Discussion Paper Series A No.771

Factor Investing with Delays

Alexander Dickerson
(UNSW Business School)
Yoshio Nozawa
(University of Toronto)
and
Cesare Robotti
(Warwick Business School)

July 2025

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan

Factor Investing with Delays*

Alexander Dickerson[†] Yoshio Nozawa[‡] Cesare Robotti[§]

July 23, 2025

Abstract

We present a tractable framework for evaluating the cost of delays induced by infrequent trading in the corporate bond market. Using 341 corporate bond factors from OpenBondAssetPricing.com and machine learning models trained on their underlying signals, we demonstrate that, before transaction costs, 51 factors outperform the bond market. However, this number drops to nearly zero after accounting for trading frictions because the cost of delay is amplified for highly profitable factors. Trading a subset of liquid bonds does not eliminate this cost because liquidity is hard to predict and sales delays cannot be avoided, underscoring the critical impact of delay costs.

JEL Classification: G12, G13

Keywords: Corporate Bonds, Liquidity, Market Efficiency, Fixed-Income Securities, Credit Risk, Machine-Learning

---

*We thank Kevin Aretz, Pedro Barroso (discussant), Andrew Chen, Mike Chernov, Jaewon Choi, Oleg Chuprinin, Sean Foley (discussant), Nils Friewald, Christian Julliard, Ai He (discussant), Grace Hu (discussant), Alexey Ivashchenko (discussant), Jane Li, Hai Lin (discussant), Florian Nagler, Savina Risova, Norman Schürhoff, Sean Shin, Konstantin Sokolov (discussant), Sam Wang, Rossen Valkanov, Xintong Zhan (discussant), Alex Zhou, and seminar and conference participants at the ABFER, Atlanta Fed, ESSFM 2024 at Gerzensee, FIRS, Northern Finance Association, Southern Methodist University, University of Wisconsin-Madison, Temple University, Texas A&M University, University of British Columbia, University of Georgia, University of Illinois at Urbana-Champaign, University of Kansas, University of New South Wales, University of North Carolina at Chapel Hill, University of Technology Sydney, Monash University, University of Toronto, University of California (UCLA), World Symposium for Investment Research, Bristol Financial Markets Conference, University of York Asset Pricing Workshop, 13th FIRN Annual Conference, Wellington Finance Summit, KAIST-KU Virtual Finance seminar, and Robeco Quant Fixed Income team for helpful comments. We also thank Julien Vincent for cloud computing support. A part of this research was conducted while Nozawa visited the Institute of Economic Research at Hitotsubashi University.

†UNSW Business School, alexander.dickerson1@unsw.edu.au.

‡University of Toronto, yoshio.nozawa@rotman.utoronto.ca.

§Warwick Business School, cesare.robotti@wbs.ac.uk

Conflict-of-Interest Disclosure Statement

Alexander Dickerson

I have nothing to disclose.

Yoshio Nozawa

I have nothing to disclose.

Cesare Robotti

I have nothing to disclose.

1 Introduction

Factor investing research has a long tradition dating back to Fama and French (1993). Typically, researchers evaluate proposed strategies using historical data, assuming that the representative investor is endowed with abundant trading opportunities and can trade instantaneously, without delay. This assumption may be justified in continuously traded, limit order book-driven equity markets, where hedge funds and other active investors trade frequently. However, this assumption is rendered questionable for illiquid assets, such as corporate bonds, where trading occurs in fragmented, over-the-counter (OTC) markets. These markets are characterized by relatively sparse trade frequency and the resulting trading delays are likely to significantly alter the performance of factor strategies.

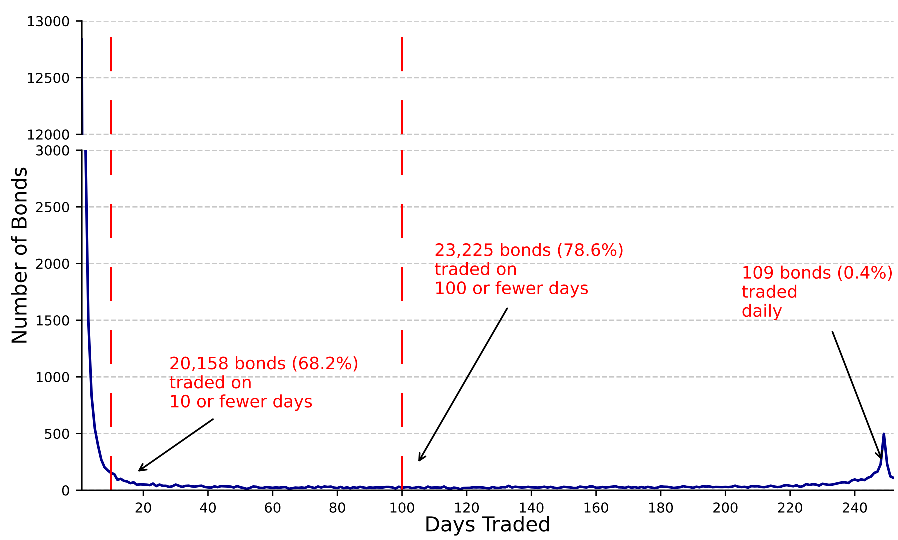


Figure 1: Distribution of Bonds by Number of Trading Days over 2022:01–2022:12.

As Fig. 1 illustrates, of the 29,543 publicly traded bonds in 2022, ∼70% of them traded on 10 business days or *less*. Only 109 bonds (0.4% of the sample) traded daily. Despite trade infrequency being a dominant characteristic of the corporate bond market, its implications for factor investing remain largely unexplored. In this paper, we fill this gap and comprehensively evaluate the performance of corporate bond factors after accounting for delays. Delays are costly because the predictive value of signals used in factor investing decays over time. With the delay, investors end up trading with stale signals or not trading at all, which lowers the average returns of the factors. The difference in average returns between factors with and without delays is the 'cost of delay'.

To transparently characterize the cost of delay, we simulate data on returns and delays, where delays are generated from an exponential distribution. The simulation enables us to quantify the cost of delay given a set of parameters, including the portfolio turnover rate and average excess returns of the factor. In particular, we show that the trading intensity parameter of the exponential distribution, which measures the probability of a trade occurring in a given period, is key to determining the delay cost. Higher trade intensity reduces the possibility of trade execution failure and thus the cost of delay. Consequently, measuring delay costs crucially depends on estimates of trade intensity.

Our goal is to estimate the delay cost for an active factor investor. Thus, instead of taking the historical trade frequency in Fig. 1 as given, we take into account the difference between the trading intensity of active factor investors and that of buy-and-hold investors who both generate the data. In addition, our procedure accounts for potential trade-offs between delay and bid-ask spreads facing investors who optimally choose to delay, and provides a lower bound for total transaction costs.[1]

Specifically, in our main results, we use two estimates of trade intensity: our own estimate based on historical transaction data in TRACE, and another estimate based on order-level data from an electronic trading platform, MarketAxess, provided by Kargar, Lester, Plante, and Weill (2025). We estimate trade intensity by running predictive regressions of the trade dummy on Li and Yu (2025)'s investor composition measure, which represents the

[1]In many OTC markets, including corporate bonds, large client orders enjoy lower half-spreads relative to smaller orders (Edwards, Harris, and Piwowar, 2007). This phenomenon is known as the 'size discount'.

average trading activity of the holders of a given bond. Our model-based trade intensity is computed using the 99$^{\text{th}}$ percentile values of the investor composition measure to obtain the most conservative measure of delay. This intensity estimate represents the delay facing active factor investors, who are most likely to demand high immediacy.

Armed with the estimated costs of delays and bid-ask spreads, we study the performance of corporate bond factor investing using 341 characteristics based on the Intercontinental Exchange (ICE) and TRACE corporate bond pricing data. We make our data library publicly available on `openbondassetpricing.com/machine-learning-data`. To complement the publicly available data offering, and to aid replicability in the field, our factors can be reproduced with the newly developed `PyBondLab` package in Python.[2] We use this package to calculate gross average returns and portfolio turnover rates. We then assign the simulated delay cost to each of the 341 factors by adjusting the simulation parameters based on the factor's portfolio turnover rate and average excess returns. Total transaction costs are the sum of the delay costs and the standard measure of bid-ask spread costs, which is obtained by multiplying the half spread by the portfolio turnover rate.

We have four main findings. First, pre-transaction costs, 58 of the 341 factors yield statistically significant average returns, with 42 generating significant bond market CAPM alpha. For ease of interpretation, we group the 58 bond factors into six categories based on the characteristics they are formed on: Equity Momentum, Equity Reversal, Investment, (Credit) Spreads, Value/Profits, and Volatility/Risk. Spread-based factors yield the highest

[2]This package is made available to the public and can be installed on any machine using pip (`pip install PyBondLab`) to produce the factors used in this study. The package includes functionality to compute bond portfolios rebalanced over any given horizon, as well as turnover, which correctly accounts for endogenous changes in portfolio value. Furthermore, various tools are provided to dampen portfolio turnover, including a buy-hold-spread (known as banding) and staggered holding periods. Detailed tutorials, examples, and documentation can be found on the `PyBondLab` download webpage. The software can be used in conjunction with the publicly available factor characteristic data and bond returns that we make available on `openbondassetpricing.com/machine-learning-data`. In addition, the time-series of our factor returns (341 of them) and monthly turnovers that are used in the main results (ICE returns) and robustness (TRACE returns) are available for download here.

average CAPM alpha (0.44% monthly), followed by Equity Reversal factors (0.43% monthly). Across all 58 significant factors, the average CAPM alpha is 0.28%.

Second, the delay and bid-ask spread costs are both economically large, and mostly eliminate these alphas. Starting with the Spread category, the delay cost is 0.11% per month for large trades, which accounts for about one quarter of its gross (pre-transaction cost) alpha. Our second-best performing category, Equity Reversal, yields substantially higher delay costs at 0.22% per month, which exceeds half of its gross alpha. Notably, these delay costs exceed both the bid-ask spread costs (0.09% and 0.20%) associated with these two categories, highlighting the significance of delays for the best-performing factors. The delay cost is also nontrivial for the Equity Momentum (0.11%) and Volatility/Risk (0.06%) categories. On the other hand, it is negligible for the Investment (0.01%) and Value/Profit (0.02%) categories that are based on slower moving Compustat accounting data.

The impact of delays is greatest for the Equity Reversal category because the factors therein are highly profitable based on signals that change quickly month-to-month. Unlike bid-ask spread costs, the delay cost is greater for more profitable factors because the opportunity cost of not trading immediately upon observing the signal is larger.[3] The Reversal factors' high turnover rate exacerbates this cost. With smaller trade sizes, delay costs decrease, but this reduction comes at the expense of increased bid-ask spreads. Taken together, we find that the average CAPM alpha of the 58 factors drops to 0.09% and 0.05% using large and small trades, with the number of significant factors of 2 and 1, respectively. The delay costs using the trading intensity estimates of Kargar et al. (2025) lead to a very similar conclusion. After accounting for delay and bid-ask spread costs, the average CAPM alpha is 0.09% and 0.04% for large and small trades, respectively.

Third, standard machine-learning models that optimally combine the 341 characteristics

[3]As part of our main results, we empirically verify that average factor returns and alphas exhibit more rapid decay for factors with higher profitability.

exacerbate the cost of delay. We follow Gu et al. (2020) and train nine machine learning (ML) models based on the 341 characteristics and compute model-based expected returns at the bond level. The long-short strategies based on the expected returns as signals are our machine learning based factors.

These ML models generate impressive out-of-sample performance before transaction costs, with an average CAPM alpha of 0.82% (9.89% per annum) and t-statistics above 4. Moreover, all of the ML-based factors generate significant bond CAPM alphas after subtracting bid-ask spread costs. However, the models tend to place a large weight on signals that move quickly and have a high predictive ability, such as past stock returns of the bond issuer. As a result, the cost of delay for the ML model-based factor is high, ranging from 0.17% to 0.36% for large trades and 0.09% to 0.19% for small trades. After adjusting for the cost of delay, the CAPM alphas of the ML-based factors become only marginally significant, averaging 0.33% for large trades and 0.35% for small trades, a significant decline from the pre-cost alpha of 0.82%. The large decline in alpha showcases the importance of accounting for delays, particularly for factors with impressive performance before transaction costs.

Finally, various execution strategies designed to reduce delays fail to eliminate transaction costs. We examine three approaches. First, we partition large $2 million orders into smaller tranches. This strategy yields insignificant cost reductions, as widening bid-ask spreads from trading in smaller size offset the benefits obtained from reduced execution delays. Second, inspired by Chaudhary et al. (2023), we treat corporate bonds in each portfolio as close substitutes and trade only those that offer faster execution. However, this method also fails to significantly reduce delay costs because sales delays cannot be fully eliminated, and the factor Sharpe ratio declines due to poor portfolio diversification from owning fewer bonds. Third, we extend the holding period to three months and introduce inertia in portfolio rebalancing, as proposed in Novy-Marx and Velikov (2019). This approach also fails to improve net-of-cost factor returns, as the reduced delay costs are offset by declining factor performance from a slower rate of rebalancing.

Our paper contributes to the rapidly growing literature that evaluates (and re-evaluates) the performance of factor investing in the corporate bond market (e.g., Bali et al. 2020; Augustin et al. 2020; Kelly et al. 2023; Sandulescu 2022; van Binsbergen et al. 2025; Dickerson et al. 2023; Dick-Nielsen et al. 2023). The paper closest to ours is Ivashchenko and Kosowski (2024), which studies the performance of nine factors after accounting for transaction costs. Our paper differs from Ivashchenko and Kosowski (2024) in that we highlight the novel delay costs faced by investors and employ a comprehensive factor data library in testing the performance of factor models.[4]

This paper also relates to the extensive literature measuring illiquidity and transaction costs in the corporate bond market (e.g., Edwards et al. 2007; Chen et al. 2007; Feldhütter 2010; Bao et al. 2011; Schestag et al. 2016; Bao et al. 2018; Bessembinder et al. 2018; Dick-Nielsen and Rossi 2018; O'Hara and Zhou 2021; Hendershott et al. 2021, 2022; Wu 2022; Choi et al. 2024; Pinter et al. 2024). In particular, Kargar et al. (2025) measures execution delays at the lower bound using relatively liquid corporate bonds traded on electronic exchanges.[5] Other closely related papers include O'Hara et al. (2018), who examine market power in determining corporate bonds' half spreads, as well as Goldstein and Hotchkiss (2020) and Reichenbacher and Schuster (2022), who argue that observed transaction costs strongly depend on transaction size and dealers' strategic inventory management. However, none of these papers quantify the impact of trading delays in evaluating trading strategies.[6]

Our paper aims to provide a set of best practices in accounting for transaction costs

[4]In our companion paper, Dickerson, Nozawa, and Robotti (2024), we measure delays using an alternative methodology. However, this method has the limitation of taking the observed trade frequency as given. In the current paper, we overcome this limitation by adopting a simulation-based approach that offers greater flexibility in analyzing trading patterns. Either way, the main results are closely aligned: pinning down a profitable net-of-cost corporate bond factor is challenging.

[5]More broadly, there is a strand of literature that studies the role of liquidity and dealer inventory in explaining credit spreads and bond risk premiums. This body of research includes Lin et al. (2011); Friewald and Nagler (2019); He et al. (2019); Goldberg and Nozawa (2021); Eisfeldt et al. (2023).

[6]In the CLO market, Hendershott et al. (2024) measure delays after accounting for trade failures.

in the study of the cross-section of corporate bond returns. Table 1 summarizes recent papers examining factor investing in corporate bond markets and reports the holding period length (in months) required to obtain the main results. The majority of prior research assumes one-month rebalancing with immediate order execution, an assumption that appears inconsistent with the infrequent trading activity observed in corporate bond markets, as illustrated in Fig. 1. Moreover, even for the papers that account for transaction costs, they only consider a variant of bid-ask spreads, not delay costs.[7] This paper reevaluates bond factor performance by highlighting a 'paradox' in realistic factor implementation. While factors based on fast-moving, cross-sectionally predictive signals generate substantial gross profits, these same factors face the largest execution challenges due to limited liquidity in corporate bond markets.

The remainder of the paper is organized as follows: Section 2 describes our dataset; Section 3 provides detailed methods for calculating delay costs using simulated data; Section 4 estimates the trade intensity and expected delays; Section 5 provides the evaluation of the factor performance; Section 6 assesses the machine learning models; Section 7 extends the baseline results with more advanced factor investment strategies; and Section 8 provides concluding remarks.

2 Data

2.1 Data for Constructing Bond Factors

Our datasets include daily bond data for the constituent bonds of the Bank of America (BAML) Investment Grade and High Yield indices as made available via the Intercontinental Exchange (ICE). Factor returns are computed using quote-based ICE data, which does not

[7] The importance of transaction costs in other asset classes, such as stocks and options, are documented in Novy-Marx and Velikov (2015), Novy-Marx and Velikov (2019), Chen and Velikov (2023), Detzel et al. (2023), Avramov et al. (2023), and Muravyev et al. (2024).

have missing observations even if a bond is not traded exactly on month end. This feature allows us to estimate the portfolio turnover rate accurately.[8] In robustness exercises, we also compute returns with the TRACE transaction-based data and observe a very closely aligned set of results.[9] We source equity and accounting data from CRSP and Compustat. We filter the data using standard approaches prescribed by the literature that are explicitly described in Internet Appendix A. To form bond factors and train the machine learning models, we construct commonly used bond and equity variables used in the literature and then merge these to the equity-based characteristics from Chen and Zimmermann (2022, CZ), and Jensen, Kelly, and Pedersen (2023, JKP).[10] This data combines several monthly bond and stock-based characteristics that have been shown to predict one-month-ahead corporate bond excess returns. Our data includes 341 characteristics, of which 40 are bond-based characteristics and 301 are equity-based, covering the majority of corporate bond return predictors used in prior research. This data is publicly available for download on `openbondassetpricing.com/data`.

Detailed descriptions for the construction of our 53 custom-made bond and stock characteristics are provided in Table A.1 of the Internet Appendix. Extensive documentation for the CZ and JKP equity characteristics is available on the respective authors' websites. Missing characteristic data is set equal to its cross-sectional median at each month t. All characteristics are cross-sectionally ranked and then scaled to lie in the interval $[-1,1]$. Overall, the data used to train the ML models with the 341 stock and bond characteristics comprises 19,768 bonds issued by 2,110 firms over the sample period from January 1998 to December 2022.

Because we need return data to construct factors and train the machine learning models,

[8] Missing transaction data induces additional portfolio turnover because the bond drops out of the tradable universe at the portfolio formation date, artificially inflating turnover.

[9] These results are similar if we use the WRDS version of TRACE or our own version which has been checked for potential data errors.

[10] Available on `openassetpricing.com` and WRDS, respectively.

all factors, whether characteristic-based or ML-based, span the sample period from August 2002 to December 2022 ($T = 244$). That is, the sample prior to that is used only to generate signals and train and validate the ML models.

2.2 Data for Trade Intensity Estimates

We use Enhanced TRACE to estimate trading intensity and half spreads. We filter the TRACE data by removing trades that are i) when issued, ii) with special conditions, iii) locked in, iv) have more than two days to settlement. We also remove observations with extreme price reversals: if the product of two consecutive logarithmic price changes is less than -0.25 (e.g., a 50% increase followed by a -50% decrease), then the observations in the middle are dropped. To estimate trade intensity, we use only dealer-customer transactions.[11] Finally, we use eMAXX for bond ownership information.

To measure the frequency at which bonds are traded, we construct a panel dataset with an indicator dummy for the incidence of customer buys and sells. We treat missing customer buy/sell trade observations in TRACE as no-transactions (zero trades), assigning zero to the indicator. To distinguish no-transactions from missing observations, we first define the set of bonds in the sample, which is the intersection of the bonds that exist in ICE, the WRDS bond database (which is constructed from TRACE), and eMAXX. For each bond in the intersection, we create a contiguous monthly time series based on the first and last month the bond appears in ICE. We then merge the trading information in TRACE with this resampled panel data to determine the month in which the bond exists but is not traded. After the merge, our sample has 814,145 bond-month observations for 14,418 bonds. The sample period is August 2002 to September 2021, with the sample end determined by the availability of the eMAXX data.

[11] This filter is similar to Bessembinder et al. (2008) but our band is wider than theirs (+/− 20%) to minimize the risk of deleting correct observations.

3 Simulated Cost of Trade Delays

3.1 Cost of Delays

In this section, we estimate the cost of delays using simulated data. The simulation method is necessary to distinguish no-trade days in TRACE that result from investor inactivity from those that occur due to dealers' unwillingness to provide liquidity. It also demonstrates that the key determinants of the delay cost are the cross-sectional predictive ability of the underlying characteristic (signal) used to form the factor and the speed of the signal decay. In the empirical analysis, we use these determinants to assign this simulated cost to the factors discovered in the data.

To this end, we consider a stylized model of factor investing. We suppose that bond i's gross return consists of predictable components and unpredictable shocks,

$$r_{i,t+1} = \beta_{i,t} f_{t+1} + u_{i,t+1}, \tag{1}$$

$$\beta_{i,t+1} = \rho \beta_{i,t} + v_{i,t+1}, \tag{2}$$

where $u_{i,t+1} \sim_{i.i.d.} \mathcal{N}(0, \sigma_u^2)$ and $v_{i,t+1} \sim_{i.i.d.} \mathcal{N}(0, \sigma_v^2)$. The variable $\beta_{i,t}$ is an observable signal for investors. We have a systematic risk factor $f_t \sim_{i.i.d.} \mathcal{N}(E[f], \sigma_f^2)$ in the dynamics whose risk premium determines the Sharpe ratio of the strategy. While we obtain the main results assuming constant expected returns, we consider time-varying premia, $E_t[f]$, as an extension in Section 7.4. The fact that the unconditional mean of $\beta_{i,t}$ is zero does not affect the analysis because we focus on its cross-sectional variation.

We simulate the data using the system (1) and (2) for $i = 1, \ldots, 1,000$ (a cross-section of 1,000 bonds) and $T = 244$ months. We repeat this process 1,000 times: each time, we draw a cross section of the initial $\beta_{i,0}$ from a normal distribution with mean zero and standard deviation of $\sigma_v/\sqrt{1-\rho^2}$. Each month, we sort the bonds into quintiles based on the observed

$\beta_{i,t}$. We focus on two portfolios: one consisting of bonds in the top quintile (P(5)) and the other consisting of bonds in the bottom quintile (P(1)). The long-short strategy return, which takes long positions in P(5) and short positions in P(1), is our factor.

As $\beta_{i,t}$ changes, we need to rebalance bonds to keep track of which bonds belong to which portfolio. When ρ is close to one, $\beta_{i,t}$ is stable over time, so the bonds move across portfolios less frequently. A lower value of ρ renders the movement more frequent. This also implies that the expected returns on the long-short strategy decay more quickly. In contrast, as σ_v increases, the difference in betas between the bonds in P(5) and those in P(1) increases, raising the factor's expected returns. We will assess the effects of changing ρ and σ_v in the next section.

Each month, initiating or closing positions on bond i can be delayed by $\tau_{i,t}$ days. This implies that the factor investor needs to trade bond i, but she is unable to enter or dispose of the position because dealers cannot source the bond for same-day execution. To draw random delays, we assume that on a given day, a transaction occurs independently with an intensity λ. Then, delay $\tau_{i,t}$ follows an exponential distribution

$$\tau_{i,t} \sim_{i.i.d.} \text{Exp}(\lambda), \tag{3}$$

with a mean delay of $1/\lambda$ days. Though delay $\tau_{i,t}$ is bond specific, all bonds are ex-ante identical as they share the same trade intensity λ.

Given the central role of λ in determining the delay cost, it merits a few remarks. First, at this stage, we treat λ as a parameter that is fixed across bonds and over time. However, we allow it to vary over time as a part of the extension in Section 7.4. Second, it is exogenously given. In reality, traders face a trade-off between immediacy (i.e., a higher value of λ) and trade costs (higher bid-ask spreads) and optimize their choice (see, for example, Baruch et al. 2017). This will be accounted for when we empirically estimate λ, which will be done in

Section 4.

We consider buy and sell orders to fail if the delay is more than 21 days (i.e., a month). Once an order fails (i.e., $\tau_{i,t} > 21$), we reset the order at the end of the month and re-examine the delay over the next month (i.e., $\tau_{i,t+1}$), taking into account the updated signal $\beta_{i,t+1}$. This captures the idea that once the investor realizes that the delay is excessive, she can cancel the order based on the updated information. Consistent with real-world OTC trading dynamics, a long delay renders the trading signal obsolete, leading to an execution failure.

Trading failure drives a wedge between ideal portfolios where one purchases (sells) all bonds in the fifth (first) quintile and the actual portfolios with trade failures. Let $\mathbf{Q}(\boldsymbol{\pi})_t$ be an ideal set of bonds in quintile π based on signal $\beta_{i,t}$ in month t and $\mathbf{P}(\boldsymbol{\pi})_t$ be an actual set of bonds in the corresponding portfolio in month t. The two sets differ from one another because some bonds in the top (bottom) 20% may not be included in the portfolio if an excessive delay leads to a failed purchase. In addition, other bonds that are not in the top (bottom) 20% may be included in the portfolios due to sales failures. It is important to note that the incidence of failure depends not only on the draw of $\tau_{i,t}$, but also on whether or not the bond is already in the portfolio in the previous month.

Taking these delays and the possibility of trade failure into account, a net return on the bonds in portfolio π is,

$$r_{i,t+1}^{Net} = \begin{cases} r_{i,t+1} & \text{if } i \in \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \in \mathbf{Q}(\boldsymbol{\pi})_t, \\ r_{i,t+1}(1-\tau_{i,t}/21) & \text{if } i \notin \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \in \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} \le 21, \\ 0 & \text{if } i \notin \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \in \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} > 21, \\ r_{i,t+1}(\tau_{i,t}/21) + r_{Q(\pi),t+1}(1-\tau_{i,t}/21) & \text{if } i \in \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \notin \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} \le 21, \\ r_{i,t+1} & \text{if } i \in \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \notin \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} > 21. \end{cases} \tag{4}$$

The first two lines represent the intended positions for the bonds. These bonds are included in the portfolio either because the investor already holds them (the first line) or has successfully acquired them (the second line). In the case of a bond purchase, a delay of $\tau_{i,t}$ days reduces the return. The third line accounts for the scenario of trade failure in initiating a new position, resulting in a zero rate of return. Since the failure reduces the return for both P(5) and P(1), replacing 0 with other values (such as the one-month U.S. T-Bill rate of return) does not materially change the simulation results. The fourth and fifth lines pertain to bonds present in the portfolio from the previous month, which the investor retains due to delayed sales with a $\tau_{i,t}$-day delay (the fourth line) or unsuccessful sales attempts (the fifth line).[12] Then, the portfolio return after delay is $r^{Net}_{P(\pi),t+1} = \frac{1}{N_{P(\pi),t}} \sum r^{Net}_{i,t+1}$, where the average is taken over all bonds that corresponds to one of the five cases in Eq. (4).[13] In contrast, the gross portfolio return is $r_{Q(\pi),t+1} = \frac{1}{200} \sum_{i \in \mathbf{Q}(\boldsymbol{\pi})} r_{i,t+1}$.

For each draw of signal and delay, we record the month-t inventory for each bond in portfolio π as follows:

$$\mathbb{1}(i \in \mathbf{P}(\boldsymbol{\pi})_t) = \begin{cases} 1 & \text{if } i \in \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \in \mathbf{Q}(\boldsymbol{\pi})_t, \\ 1 & \text{if } i \notin \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \in \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} \leq 21, \\ 0 & \text{if } i \notin \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \in \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} > 21, \\ 0 & \text{if } i \in \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \notin \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} \leq 21. \\ 1 & \text{if } i \in \mathbf{P}(\boldsymbol{\pi})_{t-1} \text{ and } i \notin \mathbf{Q}(\boldsymbol{\pi})_t \text{ and } \tau_{i,t} > 21. \end{cases} \tag{5}$$

When computing month $t + 2$ portfolio returns in Eq. (4), we use this inventory record of

[12] In the fourth line, we assume that after a successful sale, the investor is allowed to invest the sale proceeds in the ideal portfolio, $Q(\pi)$. This condition ensures that, as $\tau \to 0$, the portfolio returns with delay converge to the portfolio returns without delay.

[13] When averaging bond returns, we implicitly assume that, at the end of each month, investors can rebalance their portfolios at no cost to ensure equal weighting. This assumption is made for tractability. Otherwise, the weights of newly purchased bonds would differ depending on the amount of cash available at the beginning of each month, influenced by past sales of existing bonds. To avoid this complexity, we assume that each month the investor will have an equally weighted portfolio. Additionally, we assume risk-free financing at a zero percent rate is available to cover the (usually small) difference in the number of bonds between P(5) and P(1).

bonds in $\mathbf{P}(\boldsymbol{\pi})_t$ together with the updated signal in month $t+1$ and the resulting ideal quintile assignments $\mathbf{Q}(\boldsymbol{\pi})_{t+1}$. Similarly, inventory record $\mathbf{P}(\boldsymbol{\pi})_{t+1}$, a draw for $\tau_{i,t+2}$, and $\mathbf{Q}(\boldsymbol{\pi})_{t+2}$ determine a return in $t+3$. By iterating this process, in each sample, we keep track of which bonds belong to which portfolio each month, taking into account the trading successes and failures.

In the simulation, whether the investor takes action (buy or sell) or maintains existing positions significantly affects factor performance. For example, suppose a bond is in $\mathbf{Q(5)}$ for three consecutive months in a sample, $t, t+1$, and $t+2$. In some scenarios, an investor successfully buys the bond in month $t+1$ (i.e., $\tau_{i,t} < 21$). She can then ignore the draws of delays in the following months, $t+1$ and $t+2$. This is because once the investor has purchased the bond, she does not have to worry about delays in the subsequent month until she has to sell it. In alternative scenarios, the purchase attempt fails in month $t+1$, requiring the investor to examine both the delay draw and underlying signal in the subsequent month ($\tau_{i,t+1}$) to determine whether the bond can be purchased. If the purchase remains unsuccessful, the investor continues checking successive delay draws until either the trade attempt succeeds or the bond exits quintile $\mathbf{Q(5)}$.

