6.11)
bbtm	0.38	0.28	0.19	0.10	0.17	0.09	0.19	0.18	0.21	0.19
	(4.02)	(4.11)	(2.50)	(1.86)	(2.32)	(1.64)	(6.05)	(6.58)	(5.64)	(5.56)
dcs6	−0.70	−0.65	−0.26	−0.22	−0.25	−0.22	−0.44	−0.43	−0.45	−0.43
	(−7.70)	(−8.08)	(−4.63)	(−4.98)	(−4.59)	(−4.92)	(−7.56)	(−7.44)	(−8.21)	(−8.25)
val_ipr	0.62	0.57	0.22	0.18	0.22	0.20	0.40	0.39	0.39	0.37
	(6.02)	(7.27)	(3.50)	(4.25)	(3.58)	(4.30)	(7.41)	(7.47)	(7.38)	(7.85)
val_hz	0.50	0.35	0.21	0.07	0.19	0.06	0.29	0.28	0.31	0.29
	(4.80)	(3.68)	(2.67)	(1.14)	(2.68)	(1.02)	(5.78)	(5.91)	(5.68)	(5.49)
str	−0.65	−0.61	−0.10	−0.07	−0.11	−0.10	−0.55	−0.54	−0.54	−0.52
	(−7.68)	(−7.12)	(−1.89)	(−1.36)	(−1.91)	(−1.69)	(−8.37)	(−8.51)	(−8.48)	(−8.77)
Panel B: Investment Grade										
ytm	0.49	0.22	0.22	−0.04	0.19	−0.06	0.27	0.25	0.31	0.28
	(3.77)	(2.44)	(2.10)	(−0.67)	(1.85)	(−1.15)	(5.33)	(5.44)	(4.73)	(4.36)
cs	0.62	0.43	0.21	0.04	0.18	0.02	0.40	0.39	0.43	0.41
	(5.00)	(3.44)	(2.41)	(0.57)	(2.27)	(0.31)	(6.07)	(5.93)	(5.99)	(5.36)
bbtm	0.38	0.28	0.16	0.08	0.15	0.08	0.21	0.20	0.22	0.20
	(4.09)	(3.99)	(2.57)	(1.51)	(2.58)	(1.57)	(5.56)	(5.85)	(4.73)	(4.59)
dcs6	−0.74	−0.68	−0.25	−0.21	−0.26	−0.22	−0.48	−0.47	−0.48	−0.47
	(−8.00)	(−7.99)	(−4.88)	(−4.70)	(−5.14)	(−5.40)	(−7.32)	(−7.56)	(−7.90)	(−7.59)
val_ipr	0.68	0.62	0.23	0.18	0.23	0.18	0.45	0.44	0.45	0.43
	(7.07)	(7.89)	(4.39)	(4.70)	(4.46)	(4.77)	(7.54)	(7.45)	(7.22)	(7.39)
val_hz	0.53	0.35	0.18	0.02	0.16	0.01	0.35	0.33	0.38	0.35
	(4.41)	(2.97)	(2.09)	(0.21)	(1.99)	(0.09)	(5.78)	(5.98)	(5.51)	(5.21)
str	−0.69	−0.64	−0.11	−0.07	−0.10	−0.08	−0.58	−0.57	−0.58	−0.56
	(−7.06)	(−6.30)	(−1.80)	(−1.13)	(−1.61)	(−1.20)	(−8.57)	(−8.17)	(−8.57)	(−8.56)
Panel C: Non-Investment Grade										
ytm	0.90	0.79	0.45	0.35	0.33	0.27	0.45	0.44	0.57	0.52
	(5.11)	(5.06)	(3.22)	(3.44)	(2.32)	(2.45)	(5.50)	(5.19)	(5.63)	(5.93)
cs	0.87	0.82	0.32	0.27	0.23	0.23	0.55	0.55	0.64	0.60
	(5.24)	(5.14)	(2.78)	(3.15)	(1.56)	(1.82)	(5.53)	(5.04)	(5.52)	(5.64)
bbtm	0.70	0.65	0.36	0.31	0.36	0.32	0.34	0.34	0.33	0.33
	(4.14)	(3.86)	(2.23)	(2.31)	(2.15)	(2.26)	(5.12)	(4.89)	(4.73)	(4.52)
dcs6	−0.94	−0.93	−0.29	−0.26	−0.31	−0.31	−0.66	−0.66	−0.64	−0.61
	(−7.44)	(−7.58)	(−2.68)	(−2.88)	(−3.21)	(−3.79)	(−6.34)	(−5.93)	(−7.87)	(−8.04)
val_ipr	0.62	0.64	0.20	0.26	0.14	0.19	0.42	0.38	0.48	0.45
	(3.41)	(3.65)	(1.46)	(1.93)	(1.00)	(1.25)	(6.14)	(5.84)	(6.05)	(6.40)
val_hz	0.57	0.49	0.25	0.14	0.25	0.21	0.33	0.34	0.32	0.28
	(4.37)	(4.26)	(2.35)	(1.72)	(3.04)	(2.34)	(4.24)	(3.89)	(3.83)	(3.76)
str	−0.91	−0.94	−0.20	−0.21	−0.11	−0.19	−0.72	−0.72	−0.80	−0.75
	(−9.13)	(−8.50)	(−2.75)	(−2.96)	(−1.34)	(−3.15)	(−7.46)	(−7.27)	(−7.59)	(−8.49)

Table IA.XIV: Latent implementation bias in illiquidity factors.

The table reports premia and CAPMB alphas for five illiquidity factors under standard and bias-adjusted approaches. Panel A sorts bonds into deciles each month, going long P10 and short P1. Panel B uses within-firm sorts. μ is the mean monthly return (%), α is the CAPMB alpha (%). Bias (1)−(2) compares standard versus signal-adjusted. Bias (1)−(3) compares standard versus return-adjusted. t-statistics (Newey-West, lags $= \lfloor T^{0.25} \rfloor$) in parentheses. Sample: 2002-09 to 2024-12, T=268.