Finally, the factor for each simulation is computed as the difference between the fifth (P(5)) and first (P(1)) portfolios. The cost of delay in simulation m is the difference between the average gross and net returns in the path,

$$\overline{C}_{Delay,m}(\lambda, \rho, \sigma_v) = \frac{1}{T}\sum_t (r_{Q(5),t,m} - r_{Q(1),t,m}) - \frac{1}{T}\sum_t (r^{Net}_{P(5),t,m} - r^{Net}_{P(1),t,m}), \qquad (6)$$

and our simulated cost of delay is the average over the simulated data $m = 1, \ldots, 1,000$. The difference between the gross and net factor returns reflects the opportunity cost of delay. The cost arises because the investor cannot exactly implement the profitable strategy of buying bonds with high expected returns and selling those with low expected returns (given the signal).

3.2 Portfolio Turnover Rate and Bid-Ask Spread Cost

The key characteristic that determines transaction costs is the portfolio turnover rate. This is calculated as the sum of the absolute values of the month-to-month changes in portfolio weights, adjusting for endogenous changes in portfolio weights due to changes in bond values,[14]

$$TO_{\pi,t+1} = \sum_{i\in N_t} \left| w_{i,\pi,t+1} - \frac{1+r_{i,t+1}}{1+r_{\pi,t+1}} w_{i,\pi,t} \right|, \tag{7}$$

$$\overline{TO}_f = \frac{1}{2} \sum_{\pi\in 1,5} \frac{1}{T} \sum_t TO_{\pi,t+1}.$$

In addition to delays, our framework includes a standard cost of bid-ask spreads that investors pay each time they trade. We compute two types of portfolio turnover rates in each simulation: one based on the weights in portfolio **P**, denoted $\overline{TO}_{Sim,P}$, and the other based on the weights in the ideal portfolio **Q**, denoted $\overline{TO}_{Sim,Q}$. Due to delays, the former tends to be lower than the latter, $\overline{TO}_{Sim,P} < \overline{TO}_{Sim,Q}$. The product of the after-delay portfolio turnover rate, $\overline{TO}_{Sim,P}$, and the half spreads provides the bid-ask spread costs,

$$\overline{C}_{Bid-Ask,Sim} = 2 \times \overline{TO}_{Sim,P} \times h, \tag{8}$$

where h is a half spread. The turnover rate computed without any simulated delays, $\overline{TO}_{Sim,Q}$, is used to map the simulated transaction costs to our empirical factors, because in the data, the factor turnover rate is computed without accounting for delays.

[14]The portfolio turnover rate, calculated by adding the absolute changes in portfolio weights defined in Eq. (7), is between 0 and 200%. The maximum turnover of 200% rather than 100% can be illustrated with a simple two-bond example. In one month the weights for the two bonds are (1,0) and in the next month they change to (0,1). Ignoring the price changes, the turnover is $|0-1|+|1-0| = 200\%$. The investor pays a half spread both when she sells the first bond and when she buys the second bond.

3.3 Simulation Results

In this section, we present the cost estimates while remaining agnostic about the true parameters of the model, in particular, about the trade intensity, λ. By considering a wide range of parameters, we can better understand the nature and magnitude of the delay costs in factor investing. To this end, we vary $\rho = \{0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99, 0.99999\}$, which captures the speed of variation in the signal, and $\lambda = \{1, 1/5, 1/21, 1/42, 1/252\}$, which captures the trade intensity. For the other parameters, we set $E[f] = 0.30$, $\sigma_f = 1.67$, and $\sigma_u = 3.35$.[15] In this illustration, we set the half spread to zero to focus on the cost of delay.

Panel A of Fig. 2 plots the cost of delay as a function of the portfolio turnover rate (which negatively depends on ρ) and expected delay ($1/\lambda$ days). In this figure, we fix $\sigma_v = 0.6\sqrt{1-\rho^2}$, which corresponds to average factor returns of 0.51% per month. Panel A of Table 2 provides the specific values plotted in the figure.

The cost of delay increases with the portfolio turnover rate and decreases with the trading intensity, λ. Focusing on the case where a trade occurs on average once a month $1/\lambda = 21$, the figure shows that the cost of delay increases with the turnover rate. The cost of delay is about 0.04% per month when the turnover rate is 19%, but increases above 0.30% when the turnover rate is above 150%. This cost is substantial when compared to the gross return of 0.51% per month, as more than half of the gross return is lost due to delays and trade execution failures. For an extremely illiquid bond that trades once a year, the cost of delay increases more steeply with the turnover rate, but tapers at the level of gross returns because the delay cost cannot exceed the gross returns. As expected, for the bonds that trade with a one-day delay ($\lambda = 1$), the cost of delay is nearly zero.

[15] $E[f]$ and σ_f are chosen to match the mean and standard deviation of the corporate bond market factor. To obtain σ_u, we run the bond market CAPM regression for each bond in our sample and calculate the average R-squared, which is 0.368. We choose σ_u such that a long-short strategy of a bond in P(5) and a bond in P(1) has an R-squared of 0.368.

Next, we examine how gross returns affect the cost of delay. To do so, we fix the trading intensity parameter $\lambda = 1/21$ (i.e., a bond is traded on average once a month) and vary σ_v such that the factor returns before delay range from 0.009% to 0.85% per month. A higher value of σ_v effectively increases the predictive power of the signal, generating a higher gross return for the long-short strategy. Panel B of Fig. 2 presents the cost of delay as a function of portfolio turnover and σ_v, and Panel B1 of Table 2 provides the plotted values and gross factor returns that correspond to each value of σ_v. The figure shows that the cost increases with σ_v and highlights the fact that missing out on immediate order execution is costly when the signal is more valuable. Therefore, the delay cost is distinct from standard transaction costs based on bid-ask spreads, which do not depend on gross factor returns.

These cost estimates are quite general and the one-factor structure of the returns with a normal distribution is not a restrictive assumption. We can include multiple factors and heteroskedastic shocks with a skewed distribution, but ultimately, what matters is the average return difference between the long and short portfolios. The characterization of the delay cost above highlights the fact that the cost of delay is linked to a factor, rather than to a bond, through the return predictability of the underlying characteristic used to construct the factor and the factor's turnover. This observation allows us to assign the delay cost to the factors constructed using real data.

3.4 Liquid Bond Substitution

Chaudhary et al. (2023) demonstrate that corporate bonds with similar characteristics serve as close substitutes for one another. Consequently, investors may mitigate delay costs by concentrating their trading activity on a subset of highly liquid bonds. To quantitatively assess the importance of trading substitutes, we allow the investor to use real-time trade information to minimize potential delays. Since all bonds within a quintile are similar to each other (in-terms of the underlying signal) and the delay is independent of the bonds, we assume that the investor allocates her capital to the first n bonds traded in month $t+1$ instead of spreading it across all bonds that join the quintile.

To reduce the delay, we modify the portfolio rebalancing procedure as follows. Each month and for each quintile, we determine the number of bonds she must buy, denoted by $n_{MustBuy}$, as the smaller of the following two values: (i) n, the target number of bonds to buy; or (ii) the number of bonds that enter the quintile at the end of month t. We then rank these newly entered bonds by their realized delay, $\tau_{i,t}$, and allocate capital to the $n_{MustBuy}$ bonds with the lowest delay. This adjustment enables investors to act on buying opportunities as they arise based on real-time trading conditions. Importantly, this change only applies to the buying procedure. When a bond exits the quintile, the selling strategy remains unchanged since the investor must sell the bond she has, and partially selling the position would only worsen the delay.[16]

We set the target number of buys, n, equal to 10, 20, 50, 100, and 200 (i.e., the base case) and calculate the simulated factor returns. Panels A and B of Fig. 3 show the average and standard deviation of the factor returns with delays. Comparing the base case of $n = 200$ with that of $n = 10$ in Panel A, the average factor returns with delay are higher for $n = 10$ than for $n = 200$: the cost of delay is reduced as the investor only trades at most the first 10 bonds that trade in each month. As expected, the cases with $n = 20, 50, 100$ are in between the two extreme cases. In Panel B, the standard deviation for the factors with $n = 10$ is higher than the base case of $n = 200$ due to lower diversification benefits. Indeed, Panel D shows that the average number of bonds in the portfolio is significantly lower when n is set

[16] For example, consider an investor who must allocate \$100 million to create a portfolio in month t and liquidate her position in month $t+1$. There are 100 bonds that can be included in the portfolio. The investor could invest \$1 million in all 100 bonds or reduce delays by buying \$10 million of each of the 10 bonds that trade first. However, this idea will not work when she exits the position in month $t+1$ because she must sell the ten bonds she has purchased. If she sells fewer than 10 bonds, she can recover only $n \times 10\%$ of the investment proceeds and will end up with unwanted inventory, which is equivalent to a sales failure. Thus, despite the possibility of delays, it is optimal for her to try to sell everything.

equal to 10 or 20 than the baseline. We note that even when n is set equal to 10, the number of bonds in each portfolio is more than 10 because n controls the number of bonds to be traded each month, and the investor can accumulate the positions over time when the signal is stable.

The annualized Sharpe ratio is plotted in Panel C, showing that it is lower for the case of $n = 10$ than for the base case. This result indicates that the loss of diversification benefits dominates the benefit of reduced delay costs; thus, the Sharpe ratio decreases as we reduce the number of bonds in the portfolio.

The Sharpe ratio is lower with a lower target number of bonds to be bought partly because buying only a part of the portfolio does not eliminate the cost of sales delays. In Panels E and F, we plot the trade failure rate, which is defined as the number of bonds with delays $\tau_{i,t} > 21$ to the total number of bonds that must be traded. For buy trades (Panel E), setting $n = 10$ effectively eliminates the probability of trade failure. Given that investors are required to purchase only a small number of bonds, they can always fulfill their entire allocation requirement.[17] In contrast, Panel F shows that the sales failure rate does not depend on n. Since delays constitute independent random variables and investors must liquidate their *existing* portfolio holdings, constraining n to fewer than 200 bonds does not affect the sales delay. As a result, buying a small number of bonds only partly reduces the cost of delay while inflating the portfolio volatility. Our results suggest that using substitutes cannot eliminate the cost of delays and potentially worsens the factor risk-adjusted performance.

The results thus far are not specific to corporate bond markets. In any asset class where delays are not negligible relative to the speed of changing signals, one can refer to Table 2

[17] For $n = 20$ to 100, the fail rate is a decreasing function of the turnover rate. When the turnover rate is high, the investor must buy almost 200 bonds every month and limiting the order size to buy $n < 200$ bonds makes the order easier to be fulfilled. If, on the other hand, the investor has to buy only 50 bonds due to a lower turnover rate, limiting n to 50 means that she still has to buy everything. In such a case, limiting n to less than 200 does not reduce the rate of trade failure.

for the cost of delays.

4 Estimating Trade Intensity

4.1 Estimation Framework

To quantify the delay cost, we require an estimate of the trade intensity, denoted by λ in Eq. (3), facing an 'active' factor investor, who needs to rebalance portfolios frequently and requires fast trade execution. This trade intensity differs from the observed frequency of historical trades because the data reflects trading by a multitude of institutional investors such as insurance firms and pension funds, who tend to buy and hold (Koijen and Yogo, 2023; Bretscher et al., 2024). These investors have a low demand for fast trade execution and primarily trade via a slow-moving mandate that is often independent of signals related to factor investing. Thus, we simulate a counterfactual scenario in which inactive investors are replaced with hypothetical active investors pursuing factor investing strategies, such as hedge funds and mutual funds. In this counterfactual, both trade intensity and the compensation for the intensity, namely, half spreads can differ from the data. They can be estimated by specifying liquidity supply and demand curves.

In our setup, we consider the probability of trades in a given time period as the quantity of a service provided by dealers. The market is segmented by trade size, with large investors executing large trades and small investors executing small trades. To compensate the dealers for the production of this service, customers pay half spreads, $h_{i,t}$.[18] More formally, for each

[18]In the canonical model of Duffie et al. (2005), dealers provide immediacy to investors and charge bid-ask spreads. Thus, at a very high level, our framework is related to theirs. In their model, however, trading delays are determined by search delays of fixed intensity, and bid-ask spreads are set to allocate rent through bargaining. Our specification below deviates from this setup by allowing the dealer to expend costly effort to increase immediacy, resulting in an upward-sloping liquidity supply curve.

trade size k, the liquidity demand for bond i in month t is given by

$$Prob_{i,t,k}[\mathbb{1}_{i,t+1,k} = 1] = \alpha_k h_{i,k,t} + \delta_k D_{i,t} + \zeta_k X_{i,t}. \tag{9}$$

The parameter α_k represents the sensitivity of customers' trade probability to half spreads. $D_{i,t}$ is an idiosyncratic demand shifter not captured by the bond characteristics in vector $X_{i,t}$.

For the size-specific probability of transactions in Eq. (9), the trade indicator function $\mathbb{1}_{i,t,k}$ is defined as a dummy variable equal to one if a trade with size above threshold k (such as \$100,000 or \$2 million) occurs at least once per month. This is because, when dealers are willing to accommodate large trades, they will also take smaller trades. The dummy variable is defined at the monthly rather than daily level because at the daily sampling frequency, trades tend to cluster, violating the underlying trade independence assumption of the exponential distribution.

We can similarly consider a potentially upward-sloping liquidity supply curve, reflecting increasing marginal costs to the dealer.[19] These curves are plotted in Fig. 4. With the upward-sloping supply curve, there is a tradeoff between immediacy and half spreads and an upward shift in the demand curve will increase both the equilibrium probability of trades and half spreads (point E^{CF}). Estimating this new equilibrium requires estimates of the slopes of the supply and demand curves, which depend crucially on identification assumptions.

In this paper, we avoid this complexity and focus on the lower bound estimate of transaction costs. These lower bound estimates are obtained by assuming that the supply curve is inelastic to half spreads. In Fig. 4, this supply function is depicted as the vertical line. The new equilibrium (point E^{LB}) provides lower bound estimates for transaction costs be-

[19]There can be several ways to microfound such a supply curve. For example, Li and Schürhoff (2019) show that investors face tradeoffs between bid-ask spreads and delays because dealers at the center of the network offer fast execution but at a higher cost, while those at the periphery offer lower costs with more delays.

cause half spreads do not increase due to the upward shift in the demand curve, and the probability of trade is higher than the level that would be observed if the dealer responded to the demand increase by raising half-spreads. Therefore, this counterfactual provides the lowest estimated transaction costs among the possible alternatives that consider the trade-off between trade intensity and half spread.

Estimating the probability of transactions is equivalent to estimating delays. Once we estimate the trade probability, we can use the exponential distribution and back out λ from the average probability of trade, given a trade size. Specifically, we set λ to satisfy the following equation,

$$\overline{Prob}_k[\mathbb{1}_{i,t+1,k} = 1] = 1 - e^{-21\times\lambda_k}, \tag{10}$$

where $\overline{Prob}_k[\mathbb{1}_{i,t+1,k} = 1]$ is the estimated probability averaged over the sample for trade size k.

4.2 Estimating the Demand Shift Due to Changes in Investor Base

To quantify the shift in the demand curve, $D_{i,t}$, we use the logarithm of the investor composition measure of Li and Yu (2025), denoted as $InvComp$. This variable measures the average activeness of the bond holders and is constructed as the net transactions of investors averaged at the bond level.[20] The idea is that the frequency with which investors receive a

[20]Specifically, for investor j in quarter q, the net transaction is

$$nt_{j,q} = \frac{|\sum_i holding_{i,j,q} - \sum_i holding_{i,j,q-1}|}{\sum_i holding_{i,j,q-1}},$$

and we use the four-quarter average,

$$NT_{j,q} = \frac{1}{4}\sum_{k=0}^{3} nt_{j,q-k}.$$

Finally, we aggregate at the bond level to obtain the investor composition measure,

$$InvComp_{i,q} = \frac{\sum_j holding_{i,j,q} \times NT_{j,q}}{\sum_j holding_{i,j,q}}.$$

liquidity shock or an information update is persistent, and thus the investor with high past net transactions has a higher future liquidity demand, and the bond held by these investors is more likely to be traded.

To obtain an unbiased estimate of δ in Eq. (9), we need $\log InvComp$ to be independent of liquidity supply shocks that also affect the probability of trade. However, this is unlikely to be true since investors can predict which bonds dealers will provide liquidity for in the future, and active investors will self-select into owning those bonds. To account for this potential endogeneity, we use two proxies for liquidity supply shocks as instruments for $h_{i,t}$, including the dealer capital ratio of He et al. (2017), denoted as $CAP_{i,t}$, and the 28-day inventory changes to capture the inventory pressure on dealers, denoted as $\Delta Inventory_{i,t}$.[21] We are interested in estimating δ but not α and thus replace $h_{i,t}$ in Eq. (9) with $CAP_{i,t}$ and $\Delta Inventory_{i,t}$ to estimate δ. We run a forecasting regression of transactions in month $t+1$,

$$\mathbb{1}_{i,t+1,k} = \alpha_{0,k} + \alpha_{0,k} CAP_t + \alpha_{0,k} \Delta Inventory_{i,t} + \delta_k \log InvComp_{i,t} + \zeta_k X_{i,t} + e_{i,t+1,k}, \quad (11)$$

and compute the fitted value $\widehat{Prob}_{i,t,k}[\mathbb{1}_{i,t+1,k} = 1 | \log InvComp_{i,t}]$. When running the regression, we standardize the right-hand-side variables for ease of interpretation. Standard errors are double clustered at the bond and month level.

Panel A of Table 3 reports the coefficient estimates, adjusted R^2 values, and the associated t-statistics. We find that the measure of investor composition positively predicts the incidence of bond transactions, suggesting that investors who experience more frequent liquidity shocks are more likely to demand liquidity and to trade the bonds in their portfolio. The coefficients on the control variables also satisfy economic intuition: a bond with a lower credit rating (higher numerical value of the rating), shorter maturity, and larger size tends

We use $InvComp_{i,q}$ to predict transactions in each month in quarter $q+1$.

[21] To compute the change in inventory, we calculate the difference in volume between the total customer buy over the previous 28 calendar days and the total customer sells over the same period for each bond. For this calculation, we use all trade sizes.

to trade more frequently.

Based on the regression estimates, we consider the probability of trading that a factor investor would experience. Since the investor composition measure is not the same as the investor portfolio turnover rate, we do not know exactly the value of $\log InvComp$ for a factor investor who rebalances her portfolio monthly. Thus, we set $\log InvComp$ equal to the 99^{th} percentile and calculate $\widehat{Prob}_{i,t,k}[\mathbb{1}_{i,t+1,k} = 1 | \log InvComp^{99th}]$.[22] The 99^{th} percentile value (2.278) is higher than the sample average of $\log InvComp$ for active mutual funds (-4.30), providing us with a more conservative (i.e., higher) estimate for the probability of trade. We then compute the fitted value of the regression in Eq. (11), keeping the coefficients unchanged, and report the average "hypothetical" probability of trading at the bottom of Panel B of Table 3. The hypothetical probability of a sell order, estimated at 95.3% and 70.8% for \$100,000 and \$2 million trades respectively, is higher than the observed historical frequency of trades (87.7% and 58.8%). This implies that a trade delay facing active investors is less than what would be estimated by simply taking the data as given. Converting the monthly probability of trades into trade intensities using Eq. (10), the estimated intensities for \$100,000 and \$2 million transactions are 0.145 ($1/\lambda = 6.9$ days) and 0.059 ($1/\lambda = 17.1$ days), respectively.[23]

Finally, we obtain the intensity estimates from Kargar et al. (2025). Since their delay estimates are conditional on bond and trading characteristics, we compute the delay for each combination of attributes, such as a bond's age, turnover, maturity, and amount outstanding, and average them. This intensity estimate is 0.111 ($1/\lambda = 9.1$ days) for \$100,000 trades

[22] Table A.2 of the Internet Appendix reports the summary statistics of the panel data used for this regression estimates.

[23] Our results are not sensitive to the specification of the investor liquidity demand variable. In Tables A.3, A.4, and A.5 of the Internet Appendix, we conduct robustness checks by running logit regressions instead of linear regressions, allowing for the nonlinear dependence of the probability of trading on investor composition, and introducing the interaction between investor composition and bond characteristics. The resulting counterfactual trade intensities are not materially different from those in Table 3.

(odd lot) and 0.070 ($1/\lambda$ = 14.4 days) for \$2 million trades (round lot).[24] Compared to our estimates, their intensity for small trades is lower than ours, while that for large trades is higher because they focus on liquid bonds traded on electronic platforms using the most recent data since 2017. In the next section, we compare the cost estimates based on these two estimates and show that the difference in trade intensity estimates does not materially affect the assessment of factor investing performance.

5 Performance of the Corporate Bond Factors

In this section, we comprehensively examine the performance of corporate bond factors after transaction costs by applying our simulation-based methodology outlined in Sections 3 and 4.

5.1 Analytical Framework

Factor Construction. To generate the 341 corporate bond factors, we compute monthly bond total returns in the standard manner: $r_{i,t+1} = \frac{P_{i,t+1}+AI_{i,t+1}+C_{i,t+1}}{P_{i,t}+AI_{i,t}} - 1$, where $P_{i,t+1}$ is the clean price of bond i in month $t+1$, $AI_{i,t+1}$ is the accrued interest, and $C_{i,t+1}$ is the coupon payment, if any.

In the main results, the factors are assumed to be rebalanced monthly, based on characteristic based signals observed by the factor investor at month-end.[25] We use bond excess

[24] We start from their Table 8 and enter the estimates into their Eq. (2) to obtain the delay in days for the number of failures of one. Then the intensity λ is obtained as 1/(delay in days). Specifically, for each period (Covid or non-Covid), credit rating, and customer attribute, we compute 2^5 combinations of attributes, including trade direction, age, turnover, maturity, and amount outstanding and take the equal-weighted average delay conditional on the number of failure being one. Then, the estimated delay is averaged across rating and customer attributes with the respective weights in our bond data for each mode of trade (MarketAxess, Voice, or Both), which is then equally averaged. This intensity estimate is 0.115 for \$100,000 trades (odd lot) and 0.072 for \$2 million trades (round lot) in booms, and 0.068 and 0.042 during recessions (the Covid episode). Finally, the intensity is obtained by averaging the two estimates using the probability of booms (0.91) and recessions (0.09) during the sample period.

[25] Extensions to the month-end assumption are discussed in Section 7.

returns, $r^e_{i,t+1}$, defined as the bond total returns minus the one-month risk-free rate of return from Kenneth French's webpage. Following the literature (Novy-Marx and Velikov 2015, Chen and Velikov 2023), we form the high-minus-low factor strategy using deciles with (bond) market capitalization as the weights. The long (short) position is assumed to be decile ten (one) such that all factors are sign-corrected to be increasing in expected returns. We also compute the monthly turnover of each factor based on Eq. (7).

Factor Classification. In order to present our main results in a tractable manner, we only consider corporate bond factors that generate a statistically significant average return at the 5% significance level before transaction costs, leaving us with 58 factors (17% of the 341).[26] In the spirit of Jensen et al. (2023), we then cluster these factors into 6 distinct groups including (i) Equity Momentum (EquityMomentum), (ii) Equity Short-Term Reversal (EquityReversal), (iii) Firm Investment and Accruals (Investment), (iv) Corporate bond yields, prices and credit spreads (Spreads), (v) Firm Value/Profitability (Value/Profit), and lastly, (vi) Bond and Stock Volatility/Risk-Based Characteristics (Volatility/Risk). We present the factor mnemonics, short name, the clusters, and associated references in Table A.6 of the Internet Appendix.

Risk Adjustments. To compute factor alphas, we use the net of cost bond market factor computed using liquid bond market ETFs. We estimate the net bond CAPM (CAPMB) alpha by running time series regressions of the factor i's excess returns on the corporate bond market factor:

$$f_{i,t} = \alpha + \beta\, MKTB_{Net,t} + \varepsilon_{i,t}, \tag{12}$$

[26] All factors are sign-corrected to have a positive mean. The number of significant factors we document is larger than Dick-Nielsen et al. (2023), who examine 151 factors (41 significant) with a different sample and time period.

where $MKTB_{Net,t}$ is the net excess return of BlackRock's corporate bond exchange-traded funds (ETFs), averaged between the investment-grade ETF (Ticker: LQD) and the high-yield ETF (Ticker: HYG) using the total market value of corporate bonds in each respective rating category as the weights.[27] To account for autocorrelation in the returns, we adjust the standard errors using the Newey and West (1987) procedure with 12 lags.

Half Spread Estimates. To measure half spreads, we follow O'Hara and Zhou (2021) and compute the half spread for each trade. Let $\bar{P}$ denote the transaction price of the most recent interdealer trade preceding a customer-dealer trade. The cost of the customer-dealer trade for bond i in transaction ν is then given by

$$h_{i,\nu} = (\log P_{i,\nu} - \log \bar{P}_{i,\nu}) \times I_{i,\nu}, \tag{13}$$

where $I_{i,\nu}$ equals 1 if the trade is a customer buy and -1 if the trade is a customer sell. The reference trade is allowed to occur within a five-business-day window preceding the customer-dealer trade. If no interdealer trades occur within this window, the half spread for that trade is treated as missing. The daily half-spread is then calculated as the volume-weighted average of transactions with size above k on a given day, and the monthly measure is obtained by averaging daily half-spreads across all days within a month. Half spreads are winsorized at the 0.5 and 99.5 percentiles to reduce the influence of outliers.[28]

[27] We use the ETF returns because they reflect the real cost of buying and holding the bond market portfolio. Therefore, ETF returns provide a fair benchmark to evaluate the performance of trading strategies net of costs. The detailed construction methodology of the combined net of cost market factor is provided in Appendix B. We find that the average net excess returns on our ETF-based market factor is 0.32% per month, while the corresponding value for the value-weighted market bond portfolio of Dickerson et al. (2023) is 0.36% (gross) over the same period. The lower value of the ETF-based market factor suggests that even holding the market is marginally costly for investors.

[28] Fig. 5 plots average half spreads as a function of trade size. Consistent with Edwards et al. (2007), half spreads decline sharply with trade size. This pattern is explained by the bargaining power of large customers (Duffie et al., 2005) and the strategic behavior of dealers who pre-arrange large trades (Goldstein and Hotchkiss, 2020; Choi et al., 2024).

Factor Net Returns. For net returns, the delay and bid-ask spread costs are assigned to each factor based on their average turnover rate and gross returns. Specifically, we first estimate delay costs by interpolating over the simulated delay cost grids corresponding to different trade intensities and turnover levels, as reported in Panels B2 and B3 of Table 2.[29] Second, we calculate bid-ask spread costs using average observed half spreads and turnover rates after accounting for delays, interpolating across trade sizes and turnover grids (as reported in Panel C of Table 2) to obtain factor-specific estimates. Third, we combine the delay and bid-ask spread costs to derive the total trading cost for each factor. Finally, we compute the net of cost average factor return and alpha by subtracting the (time-invariant) total trading cost from the gross return and alpha. A detailed description of this procedure is provided in Section C of the Internet Appendix.

5.2 Empirical Results: Net of Cost Performance

We examine the performance of the selected 58 corporate bond factors before and after transaction costs. Panel A of Table 4 reports the gross bond CAPM alpha of the 58 factors, averaged within each category and across all factors. On average, the factors earn a gross CAPM alpha of 0.28% per month with an average t-statistic of 2.48. Moreover, 42 out of the 58 factors earn a statistically significant alpha. Thus, consistent with Dick-Nielsen et al. (2023), there is a wide range of factors that perform well after adjusting for market risk. Across categories (clusters), the best-performing factor-types are Spreads (CAPM alpha = 0.44%) and Equity Reversal (0.43%), followed by Equity Momentum (0.29%). Table A.7 of the Internet Appendix reports the performance of each individual factor by category. Among the 58 factors, the best performer is the equity short-term reversal factor (ret10; Chordia et al. 2017), from the Equity Reversal category, earning a CAPM alpha of 0.69%. The second best factor is based on the previous month's corporate bond price (bondprice: Bartram et al.

[29] Panel D of Table 2 reports the gross and net turnover rates.

2025) with a CAPM alpha of 0.63%. Similar to the equity factor zoo, other significant bond factors stem from value and momentum (see, e.g., Gebhardt et al. 2005a and Choi and Kim 2018).[30]

The CAPM alphas largely disappear after adjusting for transaction costs. Panels B1 and B2 of Table 4 use the estimates of delay costs $\overline{C}_{Delay}$ and bid-ask costs $\overline{C}_{Bid-Ask}$ discussed in Section 5.1 to compute total costs and net CAPM alphas for small ($100,000) and large ($2 million) trades, respectively. After accounting for transaction costs, the average CAPM alpha across the 58 factors declines to 0.05% and 0.09% for small and large trades, respectively. For small trades, the average bid-ask spread cost is 0.19% and the cost of delay is 0.04%. For large trades, the bid-ask spread cost decreases to 0.12%, while the cost of delay increases to 0.08%. Given that the bid-ask spread cost decreases more than the increase in the cost of delay, the total cost is lower for large trades than for small trades. Nevertheless, the net CAPM alpha for large trades is only 0.09% per month, suggesting that it is difficult for investors to exploit these anomalies in practice. Only two factors earn statistically significant CAPM alphas with large trades, while one factor earns a significant alpha with small trades (bond kurtosis in the Volatility and Risk category; see Table A.7 of the Internet Appendix).[31]

Delay costs are particularly significant for the Equity Momentum and Equity Reversal categories, which are associated with rapidly decaying signals. For large trades, the average delay costs are 0.11% (Equity Momentum) and 0.22% (Equity Reversal) compared to 0.01% (Investment) and 0.02% (Value/Profit), which are based on slower moving Compustat accounting signals. The Equity Reversal factor category has the highest average cost of delay (0.22%) due to its high turnover and average excess return, which is higher than its bid-ask

[30] As an additional exercise, Fig. A.1 of the Internet Appendix employs a 3×3 conditional rating $\times$ signal double tercile sort and reveals similar results; however, the double tercile sort produces substantially lower average gross factor returns with only marginally reduced turnover. The average monthly returns across the union of the 67 significant factors (58 for deciles and 42 for terciles) are 0.31% and 0.15%, respectively. Average turnover decreases from 70% (deciles) to 54% (terciles).

[31] Bond kurtosis is calculated using the last 60 months of returns according to Bai et al. (2016).

spread cost of 0.20%. The Spreads factor category, which comprises factors formed on bond yields, prices, and credit spreads (fast-moving market data), also yields an average delay cost of 0.11% for large trades, exceeding the associated bid-ask cost of 0.09%. Since these factors generate high average excess returns, the opportunity cost of not trading upon immediately observing the underlying signal is very high. In contrast, the cost of delay is negligible for the Investment and Value/Profit categories because these signals are slow moving (updated every quarter) and generate relatively low returns.

Fig. 6 presents the performance and transaction cost analysis for the 58 individual factors. Panel A plots the gross CAPM alphas with associated two-standard error bars, while Panels B and C present the breakdown of total costs between bid-ask spread and delay costs for small and large trades. The plot confirms the relative importance of the cost of delays for highly profitable factors in the Equity Momentum, Equity Reversal, and Spread categories, especially for large trades. Delay costs associated with these factor categories constitute a substantially larger share of total trading costs because their underlying signals change rapidly.[32]

Panels C1 and C2 of Table 4 report the net CAPM alphas, where the cost of delay is computed using the delay estimates of Kargar et al. (2025) with intensity parameter $\lambda_{\$100,000} = 0.111$ and $\lambda_{\$2M} = 0.070$.[33] Using Kargar et al. (2025)'s delay estimates, the average net CAPM alphas are 0.042% for small trades and 0.094% for large trades, very similar to our corresponding estimates of 0.050% and 0.087% reported in Panels B1 and B2 of Table 4.[34]

[32] The results are also qualitatively similar when using the duration-adjusted returns of van Binsbergen et al. (2025) (see Table A.8 of the Internet Appendix) and TRACE data (see Table A.9 of the Internet Appendix).

[33] Fig. A.2 of the Internet Appendix shows the transaction cost breakdown using the delay estimates of Kargar et al. (2025).

[34] The delay costs exceed the borrowing cost for short sales. To show this, we account for costs associated with short-selling corporate bonds by relying on the indicative borrowing cost from Markit discussed in Section D of the Internet Appendix. We subtract the fee from the short leg of the trade and report the resulting performance of factors net of fees in Table A.10 of the Internet Appendix. The fee averages about

6 Combining Characteristics With Machine Learning Models

The previous section demonstrated that transaction costs and delays largely eliminate individual bond factor profitability. However, this does not imply that a factor formed with an optimal combination of signals is dominated by transaction costs (DeMiguel, Martin-Utrera, Nogales, and Uppal, 2020). We now explore the application of machine learning models for forming bond factors after accounting for transaction costs.

6.1 Estimating the Machine Learning Models

Following Gu, Kelly, and Xiu (2020), we model corporate bond returns as $r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1}$, where $E_t(r_{i,t+1}) = g^*(z_{i,t})$. We seek to estimate $g^*(\cdot)$ as a flexible function of the 341-dimensional predictor vector $z_{i,t}$ that maximizes out-of-sample explanatory power. All of our model estimates minimize the mean squared prediction errors (MSE). Using the estimated expected returns $\hat{g}^*(z_{i,t})$, bonds are sorted monthly into ten value-weighted portfolios. The ML-based factors are long decile ten (high expected return) and short decile one (low expected return).

In total, we consider six linear and non-linear machine learning models including penalized linear regressions: Lasso (LASSO), Ridge (RIDGE) and Elastic Net (ENET); non-linear regression tree ensembles including random forests (RF) and extreme trees (XT); and feed forward neural networks (NN). In addition, we form the linear ensemble model (LENS), the nonlinear ensemble model (NENS), and the ensemble across all models (ENS), which is the equally-weighted average across the respective models' one-month ahead predictions (Rapach, Strauss, and Zhou, 2010). We provide extensive details related to the cross-validation

0.03% per month. To investigate the impact of short sales constraints, we reproduce the main results using a long-only strategy. The impact of delay costs reported in Table A.11 of the Internet Appendix is very similar to our main results.

and training of the respective models in Section E of the Internet Appendix.

6.2 Performance of Machine-Learning Models After Costs

Fig. 7 presents the gross alphas and net alphas for small trades (Panel A) and large trades (Panel B), along with their 95% confidence intervals for the ML factors. The models successfully combine the signals to generate high gross alphas, ranging from 0.62% (RF) to 0.90% (XT), which are also highly statistically significant. These values are also reported in Table 5. On the other hand, the portfolio turnover rate is relatively high, ranging from 97% to 126% per month.

When accounting for transaction costs, we compute two versions of net alpha, one in which we subtract only the bid-ask spread cost, $\overline{C}_{Bid-Ask}$, from the gross factor alpha, and another in which we subtract the total cost to trade, $\overline{TC}$. We first focus on the net alphas that take into account bid-ask spreads only. The figure shows that the ML models generate large CAPM alphas after accounting for bid-ask spreads. For trade sizes of \$100,000, the alpha ranges from 0.34% to 0.58% per month, while for trade sizes of \$2 million, it ranges from 0.45% to 0.70% per month. Importantly, all alphas are statistically significant. This implies that using bid-ask spreads alone as a measure of transaction costs would lead to the conclusion that investors can profit from the models' forecasts. With an average alpha exceeding 0.6% per month for large trades, the evidence in favor of ML models is quite striking.[35]

Next, we focus on the alphas net of both the bid-ask spread and delay costs. From the figure, delay costs are substantial, ranging from 0.09% to 0.19% for small trades and from 0.17% to 0.36% for large trades. For large trades, the delay cost exceeds the bid-

[35]The results reported above are very robust, and do not depend on the particular estimates of the trading intensity which underpins the transaction delay or the portfolio formation methodology (deciles or conditional tercile sorts). In Fig. A.3 of the Internet Appendix, we employ a 3×3 conditional rating double tercile sort and observe similar results, although the double tercile sort produces worse average gross returns.

ask spread cost for all nine models. After adjusting for delay costs, the net CAPM alphas remain positive, but four factors (with small trades) and six factors (with large trades) lose statistical significance, while the remainder are only marginally significant. For large trades, the extreme tree (XT) model performs best, yielding a net CAPM alpha of 0.41% ($t = 2.31$), which represents less than half of its gross alpha of 0.87% ($t = 4.89$).

ML models incur significant delay costs because they maximize average gross returns without considering portfolio turnover rates. Consequently, these models heavily weight rapidly changing signals (e.g., short-term reversals), leading to high turnover rates and gross alphas, which amplify delay costs. For machine learning models with illiquid assets, penalizing high turnover rates is essential, and in this regard, the methodologies of Bredendiek, Ottonello, and Valkanov (2023) and Jensen et al. (2025) offer a promising approach to addressing transaction costs within a machine learning portfolio construction framework.

7 Extensions of Our Framework

7.1 Different Investment Horizons and Banding

Traditional portfolio sorting procedures may overstate factor trading costs. Investment managers can employ cost-mitigation techniques while preserving expected returns. We examine two approaches using `PyBondLab`: staggered rebalancing, where only a fraction of the position is traded each month (Jegadeesh and Titman, 1993),[36] and banding, which implements thresholds for initiating versus maintaining positions (Novy-Marx and Velikov, 2015). The latter creates portfolio inertia that can reduce trading frequency while maintaining strategy performance.

[36] In any month $t + 1$, the factor return is the equally weighted average month $t + 1$ return of the factor strategy implemented in the prior month and up to H months earlier, where H is the staggered holding period. For example, if $H = 3$, this mechanically limits the one-sided monthly portfolio turnover to be $\leq 33.33\%$.

Staggered partial rebalancing. Panel A1 of Table 6 reports the average gross CAPM alphas and t-statistics of the six factor clusters when using a staggered three-month holding period. The final column reports the averages across all clusters. Panels A2 and A3 report the average net of cost alphas, bid-ask spread, and delay costs. As expected, the slower rebalancing frequency substantially reduces delay costs, which now average 0.01% per month for small trades and 0.03% per month for large trades, compared to the corresponding values of 0.04% and 0.08% under monthly rebalancing reported in Table 4.[37]

However, this benefit is accompanied by rapid signal deterioration, leading to diminished return predictability. Comparing the results with a one-month holding period in Panel A of Table 4, the average gross alpha falls from 0.28% to 0.20% per month, while the turnover rate falls from 70% to 34%. Since these effects roughly cancel with each other, the net performance changes very little. For example, the average net alpha for trade sizes of $100,000 and $2 million with the three-month holding period are 0.07% and 0.10%, respectively, which are only slightly higher than the corresponding values (0.05% and 0.09%) with the one-month holding period. The number of significant factors net of costs declines to 1 (Bond Kurtosis). Thus, infrequent rebalancing alone is insufficient to generate profitable factors after accounting for half-spread and delay costs.

Banding. Novy-Marx and Velikov (2015, 2019) demonstrate that banding is the most successful method for reducing portfolio turnover in the equity market. Following their approach, we implement a band width of two around the first and tenth portfolios, allowing bonds to remain in their respective portfolios while their signals stay within the first or last three deciles. This approach maintains positions despite minor signal fluctuations, thereby

[37] Fig. A.4 of the Internet Appendix plots delay costs across holding periods ranging from one to twelve months, demonstrating that these costs decline substantially when the investment horizon extends beyond three months. These results suggest that future research seeking to identify novel factors in corporate bond markets should evaluate whether a factor's predictive power persists beyond the standard one-month holding period.

reducing portfolio turnover.

Panel B of Table 6 reports the average of the gross and net CAPM alphas with banding. The average net CAPM alphas for small and large trades are 0.05% and 0.08%, which are again similar to the baseline results reported in Table 4. However, banding improves the right tail of the distribution: now, the number of factors with significant net alphas increases to 2 (trade size of \$100,000)[38] and 5 (trade size of \$2 million)[39] from the corresponding values of 1 and 2 in Table 4, respectively. In summary, neither method of reducing portfolio turnover greatly improves net CAPM alphas.

7.2 Liquid Bond Substitution – Double-Sort Approach

In Section 3.4, we rely on simulations to determine whether trading substitutes can speed up execution, assuming that all bonds have an ex-ante identical trade intensity, which prevents sales delays from being eliminated. In reality, some bonds might be persistently liquid and investors may select subsamples of ex-ante liquid bonds to reduce delays. To investigate this possibility, we employ a conditional double-sort approach, first dividing bonds into two groups based on month t trading volume scaled by par amount outstanding, or month t amount outstanding. Within the high-liquidity subsample, we form ten value-weighted portfolios and construct a long-short strategy (long in portfolio ten, short in portfolio one).

We re-estimate the trade intensity parameter λ using a dummy for above-median trading volume or a dummy for above-median amount outstanding as predictors in Eq. (11). In the latter case, however, we drop the (continuous) amount outstanding from the set of control variables. Results in Table 7 show that both bonds with above-median volume and above-median size have higher customer sell probabilities and shorter delays.[40] For example, using

[38]These signals are Bond Kurtosis and Bond Price.

[39]These signals are Bond Kurtosis, Bond Price, Announcement Return (bond PEAD), ΔFinancial Liabilities (fnlgr1a), and ΔNet Financial Assets (nfnagr1a).

[40]To conserve space, the results for customer buy probabilities, which are very similar, are reported in

the subsample of bonds with high past trading volumes, the customer sell order intensity is 0.183 and 0.070 for small and large trades, compared to 0.145 and 0.059 in Table 3.[41]

We simulate costs using these higher trade intensity estimates to compute net CAPM alphas using the subsample of liquid bonds. In Panel C of Table 6, we use the subsample of bonds with above-median volume. We find that the gross CAPM alpha is higher (0.32% vs. baseline 0.28%), but turnover increases to 99% from 70% because the investor must rebalance frequently to ensure she always holds only liquid bonds to reduce sales delays. As a result, the net CAPM alpha averaged across 58 factors declines to −0.002% and −0.04% for small and large trades, respectively. On the other hand, in Panel D, using bonds with above-median amount outstanding helps reduce delay costs more effectively. However, the gross CAPM alpha declines to 0.26% for this subsample, resulting in net CAPM alphas of 0.07% and 0.09% for small and large trades, which are both similar to our main results. In summary, trading only liquid bonds, though intuitive, does not improve the CAPM alpha net of costs.

7.3 Splitting Large Orders into Smaller Pieces

In the stock market, it is common for traders to split a large order into smaller pieces to reduce price impact. In limit order book markets, trade size serves as an important signal of the trader's potential private information. In OTC markets, the identity of the investor is known to the dealer, so splitting orders does not necessarily reduce price impact, as shown in Fig. 5. Splitting a large order into smaller pieces, however, can increase the speed of transactions and reduce the cost of delay.

To quantify the trade-offs between costs related to the half-spread and immediacy, we consider a simulation in which the trader is allowed to split a $2 million trade into κ equal-

Table A.12 of the Internet Appendix.

[41] Using the subsample of bonds with above-median amount outstanding, the estimated hypothetical probability exceeds one, thus in this case we set the $\lambda = 9,999$ in simulation such that delays are minimized.

sized pieces. Each trade faces a delay that is independently drawn from the exponential distribution with the trade intensity estimated for its size based on Eq. (11). For example, for $\kappa = 2$, we estimate λ for trade size \$1 million (=\$2 million/2) and draw delays for two trades in a month. The monthly return is the average of the returns arising from these two trades. In this simulation, we choose two values of gross returns (high, 0.51%; low, 0.09%) and two values of turnover rate (high, 135%; low, 38%) from Table 2.

Fig. 8 plots the simulated bid-ask spread and delay costs for $\kappa = 1, 2, 4, 10, 20$. As expected, the delay cost decreases as we split the trade into more pieces. However, each trade has higher half spreads, resulting in higher bid-ask spread costs. For example, consider the case with high gross returns and low turnover rate in Panel D. When $\kappa = 4$, the total cost is 0.139%, which is slightly lower than the cost of 0.148% when not splitting. If we split the order into smaller pieces, the cost increases. For $\kappa = 20$, each order size is \$100,000, and the total cost increases to 0.165% due to the higher half spreads. When the gross returns are low at 0.09% (Panels A and C), splitting the order merely increases the total cost as the half spread cost increases sharply.

This result illustrates the fundamental tension between trade immediacy and the half-spread cost of trading corporate bonds. Trading in smaller sizes implies a smaller chance of execution failure but is commensurate with a higher half-spread.

7.4 Effect of Time-Varying Delays

Our main results are based on a constant trade intensity λ. In this section, we relax this assumption and examine the effect of the cyclical pattern of factor performance on delay costs. According to Kargar et al. (2025), trade intensity tends to decline during recessions, leading to the wider divergence between ideal and actual bond portfolios. Meanwhile, the cyclical behavior of factor performance varies substantially across factors. For example, factors that likely capture time-varying risk premia, such as the credit spread factor, exhibit high cyclicality with elevated returns during economic expansions and depressed returns during recessions. In contrast, factors that likely represent mispricing, such as the post-earnings announcement drift factor of Nozawa et al. (2023), are countercyclical, generating higher average returns during recessions than during economic expansions.

Empirically, we find that the majority of factors with statistically significant average returns fall into the 'mispricing' category. Fig. A.5 of the Internet Appendix plots the average annualized return difference between factor premia earned during NBER recession periods versus non-NBER periods. For many of the signals, the factor premium is earned mostly over a very small part of the sample, during periods of market turmoil, precisely when trade immediacy is likely to be more constrained.

To examine how factor cyclicality affects delay costs, we employ a two-state Markov regime-switching model calibrated to the historical duration of economic expansions and recessions.[42] Specifically, we set the transition matrix between good (State 1) and bad (State 2) states to

$$\mathbf{Pr} = \begin{pmatrix} 0.99 & 0.10 \\ 0.01 & 0.90 \end{pmatrix}, \tag{14}$$

which implies the stationary distribution of $(0.9095, 0.0905)$. In each state, we use the trade intensity estimates from Kargar et al. (2025), as their results demonstrate that trade intensity declines significantly during adverse market conditions (such as the COVID-19 period in March 2020) relative to normal market states.

We assume that the expected factor return $E_t[f]$ varies across states such that $E_1[f] - E_2[f] = \Delta E[f]$, where $\Delta E[f] \in \{-1.5, -1, -0.5, 0, 0.5\}$. In each case, $E_1[f]$ and $E_2[f]$ are set such that the unconditional expected return matches the baseline of 0.51%. We vary

[42]During our sample period, the average duration of NBER recessions is $(18+2)/2 = 10$ months, while that of economic expansions is $(73 + 128)/2 = 100.5$ months. This yields an unconditional recession probability of $10/110.5 = 0.0905$, which matches the parameters in our transition matrix.

$\Delta E[f]$ to capture the broad spectrum of cyclical factor behavior.

In Fig. 9, we plot the delay cost as a function of $\Delta E[f]$. The plot shows that procyclical factors with $\Delta E[f] > 0$ have a lower delay cost than countercyclical factors. This result reveals an interesting property of delay costs: when the signal's forecast is incorrect, the delay helps avoid executing unprofitable trades. Thus, if a factor exhibits cyclical behavior, lower trade intensity (increased delay) during recessions prevents investors from entering positions that would subsequently generate low returns. This explains why the delay cost for $\Delta E[f] = 0.5$ is lower than in the base case of $\Delta E[f] = 0$. In contrast, mispricing factors with $\Delta E[f] < 0$ incur higher delay costs because delays become more severe precisely when these factors generate high returns. Since 37 of our 58 factors are of the 'mispricing' type with $\Delta E[f] < 0$, these results indicate that time-varying delays likely exacerbate delay costs for the majority of factors in our sample.[43]

8 Conclusion

Both researchers and investors alike are attracted to factors that generate high average returns. Thus, it is tempting to frequently rebalance portfolios to chase fast-moving signals and generate impressive results on a gross-return basis. However, this practice deviates substantially from the original work of Fama and French (1993), who rebalance portfolios once a year to ensure that the factors represent real-life, implementable trading strategies.

In this article, we propose a novel method to quantify the cost of delays that factor investors would face in an illiquid, over-the-counter market. We show that, unlike the transaction costs arising from bid-ask spreads, the delay cost is more severe for factors with higher

[43] In this exercise, we keep the leverage of the investor constant over time. With time-varying systematic risk premiums, we can consider a dynamic trading strategy to increase leverage when $E_t[f_{t+1}]$ is high. While this is another potential source of trading cost, we focus on the static case to present the cost estimates by following the literature on the cross-section of corporate bonds (e.g., Dickerson et al. (2023); Dick-Nielsen et al. (2023)).

gross performance, which are precisely those researchers tend to discover. Moreover, large trades, which have lower bid-ask spreads, incur a greater cost of delay. This cost cannot be removed by simply trading a subset of liquid bonds because liquidity is difficult to predict, and thus sales delays cannot be eliminated.

Our results suggest that delay costs significantly reduce the number of factors that generate bond CAPM alphas. Machine-learning-based factors are severely affected by delay costs because the underlying models generate valuable return predictions that decay quickly. This implies the cost of delays, or analogously, the cost of not trading with immediacy is elevated because these models require frequent trading. Currently, the actual trade frequency in the U.S. corporate bond market is not high enough to justify monthly rebalancing to chase such signals. Since delay costs can be minimized for signals that depend on quarterly-updated financial statements, we recommend evaluating the performance of corporate bond factors over a holding period longer than one month.

References

Augustin, Patrick, Linxiao F. Cong, Ricardo L. Aliouchkin, and Roméo Tédongap, 2020, Downside risk and the cross-section of corporate bond returns, Working Paper.

Avramov, Doron, Si Cheng, and Lior Metzker, 2023, Machine learning vs. economic restrictions: Evidence from stock return predictability, *Management Science* 69, 2587–2619.

Bai, Jennie, Turan G. Bali, and Quan Wen, 2016, Do the distributional characteristics of corporate bonds predict their future returns?, Working Paper.

Bai, Jennie, Turan G. Bali, and Quan Wen, 2019, RETRACTED: Common risk factors in the cross-section of corporate bond returns, *Journal of Financial Economics* 131, 619–642.

Bai, Jennie, Turan G. Bali, and Quan 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.

Bali, Turan G., Heiner Beckmeyer, and Amit Goyal, 2023, A joint factor model for bonds, stocks, and options, Working Paper.

Bali, Turan G., Amit Goyal, Dashan Huang, Fuwei Jiang, and Quan Wen, 2020, Predicting corporate bond returns: Merton meets machine learning, Working Paper.

Bali, Turan G., Avanidhar Subrahmanyam, and Quan Wen, 2021a, Long-term reversals in the corporate bond market, *Journal of Financial Economics* 139, 656–677.

Bali, Turan G., Avanidhar Subrahmanyam, and Quan Wen, 2021b, The macroeconomic uncertainty premium in the corporate bond market, *Journal of Financial and Quantitative Analysis* 56, 1653–1678.

Baltussen, Guido, Frederik Muskens, and Patrick Verwijmeren, 2024, The cross-section of corporate bond returns, Working Paper.

Bao, Jack, Maureen O'Hara, and Xing (Alex) Zhou, 2018, The Volcker Rule and corporate bond market making in times of stress, *Journal of Financial Economics* 130, 95–113.

Bao, Jack, Jun Pan, and Jiang Wang, 2011, The illiquidity of corporate bonds, *Journal of Finance* 66, 911–946.

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

Baruch, Shmuel, Marios Panayides, and Kumar Venkataraman, 2017, Informed trading and price discovery before corporate events, *Journal of Financial Economics* 125, 561–588.

Bessembinder, Hendrik, Stacey Jacobsen, William Maxwell, and Kumar Venkataraman, 2018, Capital commitment and illiquidity in corporate bonds, *Journal of Finance* 73, 1615–1661.

Bessembinder, Hendrik, Kathleen M. Kahle, William F. Maxwell, and Danielle Xu, 2008, Measuring abnormal bond performance, *Review of Financial Studies* 22, 4219–4258.

Bredendiek, Maximilian, Giorgio Ottonello, and Rossen I. Valkanov, 2023, Implementable corporate bond portfolios, Working Paper.

Bretscher, Lorenzo, Lukas Schmid, Ishita Sen, and Varun Sharma, 2024, Institutional corporate bond pricing, *Review of Financial Studies* forthcoming.

Cao, Jie, Amit Goyal, Xiao Xiao, and Xintong Zhan, 2023, Implied volatility changes and corporate bond returns, *Management Science* 69, 1375–1397.

Ceballos, Luis, 2023, Inflation volatility risk and the cross-section of corporate bond returns, Working Paper.

Chaudhary, Manav, Zhiyu Fu, and Jian Li, 2023, Corporate bond multipliers: Substitutes matter, Working Paper.

Chen, Andrew Y., and Mihail Velikov, 2023, Zeroing in on the expected returns of anomalies, *Journal of Financial and Quantitative Analysis* 58, 968–1004.

Chen, Andrew Y., and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 11, 207–264.

Chen, Long, David A. Lesmond, and Jason Wei, 2007, Corporate yield spreads and bond liquidity, *Journal of Finance* 62, 119–149.

Chen, Xi, Junbo Wang, and Chunchi Wu, 2022, Jump and volatility risk in the cross-section of corporate bond returns, *Journal of Financial Markets* 60, 100733.

Choi, Jaewon, Yesol Huh, and Sean Seunghun Shin, 2024, Customer liquidity provision: Implications for corporate bond transaction costs, *Management Science* 70, 187–206.

Choi, Jaewon, and Yongjun Kim, 2018, Anomalies and market (dis)integration, *Journal of Monetary Economics* 100, 16–34.

Chordia, Tarun, Amit Goyal, Yoshio Nozawa, Avanidhar Subrahmanyam, and Qing 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, Kee H., Junbo Wang, and Chunchi Wu, 2019, Volatility and the cross-section of corporate bond returns, *Journal of Financial Economics* 133, 397–417.

Dang, Thuy Duong, Fabian Hollstein, and Marcel Prokopczuk, 2023, Which factors for corporate bond returns?, *Review of Asset Pricing Studies* 13, 615–652.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal, 2020, A transaction-cost perspective on the multitude of firm characteristics, *Review of Financial Studies* 33, 2180–2222.

Detzel, Andrew, Robert Novy-Marx, and Mihail Velikov, 2023, Model comparison with transaction costs, *Journal of Finance* 78, 1743–1775.

Dick-Nielsen, Jens, Peter Feldhütter, Lasse Heje Pedersen, and Christian Stolborg, 2023, Corporate bond factors: Replication failures and a new framework, Working Paper.

Dick-Nielsen, Jens, and Marco Rossi, 2018, The cost of immediacy for corporate bonds, *Review of Financial Studies* 32, 1–41.

Dickerson, Alexander, Christian Julliard, and Philippe Mueller, 2025, The corporate bond factor zoo, Working Paper.

Dickerson, Alexander, Philippe Mueller, and Cesare Robotti, 2023, Priced risk in corporate bonds, *Journal of Financial Economics* 150, article 103707.

Dickerson, Alexander, Yoshio Nozawa, and Cesare Robotti, 2024, The low frequency trading arms race: Machines versus delays, Working Paper.

Duan, Tinghua, Frank Weikai Li, and Quan Wen, 2025, Is carbon risk priced in the cross-section of corporate bond returns?, *Journal of Financial and Quantitative Analysis* 60, 1–35.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, 2005, Over-the-counter markets, *Econometrica* 73, 1815–1847.

Edwards, Amy, Lawrence E. Harris, and Michael S. Piwowar, 2007, Corporate bond market transaction costs and transparency, *Journal of Finance* 62, 1421–1451.

Eisfeldt, Andrea L., Bernard Herskovic, and Shuo Liu, 2023, Interdealer price dispersion, Working Paper.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Feldhütter, Peter, 2010, The same bond at different prices: Identifying search frictions and selling pressures, *Review of Financial Studies* 25, 1155–1206.

Friewald, Nils, and Florian Nagler, 2019, Over-the-counter market frictions and yield spread changes, *Journal of Finance* 74, 3217–3257.

Friewald, Nils, and Florian Nagler, 2024, Dealer inventory and the cross-section of corporate bond returns, *Economics Letters* 239, 111710.

Gebhardt, William R., Soeren Hvidkjaer, and Bhaskaran Swaminathan, 2005a, The cross-section of expected corporate bond returns: Betas or characteristics?, *Journal of Financial Economics* 75, 85–114.

Gebhardt, William R., Soeren Hvidkjaer, and Bhaskaran Swaminathan, 2005b, Stock and bond market interaction: Does momentum spill over?, *Journal of Financial Economics* 75, 651–690.

Goldberg, Jonathan, and Yoshio Nozawa, 2021, Liquidity supply in the corporate bond market, *Journal of Finance* 76, 755–796.

Goldstein, Michael A., and Edith S. Hotchkiss, 2020, Providing liquidity in an illiquid market: Dealer behavior in US corporate bonds, *Journal of Financial Economics* 135, 16–40.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Haesen, Daniel, Patrick Houweling, and Jeroen van Zundert, 2017, Momentum spillover from stocks to corporate bonds, *Journal of Banking and Finance* 79, 28–41.

He, Xin, Guanhao Feng, Junbo Wang, and Chunchi Wu, 2025, Predicting individual corporate bond returns, *Journal of Banking and Finance* 171, 107372.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

He, Zhiguo, Paymon Khorrami, and Zhaogang Song, 2019, Commonality in credit spread changes: Dealer inventory and intermediary distress, Working Paper.

Hendershott, Terrence, Roman Kozhan, and Vikas Raman, 2020, Short selling and price discovery in corporate bonds, *Journal of Financial and Quantitative Analysis* 55, 77–115.

Hendershott, Terrence, Dan Li, Dmitry Livdan, Norman Schuerhoff, and Kumar Venkataraman, 2022, Quote competition in corporate bonds, Working Paper.

Hendershott, Terrence, Dan Li, Dmitry Livdan, and Norman Schürhoff, 2024, When failure is an option: Fragile liquidity in over-the-counter markets, *Journal of Financial Economics* 157, 103859.

Hendershott, Terrence, Dmitry Livdan, and Norman Schuerhoff, 2021, All-to-all liquidity in corporate bonds, Working Paper.

Houweling, Patrick, and Jeroen van Zundert, 2017, Factor investing in the corporate bond market, *Financial Analysts Journal* 73, 100–115.

Huang, Jing-Zhi, Nan Qin, and Ying Wang, 2024, Breadth of ownership and the cross-section of corporate bond returns, *Management Science* 70, 5709–5730.

Israel, Ronen, Diogo Palhares, and Scott Richardson, 2018, Common factors in corporate bond returns, *Journal of Investment Management* 16, 17–46.

Ivashchenko, Alexey, 2024, Corporate bond price reversals, *Journal of Financial Markets* 68, 100880.

Ivashchenko, Alexey, and Robert Kosowski, 2024, Transaction costs and capacity of systematic corporate bond strategies, *Financial Analysts Journal* 80, 53–80.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jensen, Theis I., Bryan T. Kelly, Semyon Malamud, and Lasse H. Pedersen, 2025, Machine learning and the implementable efficient frontier, *Review of Financial Studies* forthcoming.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *Journal of Finance* 78, 2465–2518.

Jostova, Gergana, Stanislava Nikolova, Alexander Philipov, and Christof W. Stahel, 2013, Momentum in corporate bond returns, *Review of Financial Studies* 26, 1649–1693.

Kargar, Mahyar, Benjamin Lester, Sébastien Plante, and Pierre-Olibier Weill, 2025, Sequential search for corporate bonds, *Journal of Finance* forthcoming.

Kelly, Bryan, Diogo Palhares, and Seth Pruitt, 2023, Modeling corporate bond returns, *Journal of Finance* 78, 1967–2008.

Koijen, Ralph S. J., and Motohiro Yogo, 2023, Understanding the ownership structure of corporate bonds, *American Economic Review: Insights* 5, 73–91.

Kyle, Albert S., and Anna A. Obizhaeva, 2016, Market microstructure invariance: Empirical hypotheses, *Econometrica* 84, 1345–1404.

Li, Dan, and Norman Schürhoff, 2019, Dealer networks, *Journal of Finance* 74, 91–144.

Li, Jian, and Haiyue Yu, 2025, Investor composition and the liquidity component in the U.S. corporate bond market, *Journal of Finance* forthcoming.

Lin, Hai, Junbo Wang, and Chunchi Wu, 2011, Liquidity risk and expected corporate bond returns, *Journal of Financial Economics* 99, 628–650.

Muravyev, Dmitriy, Neil D. Pearson, and Joshua M. Pollet, 2024, Anomalies and their short-sale costs, Working Paper.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Novy-Marx, Robert, and Mihail Velikov, 2015, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104–147.

Novy-Marx, Robert, and Mihail Velikov, 2019, Comparing cost-mitigation techniques, *Financial Analysts Journal* 75, 85–102.

Nozawa, Yoshio, Yancheng Qiu, and Yan Xiong, 2023, Disagreement and price drifts in the corporate bond market, Working Paper.

O'Hara, Maureen, and Xing Alex Zhou, 2021, The electronic evolution of corporate bond dealers, *Journal of Financial Economics* 140, 368–390.

O'Hara, Maureen, Yihui Wang, and Xing Zhou, 2018, The execution quality of corporate bonds, *Journal of Financial Economics* 130, 308–326.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12, 2825–2830.

Pinter, Gabor, Chaojun Wang, and Junyuan Zou, 2024, Size discount and size penalty: Trading costs in bond markets, *Review of Financial Studies* 37, 2156–2190.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.