	Unadjusted (1)		Adj. Signal (2)		Adj. Return (3)		Bias (1)−(2)		Bias (1)−(3)	
Factor	μ	α	μ	α	μ	α	$\Delta\mu$	$\Delta\alpha$	$\Delta\mu$	$\Delta\alpha$
Panel A: Single-Sort										
ilq	0.29	0.20	0.28	0.19	0.25	0.17	0.02	0.01	0.05	0.03
	(2.04)	(1.64)	(1.98)	(1.67)	(1.96)	(1.51)	(0.79)	(0.36)	(1.73)	(1.33)
ami	0.35	0.12	0.29	0.07	0.28	0.07	0.06	0.06	0.07	0.05
	(1.90)	(1.15)	(1.64)	(0.68)	(1.77)	(0.78)	(3.37)	(3.07)	(1.73)	(1.38)
cs_sprd	0.37	0.16	0.38	0.16	0.39	0.17	−0.01	−0.01	−0.02	−0.01
	(3.08)	(1.59)	(3.04)	(1.59)	(3.35)	(1.84)	(−0.99)	(−0.82)	(−1.04)	(−0.83)
ar_sprd	0.52	0.20	0.53	0.19	0.54	0.21	−0.00	0.00	−0.01	−0.01
	(2.79)	(1.44)	(2.77)	(1.42)	(2.95)	(1.55)	(−0.47)	(0.18)	(−0.54)	(−0.37)
spd_rel	0.23	0.10	0.23	0.11	0.28	0.16	−0.01	−0.00	−0.06	−0.05
	(1.71)	(0.99)	(1.70)	(1.06)	(2.17)	(1.56)	(−0.35)	(−0.19)	(−1.94)	(−2.33)
Panel B: Within-Firm Sort										
ilq	0.09	0.02	0.09	0.03	0.09	0.03	−0.00	−0.01	−0.00	−0.01
	(1.68)	(0.63)	(1.82)	(0.86)	(1.81)	(0.84)	(−0.29)	(−0.77)	(−0.28)	(−0.72)
ami	0.11	0.02	0.08	−0.01	0.08	0.01	0.03	0.03	0.02	0.01
	(1.98)	(0.70)	(1.50)	(−0.39)	(1.83)	(0.49)	(2.84)	(2.63)	(1.63)	(0.53)
cs_sprd	0.07	0.02	0.08	0.03	0.09	0.02	−0.01	−0.01	−0.01	−0.00
	(2.59)	(1.07)	(2.94)	(1.41)	(2.63)	(1.22)	(−1.33)	(−1.65)	(−1.42)	(−0.61)
ar_sprd	0.10	−0.03	0.10	−0.03	0.11	−0.01	0.00	0.00	−0.01	−0.01
	(1.82)	(−0.90)	(1.79)	(−1.04)	(2.00)	(−0.50)	(0.25)	(0.18)	(−1.05)	(−0.99)
spd_rel	0.06	−0.02	0.07	−0.01	0.08	0.00	−0.01	−0.01	−0.01	−0.02
	(1.48)	(−0.83)	(1.47)	(−0.41)	(1.83)	(0.05)	(−0.84)	(−1.10)	(−1.77)	(−2.26)

Table IA.XV: Look-ahead bias decomposition by leg (mean return).

The table reports mean returns across Long, Short, and Long-Short legs with ($\tilde{\mu}$) and without (μ) asymmetric ex-post return winsorization. Bias is the difference between the infeasible and feasible factor means. Panel A covers factors sensitive to left-tail winsorization (0.50^{th} percentile). Panel B covers factors sensitive to right-tail winsorization (99.50^{th} percentile). Factors have not been sign-corrected to have positive means. t-statistics use Newey-West standard errors with lags = $\lfloor T^{0.25} \rfloor$. Sample: 2002-08 to 2024-12, T=269.