Reichenbacher, Michael, and Philipp Schuster, 2022, Size-adapted bond liquidity measures and their asset pricing implications, *Journal of Financial Economics* 146, 425–443.

Sandulescu, Mirela, 2022, How integrated are corporate bond and stock markets?, Working Paper.

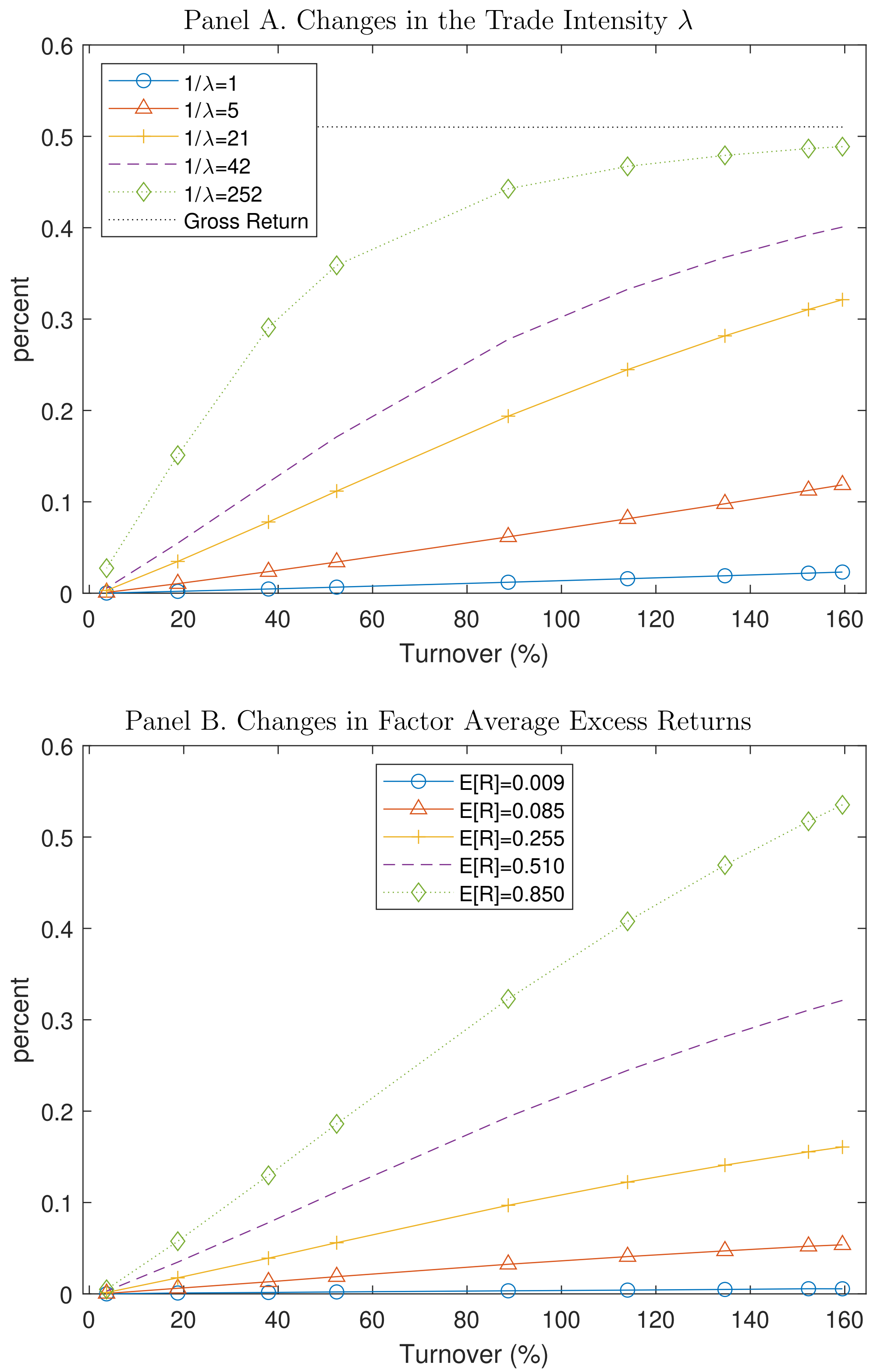
Schestag, Raphael, Philipp Schuster, and Marliese Uhrig-Homburg, 2016, Measuring liquidity in bond markets, *Review of Financial Studies* 29, 1170–1219.

Tao, Xinyuan, Bo Wang, Junbo Wang, and Chunchi Wu, 2022, Economic policy uncertainty and the cross-section of corporate bond returns, *Journal of Fixed Income* 32, 6–44.

van Binsbergen, Jules H., Yoshio Nozawa, and Michael Schwert, 2025, Duration-based valuation of corporate bonds, *Review of Financial Studies* 38, 158–191.

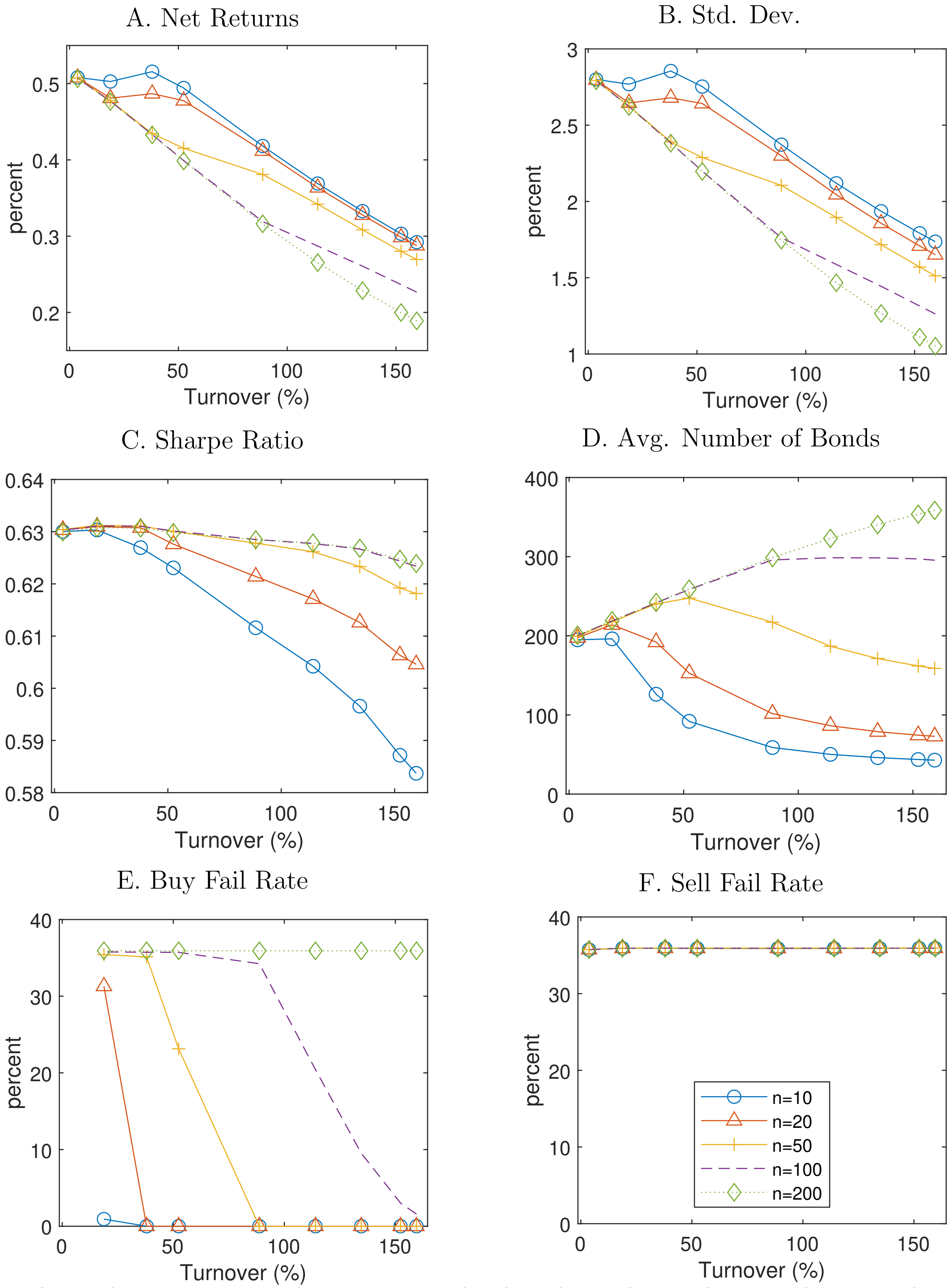
Wu, Botao, 2022, Post-crisis regulations, trading delays, and increasing corporate bond liquidity premium, Working Paper.

Figure 2: Cost of Delays (% Per Month)



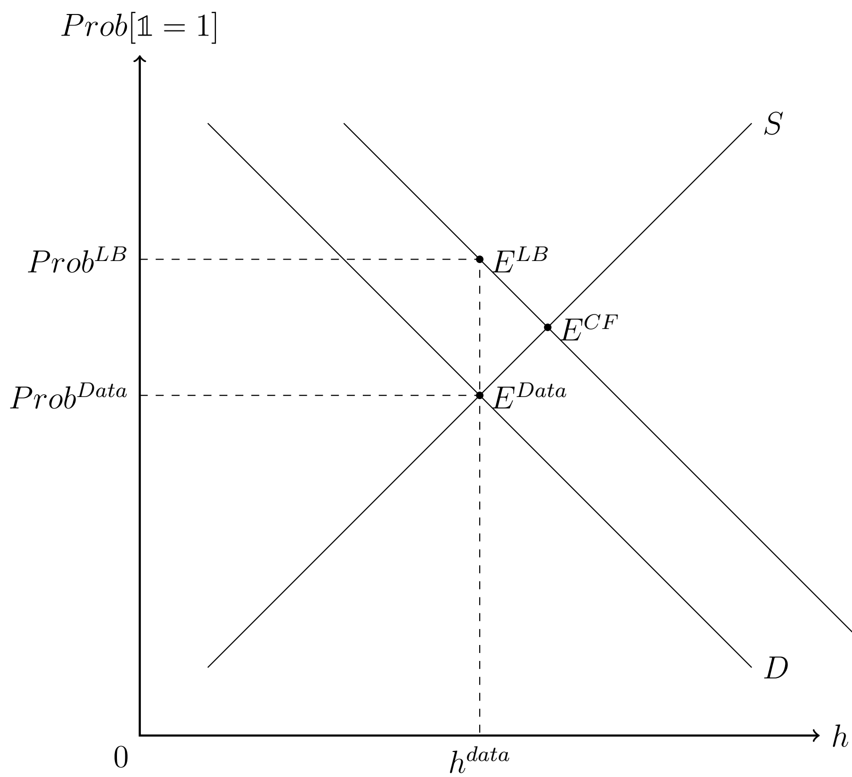
The figure plots the simulated cost of delay on the y-axis, defined as the difference in average returns between gross factors without delays and net factors with delays, as discussed in Section 3. Delays follow an exponential distribution with trade intensity parameter λ. Panel A varies the signal persistence parameter ρ to alter the portfolio turnover rate while holding the gross factor profitability parameter σ_v fixed such that average excess returns equal 0.51% per month. Panel B varies σ_v while maintaining λ at 1/21.

Figure 3: Buying Only n Bonds Among Substitutes



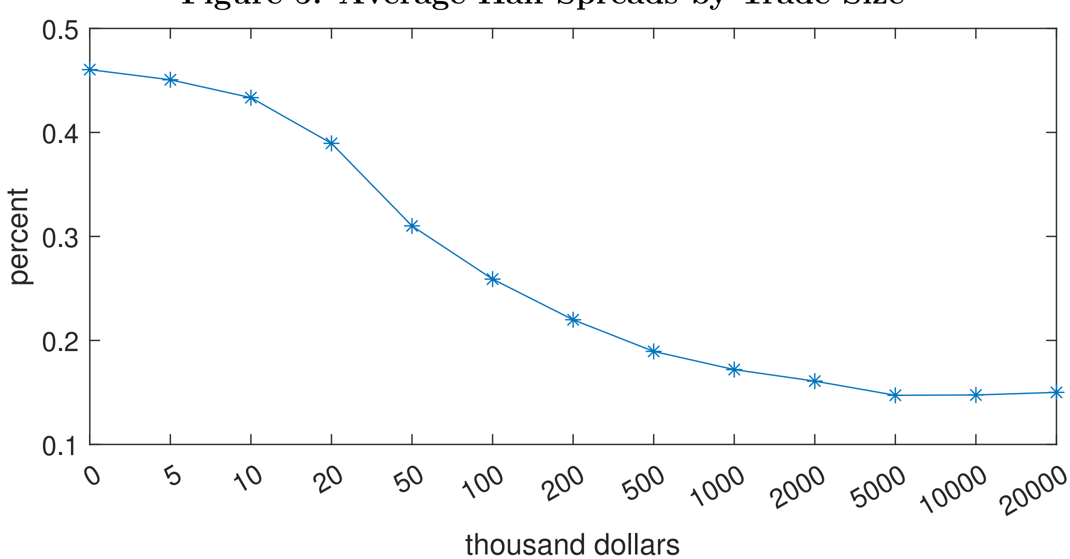
Each month, an investor purchases at most n bonds within each quintile, even when more than n bonds enter the quintile, treating these bonds as close substitutes (see Section 3.4). We set σ_v such that the average excess returns without delays equal 0.51% per month and $\lambda = 1/21$. Net returns represent average excess returns after accounting for delay costs, and Std. Dev. denotes the standard deviation of portfolio returns after delays. Avg. number of bonds represents the number of bonds in each portfolio (averaged across long and short positions) averaged over time. Buy and Sell Fail Rates indicate the percentage of bonds that cannot be traded due to delays exceeding one month relative to total bonds targeted for trading.

Figure 4: Liquidity Supply and Demand



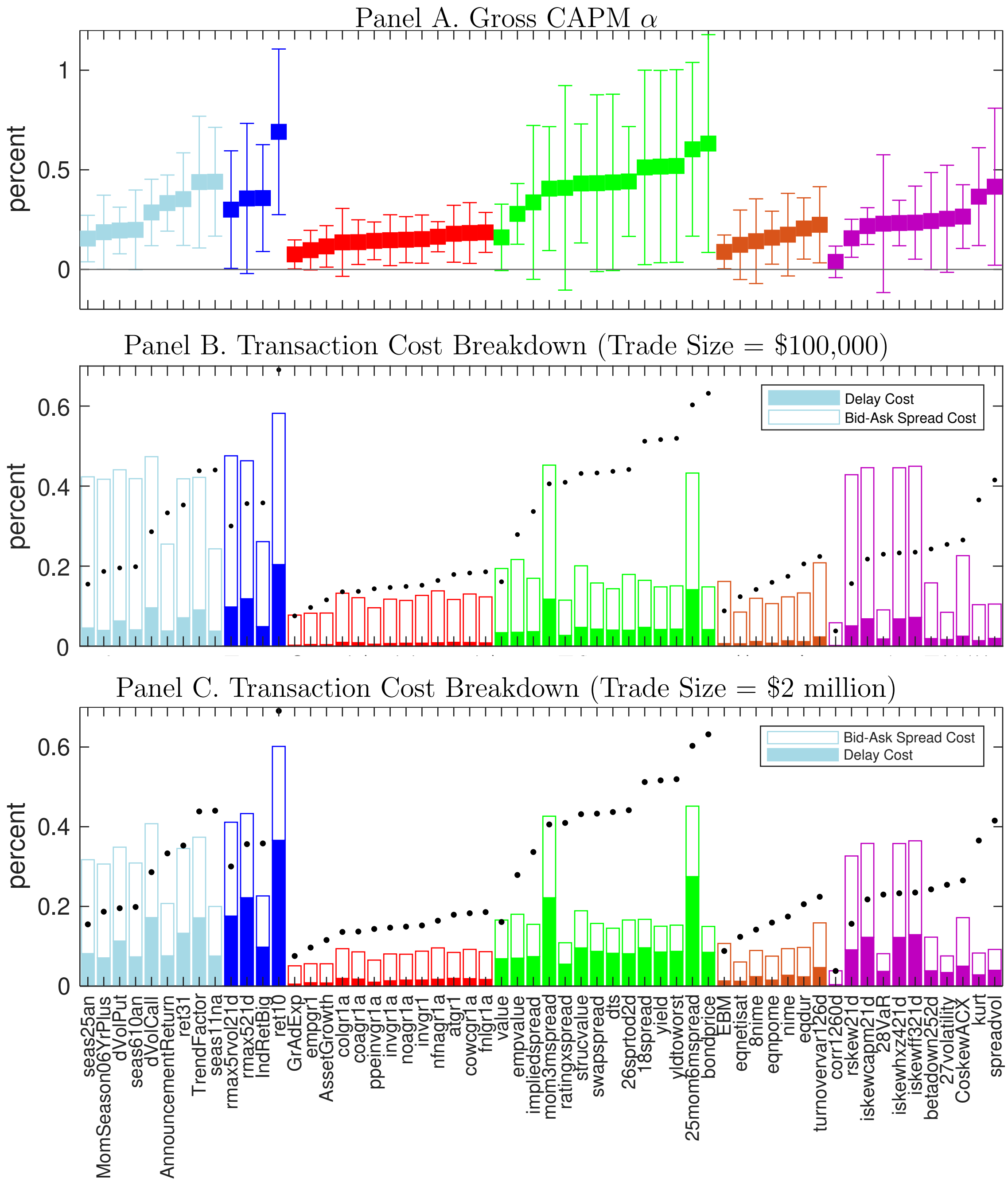
The figure illustrates the supply and demand curves for corporate bond liquidity, where the quantity is measured by the probability of transactions, $Prob[\mathbb{1} = 1]$, and the price is measured by half spreads charged by dealers, h, discussed in Section 4.1. E^{Data} denotes the equilibrium observed in the data, which is the intersection between the supply curve S and the demand curve D. D' denotes the hypothetical demand curve when an active investor is introduced and E^{LB} is the hypothetical equilibrium if the supply curve is flat. E^{CF} is the counterfactual with an active investor and dealers responding to the introduction of the active investor.

Figure 5: Average Half Spreads by Trade Size



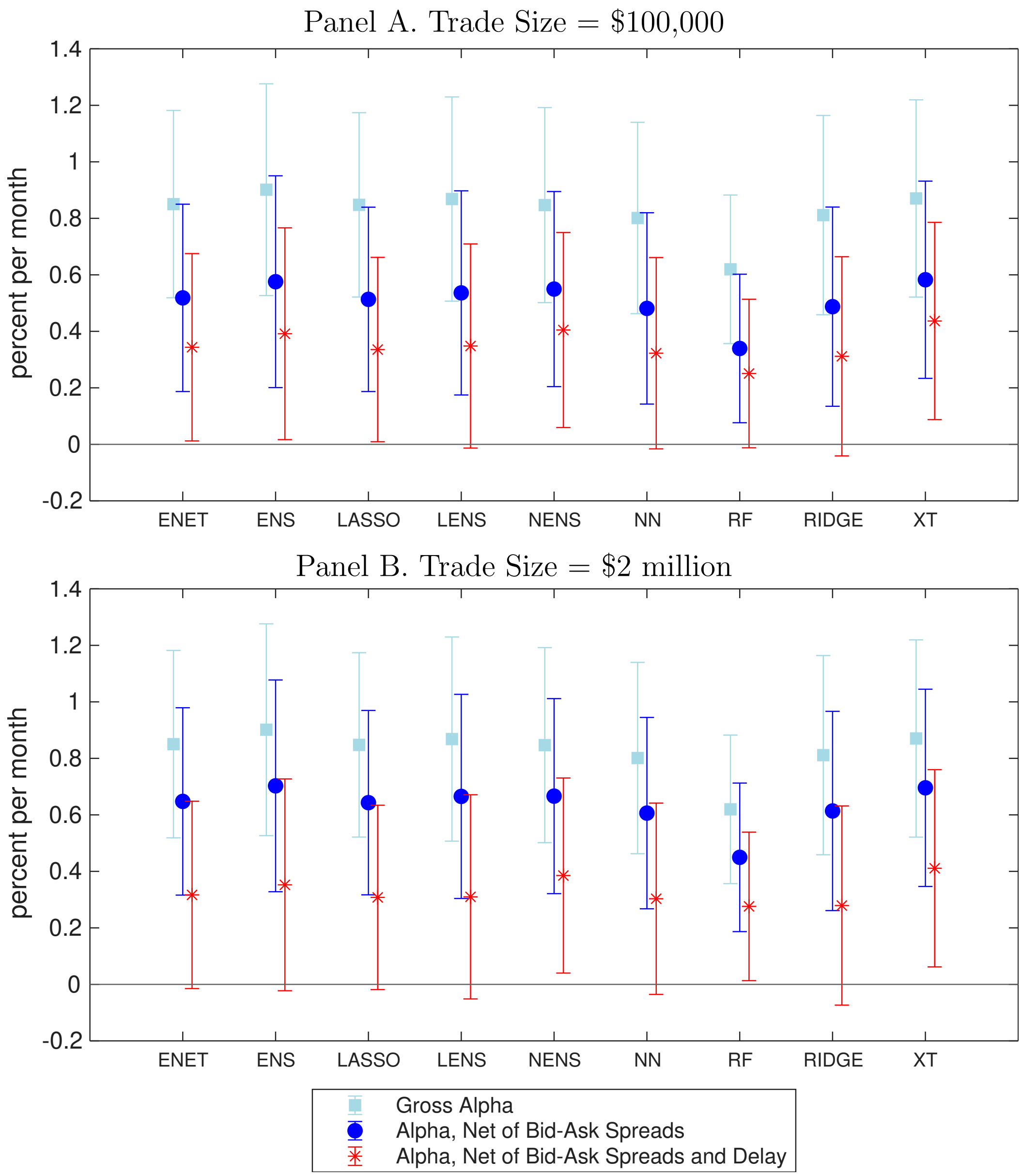
The figure plots the average half spreads for bond transactions of a given trade size from Section 5. Half spreads are defined in Eq. (13). Each day, we compute the volume-weighted average of all transactions above a trade-size threshold and then take the average within the month. The figure plots the average values in the panel data given the respective trade size.

Figure 6: CAPM α of 58 Individual Factors



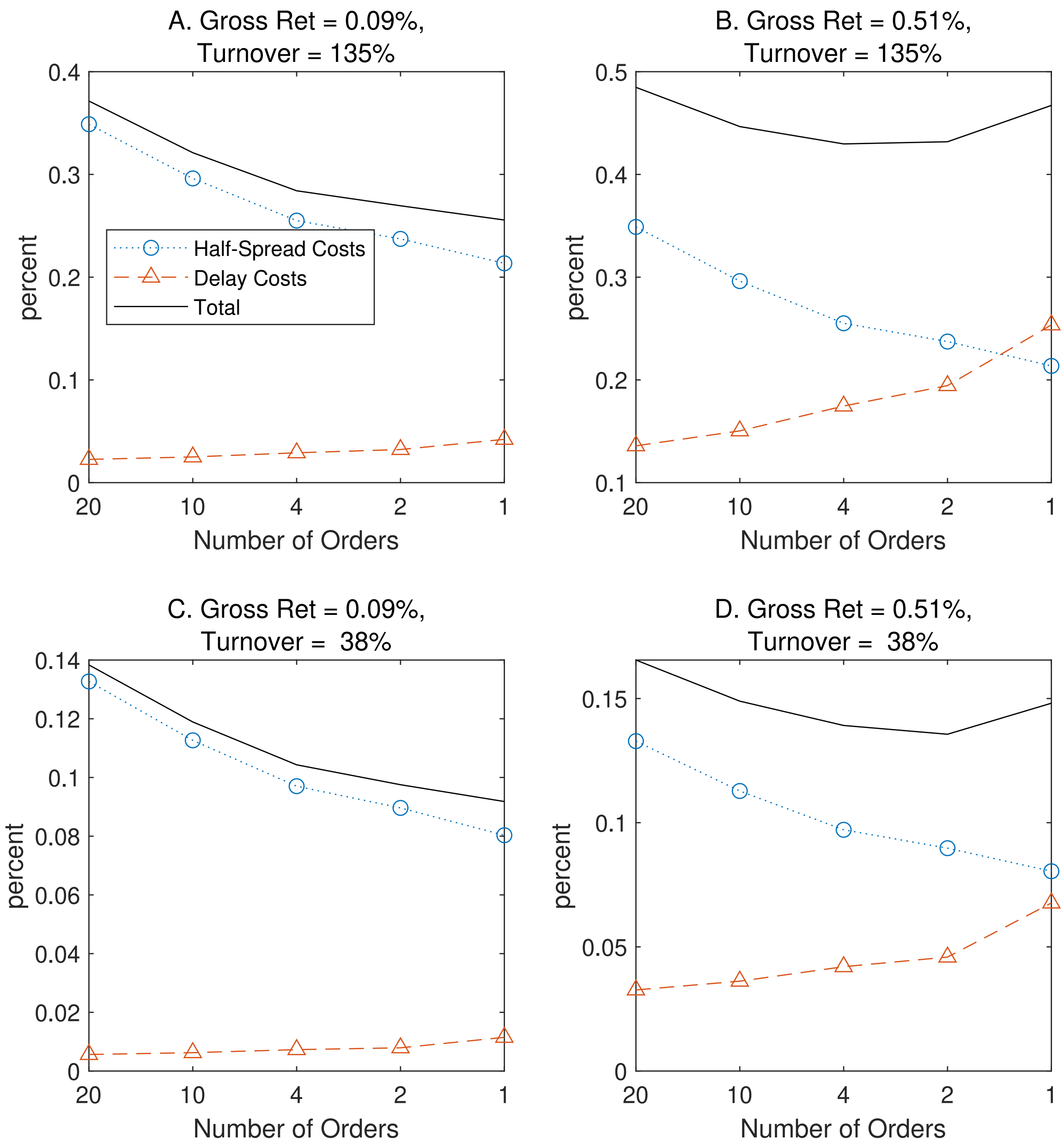
The figure plots the bond CAPM alpha (% per month) before transaction costs (Panel A), the breakdown of transaction costs using our lower bound estimates for small trades of $100,000 (Panel B) and $2 million (Panel C) discussed in Section 5.2. From the 341 candidate factors, we select 58 factors with statistically significant average excess returns and classify them into six groups, represented by different colors. The error bars in Panel A represent the two standard error bounds that account for serial autocorrelation with the Newey-West procedure using 12 lags. The dots in Panels B and C represent the gross αs for each factor. All factors are formed with the PyBondLab Python package.

Figure 7: CAPM α of the Machine Learning Model-Based Factors



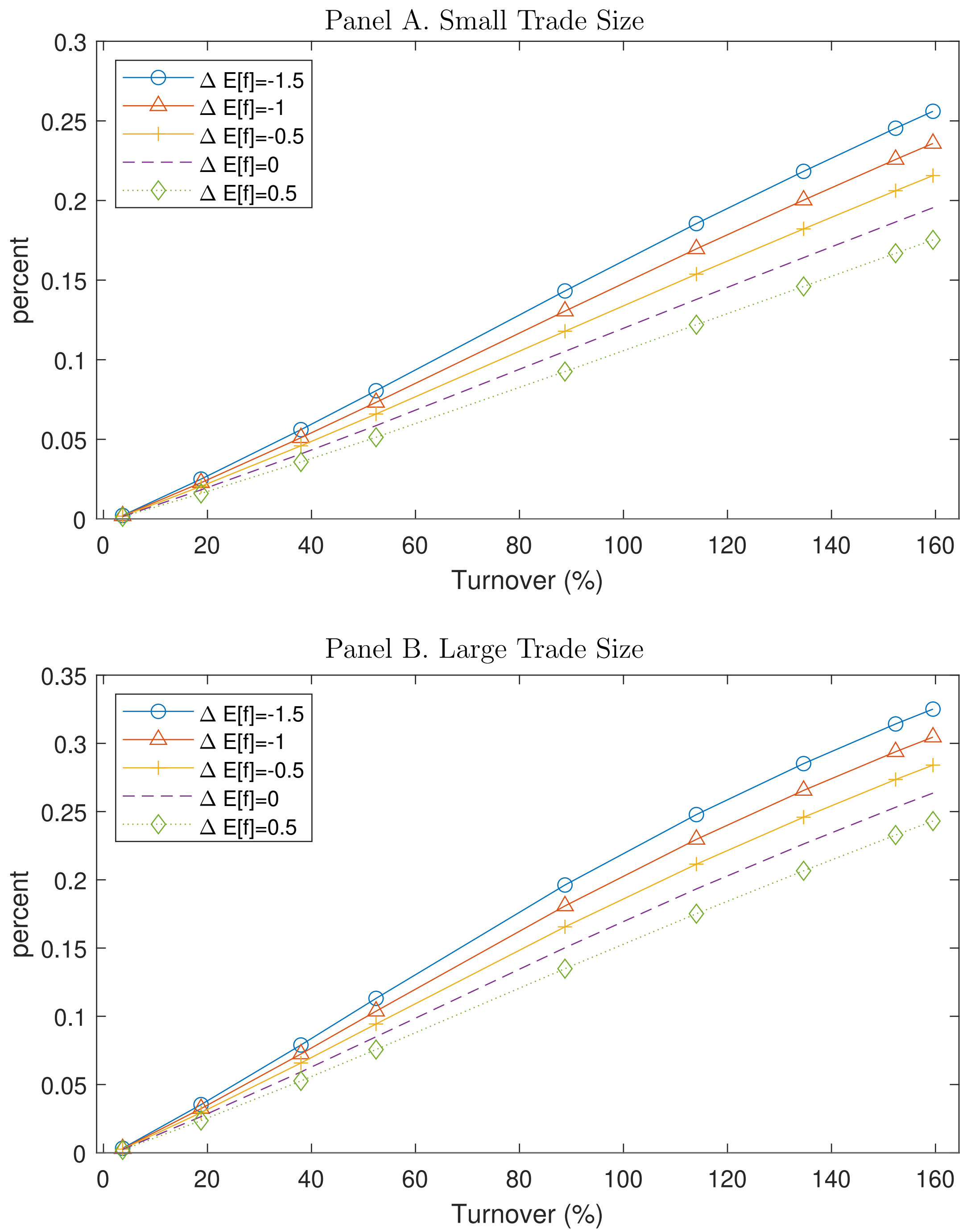
The figure plots the bond CAPM alpha (% per month) associated with the machine learning-based factors constructed in Section 6 before transaction costs, after bid-ask spread costs, and after total transaction costs (including bid-ask spread and delay costs). Panel A applies the transaction costs from our lower bound estimates for a trade size of $100,000, while Panel B uses that of $2 million. The machine learning factors are formed by sorting on the one-month ahead bond return forecasts across several models trained on the 341 stock and bond characteristics using decile sorts, where the factor initiates a long position in decile 10 and a short position in decile 1. The linear models with penalization comprise an elastic net (ENET), Lasso (LASSO), Ridge (RIDGE), and the average of the three linear return forecasts, Linear Ensemble (LENS). The nonlinear models comprise a feedforward neural network (NN), tree-based aggregation methods comprising an extremely randomized set of trees (XT) and a random forest (RF). NENS is the average of the three nonlinear return forecasts, and ENS is the average return forecast across all models. The error bars represent the two standard error bounds that account for the Newey-West 12 lags. All factors are formed with the `PyBondLab` Python package.

Figure 8: Splitting a $2 Million Order Into Smaller Pieces



The figure plots the delay and bid-ask spread costs based on the simulation with parameters $\sigma_v = \{0.10\%, 0.60\%\}$ and $\rho = \{0.3, 0.95\}$, which correspond to average excess returns of 0.09% and 0.51%, and turnover rates of 135% and 38%. We consider the strategy to split the orders of $2 million into κ pieces, as discussed in Section 7.3. The delay for each piece is independently drawn from the exponential distribution with corresponding trade sizes. For example, if $\kappa = 2$, then we use the λ (the lower bound estimate) and half spread for a trade size of $1 million (=2/2). Then, the size-specific delay is averaged across pieces to compute the total delay. If at the end of the month, the orders are partially filled, then we consider the unfilled pieces as executed exactly at the end of the month and assign zero returns. When no pieces are executed within a month, then we consider it as an execution failure and do not assume month-end execution.

Figure 9: Factor Cyclicality and the Cost of Delay



The figure plots the simulated cost of delay on the y-axis, defined as the difference in average returns between gross factors without delays and net factors with delays accounting for time-variation in the trade intensity, as discussed in Section 7.4. Delays follow an exponential distribution with trade intensity parameter λ_t, which varies across two states of the economy. The intensity estimates are from Kargar et al. (2025): 0.115 for $100,000 trades (odd lot) and 0.072 for $2 million trades (round lot) during booms, and 0.068 and 0.042 during recessions. We set the baseline factor average excess return to 0.51% per month and vary the factor cyclicality parameter $\Delta E[f]$ from -1.5% (countercyclical) to 0.5% (cyclical).

Table 1: List of Papers on the Cross-Section of Corporate Bond Returns

Article	Cost Estimates	Holding Period
Panel A. Papers Without Transaction Costs		
Bai, Bali, and Wen (2019)		1
Bai, Bali, and Wen (2021)		1
Bali, Subrahmanyam, and Wen (2021a)		1
Bali, Subrahmanyam, and Wen (2021b)		1
Ceballos (2023)		1
Chen, Wang, and Wu (2022)		1
Chung, Wang, and Wu (2019)		1
Dang, Hollstein, and Prokopczuk (2023)		1
Dick-Nielsen, Feldhütter, Pedersen, and Stolborg (2023)		1
Dickerson, Mueller, and Robotti (2023)		1
Dickerson, Julliard, and Mueller (2025)		1
Duan, Li, and Wen (2025)		12
Friewald and Nagler (2024)		0.25
Gebhardt, Hvidkjaer, and Swaminathan (2005a)		1
Gebhardt, Hvidkjaer, and Swaminathan (2005b)		1−13
Haesen, Houweling, and Zundert (2017)		6
Huang, Qin, and Wang (2024)		1
Lin, Wang, and Wu (2011)		1
Tao, Wang, Wang, and Wu (2022)		1
Panel B. Papers Incorporating Transaction Costs		
Bali et al. (2020)	Roll measure of Bao et al. (2011)	1
Bali, Beckmeyer, and Goyal (2023)	Fixed at 35bps	1
Baltussen, Muskens, and Verwijmeren (2024)	Bid-ask spreads from imputed roundtrip trades (IRT)	1
Bartram, Grinblatt, and Nozawa (2025)	Portfolio-level bid-ask spreads	1
Bredendiek, Ottonello, and Valkanov (2023)	Round-trip transaction costs	1
Cao et al. (2023)	Estimates following Edwards et al. (2007)	1
Choi and Kim (2018)	Considers transaction costs as characteristics	1, 12
Chordia et al. (2017)	Portfolio-level bid-ask spreads	1
He, Feng, Wang, and Wu (2025)	Fixed at 20 to 80bps	1
Houweling and Zundert (2017)	Maturity-rating, following Chen et al. (2007)	12
Israel, Palhares, and Richardson (2018)	Maturity-rating, following Chen et al. (2007)	1
Ivashchenko (2024)	Average 12m moving average of bond bid-ask spreads	1
Ivashchenko and Kosowski (2024)	Estimates following Kyle and Obizhaeva (2016)	1
Jostova et al. (2013)	Estimates following Edwards et al. (2007)	6
Kelly, Palhares, and Pruitt (2023)	Fixed at 19bps	1
Nozawa, Qiu, and Xiong (2023)	Bond-level bid-ask spreads	1

The table lists papers on the cross-section of corporate bond returns. Holding Period is the period over which a bond is held before rebalancing occurs, and it is measured in months.

Table 2: Simulated Cost of Delay (% Per Month)

Panel A. Fix $\sigma_v = 0.6$ and Vary λ

λ	Expected Delay	Turnover Rate (%) 3.7	18.7	37.9	52.4	88.7	114.0	134.6	152.3	159.5
1.000	1	0.000	0.002	0.005	0.007	0.012	0.016	0.019	0.022	0.023
0.200	5	0.001	0.010	0.024	0.034	0.062	0.081	0.098	0.113	0.119
0.048	21	0.003	0.035	0.078	0.112	0.194	0.245	0.282	0.311	0.321
0.024	42	0.005	0.055	0.121	0.171	0.278	0.333	0.368	0.392	0.401
0.004	252	0.027	0.151	0.291	0.359	0.443	0.467	0.479	0.487	0.489

Panel B. Fix λ and Vary σ_v

σ_v	Gross Returns	Turnover Rate (%) 3.7	18.7	37.9	52.4	88.7	114.0	134.6	152.3	159.5
B1. $\lambda = 1/21$										
0.01	0.009	0.000	0.001	0.002	0.002	0.003	0.004	0.005	0.006	0.006
0.10	0.085	0.001	0.006	0.013	0.019	0.032	0.041	0.047	0.052	0.054
0.30	0.255	0.002	0.018	0.039	0.056	0.097	0.122	0.141	0.156	0.161
0.60	0.510	0.003	0.035	0.078	0.112	0.194	0.245	0.282	0.311	0.321
1.00	0.850	0.005	0.058	0.130	0.186	0.323	0.408	0.469	0.517	0.535
B2. Own Estimates of λ, Small Trade ($100,000)										
0.01	0.009	0.000	0.000	0.001	0.001	0.002	0.002	0.002	0.003	0.003
0.10	0.085	0.000	0.003	0.006	0.008	0.015	0.019	0.023	0.026	0.027
0.30	0.255	0.001	0.007	0.017	0.024	0.043	0.057	0.068	0.078	0.082
0.60	0.510	0.001	0.015	0.033	0.048	0.086	0.114	0.136	0.155	0.163
1.00	0.850	0.002	0.024	0.055	0.080	0.144	0.189	0.226	0.259	0.272
B3. Own Estimates of λ, Large Trade ($2 million)										
0.01	0.009	0.000	0.001	0.001	0.002	0.003	0.004	0.004	0.005	0.005
0.10	0.085	0.000	0.005	0.012	0.016	0.029	0.036	0.042	0.047	0.049
0.30	0.255	0.001	0.015	0.034	0.049	0.086	0.109	0.127	0.141	0.147
0.60	0.510	0.003	0.030	0.068	0.097	0.171	0.218	0.254	0.282	0.293
1.00	0.850	0.004	0.050	0.113	0.162	0.285	0.364	0.423	0.470	0.489

Panel C. Bid-Ask Spread Costs

Trade Size	Turnover Rate (%) 3.7	18.7	37.9	52.4	88.7	114.0	134.6	152.3	159.5
Small ($100,000)	0.018	0.071	0.132	0.174	0.263	0.314	0.348	0.371	0.378
Large ($2 million)	0.011	0.044	0.080	0.105	0.159	0.191	0.214	0.230	0.236

Panel D. Net Turnover Rates

Trade Size	Turnover Rate (%) 3.7	18.7	37.9	52.4	88.7	114.0	134.6	152.3	159.5
Small ($100,000)	3.5	13.8	25.6	33.5	50.8	60.6	67.3	71.6	72.9
Large ($2 million)	3.5	13.5	25.0	32.7	49.5	59.3	66.3	71.5	73.2

Panels A and B report the simulated cost of delays from Eq. (6) in Section 3.3. λ is the trade intensity and σ_v is the volatility of the shock to the factor loading discussed in Section 3. We vary the speed of mean reversion of the signal, ρ, from 0.01 to $1 - 10^{-5}$, corresponding to monthly turnover rates ranging from 159.5% to 3.7%. Other parameters include $T = 244$, $N = 1,000$, $E[f] = 0.30$, and $\sigma_f = 1.67$, with σ_u calibrated to match the average time-series R^2 of 0.369 obtained by regressing each bond's excess returns on the bond market factor. The number of simulated paths is 1,000. Panel C reports the bid-ask spread costs defined in Eq. (8), and Panel D reports the net turnover rate, $\overline{TO}_{Sim,P}$, accounting for delays.

Table 3: Forecasting Regressions of Corporate Bond Transactions

		Customer Sell				Customer Buy			
		\$100,000		\$2M		\$100,000		\$2M	
Panel A. Regression Estimates									
$\log InvComp$		0.018	(15.08)	0.028	(12.13)	0.017	(14.61)	0.030	(12.76)
$\Delta Inventory$		−0.010	(−23.58)	−0.047	(−45.14)	0.008	(15.59)	0.044	(35.25)
CAP		0.002	(0.79)	0.016	(5.14)	0.000	(-0.12)	0.013	(4.40)
$Rating$		0.054	(26.37)	0.109	(41.40)	0.054	(26.53)	0.111	(43.05)
$Maturity$		−0.017	(−7.65)	−0.007	(−4.45)	−0.020	(−9.61)	−0.002	(−1.40)
$\log FaceValue$		0.099	(45.10)	0.206	(107.89)	0.104	(49.34)	0.203	(107.22)
$Coupon$		−0.028	(−14.25)	−0.003	(−0.81)	−0.028	(−14.11)	0.002	(0.74)
Age		−0.031	(−17.09)	−0.077	(−37.19)	−0.036	(−19.18)	−0.078	(−38.25)
Intercept		0.877	(435.77)	0.588	(188.51)	0.868	(439.30)	0.576	(187.46)
Adj. R^2		0.142		0.251		0.151		0.248	
N		814,145		814,145		814,145		814,145	
Panel B. Model-Based Estimates of Trade Intensity and Delays									
Prob. Trade	$Prob^{Data}$	0.877		0.588		0.868		0.576	
	$Prob^{CF}$	0.953		0.708		0.941		0.702	
λ	λ^{Data}	0.100		0.042		0.097		0.041	
	λ^{CF}	0.145		0.059		0.135		0.058	
Exp. Delay	$1/\lambda^{Data}$	10.014		23.672		10.362		24.442	
(days)	$1/\lambda^{CF}$	6.891		17.054		7.418		17.355	

The table reports the coefficient estimates for the regressions of the bond trading dummy on corporate bond characteristics based on Eq. (11) of Section 4.2. The variable $InvComp$ is the investor composition metric of Li and Yu (2025), which measures the activeness of bond investors at the end of the previous quarter. The variable $\Delta Inventory$ is the difference between customer buys and customer sells in the preceding 28 days, CAP is the intermediary capital ratio of He et al. (2017) in the previous month. The right-hand-side variables are standardized for ease of interpretation. Values in parentheses are t-statistics double clustered at the bond and month levels. The sample is at the monthly frequency from August 2002 to November 2022.

Table 4: Gross and Net Bond CAPM α of 58 Factors by Category

Category	Equity Momentum	Equity Reversal	Investment	Spreads	Value Profit	Volatility Risk	All
Count	9	4	13	14	7	11	58
Panel A. Gross CAPM α							
Avg. α	0.287	0.426	0.143	0.437	0.160	0.241	0.277
Avg. $t(\alpha)$	(2.95)	(2.43)	(2.54)	(2.31)	(1.95)	(2.60)	(2.48)
Avg. Turnover	142.46	136.30	29.43	47.45	35.33	83.99	69.75
$\#(t(\alpha) > 1.96)$	8	3	11	9	4	7	42
Panel B. Net CAPM α, Own Estimates of Delay							
B1. Small Trades							
Avg. α	−0.103	−0.019	0.031	0.232	0.026	0.005	0.050
Avg. $t(\alpha)$	(−1.48)	(−0.13)	(0.54)	(1.06)	(0.18)	(−0.92)	(−0.01)
Avg. BidAsk Cost	0.332	0.328	0.105	0.153	0.123	0.202	0.188
Avg. Delay Cost	0.058	0.117	0.007	0.052	0.011	0.034	0.039
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	1	1
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	0	0	0	1	0	1	2
B2. Large Trades							
Avg. α	−0.025	0.008	0.065	0.239	0.062	0.053	0.087
Avg. $t(\alpha)$	(−0.57)	(0.06)	(1.15)	(1.14)	(0.69)	(−0.17)	(0.50)
Avg. BidAsk Cost	0.205	0.203	0.064	0.093	0.074	0.125	0.115
Avg. Delay Cost	0.107	0.215	0.014	0.105	0.023	0.063	0.075
$\#(t(\alpha) > 1.96)$	0	0	0	1	0	1	2
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	1	2	4	0	1	10
Panel C. Net CAPM α, Kargar et al. (2023) Estimates of Delay							
C1. Small Trades							
Avg. α	−0.115	−0.043	0.030	0.220	0.024	−0.002	0.042
Avg. $t(\alpha)$	(−1.61)	(−0.26)	(0.52)	(1.00)	(0.16)	(−1.02)	(−0.08)
Avg. BidAsk Cost	0.332	0.328	0.105	0.152	0.122	0.202	0.188
Avg. Delay Cost	0.070	0.142	0.008	0.064	0.014	0.041	0.047
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	1	1
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	0	0	0	1	0	1	2
C2. Large Trades							
Avg. α	−0.015	0.029	0.067	0.251	0.065	0.059	0.094
Avg. $t(\alpha)$	(−0.46)	(0.17)	(1.18)	(1.21)	(0.71)	(−0.09)	(0.57)
Avg. BidAsk Cost	0.206	0.203	0.064	0.094	0.075	0.125	0.116
Avg. Delay Cost	0.097	0.194	0.012	0.092	0.020	0.057	0.067
$\#(t(\alpha) > 1.96)$	0	0	1	1	0	1	3
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	1	2	4	0	1	10

The table reports the average bond CAPM α, the associated average t-statistic, and the number of factors with a t-statistic greater than 1.96 before and after accounting for transaction costs, including bid-ask spread costs and delay costs across factors associated with the six categories from Section 5.2. Of the 341 factors that we generate, we consider the 58 factors with significant average excess returns. We regress each factor on the corporate bond market factor and estimate the intercept of the regression, α. We group the 58 factors into six categories and average the estimates within each group. The values in parentheses are the average of the t-statistics (Newey-West adjusted with 12 lags). In Panel B, we use our own estimates of trade intensity to calculate the cost of delays and present the alphas net of costs and the number of factors that remain statistically significant. In Panel C, we use the Kargar et al. (2025) delay estimates for the trade intensity parameter and compute the cost of delays for small and large trades, respectively. All factors are formed with the `PyBondLab` Python package.

Table 5: Profitability of ML Strategies Based on Simulated Cost of Transactions

Signal	Turnover	Gross α	Trade Size: \$100,000			Trade Size: \$2M		
			Net α	Cost (%)		Net α	Cost (%)	
	(%)	(%)	(%)	Delay	BA	(%)	Delay	BA
ENET	125	0.850	0.344	0.175	0.332	0.317	0.331	0.203
		(5.03)	(2.03)			(1.87)		
ENS	121	0.901	0.392	0.184	0.326	0.352	0.350	0.199
		(4.71)	(2.05)			(1.84)		
LASSO	126	0.848	0.336	0.178	0.335	0.308	0.335	0.204
		(5.09)	(2.02)			(1.85)		
LENS	125	0.868	0.348	0.188	0.332	0.310	0.356	0.203
		(4.71)	(1.89)			(1.68)		
NENS	106	0.847	0.405	0.145	0.297	0.385	0.281	0.181
		(4.81)	(2.30)			(2.19)		
NN	117	0.801	0.323	0.159	0.320	0.303	0.303	0.195
		(4.64)	(1.87)			(1.75)		
RF	97	0.620	0.251	0.089	0.280	0.276	0.174	0.170
		(4.62)	(1.87)			(2.06)		
RIDGE	120	0.811	0.312	0.176	0.324	0.279	0.335	0.198
		(4.51)	(1.73)			(1.55)		
XT	101	0.870	0.437	0.146	0.288	0.411	0.285	0.175
		(4.89)	(2.45)			(2.31)		

The table reports the bond CAPM α for the bond factors formed using the predictions of the machine learning (ML) models from Section 6 and the transaction cost estimates based on the simulations described in Section 3. The machine learning factors are formed by sorting on the one-month ahead bond return forecasts across several models trained on the 341 stock and bond characteristics using decile sorts, where the factor initiates a long position in decile 10 and a short position in decile 1. The simulated costs are assigned to each ML factor using the factor's gross returns and turnover rates. Net α is the difference between gross α and the simulated total costs. The trade intensity is our own estimate based on the counterfactual trade probability. The linear models with penalization comprise an elastic net (ENET), Lasso (LASSO), Ridge (RIDGE), and the average of the three return forecasts, Linear Ensemble (LENS). The nonlinear models comprise a feedforward neural network (NN), tree-based aggregation methods comprising an extremely randomized set of trees (XT) and a random forest (RF). NENS is the average of the three nonlinear model bond return forecasts, and ENS is the average across all models. All factors are formed with the `PyBondLab` Python package.

Table 6: Effect of Delay Cost Mitigation Techniques

Category	Equity Momentum	Equity Reversal	Investment	Spreads	Value Profit	Volatility Risk	All
Panel A. 3-Month Holding Period							
A1. Gross CAPM α							
Avg. α	0.174	0.271	0.117	0.333	0.130	0.169	0.200
Avg. $t(\alpha)$	(2.51)	(2.26)	(2.32)	(1.78)	(1.71)	(2.09)	(2.10)
Avg. Turnover	54.36	48.50	26.45	27.09	25.00	35.84	34.06
A2. Net CAPM α, Small Trades							
Avg. α	−0.018	0.086	0.016	0.214	0.033	0.036	0.069
Avg. $t(\alpha)$	(−0.99)	(0.53)	(0.29)	(1.01)	(0.34)	(−0.67)	(0.11)
Avg. BidAsk Cost	0.178	0.160	0.096	0.097	0.091	0.121	0.117
Avg. Delay Cost	0.013	0.025	0.005	0.023	0.007	0.011	0.013
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	1	1
A3. Net CAPM α, Large Trades							
Avg. α	0.039	0.124	0.048	0.228	0.061	0.073	0.102
Avg. $t(\alpha)$	(0.10)	(0.91)	(0.95)	(1.13)	(0.76)	(0.16)	(0.68)
Avg. BidAsk Cost	0.108	0.097	0.058	0.059	0.055	0.074	0.071
Avg. Delay Cost	0.027	0.050	0.010	0.047	0.013	0.022	0.027
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	1	1
Panel B. Banding							
B1. Gross CAPM α							
Avg. α	0.240	0.395	0.114	0.289	0.098	0.190	0.208
Avg. $t(\alpha)$	(2.86)	(2.46)	(2.35)	(1.95)	(1.43)	(2.44)	(2.25)
Avg. Turnover	111.07	101.25	18.70	20.80	17.78	60.67	47.08
B2. Net CAPM α, Small Trades							
Avg. α	−0.091	0.031	0.040	0.200	0.027	0.003	0.049
Avg. $t(\alpha)$	(−1.44)	(0.12)	(0.81)	(1.30)	(0.31)	(−0.68)	(0.19)
Avg. BidAsk Cost	0.291	0.279	0.071	0.073	0.067	0.164	0.137
Avg. Delay Cost	0.041	0.085	0.003	0.016	0.004	0.023	0.022
$\#(t(\alpha) > 1.96)$	0	0	0	1	0	1	2
B3. Net CAPM α, Large Trades							
Avg. α	−0.015	0.063	0.064	0.212	0.050	0.045	0.082
Avg. $t(\alpha)$	(−0.44)	(0.36)	(1.31)	(1.40)	(0.67)	(0.04)	(0.68)
Avg. BidAsk Cost	0.179	0.170	0.043	0.045	0.041	0.101	0.084
Avg. Delay Cost	0.077	0.162	0.007	0.033	0.008	0.044	0.042
$\#(t(\alpha) > 1.96)$	1	0	2	1	0	1	5

The table presents results that refer to the factor cost mitigation strategies discussed in Section 7.1. We report the average bond CAPM α, the average t-statistic of the CAPM α, and the number of factors with a t-statistic greater than 1.96 before and after accounting for transaction costs, including bid-ask spread costs and delay costs. We consider the 58 factors with significant average excess returns. We regress each factor on the corporate bond market factor and estimate the intercept of the regression, α. We group the 58 factors into six categories and average the estimates within each group. The values in parentheses are the average of the t-statistics (Newey-West adjusted with 12 lags). To compute the cost of delays, we use our own estimates of the trade intensity. In Panel A, we extend the holding period from one month to three months, employing the staggered rebalancing approach of Jegadeesh and Titman (1993). In Panel B, we hold a bond in the high or low portfolios once it enters, contingent on the bond remaining within two neighboring deciles (i.e., within a 2-decile 'band') following Novy-Marx and Velikov (2015). The bond only exits the portfolio once it drops out of this band. (Continued on the next page.)

Table 6 (Continued)

Category	Equity Momentum	Equity Reversal	Investment	Spreads	Value Profit	Volatility Risk	All
Panel C. Subsample of Bonds with Above-Median Trading Volume							
C1. Gross CAPM α							
Avg. α	0.318	0.496	0.191	0.486	0.175	0.290	0.320
Avg. $t(\alpha)$	(2.61)	(2.56)	(2.44)	(2.08)	(1.69)	(2.50)	(2.31)
Avg. Turnover	152.67	150.53	67.80	81.03	70.33	114.63	99.06
C2. Net CAPM α, Small Trades							
Avg. α	−0.087	0.022	−0.039	0.163	−0.066	−0.047	0.002
Avg. $t(\alpha)$	(−0.86)	(0.11)	(−0.61)	(0.50)	(−0.86)	(−1.05)	(−0.45)
Avg. BidAsk Cost	0.349	0.352	0.212	0.241	0.218	0.282	0.264
Avg. Delay Cost	0.056	0.122	0.017	0.083	0.023	0.055	0.054
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	0	0
C3. Net CAPM α, Large Trades							
Avg. α	−0.013	0.032	0.024	0.161	−0.008	0.001	0.044
Avg. $t(\alpha)$	(−0.20)	(0.18)	(0.24)	(0.54)	(−0.20)	(−0.45)	(0.06)
Avg. BidAsk Cost	0.216	0.218	0.129	0.147	0.133	0.174	0.162
Avg. Delay Cost	0.114	0.247	0.038	0.178	0.051	0.115	0.114
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	0	0
Panel D. Subsample of Bonds with Above-Median Amount Outstanding							
D1. Gross CAPM α							
Avg. α	0.300	0.375	0.137	0.378	0.151	0.236	0.257
Avg. $t(\alpha)$	(2.63)	(2.15)	(2.22)	(2.15)	(1.79)	(2.51)	(2.27)
Avg. Turnover	142.71	136.68	30.54	46.93	36.79	84.37	70.19
D2. Net CAPM α, Small Trades							
Avg. α	−0.034	0.044	0.028	0.225	0.023	0.032	0.067
Avg. $t(\alpha)$	(−0.57)	(0.22)	(0.44)	(1.16)	(0.17)	(−0.46)	(0.24)
Avg. BidAsk Cost	0.334	0.330	0.109	0.152	0.128	0.204	0.190
Avg. Delay Cost	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	1	1
D3. Net CAPM α, Large Trades							
Avg. α	0.015	0.036	0.061	0.221	0.058	0.068	0.092
Avg. $t(\alpha)$	(−0.06)	(0.19)	(0.99)	(1.17)	(0.63)	(0.08)	(0.60)
Avg. BidAsk Cost	0.207	0.204	0.067	0.094	0.079	0.126	0.117
Avg. Delay Cost	0.078	0.134	0.009	0.063	0.014	0.042	0.048
$\#(t(\alpha) > 1.96)$	0	0	1	0	0	1	2

In Panels C and D, we employ the subsample of bonds whose turnover rate (volume/amount outstanding) or amount outstanding, respectively, is above the median in the previous month and construct quintile portfolios to create the factors. All factors are formed with the `PyBondLab` Python package.

Table 7: Forecasting Regressions of Corporate Bond Transactions: Liquid Bonds

		Bonds with High Turnover				Large Bonds			
		\$100,000		\$2M		\$100,000		\$2M	
Panel A. Regression Estimates									
$\log InvComp$		0.015	(13.62)	0.022	(10.74)	0.019	(12.89)	0.030	(11.10)
$\Delta Inventory$		−0.011	(−24.70)	−0.050	(−47.64)	−0.009	(−25.16)	−0.044	(−37.39)
CAP		0.001	(0.33)	0.014	(4.77)	−0.007	(−2.55)	−0.002	(−0.73)
$Rating$		0.042	(21.04)	0.081	(34.20)	0.052	(23.84)	0.105	(38.71)
$Maturity$		−0.017	(−7.88)	−0.008	(−5.35)	−0.013	(−5.63)	0.000	(0.16)
$\log FaceValue$		0.089	(46.35)	0.182	(108.50)				
$Coupon$		−0.026	(−13.33)	0.003	(1.04)	−0.053	(−20.77)	−0.054	(−18.24)
Age		−0.025	(−13.46)	−0.061	(−33.90)	−0.030	(−14.83)	−0.073	(−32.22)
$HighTurnover$		0.077	(23.33)	0.181	(53.69)				
$LargeFaceValue$						0.156	(41.09)	0.327	(90.07)
Intercept		0.839	(270.78)	0.498	(148.89)	0.804	(190.55)	0.435	(110.21)
Adj. R^2		0.153		0.278		0.118		0.207	
N		814,145		814,145		814,145		814,145	
Panel B. Model-Based Estimates of Trade Intensity and Delays									
Prob. Trade	$Prob^{Data}$	0.916		0.679		0.960		0.762	
	$Prob^{LB}$	0.979		0.770		-		0.890	
λ	λ^{Data}	0.118		0.054		0.153		0.068	
	λ^{LB}	0.183		0.070		-		0.105	
Exp. Delay	$1/\lambda^{Data}$	8.488		18.503		6.515		14.613	
(days)	$1/\lambda^{LB}$	5.451		14.305		-		9.498	

The table reports the coefficient estimates for the regressions of the bond trading dummy for the customer sell side on characteristics based on Eq. (11) of Section 3.4. The variable $InvComp$ is the investor composition variable of Li and Yu (2025), which measures the activeness of bond investors at the end of the previous quarter. The variable $\Delta Inventory$ is the difference between customer buys and customer sells in the preceding 28 days, CAP is the intermediary capital ratio of He et al. (2017) in the previous month. $HighTurnover$ is a dummy which is one if the bond is above the median in terms of bond turnover rate in month t and zero otherwise. $LargeFaceValue$ is a dummy which is one if the bond is above the median in terms of amount outstanding in month t and zero otherwise. Except for the two dummy variables, the right-hand-side variables are standardized for ease of interpretation. Values in parentheses are t-statistics clustered at the bond and month levels. The − in the probability of trade estimate, $Prob^{LB}$, indicates that the estimate exceeds one. In this case, we set $\lambda = 9999$ in the simulations to minimize the delay. The sample is monthly, spanning August 2002 to November 2022.

Internet Appendix to

Factor Investing with Delays

(not for publication)

Abstract

This Internet Appendix presents supplementary material and results not included in the main body of the paper.

A Data and Variable Construction

The following sections describe the various databases that we use in the paper. Across all databases, we filter out bonds that have a time-to-maturity of less than 1 year. Furthermore, for consistency, across all databases, we define bond ratings as those provided by Standard&Poors (S&P). We include the full spectrum of ratings (AAA to D), but exclude bonds that are unrated. For each database that we consider, we *do not* winsorize or trim bond returns in any way.

A.1 Corporate Bond Databases

Mergent Fixed Income Securities Database (FISD)

FISD for academia is a comprehensive database of publicly-offered U.S. bonds. FISD provides details on market trends, deal structures, issuer capital structures, and other areas of fixed income debt research. We apply the standard filters to the FISD data as they relate to corporate bond pricing:

1. Only keep bonds that are issued by firms domiciled in the United States of America, `COUNTRY_DOMICILE == 'USA'`.
2. Remove bonds that are private placements, `PRIVATE_PLACEMENT == 'N'`.
3. Only keep bonds that are traded in U.S. Dollars, `FOREIGN_CURRENCY == 'N'`.
4. Bonds that trade under the 144A Rule are discarded, `RULE_144A == 'N'`.
5. Remove all asset-backed bonds, `ASSET_BACKED == 'N'`.
6. Remove convertible bonds, `CONVERTIBLE == 'N'`.

7. Only keep bonds with a fixed or zero coupon payment structure, i.e., remove bonds with a floating (variable) coupon, `COUPON_TYPE != 'V'`.

8. Remove bonds that are equity linked, agency-backed, U.S. Government, and mortgage-backed, based on their `BOND_TYPE`.

Bank of America Merrill Lynch (BAML) Database

The BAML data is provided by the Intercontinental Exchange (ICE) and provides daily bond price quotes, accrued interest, and a host of pre-computed corporate bond characteristics such as the bond option-adjusted credit spread (OAS), the asset swap spread, duration, convexity, and bond returns in excess of a portfolio of duration-matched Treasuries. The ICE sample spans the time period 1997:01 to 2022:12 and includes constituent bonds from the ICE Bank of America High Yield (H0A0) and Investment Grade (C0A0) Corporate Bond Indices.