	Long (%)			Short (%)			Long-Short (%)		
	$\tilde{\mu}_L$	μ_L	Bias	$\tilde{\mu}_S$	μ_S	Bias	$\tilde{\mu}_{\text{LS}}$	μ_{LS}	Bias
Panel A: Left-tail winsorized factors (0.50^{th} percentile)									
`b_dunc`	0.58	0.51	0.07	0.75	0.55	0.20	−0.17	−0.04	−0.13
	(2.47)	(2.17)	(2.28)	(3.30)	(2.19)	(2.32)	(−1.03)	(−0.22)	(−2.14)
`b_dunc3`	0.49	0.42	0.07	0.85	0.62	0.23	−0.36	−0.20	−0.16
	(2.53)	(2.11)	(2.69)	(3.09)	(2.04)	(2.04)	(−1.91)	(−1.04)	(−1.79)
`b_unc`	0.53	0.43	0.09	0.84	0.62	0.22	−0.32	−0.18	−0.13
	(2.77)	(2.20)	(2.61)	(2.81)	(2.03)	(2.00)	(−1.24)	(−0.78)	(−1.58)
`ltr48_12`	0.37	0.32	0.05	0.77	0.64	0.13	−0.40	−0.32	−0.08
	(2.52)	(2.15)	(2.17)	(3.52)	(2.75)	(2.43)	(−2.93)	(−2.21)	(−1.90)
`ltr30_6`	0.33	0.29	0.04	0.91	0.74	0.18	−0.58	−0.45	−0.14
	(2.23)	(1.91)	(2.62)	(3.22)	(2.44)	(2.41)	(−2.76)	(−1.98)	(−2.30)
`ivol_bbw`	0.99	0.69	0.30	0.13	0.12	0.00	0.87	0.57	0.29
	(2.96)	(1.90)	(2.73)	(2.35)	(2.27)	(1.05)	(2.84)	(1.68)	(2.75)
`ivol_vp`	0.89	0.63	0.26	0.11	0.11	0.00	0.78	0.52	0.26
	(3.07)	(1.95)	(2.37)	(2.62)	(2.52)	(1.17)	(2.99)	(1.76)	(2.39)
`b_dvix_vp`	0.57	0.50	0.07	0.59	0.48	0.11	−0.02	0.02	−0.04
	(2.94)	(2.55)	(3.21)	(2.79)	(2.20)	(2.28)	(−0.19)	(0.14)	(−1.36)
`b_psb_m`	0.62	0.53	0.09	0.55	0.39	0.16	0.07	0.15	−0.08
	(3.33)	(3.03)	(1.93)	(2.60)	(1.57)	(2.23)	(0.43)	(0.80)	(−2.36)
`b_amd_m`	0.46	0.38	0.08	0.75	0.62	0.13	−0.29	−0.24	−0.05
	(2.60)	(1.98)	(2.03)	(3.52)	(2.97)	(2.66)	(−1.99)	(−1.68)	(−2.93)
`var_95`	1.16	0.90	0.26	0.12	0.12	0.00	1.05	0.78	0.26
	(3.53)	(2.76)	(2.67)	(2.81)	(2.74)	(1.32)	(3.50)	(2.65)	(2.69)
`es_90`	1.36	1.01	0.35	0.11	0.10	0.00	1.26	0.91	0.35
	(3.56)	(2.67)	(2.38)	(2.53)	(2.45)	(1.21)	(3.53)	(2.59)	(2.38)
Panel B: Right-tail winsorized factors (99.50^{th} percentile)									
`mom3_1`	0.55	0.63	−0.08	0.22	0.46	−0.24	0.33	0.17	0.16
	(2.88)	(3.03)	(−2.58)	(0.82)	(1.61)	(−2.63)	(1.78)	(0.90)	(2.22)
`mom6_1`	0.49	0.54	−0.05	0.19	0.53	−0.35	0.30	0.00	0.30
	(2.65)	(2.79)	(−3.10)	(0.62)	(1.56)	(−2.28)	(1.45)	(0.01)	(2.03)
`mom12_1`	0.49	0.51	−0.02	0.22	0.65	−0.42	0.26	−0.13	0.40
	(3.34)	(3.43)	(−3.82)	(0.64)	(1.57)	(−2.29)	(0.92)	(−0.39)	(2.21)

Table IA.XVI: Look-ahead bias decomposition by leg (CAPMB alpha).

The table reports CAPMB alphas across Long, Short, and Long-Short legs with ($\tilde{\alpha}$) and without (α) asymmetric ex-post return winsorization. Bias is the difference between the infeasible and feasible factor alphas. Panel A covers factors sensitive to left-tail winsorization (0.50^{th} percentile). Panel B covers factors sensitive to right-tail winsorization (99.50^{th} percentile). Factors have not been sign-corrected to have positive means. t-statistics use Newey-West standard errors with lags = $\lfloor T^{0.25} \rfloor$. Sample: 2002-08 to 2024-12, T=269.

	Long (%)			Short (%)			Long-Short (%)		
	$\tilde{\alpha}_L$	α_L	Bias	$\tilde{\alpha}_S$	α_S	Bias	$\tilde{\alpha}_{\text{LS}}$	α_{LS}	Bias
Panel A: Left-tail winsorized factors (0.50^{th} percentile)									
b_dunc	0.14	0.05	0.08	0.36	0.09	0.27	−0.22	−0.04	−0.18
	(1.37)	(0.53)	(2.31)	(2.76)	(0.64)	(2.37)	(−1.38)	(−0.22)	(−2.20)
b_dunc3	0.07	−0.01	0.08	0.40	0.10	0.29	−0.32	−0.11	−0.21
	(0.99)	(−0.09)	(2.74)	(2.66)	(0.53)	(2.12)	(−1.87)	(−0.57)	(−1.88)
b_unc	0.15	0.05	0.11	0.35	0.06	0.29	−0.20	−0.02	−0.18
	(1.44)	(0.39)	(2.65)	(2.19)	(0.42)	(2.09)	(−0.91)	(−0.09)	(−1.68)
ltr48_12	0.01	−0.05	0.06	0.34	0.18	0.16	−0.33	−0.23	−0.10
	(0.28)	(−0.86)	(2.00)	(3.23)	(1.50)	(2.38)	(−2.57)	(−1.56)	(−1.65)
ltr30_6	−0.00	−0.05	0.05	0.43	0.20	0.22	−0.43	−0.26	−0.17
	(−0.02)	(−0.75)	(2.70)	(2.84)	(1.19)	(2.44)	(−2.37)	(−1.26)	(−2.32)
ivol_bbw	0.54	0.19	0.35	0.01	0.01	0.00	0.53	0.18	0.35
	(2.78)	(0.79)	(2.81)	(0.55)	(0.37)	(1.18)	(2.58)	(0.72)	(2.83)
ivol_vp	0.42	0.10	0.32	0.03	0.02	0.00	0.40	0.08	0.31
	(2.72)	(0.57)	(2.47)	(1.52)	(1.20)	(1.29)	(2.48)	(0.44)	(2.49)
b_dvix_vp	0.21	0.13	0.08	0.22	0.08	0.14	−0.01	0.04	−0.05
	(2.78)	(1.81)	(3.21)	(1.95)	(0.77)	(2.39)	(−0.09)	(0.38)	(−1.60)
b_psb_m	0.28	0.17	0.11	0.16	−0.05	0.22	0.12	0.23	−0.11
	(2.36)	(1.96)	(2.07)	(1.97)	(−0.39)	(2.34)	(0.85)	(1.35)	(−2.47)
b_amd_m	0.10	−0.00	0.10	0.41	0.26	0.15	−0.32	−0.26	−0.06
	(1.41)	(−0.01)	(2.19)	(3.09)	(2.42)	(2.71)	(−2.07)	(−1.74)	(−2.74)
var_95	0.58	0.26	0.33	0.02	0.02	0.00	0.56	0.24	0.32
	(3.20)	(1.58)	(2.70)	(1.24)	(1.02)	(1.50)	(3.01)	(1.42)	(2.71)
es_90	0.75	0.31	0.44	0.01	0.01	0.00	0.74	0.31	0.43
	(3.24)	(1.48)	(2.44)	(0.56)	(0.34)	(1.37)	(3.10)	(1.42)	(2.45)
Panel B: Right-tail winsorized factors (99.50^{th} percentile)									
mom3_1	0.20	0.27	−0.07	−0.25	−0.05	−0.20	0.45	0.31	0.13
	(1.78)	(2.25)	(−2.84)	(−1.47)	(−0.32)	(−2.63)	(2.26)	(1.65)	(2.10)
mom6_1	0.16	0.20	−0.04	−0.30	−0.02	−0.28	0.46	0.22	0.24
	(1.41)	(1.72)	(−3.30)	(−1.49)	(−0.09)	(−2.35)	(2.39)	(1.08)	(2.05)
mom12_1	0.19	0.21	−0.02	−0.26	0.07	−0.33	0.46	0.14	0.31
	(2.20)	(2.37)	(−4.06)	(−1.04)	(0.29)	(−2.43)	(1.60)	(0.51)	(2.33)