ICE Bond Filters. We follow van Binsbergen, Nozawa, and Schwert (2025) and take the last quote of each month to form the bond-month panel. We then merge the ICE data to the filtered FISD data. The following ICE-specific filters are then applied:

1. Only include corporate bonds, `Ind_Lvl_1 == 'corporate'`

2. Only include bonds issued by U.S. firms, `Country == 'US'`

3. Only include corporate bonds denominated in U.S. Dollars, `Currency == 'USD'`

BAML/ICE Bond Returns. Total bond returns are computed in a standard manner in ICE, and no assumptions about the timing of the last trading day of the month are made because the data is quote-based, i.e., there is always a valid quote at month-end to compute a bond return. This means that each bond return is computed using a price quote at exactly the end of the month, each and every month. This introduces homogeneity into the bond returns because prices are sampled at exactly the same time each month. ICE only provides bid-side pricing, meaning bid-ask bias is inherently not present in the monthly sampled prices, returns, and credit spreads. The monthly ICE return variable is (as denoted in the original database) `trr_mtd_loc`, which is the month-to-date return on the last business day of month t. We use this return specification (in excess of the one-month risk-free rate of return) and the bond returns in excess of a portfolio of duration-matched U.S. Treasury bond returns (denoted as `ex_rtn_mtd` in the ICE dataset) as the dependent variables to train the machine learning models. In robustness exercises, we use several versions of the TRACE data – our main results remain unchanged.

Enhanced TRACE Database

TRACE provides data on corporate bond transactions. Since we measure the profitability of factor investing from an end-user perspective, we use only dealer-customer transactions (`cntra_mp_id = 'C'`). We remove trades that are i) when-issued (`wis_fl != 'Y'`), ii) locked-in (`lckd_in_ind != 'Y'`), iii) with special conditions (`sale_cndtn_cd = '@' or sale_cndtn_cd = ''`). In addition, we restrict our sample to those with standard settlement days (`days_to_sttl_ct = '' or days_to_sttl_ct = '000' or days_to_sttl_ct = '001' or days_to_sttl_ct = '002'`).

However, some transaction records contain prices that appear to reflect clerical/recording errors. We avoid simply removing outliers in terms of prices and returns because such procedures bias the standard deviation of returns downward and inflate Sharpe ratios. Furthermore, if we simply removed very low returns, we would eliminate bonds that have defaulted, leading to potentially spurious profitability of a given factor. To avoid these problems, we apply the reversal filter of Bessembinder et al. (2008) with a wider band. That is, we examine the log price changes of a bond using two consecutive transactions. If a product of two adjacent log price changes is less than -0.25 (i.e., a 50% decline followed by a 50% increase), then we consider the price record in the middle to be an error and remove it. Finally, we follow Edwards et al. (2007) and remove transactions whose volume is more than 50% of the amount outstanding or whose volume is not an integer.

After applying these filters, we compute the average price of a bond on a day, separately for dealer buys and dealer sells.

A.2 Bond and Stock Characteristics

We describe our 53 'custom-made' bond and equity characteristics in Table A.1. Panel A describes our 37 bond-based characteristics that span the vast majority of those used in the literature on corporate bond factors. Panel B describes additional equity-based characteristics that are not included in publicly available equity repositories, but have all been used in research that attempts to predict bond returns or form bond factors. All rank-demeaned characteristics are made publicly available on `openbondassetpricing.com/data`.

For the publicly available equity databases, we follow Chen and Velikov (2023) and drop characteristics that are discrete (i.e., exchange indicators) or dominated by missing values at the stock level. When there are overlaps between the Chen and Zimmermann (2022), CZ, and Jensen et al. (2023), JKP, characteristics, we drop the CZ version of the characteristic. In total, after dropping characteristics based on the above, we are left with 137 CZ characteristics and 151 JKP characteristics.

B Net-of-Fees Corporate Bond Market Factor

We risk-adjust our net-of-cost strategies with a realistic corporate bond market factor that combines tradable passively managed investment grade and high yield exchange traded funds (ETFs). We source the BlackRock iShares iBoxx Investment Grade (ticker: LQD) and High Yield (ticker: HYG) ETF net returns from the CRSP Mutual Funds database as provided by WRDS. The LQD ETF has an inception date of 2002:06, which spans the full length of our out-of-sample period. The HYG inception date is 2007:03. To address the shorter sample period for HYG, we source high yield gross return data from the Bloomberg-Barclays (BB) High-Yield bond index. Thereafter, we estimate a simple OLS regression of the HYG net returns on the BB gross returns such that we can extrapolate values for HYG before 2007:03,

$$R_{HYG,t} = \beta_0 + \beta_{BB} \cdot R_{BB,t} + \varepsilon_t,$$

$$\widehat{R_{HYG,t}} = \underset{(-2.010)}{-0.095} + \underset{(60.13)}{0.883} \cdot R_{BB,t},$$

where $R_{HYG,t}$ and $R_{BB,t}$ are the net-of-cost and gross returns on the HYG ETF and BB High-Yield bond index over the sample period 2007:03–2023:06 ($T = 251$). The intercept, β_0, is estimated at -9.5 basis points (statistically different from zero at the 5% nominal level), which captures the fact that HYG is adversely impacted by trading costs and ETF fees. From the OLS estimation above, we set the net return value of the HYG index to $\widehat{R_{HYG}}$ before 2007:03 and to the actual net return of the HYG index thereafter. We denote this return R_{HYG}.

To generate the $MKTB_{Net}$ factor, we require appropriate weights for the representative investor to apportion their funds between HYG and LQD. To accomplish this, we source *all* bonds that are included in the Bank of America Merrill Lynch Investment Grade (C0A0) and High Yield (H0A0) corporate indices and compute their respective market capitalizations (Clean Price × Units Outstanding). The weight for each index for each month is simply the sum of the respective index market capitalization at month t divided by the total market capitalization. On average, over the sample period, the investor apportions 19.90% to the high yield index and 80.10% to the investment grade index. Finally, the $MKTB_{Net}$ factor is computed as,

$$R^{Net}_{MKTB,t+1} = (R_{HYG,t+1} \cdot \omega_{HYG,t} + R_{LQD,t+1} \cdot \omega_{LQD,t}) - R_{f,t+1},$$

where $\omega_{HYG,t}$ is the weight in the HYG ETF, $\omega_{LQD,t}$ is the weight in the LQD ETF, and $R_{f,t+1}$ is the one-month risk-free rate of return from Kenneth French's webpage.

We report summary statistics for the $MKTB_{Net}$, $MKTB_{Gross}$ (computed using the same weights as above with the Bloomberg-Barclays Investment Grade and High Yield index gross returns), and $MKTB$ available from `openbondassetpricing.com` in Table A.14.

C Computing Bid-Ask and Delay Costs for the Factors

More specifically, for each factor $f_i = 1, \ldots, 58$, we compute the bid-ask, delay, and total costs as follows:

Procedure Computation of Delay and Bid-Ask Costs for Each Factor $f_i = 1, \ldots, 58$

Input: Gross average factor returns $\overline{f}_i$, alphas $\overline{f}_{\alpha_i}$, turnovers $\overline{TO}_{f_i}$, trade intensities $\lambda_{\$100{,}000}$ and $\lambda_{\$2\text{M}}$, simulated turnover grids, and average half-spreads $\overline{h}_k$.

Output: Net-of-cost average factor return $\overline{f}_i^{Net}$ and alpha $\overline{f}_{\alpha_i}^{Net}$ for each factor f_i.

for each factor $f_i = 1, \ldots, 58$ **do**

Step 1: Simulated Delay Cost

Using the grid of simulated delay costs (e.g., $\lambda_{\$100{,}000} = 0.123$, $\lambda_{\$2\text{M}} = 0.053$) and simulated turnover from Panel B2 ($100,000 trades) or B3 ($2M trades) of Table 2, linearly interpolate to compute the delay cost $\overline{C}_{Delay,f_i}$ using $\overline{f}_i$ and $\overline{TO}_{f_i}$ computed in Step 1.

Step 2: Bid-Ask Spread Cost

For a given trade size and simulated turnover $\overline{TO}_{Sim,P}$, compute the bid-ask spread cost as:

$$\overline{C}_{Bid-Ask,Sim} = 2 \times \overline{TO}_{Sim,P} \times \overline{h}_k.$$

This generates a vector of average bid-ask costs for a given trade size across different levels of simulated turnover rates (Panel C of Table 2).

Using the average realized factor turnover $\overline{TO}_{f_i}$, linearly interpolate to compute the bid-ask spread cost $\overline{C}_{Bid-Ask,f_i}$ for f_i across the grid defined by $\overline{TO}_{Sim,Q}$.

Step 3: Total Cost to Trade

Compute the total cost to trade for each factor f_i:

$$\overline{TC}_{f_i} = \overline{C}_{Delay,f_i} + \overline{C}_{Bid-Ask,f_i}.$$

Step 4: Net of Cost Return and Alpha

Compute the net of total cost average factor return:

$$\overline{f}_i^{Net} = \overline{f}_i - \overline{TC}_{f_i}.$$

Similarly, compute the net of cost alphas.

end for

D Markit Corporate Bond Short Selling Costs

D.1 `IndicativeFee` Borrowing Costs

For additional robustness, we adjust for the cost associated with short selling corporate bonds with the Markit `IndicativeFee` variable. The exact definition of `IndicativeFee` form Markit is as follows:

> *The expected borrow cost, in fee terms, for a hedge fund on a given day. This is a derived rate using Data Explorers proprietary analytics and data set. The calculation uses both borrow costs between Agent Lenders and Prime Brokers as well as rates from hedge funds to produce an indication of the current market rate. It should not be assumed that the indicative rate is the actual rate a Prime Broker will quote or charge but rather an indication of the standard market cost.*

The fee can be interpreted as the difference between the market short term interest rate and the rebate rate paid on the cash collateral associated with shorting the underlying (for the use of the data within an asset pricing context (see Hendershott et al. 2020 and Muravyev et al. 2024). The data is available daily. We resample the data to the monthly horizon by taking the average of the `IndicativeFee` each month as the representative cost to sell short a corporate bond. Given that the data only starts in September 2006, and our sample start is August 2002, we impute and backfill the `IndicativeFee` data to our sample start. First, we merge the Markit data to our bond level data, such that the two samples align. Thereafter, for each month, we compute the equally-weighted average `IndicativeFee` grouped by rating category. The time series of the average monthly `IndicativeFee` by rating category is presented in Figure A.7. These average monthly values in basis points are as follows: 3.10 (AAA to AA−), 3.15 (A+ to A), 3.13 (A− to BBB+), 3.20 (BBB to BBB−), 4.36 (BB to B−), and 15.20 (CCC+ to D). The imputation method allows the fee to vary in the cross-section before September 2006 based on each bond month t rating category.

Descriptive Statistics. Table A.13 reports pooled bond-month summary statistics for `IndicativeFee` over the full sample period of August 2002 to November 2022. The average fee for all bonds is relatively modest at 3.6 basis points per month, compared to bid-ask spreads of 18.8 basis points for small trades and 11.5 basis points for large trades, and delay costs of 3.9 and 7.5 basis points, respectively, as reported in Panels B.1 and B.2 of Table 4. Notably, our lower bound estimates for the cost of delays exceed the average cost of short-selling corporate bonds. The average `IndicativeFee` is relatively homogeneous across rating classes, ranging between 3.14 and 4.36 basis points, except for bonds rated CCC+ and below, which exhibit costs of 12.55 basis points – approximately four times higher than other rating categories.[44] Intuitively, it is far more expensive to short-sell poorly rated corporate bonds that are close to default relative to more highly rated bonds.

D.2 Adjusting the Short Leg of the Factor for the Borrowing Cost

To adjust the returns of the short-leg of each factor for short-selling fees, we compute the average monthly `IndicativeFee` over the holding period. Thereafter, the net return on the short leg portfolio is simply the average gross return minus the average `IndicativeFee` over the holding period.

[44] The median is within an even closer range of 3.12 (for 4 categories) to 3.27 (BB+ to B− bonds).

E Machine Learning Model Estimation and Cross Validation

Estimation and cross validation. For the first estimation as of July 2002 (start of the TRACE data), we source the last 55 months of data back to January 1998, and estimate the respective ML model. We measure excess returns at t and the 341-dimensional vector of bond characteristics at $t-1$. We perform cross-validation using a 70:30 training-validation split that preserves the temporal ordering of the panel data. We then employ the vector of characteristics available at time t to produce a forecast of bond excess returns for $t+1$. These forecasts (expected returns) are available to the bond portfolio manager at time t, meaning that she can trade on them at the end of the month. Thereafter, all models are re-trained every 12-months and cross-validated every five-years with an expanding window.[45]

For all of our machine learning models, we cross-validate the model hyperparameters every five-years and re-train the model every 12-months with an expanding window. Within each window, we perform the cross-validation with a 70:30 training-validation split. For example, if we have a window of 1,000 temporally ordered observations, 1-700 are used to train the model and the remaining 300 are used for validation. We graphically depict the sample splitting strategy for the training and cross-validation in Fig. A.8. For all models except for the feed forward neural network we utilize the `sklearn` Python package (Pedregosa et al., 2011). We use the `tensorflow` Python package to estimate the neural network.

We report the respective sets of hyperparameters that we cross-validate over in Table A.15.

Linear Models with Penalties. Panel A of Table A.15 reports the hyperparameters for the linear models with penalties for the Lasso (LASSO), Ridge (RIDGE) and the Elastic Net

[45] This gives the models an advantage in that they are re-trained and re-cross-validated multiple times over our sample.

(ENET). For the LASSO-style penalty, we cross-validate over 100 possible ℓ_1 penalties that change dynamically with the sample. The 100 potential ℓ_1 penalties are set by default with `sklearn` with a logarithmic scale. The maximum penalty is set to be the smallest value such that the coefficients are all set to zero. The minimum penalty is set to be 0.001 scaled by the maximum penalty. The ℓ_2 (RIDGE) penalties are defined as 100 values between 0.0001 and 1 with a logarithmic scale. The elastic net model hyperparameters are tuned with the 100 possible ℓ_1 penalties that change dynamically with the sample and a set of ℓ_1 vs. ℓ_2 ratios.

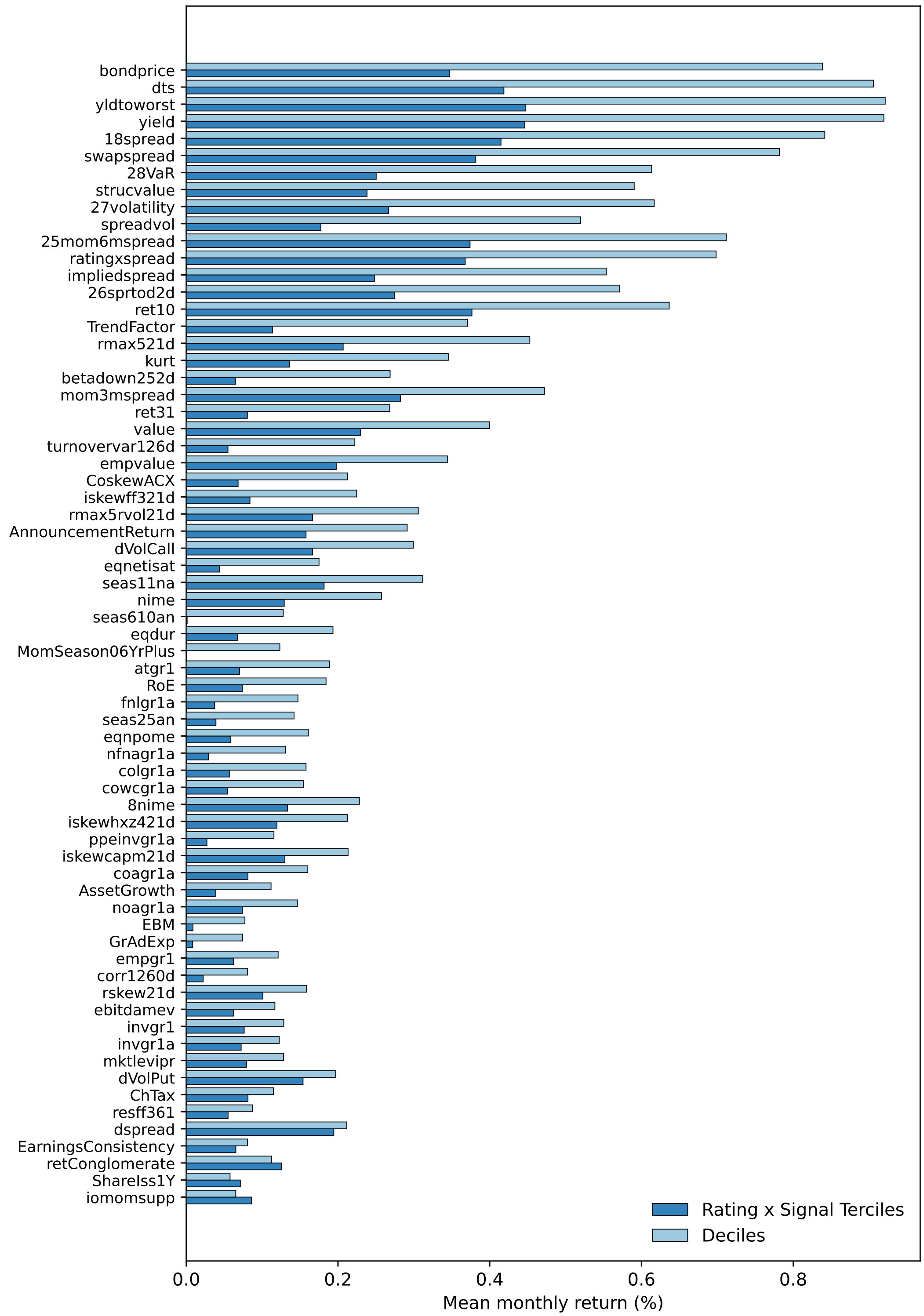
Nonlinear Tree-Based Ensembles Panel B of Table A.15 reports the hyperparameters for the tree-based nonlinear ensemble models that include the Random Forest (RF) and the Extremely Randomized Trees (XT). For both ensemble models, we use 100 estimators (trees). We also follow Gu et al. (2020) and set the maximum tree depth to be $\in [2, 4, 6]$. Thereafter, we allow the trees to consider a maximum of 5, 10, 15, or 30 features (characteristics) at each split point. Finally, at each end node of the tree (final leaf), we impose a minimum of 1, 10, or 50 samples (i.e., bond returns) in each leaf.

Feed Forward Neural Network Ensemble. Panel C of Table A.15 reports the hyperparameters for the feed forward neural network (NN). We estimate a shallow network with a single layer and 32 neurons. Since our sample starts off with a relatively smaller sample size than that of Gu et al. (2020) and other work that utilizes equity data only, we set the batch size to 1024 (with batch normalization) and the number of epochs to 100. We cross-validate over the learning rate that is $\in [0.001, 0.01]$ and an ℓ_1 penalty $\in [0.001, 0.01]$. We also implement early stopping with the 'patience' parameter set to 5. The prediction variance of each individually estimated neural network is high. In order to reduce prediction variance across estimated neural network models, at each training date we estimate 10 models with different randomly assigned initial weights. In doing so, we select the best performing 5 models based on the smallest mean squared error estimated in the validation sample at that training date. This means that at each date $t+1$, we produce five predictions from the five best performing models estimated at the training date. The overall $t+1$ prediction is the average over these five best performing models. At each training date, we then repeat this process ten times, yielding ten ensembled predictions. The final NN prediction for each month $t+1$ is the average over these ten ensembled predictions, i.e., an ensemble over the ensemble.

F Additional Results

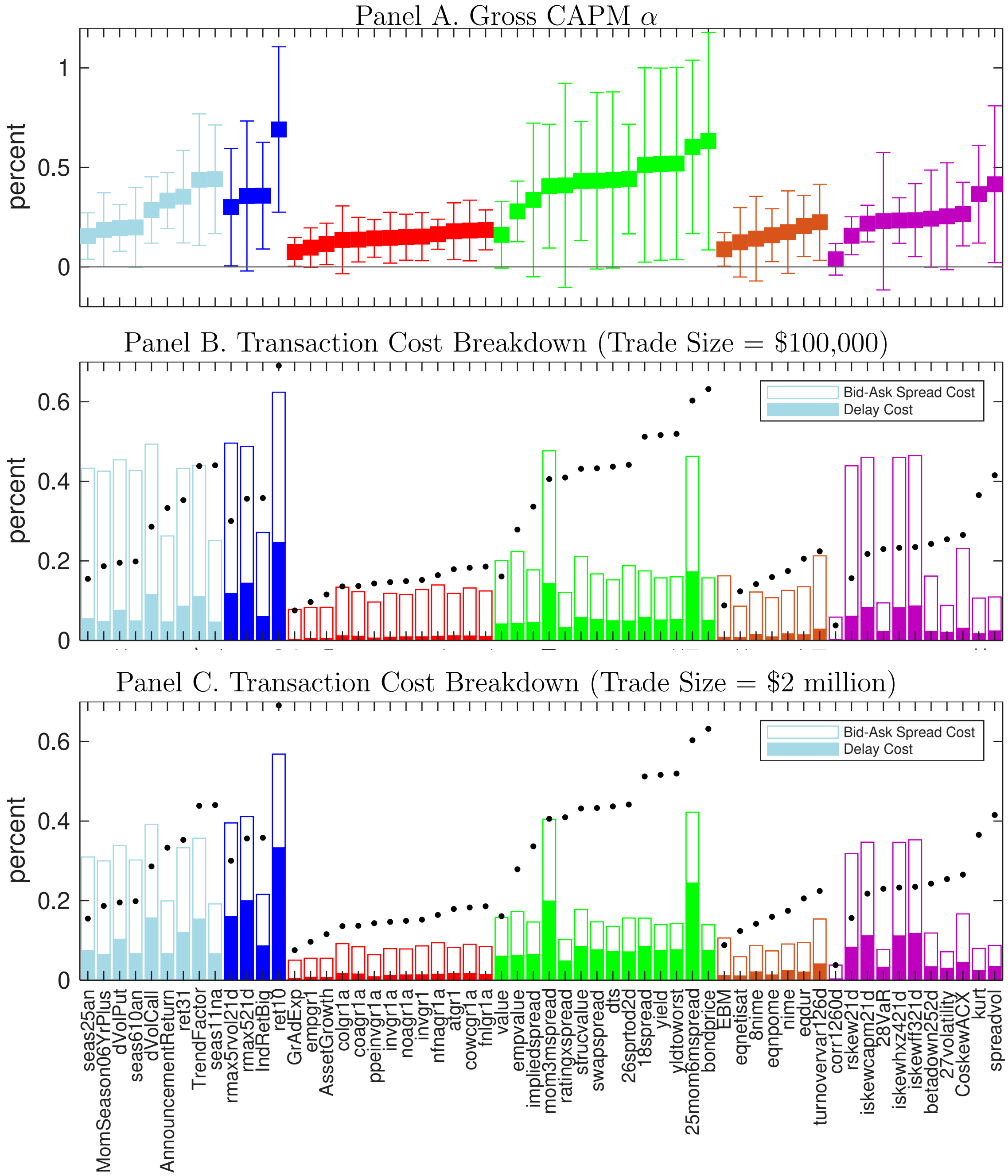
This section reports the additional figures and tables that we refer to in the main body of the paper.

Figure A.1: Deciles vs. Conditional 3×3 Rating and Signal Terciles Average Factor Returns



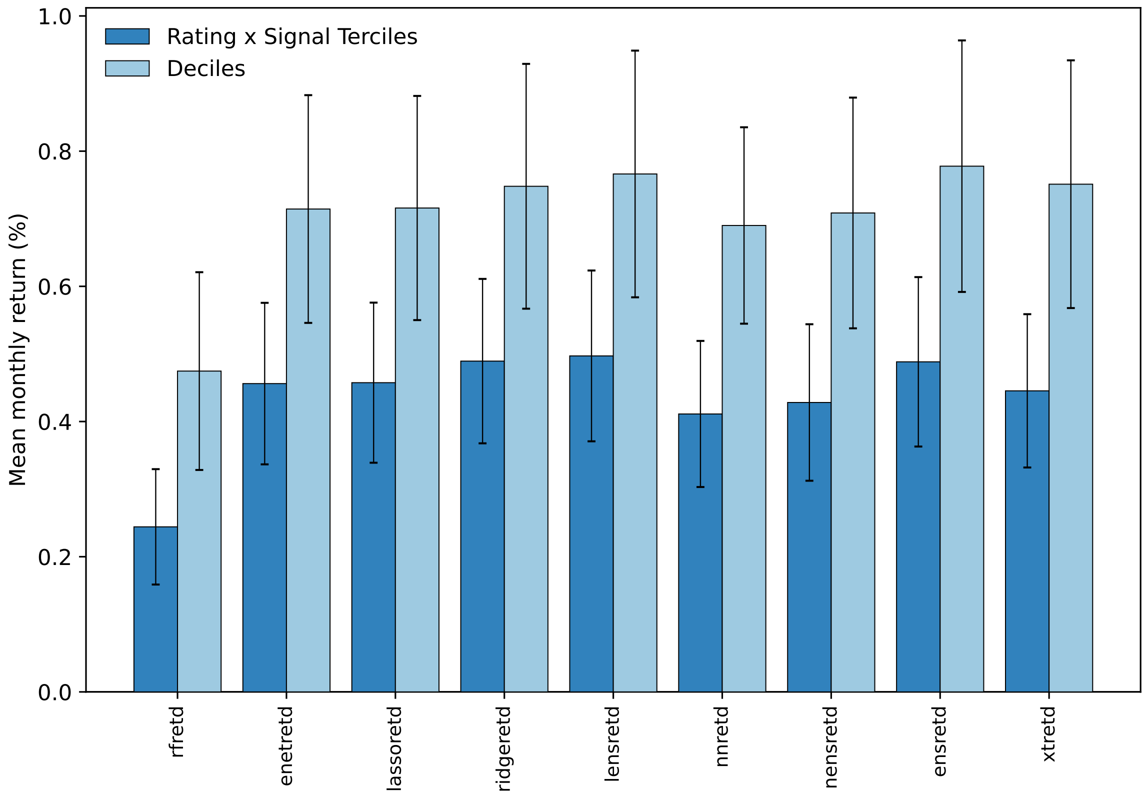
The figure plots the 67 factors that have a statistically significant average return (factor premium) at the 5% nominal level of the test. The 67 factors include those that are significant using the decile sorts (58 in total from the main text) and those that are significant using a conditional 3×3 rating tercile double sort (42 in total). The 67 factors represent the union of those factors that are statistically significant from both sorting methods. All factors are formed with the PyBondLab Python package.

Figure A.2: CAPM α of 58 Individual Factors: Cost Estimates from Kargar et al. (2025)



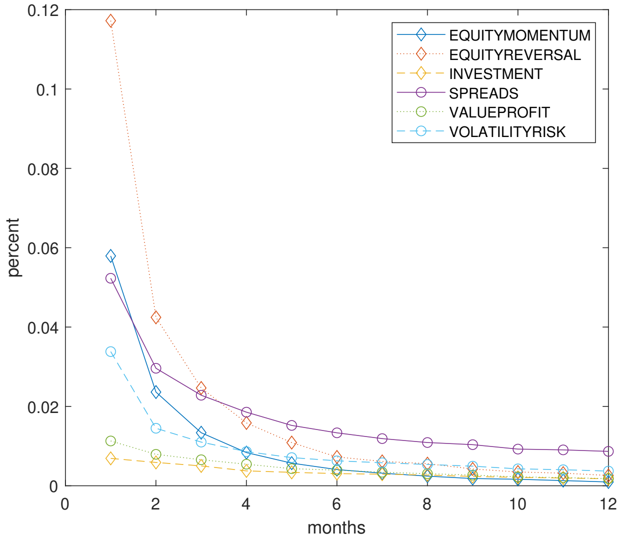
The figure plots the bond CAPM alpha (% per month) before transaction costs (Panel A), the breakdown of transaction costs using the delay estimates of Kargar et al. (2025) for small trades of $100,000 (Panel B) and $2 million (Panel C), as discussed in Section 5.2. From the 341 candidate factors, we select 58 factors with statistically significant average excess returns and classify them into six groups, represented by different colors. The error bars in Panel A represent the two standard error bounds that account for serial autocorrelation with the Newey-West procedure using 12 lags. The dots in Panels B and C represent the gross αs for each factor. All factors are formed with the PyBondLab Python package.

Figure A.3: Deciles vs. Conditional 3×3 Rating and ML Prediction Terciles Average Factor Returns

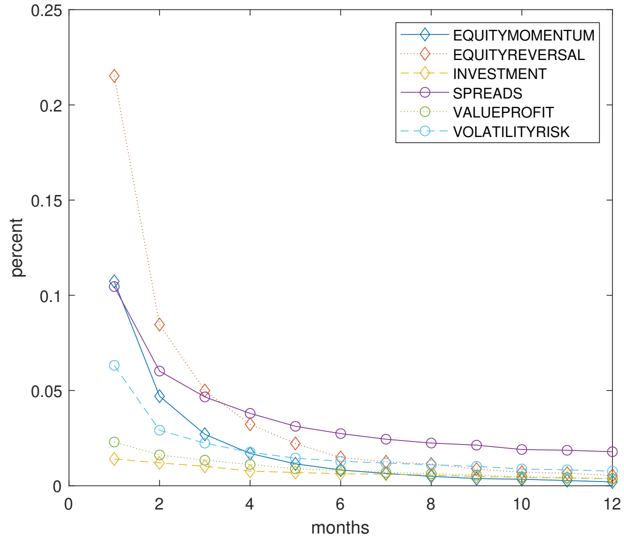


The figure plots the 9 machine learning (ML)-based factors formed with deciles (as in the main text) and with a conditional 3×3 rating double sort. Standard error bars are adjusted using the Newey-West procedure with 12 lags. All factors are formed with the `PyBondLab` Python package.