Table IA.XVII: Filter path improvement analysis by factor cluster, filter type, and tail location.

The table reports the count and percentage of filter configurations that improve a factor's CAPMB alpha relative to the unfiltered baseline. A filter improves if $t_\alpha > 1.96$, $t_\alpha > t_{\alpha,\text{baseline}}$, and $\alpha > \alpha_{\text{baseline}}$. Panel A reports improvement counts n and rates $p = n/N_{\text{total}} \times 100$. Panel B reports total filter paths (denominator). Panel C reports improving filter paths (numerator).

	Left Tail			Right Tail			Both Tails		
Signal Group	`trim`	`price`	`bounce`	`trim`	`price`	`bounce`	`trim`	`price`	`bounce`
Panel A: Improving filter configurations n (p%)									
Spreads, Yields, Size	11 (1%)	15 (3%)	5 (1%)	43 (5%)	42 (8%)	2 (0%)	15 (2%)	16 (3%)	0 (0%)
Value	100 (21%)	7 (2%)	5 (2%)	38 (8%)	6 (2%)	8 (3%)	73 (15%)	7 (2%)	0 (0%)
Momentum & Reversal	93 (5%)	35 (3%)	37 (3%)	48 (2%)	11 (1%)	48 (4%)	57 (3%)	36 (3%)	74 (6%)
Illiquidity	12 (1%)	23 (3%)	36 (5%)	74 (6%)	10 (1%)	8 (1%)	38 (3%)	21 (3%)	22 (3%)
Volatility & Risk	0 (0%)	0 (0%)	3 (0%)	30 (2%)	0 (0%)	0 (0%)	13 (1%)	0 (0%)	7 (1%)
Market Risk	7 (1%)	0 (0%)	17 (3%)	0 (0%)	4 (1%)	4 (1%)	9 (1%)	0 (0%)	5 (1%)
Credit & Default Risk	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Volatility & Liquidity Risk	34 (3%)	19 (2%)	16 (2%)	60 (5%)	31 (4%)	29 (4%)	47 (4%)	19 (2%)	17 (2%)
Macro & Other Risk	17 (1%)	5 (0%)	1 (0%)	1 (0%)	2 (0%)	6 (1%)	1 (0%)	5 (0%)	1 (0%)
Panel B: Total filter paths (denominator)									
Spreads, Yields, Size	864	540	540	864	540	540	864	540	540
Value	480	300	300	480	300	300	480	300	300
Momentum & Reversal	2016	1260	1260	2016	1260	1260	2016	1260	1260
Illiquidity	1248	780	780	1248	780	780	1248	780	780
Volatility & Risk	1536	960	960	1536	960	960	1536	960	960
Market Risk	864	540	540	864	540	540	864	540	540
Credit & Default Risk	384	240	240	384	240	240	384	240	240
Volatility & Liquidity Risk	1248	780	780	1248	780	780	1248	780	780
Macro & Other Risk	1728	1080	1080	1728	1080	1080	1728	1080	1080
Panel C: Improving filter paths (numerator)									
Spreads, Yields, Size	11	15	5	43	42	2	15	16	0
Value	100	7	5	38	6	8	73	7	0
Momentum & Reversal	93	35	37	48	11	48	57	36	74
Illiquidity	12	23	36	74	10	8	38	21	22
Volatility & Risk	0	0	3	30	0	0	13	0	7
Market Risk	7	0	17	0	4	4	9	0	5
Credit & Default Risk	0	0	0	0	0	0	0	0	0
Volatility & Liquidity Risk	34	19	16	60	31	29	47	19	17
Macro & Other Risk	17	5	1	1	2	6	1	5	1

Table IA.XVIII: Cross-sectional portfolio size by design dimension.

The table reports cross-sectional portfolio size across long-short legs for the methodological uncertainty design grid. Average, Median, Minimum (Min), and 5^{th} percentile (pct) are computed from the time-series distribution of portfolio size within each strategy and then averaged across strategies. %Low reports the average within-strategy share of months with fewer than 20 bonds. Tail-size bins classify strategies according to their time-series average combined tail size. Statistics are pooled over all 108 factors and all economically admissible portfolio constructions.

	Average	Median	Min	5^{th} pct	%Low	N_{paths}
Long–Short						
Portfolio granularity						
Terciles	1261.2	1168.1	515.3	656.9	0.0	6048
Quintiles	771.7	713.1	313.6	403.0	0.0	6048
Deciles	403.9	374.0	159.9	209.0	0.5	6032
Rating subsample						
All	1146.6	1044.3	478.6	616.6	0.0	7776
IG	856.0	801.6	339.6	410.1	0.0	5184
NIG	266.6	262.6	95.7	145.2	0.5	5168
Maturity bucket						
All	1618.0	1496.9	668.3	851.3	0.0	4536
Short	645.1	585.6	270.7	343.9	0.0	4536
Intermediate	544.9	535.3	222.0	284.1	0.0	4536
Long	441.2	389.1	157.4	212.6	0.6	4520
Breakpoint universe						
Full universe	734.4	681.4	293.7	374.5	0.2	7762
IG bonds	1195.2	1076.8	506.2	645.9	0.0	2592
Large bonds	763.2	714.3	306.9	397.5	0.1	7774
Tail size bins (by average LS portfolio size)						
<200	88.7	84.0	34.1	51.0	1.2	2258
200–600	374.9	352.6	140.1	190.0	0.0	7138
>600	1357.6	1251.4	561.2	710.0	0.0	8732
All specifications	812.6	752.1	329.7	423.2	0.2	18128

Table IA.XIX: Portfolio construction improvement analysis by factor cluster and methodology dimension.

The table reports the count and percentage of portfolio construction configurations that improve a factor's CAPMB alpha relative to the baseline. A configuration improves if $t_\alpha > 1.96$, $t_\alpha > t_{\alpha,\text{baseline}}$, and $\alpha > \alpha_{\text{baseline}}$. Panel A reports improvement counts n and rates $p = n/N_{\text{total}} \times 100$. Panel B reports total paths (denominator). Panel C reports improving paths (numerator).

	Rating			Maturity				Breakpoint Universe		
Signal Group	All	IG	NIG	All	Short	Mid	Long	All	IG	Large
Panel A: Improving specifications n (p%)										
Spreads, Yields, Size	59 (10%)	49 (11%)	10 (2%)	28 (9%)	29 (8%)	7 (2%)	54 (14%)	56 (9%)	18 (8%)	44 (7%)
Value	52 (16%)	39 (16%)	33 (14%)	12 (7%)	36 (17%)	2 (1%)	74 (36%)	55 (17%)	20 (17%)	49 (14%)
Momentum & Reversal	86 (6%)	32 (3%)	60 (6%)	21 (3%)	61 (7%)	13 (1%)	83 (9%)	74 (5%)	19 (4%)	85 (6%)
Illiquidity	91 (11%)	29 (5%)	35 (6%)	21 (4%)	48 (9%)	15 (3%)	71 (13%)	66 (8%)	33 (11%)	56 (6%)
Volatility & Risk	33 (3%)	16 (2%)	51 (7%)	12 (2%)	25 (4%)	20 (3%)	43 (6%)	47 (4%)	12 (3%)	41 (4%)
Market Risk	21 (4%)	0 (0%)	24 (6%)	1 (0%)	24 (6%)	8 (2%)	12 (3%)	21 (4%)	4 (2%)	20 (3%)
Credit & Default Risk	7 (3%)	0 (0%)	0 (0%)	3 (2%)	3 (2%)	0 (0%)	1 (1%)	2 (1%)	5 (5%)	0 (0%)
Volatility & Liquidity Risk	42 (5%)	38 (6%)	10 (2%)	34 (7%)	2 (0%)	6 (1%)	48 (9%)	30 (3%)	19 (6%)	41 (4%)
Macro & Other Risk	36 (3%)	21 (2%)	30 (3%)	13 (2%)	15 (2%)	14 (2%)	45 (6%)	33 (3%)	11 (3%)	43 (3%)
Panel B: Total specifications (denominator)										
Spreads, Yields, Size	594	432	428	324	378	378	374	590	216	648
Value	330	240	238	180	210	210	208	330	120	358
Momentum & Reversal	1386	1008	1000	756	882	882	874	1378	504	1512
Illiquidity	858	624	624	468	546	546	546	858	312	936
Volatility & Risk	1056	768	768	576	672	672	672	1056	384	1152
Market Risk	594	432	432	324	378	378	378	594	216	648
Credit & Default Risk	264	192	192	144	168	168	168	264	96	288
Volatility & Liquidity Risk	858	624	624	468	546	546	546	858	312	936
Macro & Other Risk	1188	864	862	648	756	756	754	1186	432	1296
Panel C: Improving specifications (numerator)										
Spreads, Yields, Size	59	49	10	28	29	7	54	56	18	44
Value	52	39	33	12	36	2	74	55	20	49
Momentum & Reversal	86	32	60	21	61	13	83	74	19	85
Illiquidity	91	29	35	21	48	15	71	66	33	56
Volatility & Risk	33	16	51	12	25	20	43	47	12	41
Market Risk	21	0	24	1	24	8	12	21	4	20
Credit & Default Risk	7	0	0	3	3	0	1	2	5	0
Volatility & Liquidity Risk	42	38	10	34	2	6	48	30	19	41
Macro & Other Risk	36	21	30	13	15	14	45	33	11	43