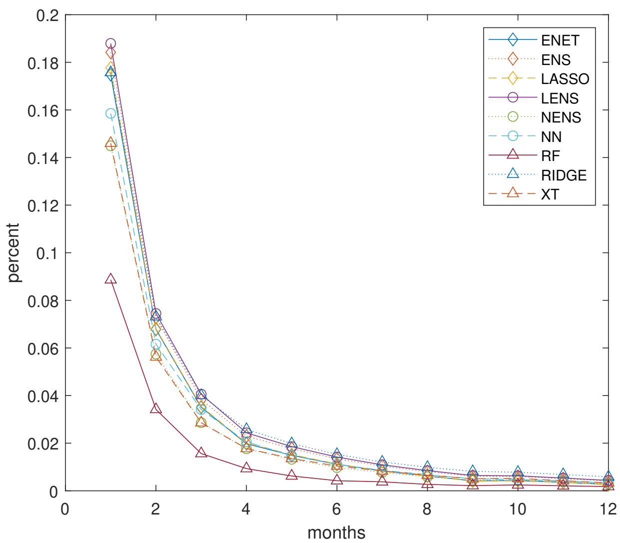
Figure A.4: Cost of Delays with Longer Rebalancing Frequency



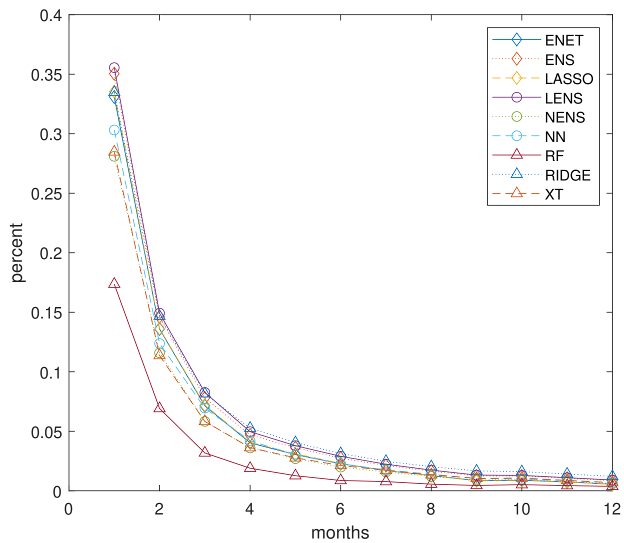
(a) Individual Factors, Small Trades



(b) Individual Factors, Large Trades



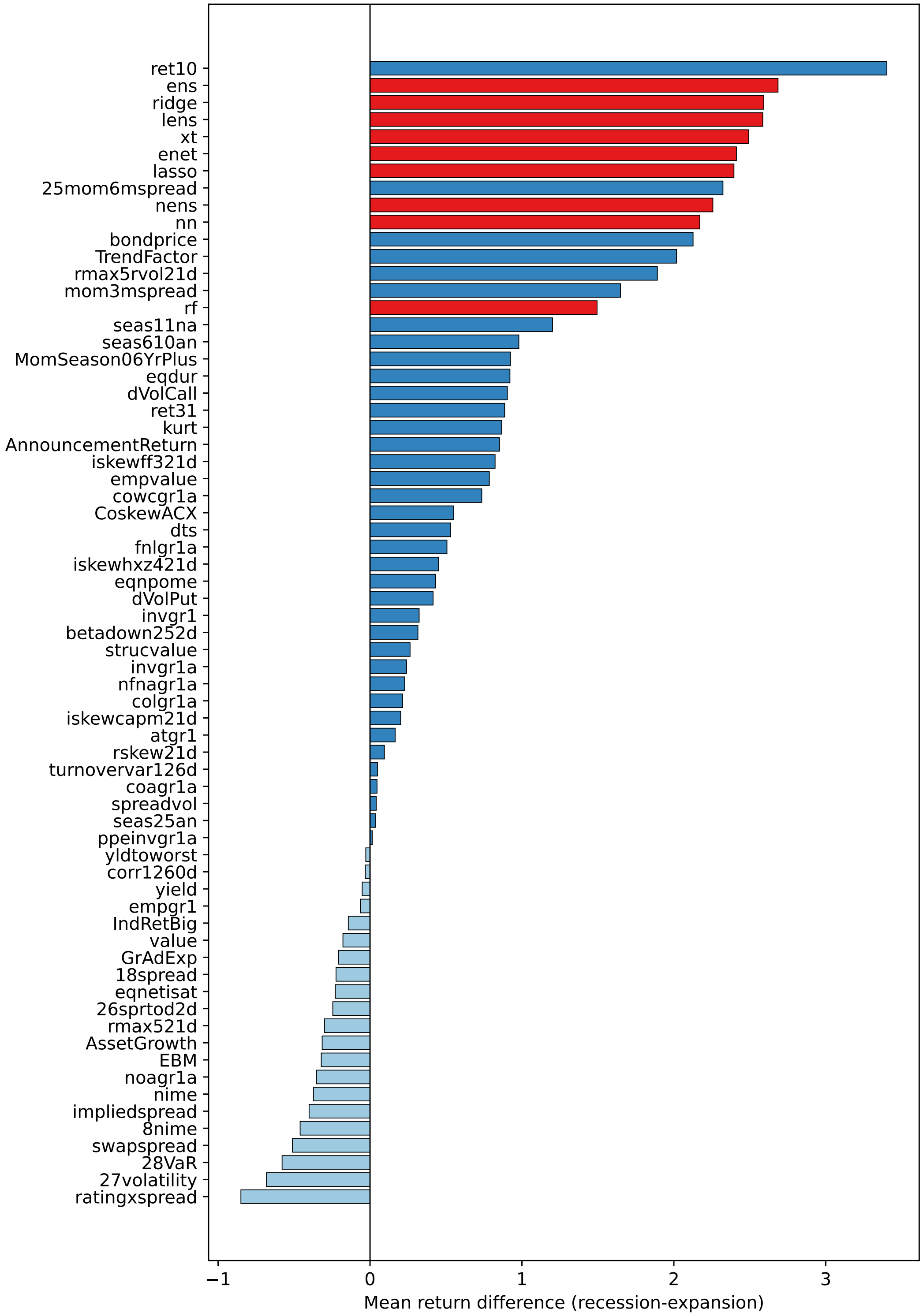
(c) ML Models, Small Trades



(d) ML Models, Large Trades

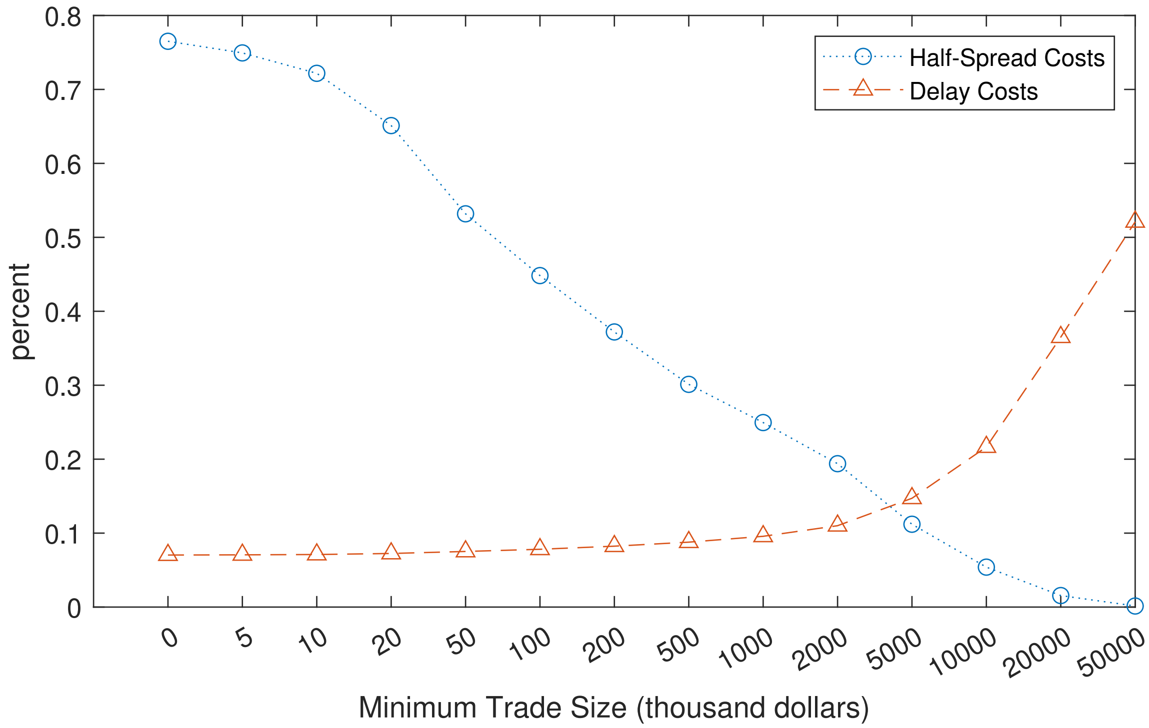
The figure plots the delay costs as a function of the portfolio rebalancing frequency from 1 to 12 months. Every month, we rebalance the portfolios and hold the bonds in the portfolio for $H = 1, \ldots, 12$ months. In month $t+1$, we take the average of H monthly returns on the portfolios formed in months t, $t-1$, ..., $t-H+1$. The portfolio turnover rate is computed using Eq. (7), which is used to assign the cost of delay. Small trades are the ones with trade size of \$100,000 and large trades are those with trade size of \$2 million. All factors are formed with the PyBondLab Python package.

Figure A.5: Average Factor Return Differences – During NBER Recessions Minus Non-Recession Periods



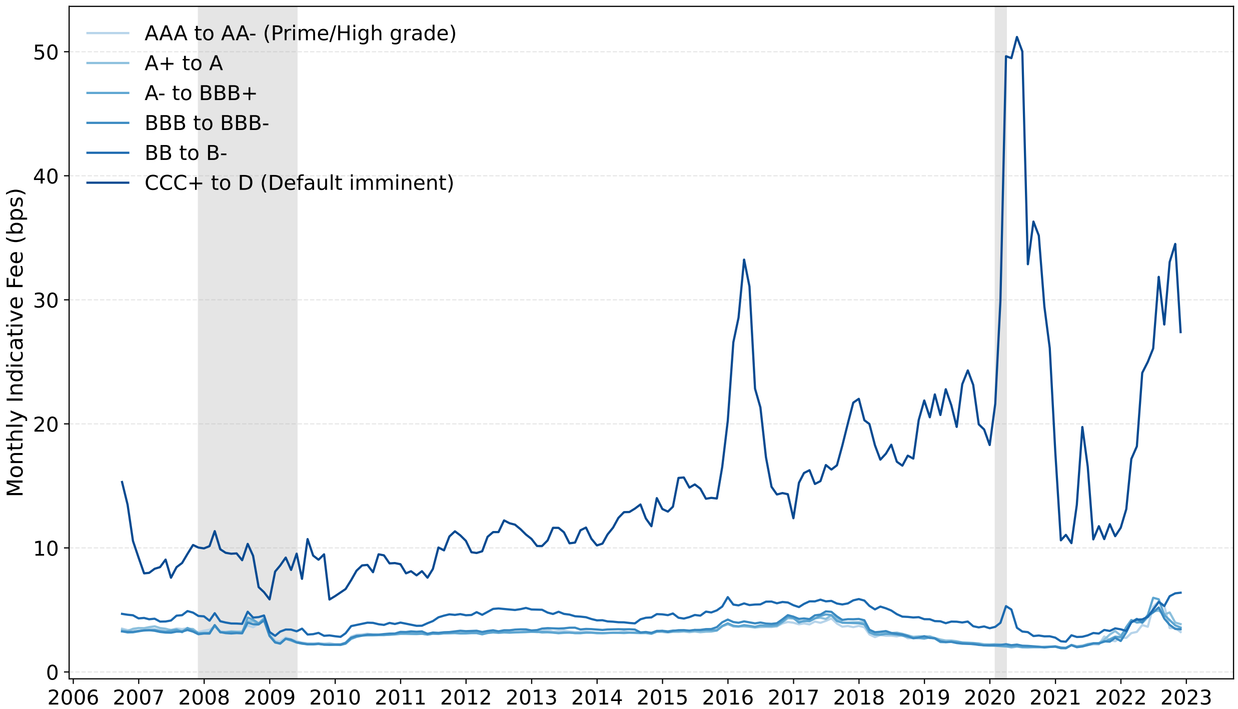
The figure plots the annualized average return differences between NBER and non-NBER recession periods for the 58 characteristic-based and 9 machine learning (ML)-based decile-sorted factors. The NBER recession dates span 20 months (8.20%) of the sample period over 2002-08:2022-11. All factors are formed with the PyBondLab Python package.

Figure A.6: Trade Size and Transaction Costs



The figure plots the delay and half-spread costs based on simulations with parameters $\sigma_v = 0.20\%$ and $\rho = 0.95$. The delay is drawn from the exponential distribution with an intensity (lower bound) estimated for each trade size. The half spread is the median half-spread for each trade size and the half-spread cost accounts for portfolio turnovers.

Figure A.7: Markit Short Selling Fees (Indicative Fee) Over Time



The figure presents the time series of the average (equally-weighted) Markit `IndicativeFee` over time across S&P bond rating categories. The averages are monthly and presented in basis points. The `IndicativeFee` variable is the net buy side fee paid to borrow the underlying bond. Specifically, it is defined as the interest rate on cash funds minus the rebate rate (that is paid for collateral) and is directly provided by Markit. This fee (cost) is used to adjust the short-leg of the bond factors. The data spans the sample period September 2006 to November 2022. `IndicativeFee` data are unavailable for all bonds prior to September 2006, Therefore, we impute these missing values using the average `IndicativeFee` within each rating category, enabling the sample to begin in August 2002 and align with the sample we employ for our main results.

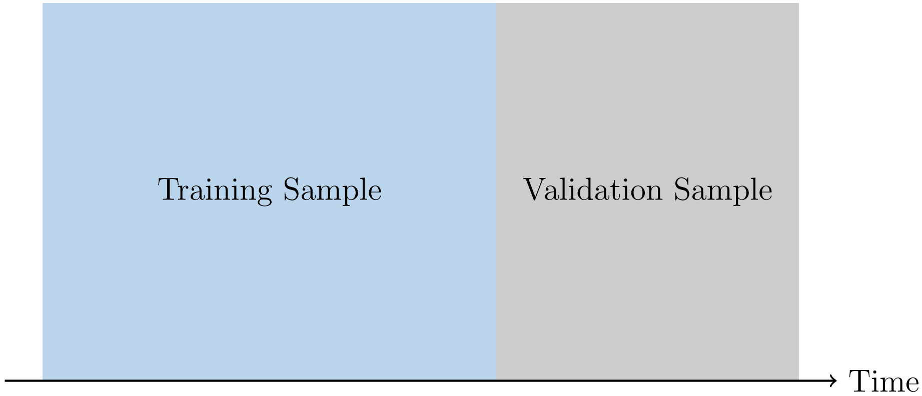
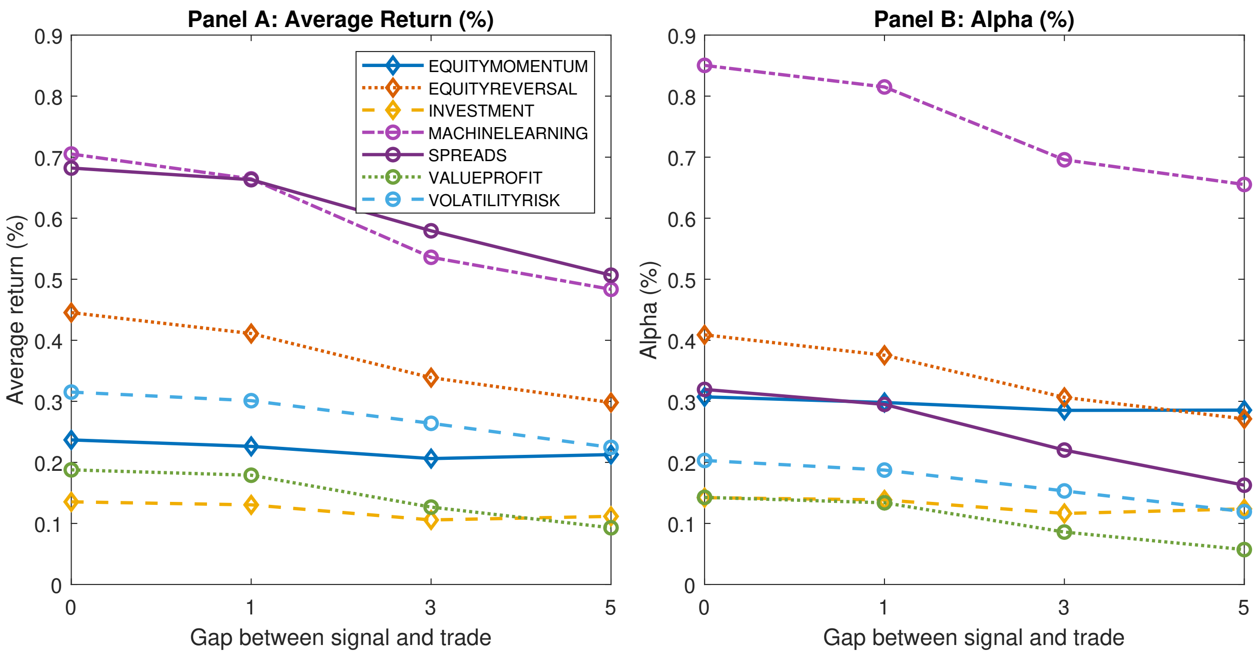


Figure A.8: Sample Splitting for Cross-Validation of Model Hyperparameters.

The figure shows the sample splitting scheme used for cross-validation of the machine learning model hyperparameters for the various machine learning models we consider. The forecasting exercise involves an expanding window that starts in January 1998. The initial window spans 1998:01–2002:07 ($T = 55$), and then expands forward each and every month until the sample end on 2022:12. The first (last) out-of-sample forecast is made in 2002:07 (2022:11) for the following month 2002:08 (2022:12). Hence, the out-of-sample ML portfolio returns commence in 2002:08 and end in 2022:12, $T = 245$. For each window, the blue area represents the training sample and the grey area represents the validation sample. The former consists of the first 70% of the observations while the latter consists of the final 30% of the observations. The training and the validation samples are contiguous in time and not randomly selected in order to preserve the temporal dependence of the data.

Figure A.9: Alpha Decay Across Factor Clusters



Alpha decay across exogenously imposed gaps between signal observation and bond prices used for trade. The figure presents the average monthly returns (Panel A) and CAPM alphas (Panel B) across factor clusters. The x-axis represents the "gap" between the observation of the month-end signal at month t and the price used to enter into a position in the bond g-days later in month $t+1$. That is, we estimate returns using the price $g = 1, 3$ and 5 days into month $t+1$ and the month-end price on the last business day of month $t+1$. The signal is assumed to be observed on the last business day of month t. When $g = 0$, we assume concurrent observation of the cross-section of signals and prices used to compute the monthly returns. The y-axis captures the average returns/alphas across this gap. Factor returns are scaled by so that they capture a representative month's return: $R_{\text{adj}} \approx R_k \times Days_{t+1}/(Days_{t+1} - g)$, where $Days_{t+1}$ are the number of business days in month $t+1$. The within month returns are computed with the BAML/ICE data.

Table A.1: List of the Corporate Bond and Stock Characteristics.

Num.	ID	Characteristic Name and Description	Reference	Source
		Panel A: Bond Characteristics Available on openbondassetpricing.com/data		
1	age	Bond age. The number of years the bond has been in issuance scaled by the tenor of the bond.	Israel et al. (2018)	BAML/ICE
2	ave12mspread	Rolling 12-month moving average of bond option adjusted credit spreads skipping the prior month	Elkamhi et al. (2021)	BAML/ICE
3	bond_mom_ipr	6-month bond credit momentum skipping the prior month. Demeaned with duration-times-spread.	Israel et al. (2018)	BAML/ICE
4	bond_price	Bond price. Clean price of the bond.	Bartram et al. (2023)	BAML/ICE
5	convexity	Bond convexity. Convexity of the bond.	–	BAML/ICE
6	coupon	Bond coupon. The annualised bond coupon payment in percent (%)	Chung et al. (2019)	BAML/ICE
7	dspread	First difference in bond option adjusted credit spread.	–	BAML/ICE
8	dts	Duration-times-spread. Annualized bond duration multiplied by the bond option adjusted credit spread.	Dor et al. (2007)	BAML/ICE
9	duration	Bond duration. The derivative of the bond value to the credit spread divided by the bond value, and is calculated by ICE	Israel et al. (2018)	BAML/ICE
10	emp_value	Empirical bond value. Defined as the percentage difference between the actual credit spread and the fitted ("fair") credit spread for each bond. The fitted spread is derived from cross-sectional regressions of the log of bond option adjusted credit spreads onto the log of duration, bond ratings and bond credit return volatility. Demeaned with duration-times-spread.	Israel et al. (2018)	BAML/ICE
11	faceval	Face value. The bond amount outstanding in units	Israel et al. (2018)	BAML/ICE
12	impliedspread	The fitted spread used to estimate the 33. value characteristic.	Houweling and Van Zundert (2017)	BAML/ICE
13	kurtosis	Bond kurtosis. Rolling bond excess kurtosis computed with a minimum amount of rolling observations equalling 12 which then expands up to 60-months	–	BAML/ICE
14	ltrev30_6	Bond long-term reversal (medium-term). Computed as the rolling sum of the prior 30-months of bond returns (minimum 12 monthly values) minus the rolling sum of the most recent 6-month returns (minimum 6 monthly values).	Subrahmanyam (2023)	BAML/ICE
15	ltrev48_12	Bond long-term reversal (long-term). Computed as the rolling sum of the prior 48-months of bond returns (minimum 12 monthly values) minus the rolling sum of the most recent 12-month returns (minimum 6 monthly values).	Bali et al. (2021)	BAML/ICE
16	mom3mspread	Mom. 3m log(Spread). The log of the spread 3 months earlier minus current log spread	–	BAML/ICE
17	mom6	Corporate bond momentum. The sum of the last 6-months of bond returns minus the prior month	Gebhardt et al. (2005)	BAML/ICE
18	mom6ind	Corporate bond portfolio industry momentum. The sum of the last 6-months of bond portfolio returns minus the prior month. Portfolios are formed based on the Fama-French Industry 17 classification	Kelly et al. (2021)	BAML/ICE
19	mom6mspread	Mom. 6m log(Spread). The log of the spread 6 months earlier minus current log spread	–	BAML/ICE
20	mom6xrtg	Corporate bond momentum multiplied by bond rating. The sum of the last 6-months of bond returns minus the prior month multiplied by the bond's numerical rating AAA = 1, ... , D = 22	Kelly et al. (2021)	BAML/ICE
21	rating	Bond S&P rating. Bond numerical rating. AAA = 1, ... , D = 22	Kelly et al. (2021)	BAML/ICE
22	ratingxspread	Corporate bond ratings multiplied by credit spread.	–	BAML/ICE
23	sizeb	Bond market capitalization. Computed as bond units amount outstanding multiplied by the clean price of the bond	Houweling and Van Zundert (2017)	BAML/ICE
24	skew	Bond skewness. The rolling 60-month skewness of bond total returns. We require a minimum of 12 observations, once this threshold is hit, the rolling window expands upward to 60-months	Kelly et al. (2021)	BAML/ICE

25	sprtod2d	Spread-to-Distance-to-Default. Spread-to-D2D is the option-adjusted spread, divided by one minus the CDF of the distance-to-default	Kelly et al. (2021)	CRSP/COMP
26	spread	Bond option adjusted credit spread. The option adjusted spread of the bond provided by ICE	Kelly et al. (2021)	BAML/ICE
27	spreadvol	Volatility of the first difference of the bond option adjusted credit spread. Rolling period of 24-months with a minimum required observations of 12	–	BAML/ICE
28	strevb	Short-term bond reversal. Defined as the previous months bond return	–	BAML/ICE
29	struc_value	Structural bond value. Defined as the percentage difference between the actual credit spread and the fitted ("fair") credit spread for each bond. The fitted spread is derived from cross-sectional regressions of the log of bond option adjusted credit spreads onto the log of the probability of default computed with firm-level distance-to-default. Demeaned with duration-times-spread.	Israel et al. (2018)	BAML/ICE
30	swap_spread	Bond swap spread. The swap spread of the bond provided by ICE	Kelly et al. (2021)	BAML/ICE
31	tmt	Bond time to maturity	–	BAML/ICE
32	value	Bond value. Defined as the percentage difference between the actual credit spread and the fitted ("fair") credit spread for each bond. The fitted spread is derived from cross-sectional regressions of the log of bond option adjusted credit spreads onto the 3-month change in spreads, maturity and credit ratings.	Houweling and Van Zundert (2017)	BAML/ICE
33	var	Historical 95% value-at-risk. Rolling 36-month bond total 95% value-at-risk. We require a minimum of 12 observations, once this threshold is hit, the rolling window expands upward to 36-months	Bai et al. (2019)	BAML/ICE
34	vixbeta	VIX beta. Rolling 60-month regression of bond returns on the Fama French 3-factors ($Mkt\text{-}RF$,SMB,HML), the default risk factor DEF, and the interest rate risk factor, $TERM$ and the first difference in the CBOE VIX and lagged VIX. The VIX beta in month t is the sum of the coefficient on VIX and lagged VIX. We require a minimum of 12 observations, once this threshold is hit, the rolling window expands upward to 60-months	Chung et al. (2019)	BAML/ICE
35	volatility	Bond return volatility. Rolling 36-month bond total return volatility. We require a minimum of 12 observations, once this threshold is hit, the rolling window expands upward to 36-months	Kelly et al. (2021)	BAML/ICE
36	yield	Bond yield. The yield-to-maturity of the bond provided by ICE	Gebhardt et al. (2005)	BAML/ICE
37	yld_to_worst	Bond yield-to-worst. The yield-to-worst of the bond provided by ICE	–	BAML/ICE
		Panel B: Equity Characteristics Available on openbondassetpricing.com/data		
38	book_lev	Book leverage. Shareholder's equity and long-/short-term debt (DLTTQ + DLCQ) and minority interest (MIBTQ) minus cash and inventories (CHEQ), divided by shareholder's equity minus preferred stock	Kelly et al. (2021)	COMP
39	book_prc	Book-to-price. Firm Book-to-price is the sum of shareholder's equity and preferred stock divided by equity market capitalization for the issuing firm	Kelly et al. (2021)	CRSP/COMP
40	chg_gp_at	Profitability change. The 5-year change in gross profitability	Asness et al. (2019)	COMP
41	d2d	Distance-to-default. Computed as in Bharath and Shumway (2008)	Bharath and Shumway (2008)	CRSP/COMP
42	debt_ebitda	Debt-to-EBITDA. Total debt (DLTTQ + DLCQ) divided by EBITDA (SALEQ − COGSQ − XSGAQ)	Kelly et al. (2021)	CRSP/COMP
43	eqtyvol	Equity volatility defined as the month-end value from a 180-day rolling-period	Campbell and Taksler (2003)	CRSP
44	gp_at	Profitability. Sales (REVTQ) minus cost-of-goods-sold (COGSQ), divided by assets (ATQ)	Choi and Kim (2018)	COMP
45	gross_prof_ipr	Demeaned profitability. Duration-times-spread demeaned gp_at	Israel et al. (2018)	COMP/ICE
46	me	Equity market capitalization	Choi and Kim (2018)	CRSP
47	mkt_lev	Market leverage. Market capitalization and long-/short-term debt (DLTTQ + DLCQ) and minority interest (MIBTQ) and preferred stock minus cash and inventories (CHEQ), divided by market capitalization	Kelly et al. (2021)	CRSP/COMP

48	mkt_lev_ipr	Demeaned Market leverage. Duration-times-spread demeaned mkt_lev	Israel et al. (2018)	COMP/ICE
49	ni_me	Earnings-to-price. Net income (NIQ) divided by market equity	Correia et al. (2012)	CRSP/COMP
50	operlvg	Operating leverage. Sales (SALEQ) minus EBITDA (SALEQ − COGSQ − XSGAQ), divided by EBITDA	Gamba and Saretto (2013)	COMP
51	stock_mom_ipr	Demeaned equity momentum. Duration-times-spread demeaned momentum (sum of the past 6 months, skipping the most recent month).	Israel et al. (2018)	COMP/ICE
52	totaldebt	Total firm debt (DLTTQ + DLCQ)	Kelly et al. (2021)	COMP
53	turnvol	Turnover volatility. Turnover volatility is the quarterly standard deviation of sales (SALEQ) divided by assets (ATQ). The volatility is computed over 80 quarters, with a minimum required period of 10 quarters. Thereafter, the volatility is averaged (smoothed) over the preceding 4-quarters in a rolling fashion	Kelly et al. (2021)	CRSP/COMP

The table presents information on the 53 characteristics that are generated by the authors and that are made publicly available on openbondassetpricing.com/data. The remaining characteristics are sourced from the publicly available equity data repositories of Chen and Zimmermann (2022) and Jensen et al. (2023). Panel A reports the 37 bond-only characteristics that are constructed using the BAML/ICE corporate bond database. Panel B reports the 16 equity-and-bond characteristics that are constructed using CRSP and COMPUSTAT (COMP). Additional resources and description notes for the equity openassetpricing.com-based data can be downloaded here. Documentation for the Jensen et al. (2023) data can be found here.

Table A.2: Summary Statistics of the Dataset for the Trade Intensity Estimates

Variable	N	Mean	Std.	Percentiles						
				p1	p10	p25	p50	p75	p90	p99
log invcomp	814,145	−2.161	1.049	−3.627	−3.162	−2.831	−2.391	−1.823	−0.910	2.278
inventory28	814,145	0.095	10.310	−32.135	−8.021	−1.960	0.000	2.001	8.220	31.691
capratio	814,145	0.062	0.015	0.026	0.044	0.050	0.061	0.075	0.083	0.090
halfspread	703,346	1.616	2.850	−0.341	0.274	0.596	1.131	1.936	3.122	9.645
rat	814,145	9.403	3.741	1	5	7	9	11	15	19
tau	814,145	10.660	9.935	1.126	2.370	4.047	6.795	16.219	26.186	31.981
log facevalue	814,145	6.133	0.662	4.745	5.416	5.704	6.109	6.551	7.003	7.919
coupon	814,145	5.809	2.126	1.554	3.125	4.250	5.750	7.125	8.500	11.375
age	814,145	1675.4	1471.9	80	284	618	1280	2282	3414	6998

The table presents the summary statistics of the panel data used to estimate the liquidity supply and demand curves. The unit of the analysis is at the bond-month level, and the sample is from August 2002 to November 2022 based on an intersection between the ICE and TRACE data.