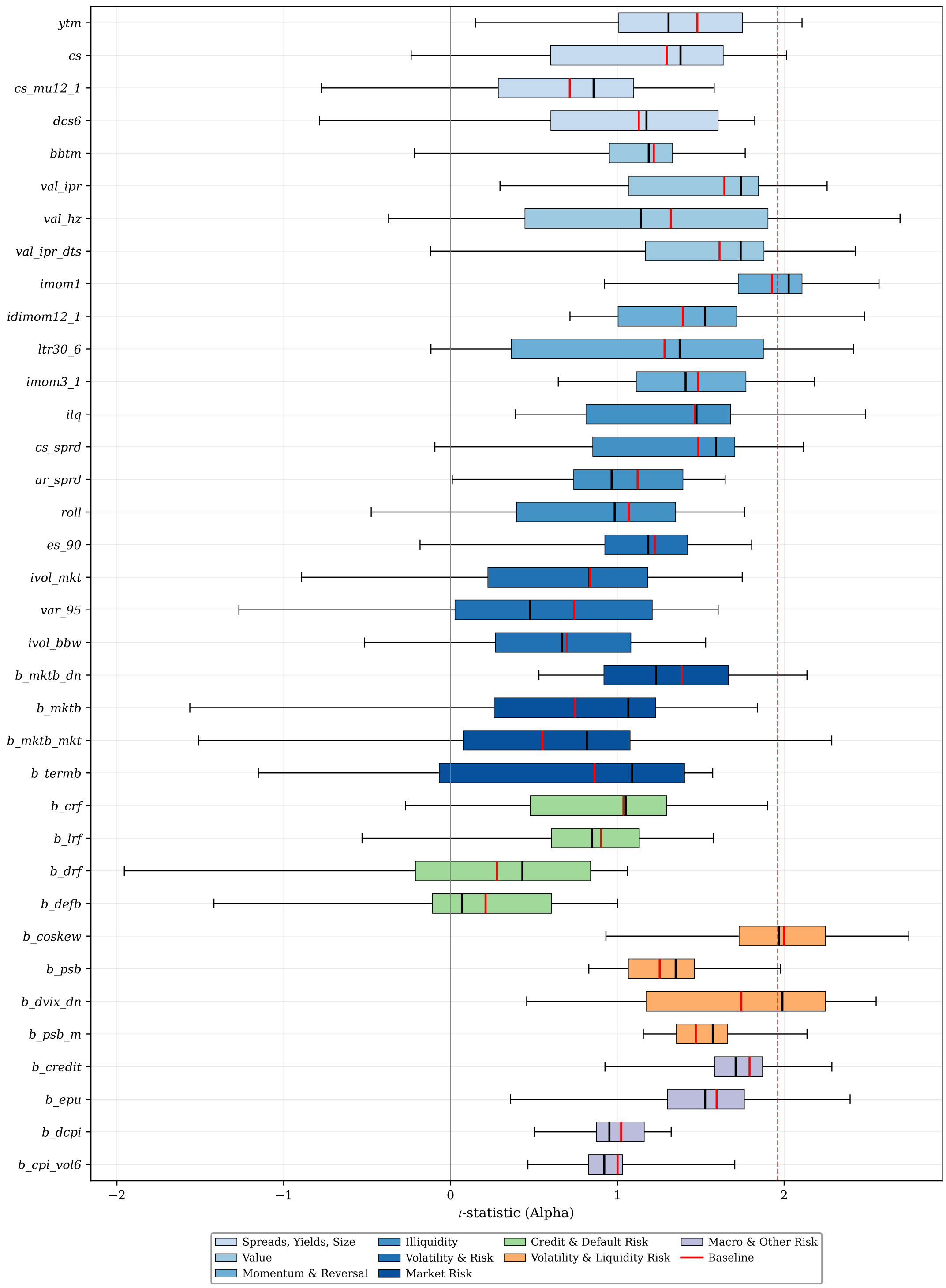

Figure IA.3: CAPMB alpha t-statistics across filter paths.

The figure shows the t-statistics for the CAPMB alpha (Newey-West, lags = $\lfloor T^{0.25} \rfloor$) in box plots for the top 4 sorting variables per group across all 648 filter paths. The boxes cover the interquartile range (IQR) and thus correspond to the non-standard errors (NSE). The lines at the ends of each box indicate the minimum and maximum of each distribution. The black vertical line inside each box shows the median t-statistic. The red solid vertical line shows the baseline t-statistic (averaged across ratings and weightings). The color scheme connects each sorting variable to its respective group.

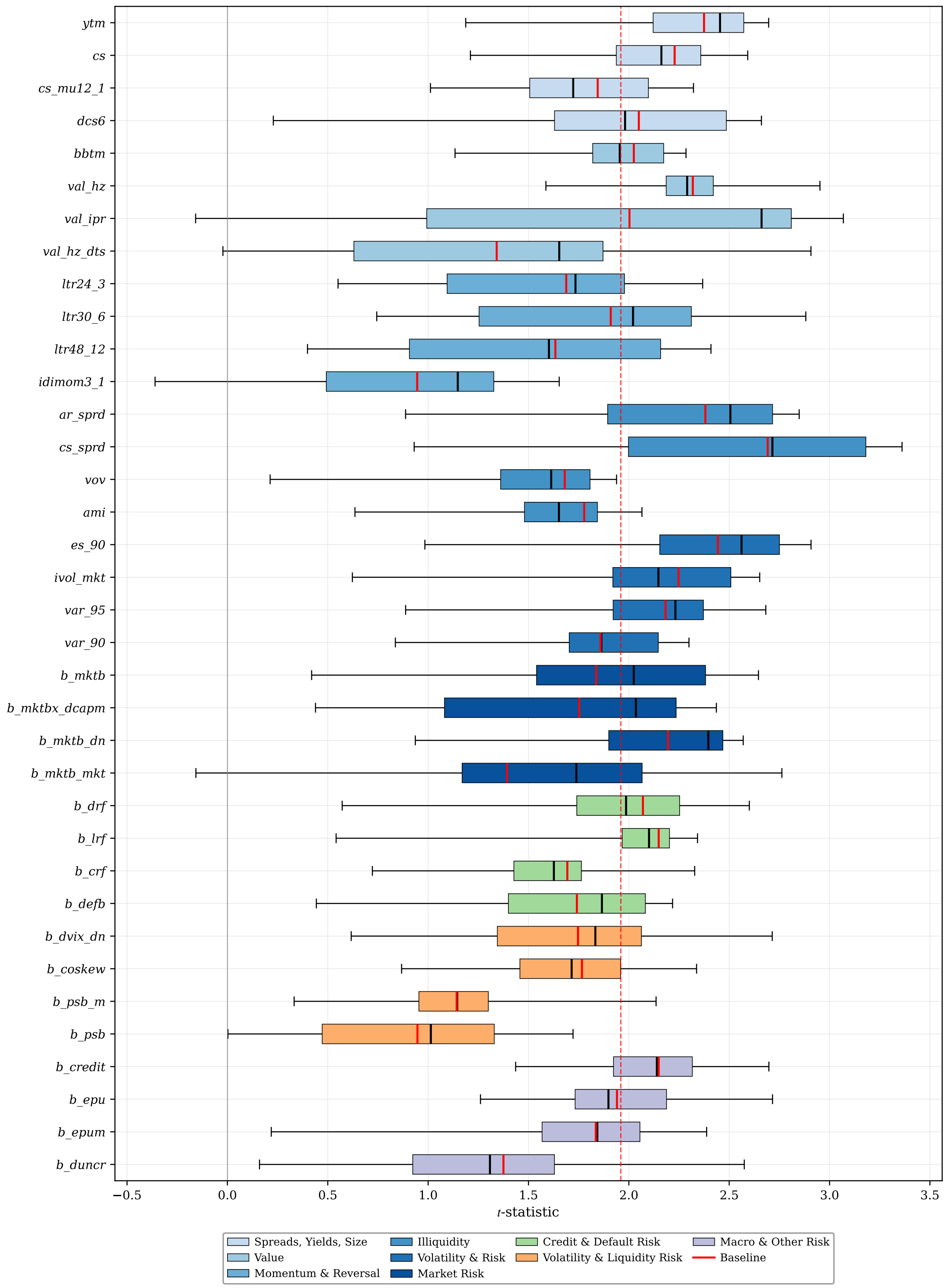

Figure IA.4: Premium t-statistics across filter paths.

The figure shows the t-statistics (Newey-West, lags = $\lfloor T^{0.25} \rfloor$) in box plots for the top 4 sorting variables per group across all 648 filter paths. The boxes cover the interquartile range (IQR) and thus correspond to the non-standard errors (NSE). The lines at the ends of each box indicate the minimum and maximum of each distribution. The black vertical line inside each box shows the median t-statistic. The red solid vertical line shows the baseline t-statistic (averaged across ratings and weightings). The color scheme connects each sorting variable to its respective group.

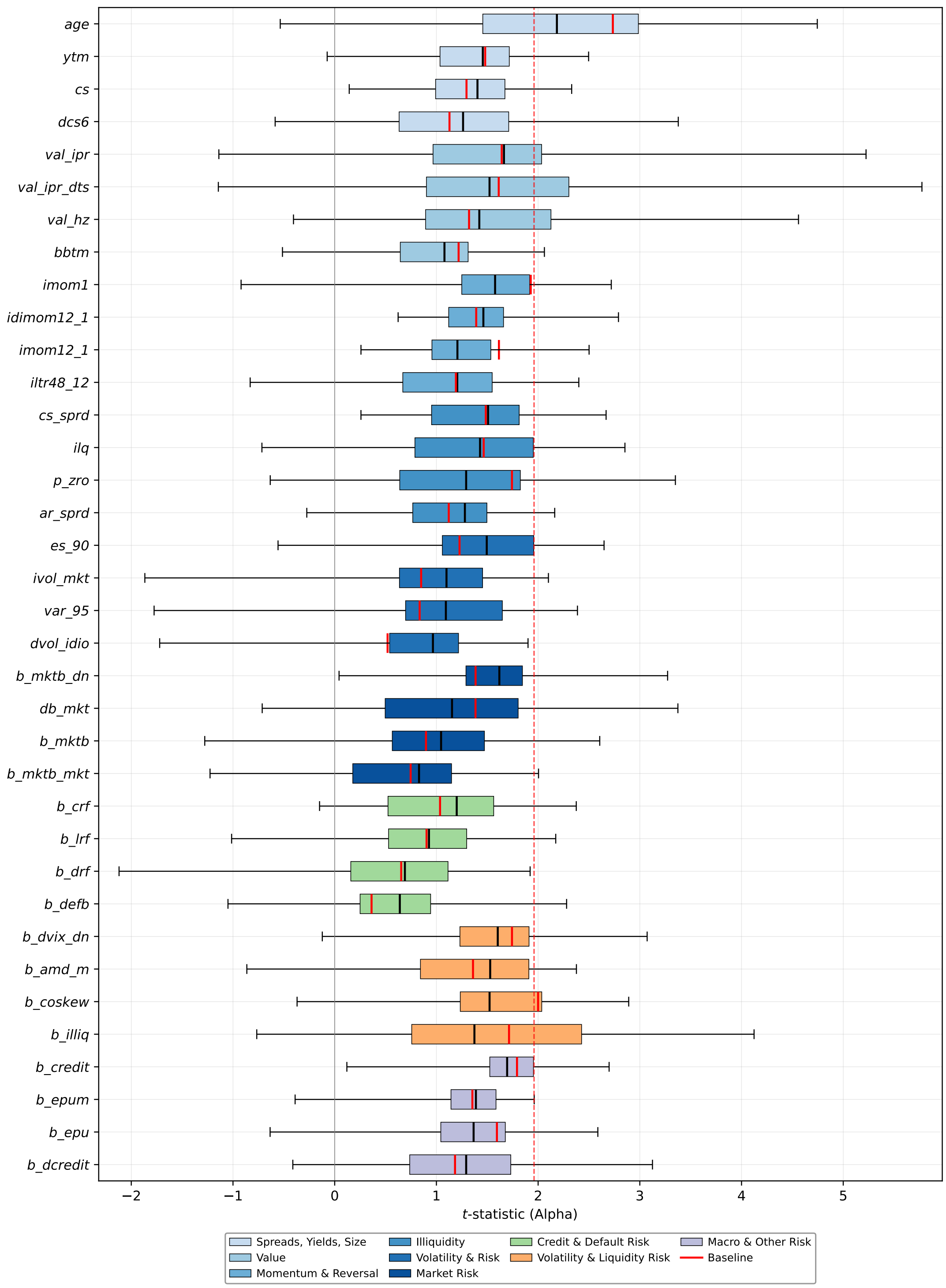

Figure IA.5: CAPMB alpha t-statistics across portfolio construction paths.

The figure shows the t-statistics for the CAPMB alpha (Newey-West, lags = $\lfloor T^{0.25} \rfloor$) in box plots for the top 4 sorting variables per group (nine groups in total) across all 168 portfolio construction paths. The boxes cover the interquartile range (IQR) and thus correspond to the non-standard errors (NSE). The lines at the ends of each box indicate the minimum and maximum of each distribution. The black vertical line inside each box shows the median t-statistic. The red solid vertical line shows the baseline t-statistic (averaged across ratings and weightings). The color scheme connects each sorting variable to its respective group.

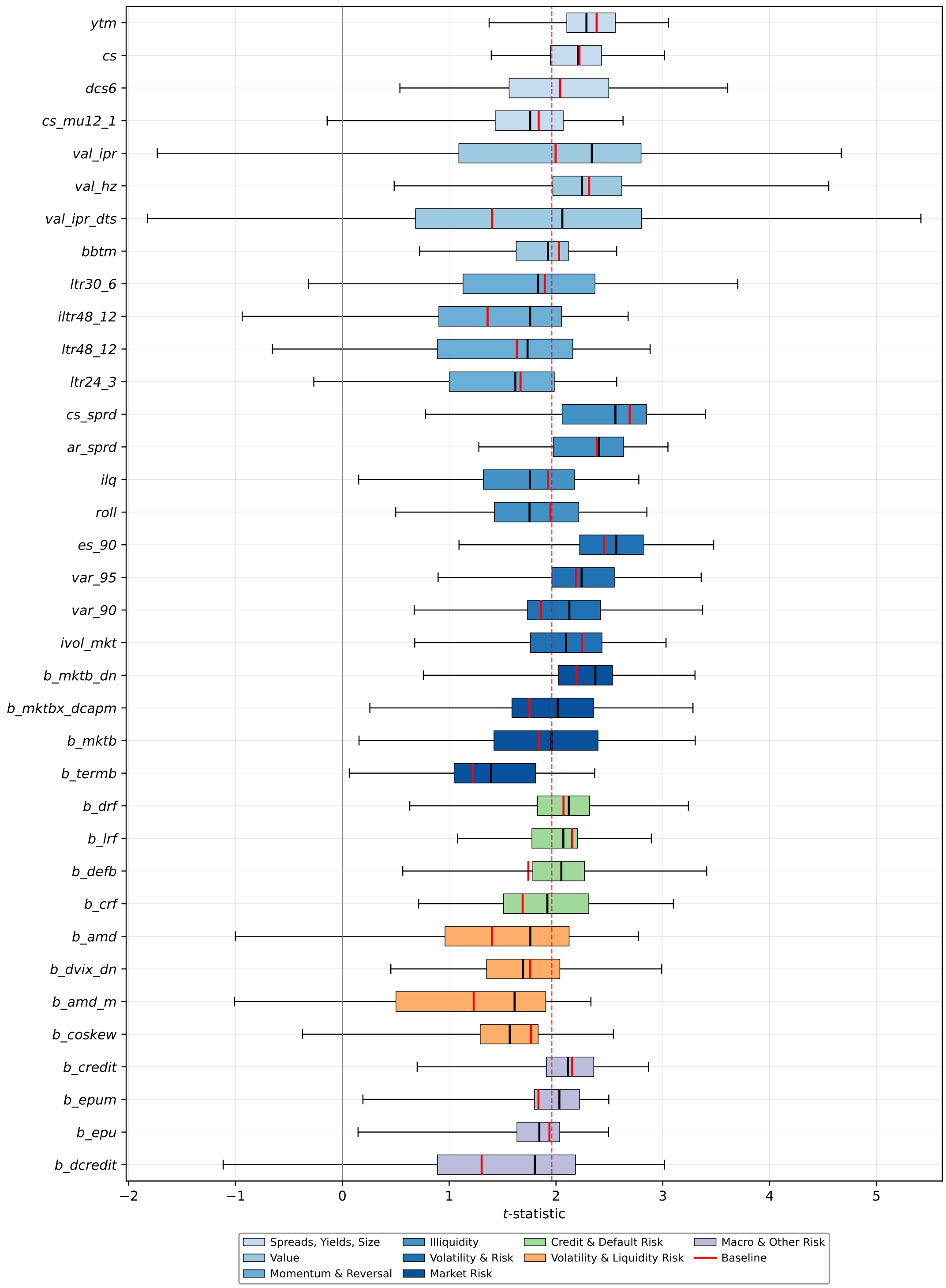

Figure IA.6: Premium t-statistics across portfolio construction paths.

The figure shows the t-statistics (Newey-West, lags = $\lfloor T^{0.25} \rfloor$) in box plots for the top 4 sorting variables per group across all 168 portfolio construction paths. The boxes cover the interquartile range (IQR) and thus correspond to the non-standard errors (NSE). The lines at the ends of each box indicate the minimum and maximum of each distribution. The black vertical line inside each box shows the median t-statistic. The red solid vertical line shows the baseline t-statistic (averaged across ratings and weightings). The color scheme connects each sorting variable to its respective group.