Table A.3: Logit Regressions of Corporate Bond Transactions

		Customer Sell				Customer Buy			
		\$100,000		\$2M		\$100,000		\$2M	
Panel A. Regression Estimates									
$\log InvComp$		0.270	(23.12)	0.158	(28.44)	0.247	(22.12)	0.163	(29.72)
$\Delta Inventory$		−0.255	(−43.01)	−0.378	(−77.92)	0.336	(60.06)	0.403	(90.75)
CAP		0.110	(12.26)	0.102	(20.13)	0.092	(10.54)	0.086	(17.07)
$Rating$		0.679	(34.66)	0.612	(60.24)	0.657	(34.90)	0.620	(61.65)
$Maturity$		−0.136	(−9.98)	−0.025	(−3.36)	−0.159	(−12.25)	0.006	(0.74)
$\log FaceValue$		1.587	(78.32)	1.239	(146.39)	1.597	(81.97)	1.222	(145.81)
$Coupon$		−0.245	(−12.69)	0.022	(2.40)	−0.231	(−12.31)	0.049	(5.28)
Age		−0.188	(−14.84)	−0.414	(−40.36)	−0.216	(−17.64)	−0.420	(−41.85)
Intercept		2.815	(150.57)	0.531	(76.08)	2.733	(152.55)	0.472	(67.46)
Pseudo R^2		0.230		0.220		0.236		0.218	
N		814,145		814,145		814,145		814,145	
Panel B. Model-Based Estimates of Trade Intensity and Delays									
Prob. Trade	$Prob^{Data}$	0.877		0.588		0.868		0.577	
	$Prob^{CF}$	0.953		0.704		0.943		0.698	
λ	λ^{Data}	0.100		0.042		0.096		0.041	
	λ^{CF}	0.145		0.058		0.137		0.057	
Exp. Delay	$1/\lambda^{Data}$	10.013		23.670		10.363		24.441	
(days)	$1/\lambda^{CF}$	6.892		17.255		7.322		17.563	

The table reports the coefficient estimates for the logit regressions of the bond trading dummy on characteristics. Variable $InvComp$ is the investor composition of Li and Yu (2025), which measures the activeness of bond investors at the end of the previous quarter. Variable $\Delta Inventory$ is the difference between customer buys and customer sells in the preceding 28 days, CAP is the intermediary capital ratio of He et al. (2017) on the previous month. The right-hand-side variables are standardized for ease of interpretation. Values in parentheses are t-statistics clustered at the bond level. The sample is monthly, spanning August 2002 to November 2022.

Table A.4: Forecasting Regression of Corporate Bond Transactions: Nonlinear Model

		Customer Sell				Customer Buy			
		\$100,000		\$2M		\$100,000		\$2M	
Panel A. Regression Estimates									
$InvComp$		0.051	(7.04)	0.016	(1.91)	0.053	(7.81)	0.008	(0.97)
$1/InvComp$		−0.008	(−13.76)	−0.009	(−16.29)	−0.008	(−14.77)	−0.009	(−16.91)
$InvComp^2$		−0.003	(−5.93)	−0.001	(−1.78)	−0.003	(−6.72)	−0.000	(−0.71)
$\log InvComp$		−0.055	(−8.93)	−0.034	(−5.19)	−0.058	(−9.77)	−0.031	(−4.63)
$\Delta Inventory$		−0.947	(−23.53)	−4.423	(−45.55)	0.788	(15.78)	4.140	(35.47)
CAP		0.298	(2.16)	1.430	(7.06)	0.190	(1.48)	1.236	(6.48)
$Rating$		0.014	(24.94)	0.028	(40.95)	0.014	(25.31)	0.028	(43.22)
$Maturity$		−0.001	(−5.77)	−0.000	(−1.20)	−0.002	(−7.58)	0.000	(1.75)
$\log FaceValue$		0.145	(42.70)	0.303	(104.82)	0.153	(46.50)	0.298	(103.80)
$Coupon$		−0.013	(−14.51)	−0.002	(−1.19)	−0.014	(−14.54)	0.001	(0.37)
Age		−0.006	(−14.05)	−0.017	(−30.81)	−0.008	(−16.02)	−0.017	(−31.70)
Intercept		−0.067	(−2.83)	−1.492	(−58.17)	−0.114	(−5.05)	−1.478	(−58.80)
Adj. R^2		0.148		0.257		0.157		0.254	
N		814,145		814,145		814,145		814,145	
Panel B. Model-Based Estimates of Trade Intensity and Delays									
Prob. Trade	$Prob^{Data}$	0.877		0.588		0.868		0.576	
	$Prob^{CF}$	0.912		0.587		0.896		0.585	
λ	λ^{Data}	0.100		0.042		0.097		0.041	
	λ^{CF}	0.116		0.042		0.108		0.042	
Exp. Delay	$1/\lambda^{Data}$	10.014		23.672		10.362		24.442	
(days)	$1/\lambda^{CF}$	8.635		23.727		9.273		23.887	

The table reports the coefficient estimates for the logit regressions of the bond trading dummy on characteristics. The model is similar to that of Table A.3, except that we add a nonlinear transformation of the investor composition measure of Li and Yu (2025) and that the right-hand-side variables are not standardized. The sample is monthly, spanning August 2002 to November 2022.

Table A.5: Forecasting Regression of Corporate Bond Transactions: Interactions between Investor Composition and Bond Characteristics

		Customer Sell				Customer Buy			
		\$100,000		\$2M		\$100,000		\$2M	
Panel A. Regression Estimates									
$\log InvComp$		0.260	(15.79)	0.173	(9.74)	0.268	(16.07)	0.182	(10.03)
$\log InvComp \times Rating$		−0.004	(−8.89)	−0.002	(−4.94)	−0.004	(−8.72)	−0.002	(−4.15)
$\log InvComp \times \log FaceValue$		−0.037	(−16.35)	−0.016	(−6.19)	−0.038	(−16.35)	−0.016	(−6.42)
$\log InvComp \times Coupon$		0.002	(3.82)	−0.005	(−4.97)	0.002	(3.06)	−0.006	(−6.09)
$\log InvComp \times Age$		0.001	(2.97)	0.000	(1.26)	0.001	(2.11)	0.001	(1.53)
$\Delta Inventory$		−0.911	(−23.77)	−4.404	(−45.10)	0.825	(16.90)	4.160	(35.94)
CAP		0.063	(0.48)	1.172	(5.66)	−0.050	(−0.41)	0.982	(4.99)
$Rating$		0.006	(6.48)	0.025	(21.01)	0.006	(6.34)	0.026	(23.84)
$Maturity$		−0.002	(−7.10)	−0.001	(−4.14)	−0.002	(−9.03)	−0.000	(−1.16)
$\log FaceValue$		0.074	(11.73)	0.278	(41.11)	0.081	(12.52)	0.272	(40.10)
$Coupon$		−0.008	(−5.46)	−0.011	(−4.52)	−0.008	(−5.59)	−0.011	(−4.54)
Age		−0.005	(−6.90)	−0.018	(−16.89)	−0.007	(−8.81)	−0.018	(−17.33)
Intercept		0.485	(10.83)	−1.198	(−24.33)	0.459	(10.19)	−1.176	(−23.47)
Adj. R^2		0.148		0.252		0.157		0.249	
N		814,145		814,145		814,145		814,145	
Panel B. Model-Based Estimates of Trade Intensity and Delays									
Prob. Trade	$Prob^{Data}$	0.877		0.588		0.868		0.576	
	$Prob^{CF}$	0.958		0.730		0.947		0.725	
λ	λ^{Data}	0.100		0.042		0.097		0.041	
	λ^{CF}	0.151		0.062		0.140		0.061	
Exp. Delay	$1/\lambda^{Data}$	10.014		23.672		10.362		24.442	
(days)	$1/\lambda^{CF}$	6.602		16.042		7.136		16.267	

The table reports the estimates of the logit regression of monthly bond transactions. The model is similar to that of Table A.3, except that we add interaction terms between the investor composition measure of Li and Yu (2025) and bond characteristics such as credit ratings ($Rating$), time to maturity ($Maturity$), log face value ($FaceValue$), coupon rate ($Coupon$), and time since issuance (Age), and that the right-hand-side variables are not standardized. The sample is monthly, spanning August 2002 to November 2022.

Table A.6: List of the Statistically Significant Corporate Bond Factors (Gross Returns).

Factor	Cluster	Reference	Sign	Source
Cluster 1: Factors Formed on Bond Yields, Credit Spreads and Bond Prices				
18spread	Spreads	Kelly et al. (2021)	1	OSBAP
25mom6mspread	Spreads	Kelly et al. (2021)	-1	OSBAP
26sprtod2d	Spreads	Kelly et al. (2021)	1	OSBAP
bondprice	Spreads	Bartram et al. (2023)	-1	OSBAP
dts	Spreads	Kelly et al. (2021)	1	OSBAP
empvalue	Spreads	Israel et al. (2018)	1	OSBAP
impliedspread	Spreads	Israel et al. (2018)	1	OSBAP
mom3mspread	Spreads	-	-1	OSBAP
ratingxspread	Spreads	Kelly et al. (2021)	1	OSBAP
strucvalue	Spreads	Israel et al. (2018)	1	OSBAP
swapspread	Spreads	-	1	OSBAP
value	Spreads	Houweling and Van Zundert (2017)	1	OSBAP
yield	Spreads	Gebhardt et al. (2005)	1	OSBAP
yldtoworst	Spreads	-	1	OSBAP
Cluster 2: Factors Formed on Equity Momentum				
AnnouncementReturn	EquityMomentum	Nozawa et al. (2024)	1	OSAP
dVolCall	EquityMomentum	Cao et al. (2023)	1	OSAP
dVolPut	EquityMomentum	Cao et al. (2023)	-1	OSAP
MomSeason06YrPlus	EquityMomentum	-	1	JKP
ret31	EquityMomentum	-	1	JKP
seas11na	EquityMomentum	-	1	JKP
seas25an	EquityMomentum	-	1	JKP
seas610an	EquityMomentum	-	1	JKP
TrendFactor	EquityMomentum	-	1	OSAP
Cluster 3: Factors Formed on Short-Term Equity Reversal				
IndRetBig	EquityReversal	-	1	OSAP
rmax5rvol21d	EquityReversal	-	1	JKP
rmax521d	EquityReversal	-	1	JKP
ret10	EquityReversal	Chordia et al. (2017)	1	JKP
Cluster 4: Factors Formed on Volatility/Risk-Based Characteristics				
27volatility	VolatilityRisk	Bai et al. (2021)	1	OSBAP
28VaR	VolatilityRisk	Bai et al. (2019)	1	OSBAP
betadown252d	VolatilityRisk	-	1	JKP
corr1260d	VolatilityRisk	-	1	JKP
CoskewACX	VolatilityRisk	-	-1	OSAP
iskewcapm21d	VolatilityRisk	-	1	JKP
iskewff321d	VolatilityRisk	-	1	JKP
iskewhxz421d	VolatilityRisk	-	1	JKP
kurt	VolatilityRisk	Bai et al. (2016)	-1	OSBAP
rskew21d	VolatilityRisk	-	1	JKP
spreadvol	VolatilityRisk	-	1	OSBAP
Cluster 5: Factors Formed on Firm Investment/Accruals				
AssetGrowth	Investment	Choi & Kim (2018)	1	OSAP
atgr1	Investment	Choi & Kim (2018)	-1	JKP
coagr1a	Investment	-	-1	JKP
colgr1a	Investment	-	-1	JKP
cowcgr1a	Investment	-	-1	JKP
empgr1	Investment	-	-1	JKP
fnlgr1a	Investment	-	-1	JKP
GrAdExp	Investment	-	1	OSAP
invgr1	Investment	-	-1	JKP
invgr1a	Investment	-	-1	JKP
nfnagr1a	Investment	-	-1	JKP
noagr1a	Investment	-	-1	JKP

ppeinvgr1a	Investment	-	-1	JKP

Cluster 6: Factors Formed on Firm Value/Profitability

8nime	ValueProfit	-	-1	JKP
EBM	ValueProfit	-	1	OSAP
eqdur	ValueProfit	-	1	JKP
eqnetisat	ValueProfit	-	1	JKP
eqnpome	ValueProfit	-	1	JKP
nime	ValueProfit	-	-1	JKP
turnovervar126d	ValueProfit	-	1	JKP

The table presents information on the 58 long-short corporate bond factors that have a statistically significant average gross return at the 5% nominal level of the tests. That is, we screen out any bond factor that does not have a Newey-West t-statistic >1.96 (estimated with 12-lags). The bond factors are sign-corrected such they all have a positive mean by construction. The first column, Factor, presents the factor mnemonic, followed by Cluster, which gives the respective cluster that factor has been grouped into. The final three columns present the relevant reference (if applicable), the sign correction, and the original source of the underlying characteristic data. Additional resources and description notes for the corporate bond signals can be found on `openbondassetpricing.com/machine-learning-data/` (OSBAP), while the equity `openassetpricing.com`-based data (OSAP) can be downloaded here. Documentation for the Jensen et al. (2023), JKP, data can be found here.

Table A.7: Bond CAPM Alphas of 58 Factors

Signal	Turnover	Gross α	Trade Size : \$100,000			Trade Size : \$2M		
			Net α	Cost (%)		Net α	Cost (%)	
	(%)	(%)	(%)	Delay	BA	(%)	Delay	BA
Equity Momentum								
seas25an	174	0.155	−0.268	0.045	0.378	−0.162	0.082	0.236
		(2.60)	(−4.50)			(−2.72)		
MomSeason06YrPlus	171	0.187	−0.230	0.039	0.378	−0.120	0.071	0.236
		(1.97)	(−2.43)			(−1.26)		
dVolPut	185	0.195	−0.245	0.063	0.378	−0.153	0.113	0.236
		(3.27)	(−4.11)			(−2.56)		
seas610an	174	0.199	−0.220	0.041	0.378	−0.110	0.073	0.236
		(1.94)	(−2.16)			(−1.08)		
dVolCall	185	0.286	−0.188	0.096	0.378	−0.122	0.172	0.236
		(3.36)	(−2.20)			(−1.43)		
AnnouncementReturn	70	0.333	0.078	0.038	0.217	0.126	0.076	0.131
		(4.64)	(1.09)			(1.75)		
ret31	134	0.353	−0.066	0.071	0.347	0.007	0.133	0.213
		(2.97)	(−0.55)			(0.06)		
TrendFactor	124	0.438	0.017	0.091	0.331	0.065	0.171	0.202
		(2.60)	(0.10)			(0.38)		
seas11na	65	0.440	0.197	0.038	0.206	0.240	0.076	0.124
		(3.16)	(1.41)			(1.72)		
Equity Reversal								
rmax5rvol21d	170	0.300	−0.175	0.098	0.378	−0.111	0.176	0.236
		(1.99)	(−1.17)			(−0.74)		
rmax521d	132	0.356	−0.107	0.118	0.345	−0.077	0.222	0.211
		(1.85)	(−0.56)			(−0.40)		
IndRetBig	68	0.358	0.097	0.049	0.212	0.132	0.098	0.129
		(2.62)	(0.71)			(0.96)		
ret10	175	0.691	0.109	0.204	0.378	0.089	0.366	0.236
		(3.26)	(0.51)			(0.42)		
Investment								
GrAdExp	20	0.076	−0.002	0.002	0.075	0.025	0.005	0.046
		(2.04)	(−0.06)			(0.66)		
empgr1	21	0.097	0.014	0.004	0.078	0.041	0.008	0.048
		(1.91)	(0.28)			(0.80)		
AssetGrowth	21	0.116	0.033	0.004	0.079	0.060	0.008	0.048
		(2.18)	(0.62)			(1.12)		

The table reports the gross and net bond CAPM alphas and associated monthly turnover rates for small (\$100,000) and large (\$2M) trade sizes, which accounts for transaction costs including bid-ask spread costs and delay costs. The 'Cost' column decomposes the total cost to trade into the Delay and bid-ask (BA) components. We use the 58 factors with significant average excess returns. We regress each factor on the corporate bond market factor (CAPMB) and estimate the intercept of the regression, α. We group the 58 factors into six categories and report the results by group. The values in parentheses are the t-statistics (Newey-West adjusted with 12 lags). We use our own estimates of trade intensity to calculate the cost of delays. All factors are formed with the PyBondLab Python package.

Table A.7 (Continued)

Signal	Turnover	Gross α	Trade Size : \$100,000			Trade Size : \$2M		
			Net α	Cost (%)		Net α	Cost (%)	
	(%)	(%)	(%)	Delay	BA	(%)	Delay	BA
Investment, Contd.								
colgr1a	35	0.136	0.004	0.010	0.123	0.042	0.019	0.075
		(1.56)	(0.04)			(0.48)		
coagr1a	32	0.137	0.016	0.009	0.112	0.051	0.018	0.068
		(2.39)	(0.28)			(0.89)		
ppeinvgr1a	25	0.143	0.048	0.005	0.091	0.078	0.010	0.056
		(2.97)	(0.98)			(1.62)		
invgr1a	31	0.147	0.029	0.007	0.111	0.066	0.013	0.067
		(2.25)	(0.45)			(1.01)		
noagr1a	30	0.149	0.035	0.007	0.107	0.069	0.015	0.065
		(2.53)	(0.59)			(1.17)		
invgr1	34	0.152	0.025	0.008	0.119	0.065	0.015	0.073
		(2.46)	(0.41)			(1.04)		
nfnagr1a	37	0.164	0.026	0.008	0.130	0.068	0.017	0.079
		(4.26)	(0.67)			(1.77)		
atgr1	30	0.179	0.063	0.010	0.107	0.095	0.019	0.065
		(2.46)	(0.86)			(1.30)		
cowcgr1a	34	0.183	0.053	0.009	0.121	0.091	0.019	0.074
		(2.35)	(0.67)			(1.17)		
fnlgr1a	32	0.186	0.063	0.008	0.115	0.099	0.017	0.070
		(3.63)	(1.22)			(1.94)		
Spreads								
value	48	0.161	−0.033	0.034	0.160	−0.005	0.069	0.097
		(1.89)	(−0.39)			(−0.06)		
empvalue	56	0.279	0.062	0.035	0.182	0.099	0.070	0.110
		(3.59)	(0.80)			(1.27)		
impliedspread	38	0.337	0.167	0.036	0.133	0.182	0.074	0.081
		(1.71)	(0.85)			(0.92)		
mom3mspread	126	0.406	−0.047	0.117	0.335	−0.021	0.222	0.205
		(2.56)	(−0.29)			(−0.13)		
ratingxspread	24	0.410	0.295	0.027	0.088	0.301	0.055	0.054
		(1.56)	(1.12)			(1.15)		
strucvalue	45	0.432	0.231	0.047	0.154	0.243	0.096	0.093
		(2.83)	(1.51)			(1.59)		
swapspread	32	0.433	0.275	0.043	0.115	0.276	0.087	0.070
		(1.91)	(1.21)			(1.22)		
dts	29	0.437	0.293	0.040	0.103	0.292	0.082	0.063
		(1.93)	(1.30)			(1.29)		
26sprtod2d	40	0.441	0.262	0.040	0.139	0.276	0.081	0.085
		(3.14)	(1.87)			(1.96)		
18spread	33	0.512	0.348	0.047	0.118	0.345	0.096	0.072
		(2.06)	(1.40)			(1.38)		
yield	30	0.516	0.368	0.042	0.106	0.366	0.086	0.065
		(2.10)	(1.50)			(1.49)		
yldtoworst	30	0.519	0.369	0.043	0.108	0.366	0.087	0.066
		(2.11)	(1.50)			(1.49)		
25mom6mspread	103	0.603	0.170	0.141	0.291	0.151	0.275	0.177
		(2.71)	(0.77)			(0.68)		
bondprice	30	0.632	0.484	0.041	0.107	0.482	0.085	0.065
		(2.27)	(1.73)			(1.73)		

Table A.7 (Continued)

Signal	Turnover (%)	Gross α (%)	Trade Size : \$100,000 Net α (%)	Cost (%) Delay	Cost (%) BA	Trade Size : \$2M Net α (%)	Cost (%) Delay	Cost (%) BA
Value and Profit								
EBM	46	0.088	−0.074	0.006	0.155	−0.019	0.013	0.094
		(2.04)	(−1.70)			(−0.44)		
eqnetisat	21	0.124	0.039	0.006	0.079	0.063	0.012	0.048
		(1.39)	(0.43)			(0.71)		
8nime	30	0.142	0.022	0.012	0.108	0.052	0.024	0.066
		(1.31)	(0.20)			(0.48)		
eqnpome	27	0.159	0.053	0.007	0.099	0.084	0.015	0.060
		(2.35)	(0.78)			(1.24)		
nime	31	0.175	0.051	0.013	0.110	0.081	0.027	0.067
		(1.65)	(0.49)			(0.76)		
eqdur	35	0.206	0.073	0.011	0.122	0.109	0.023	0.074
		(2.62)	(0.93)			(1.38)		
turnovervar126d	57	0.224	0.016	0.023	0.185	0.066	0.047	0.112
		(2.30)	(0.16)			(0.67)		
Volatility and Risk								
corr1260d	15	0.038	−0.021	0.002	0.057	−0.000	0.004	0.035
		(0.94)	(−0.51)			(−0.01)		
rskew21d	180	0.156	−0.272	0.051	0.378	−0.170	0.091	0.236
		(3.21)	(−5.59)			(−3.50)		
iskewcapm21d	180	0.218	−0.228	0.068	0.378	−0.141	0.123	0.236
		(4.62)	(−4.86)			(−2.99)		
28VaR	19	0.230	0.139	0.018	0.073	0.149	0.037	0.044
		(1.30)	(0.79)			(0.84)		
iskewhxz421d	180	0.233	−0.213	0.068	0.378	−0.125	0.122	0.236
		(4.00)	(−3.65)			(−2.14)		
iskewff321d	180	0.235	−0.215	0.072	0.378	−0.130	0.129	0.236
		(2.51)	(−2.30)			(−1.39)		
betadown252d	40	0.243	0.085	0.019	0.139	0.120	0.038	0.085
		(1.95)	(0.68)			(0.96)		
27volatility	18	0.254	0.170	0.017	0.068	0.179	0.034	0.042
		(1.85)	(1.24)			(1.30)		
CoskewACX	64	0.265	0.039	0.025	0.201	0.094	0.050	0.122
		(3.25)	(0.48)			(1.15)		
kurt	25	0.365	0.262	0.014	0.090	0.282	0.028	0.055
		(2.92)	(2.09)			(2.25)		
spreadvol	23	0.415	0.310	0.019	0.086	0.323	0.040	0.052
		(2.07)	(1.54)			(1.61)		

Table A.8: CAPM Alphas Using Duration-Adjusted Returns

Category	Equity Momentum	Equity Reversal	Investment	Spreads	Value Profit	Volatility Risk	All
Count	9	4	13	14	7	11	58
Panel A. Gross CAPM α							
Avg. α	0.269	0.404	0.152	0.465	0.167	0.239	0.281
Avg. $t(\alpha)$	(2.83)	(2.51)	(2.79)	(2.31)	(2.13)	(2.64)	(2.55)
Avg. Turnover	142.47	136.34	29.54	47.48	35.41	84.04	69.80
$\#(t(\alpha) > 1.96)$	9	3	12	8	4	8	44
Panel B. Net CAPM α, Own Estimates of Delay							
B1. Small Trades							
Avg. α	−0.117	−0.038	0.039	0.258	0.032	0.002	0.055
Avg. $t(\alpha)$	(−1.74)	(−0.26)	(0.70)	(1.15)	(0.18)	(−1.00)	(−0.02)
Avg. BidAsk Cost	0.332	0.328	0.106	0.153	0.123	0.202	0.188
Avg. Delay Cost	0.054	0.113	0.008	0.054	0.012	0.035	0.039
$\#(t(\alpha) > 1.96)$	0	0	0	1	0	1	2
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	0	0	0	2	0	2	4
B2. Large Trades							
Avg. α	−0.036	−0.007	0.072	0.265	0.068	0.049	0.092
Avg. $t(\alpha)$	(−0.75)	(−0.03)	(1.33)	(1.21)	(0.72)	(−0.25)	(0.51)
Avg. BidAsk Cost	0.205	0.203	0.064	0.093	0.075	0.125	0.115
Avg. Delay Cost	0.100	0.209	0.015	0.107	0.025	0.065	0.074
$\#(t(\alpha) > 1.96)$	0	0	1	1	0	2	4
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	1	3	4	1	2	13
Panel C. Net CAPM α, Kargar et al. (2023) Estimates of Delay							
C1. Small Trades							
Avg. α	−0.128	−0.061	0.038	0.247	0.030	−0.005	0.046
Avg. $t(\alpha)$	(−1.86)	(−0.40)	(0.68)	(1.09)	(0.15)	(−1.11)	(−0.09)
Avg. BidAsk Cost	0.332	0.328	0.105	0.152	0.122	0.202	0.188
Avg. Delay Cost	0.065	0.137	0.009	0.066	0.015	0.042	0.047
$\#(t(\alpha) > 1.96)$	0	0	0	1	0	1	2
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	0	0	0	2	0	2	4
C2. Large Trades							
Avg. α	−0.027	0.012	0.074	0.277	0.070	0.056	0.099
Avg. $t(\alpha)$	(−0.64)	(0.09)	(1.35)	(1.27)	(0.76)	(−0.16)	(0.58)
Avg. BidAsk Cost	0.206	0.203	0.065	0.094	0.075	0.125	0.116
Avg. Delay Cost	0.090	0.189	0.013	0.095	0.022	0.058	0.066
$\#(t(\alpha) > 1.96)$	0	0	1	1	0	2	4
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	1	3	4	1	2	13

The table reports the average bond CAPM α, the associated average t-statistic, and the number of factors with a t-statistic greater than 1.96 before and after accounting for transaction costs, including bid-ask spread costs and delay costs across factors associated with the six categories from Section 5.2. Of the 341 factors that we generate, we consider the 58 factors with significant average excess returns. We regress each factor's duration-adjusted returns on the corporate bond market factor and estimate the intercept of the regression, α. We group the 58 factors into six categories and average the estimates within each group. The values in parentheses are the averages of the t-statistics (Newey-West adjusted with 12 lags). In Panel B, we employ our own estimates of trade intensity to calculate the cost of delays and present the net-of-cost alphas and the number of factors that remain statistically significant. In Panel C, we employ the Kargar et al. (2025) delay estimates for the trade intensity parameter and compute the cost of delays for small and large trades respectively. All factors are formed with the `PyBondLab` Python package.

Table A.9: CAPM Alphas Using TRACE Returns

Category	Equity Momentum	Equity Reversal	Investment	Spreads	Value Profit	Volatility Risk	All
Count	9	4	13	14	7	11	58
Panel A. Gross CAPM α							
Avg. α	0.288	0.437	0.146	0.339	0.180	0.238	0.256
Avg. $t(\alpha)$	(2.80)	(2.42)	(2.15)	(1.75)	(2.20)	(2.41)	(2.23)
Avg. Turnover	144.37	138.48	34.49	51.09	40.78	88.06	73.64
$\#(t(\alpha) > 1.96)$	7	3	8	2	5	6	31
Panel B. Net CAPM α, Own Estimates of Delay							
B1. Small Trades							
Avg. α	−0.106	−0.020	0.017	0.131	0.026	−0.015	0.018
Avg. $t(\alpha)$	(−1.39)	(−0.15)	(0.22)	(0.41)	(0.21)	(−0.78)	(−0.20)
Avg. BidAsk Cost	0.335	0.332	0.121	0.163	0.139	0.214	0.199
Avg. Delay Cost	0.059	0.124	0.008	0.044	0.014	0.039	0.039
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	0	0
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	0	0	0	0	0	0	0
B2. Large Trades							
Avg. α	−0.028	0.004	0.056	0.151	0.066	0.033	0.059
Avg. $t(\alpha)$	(−0.53)	(0.00)	(0.82)	(0.59)	(0.75)	(−0.15)	(0.31)
Avg. BidAsk Cost	0.207	0.205	0.074	0.099	0.084	0.132	0.122
Avg. Delay Cost	0.109	0.228	0.016	0.089	0.029	0.073	0.075
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	0	0
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	1	0	1	0	1	5
Panel C. Net CAPM α, Kargar et al. (2023) Estimates of Delay							
C1. Small Trades							
Avg. α	−0.118	−0.045	0.016	0.122	0.024	−0.023	0.010
Avg. $t(\alpha)$	(−1.51)	(−0.29)	(0.20)	(0.36)	(0.17)	(−0.88)	(−0.27)
Avg. BidAsk Cost	0.335	0.332	0.121	0.163	0.139	0.214	0.199
Avg. Delay Cost	0.071	0.150	0.010	0.054	0.018	0.047	0.048
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	0	0
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	0	0	0	0	0	0	0
C2. Large Trades							
Avg. α	−0.018	0.025	0.058	0.161	0.069	0.040	0.067
Avg. $t(\alpha)$	(−0.42)	(0.12)	(0.84)	(0.64)	(0.78)	(−0.06)	(0.37)
Avg. BidAsk Cost	0.208	0.206	0.074	0.100	0.085	0.132	0.123
Avg. Delay Cost	0.098	0.206	0.014	0.078	0.026	0.065	0.067
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	0	0
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	1	0	0	0	1	4

The table reports the average bond CAPM α, the associated average t-statistic, and the number of factors with a t-statistic greater than 1.96 before and after accounting for transaction costs, including bid-ask spread costs and delay costs across factors associated with the six categories from Section 5.2. Of the 341 factors that we generate, we consider the 58 factors with significant average excess returns. We regress each factor's excess returns on the corporate bond market factor and estimate the intercept of the regression, α. We group the 58 factors into six categories and average the estimates within each group. The values in parentheses are the average of the t-statistics (Newey-West adjusted with 12 lags). In Panel B, we employ our own estimates of trade intensity to calculate the cost of delays and present the net-of-cost alphas and the number of factors that remain statistically significant. In Panel C, we employ the Kargar et al. (2025) delay estimates for the trade intensity parameter and compute the cost of delays for small and large trades respectively. All factors are formed with the `PyBondLab` Python package.

Table A.10: CAPM Alphas after Accounting for Borrowing Fees

Category	Equity Momentum	Equity Reversal	Investment	Spreads	Value Profit	Volatility Risk	All
Count	9	4	13	14	7	11	58
Panel A. Gross CAPM α							
Avg. α	0.287	0.426	0.143	0.437	0.160	0.241	0.277
Avg. $t(\alpha)$	(2.95)	(2.43)	(2.54)	(2.31)	(1.95)	(2.60)	(2.48)
Avg. Turnover	142.46	136.30	29.43	47.45	35.33	83.99	69.75
$\#(t(\alpha) > 1.96)$	8	3	11	9	4	7	42
Panel B. CAPM α After Borrowing Cost							
Avg. α	0.251	0.392	0.111	0.405	0.127	0.209	0.244
Avg. $t(\alpha)$	(2.54)	(2.23)	(1.95)	(2.12)	(1.53)	(2.19)	(2.10)
Avg. Borrowing Cost	0.036	0.034	0.033	0.032	0.032	0.032	0.033
$\#(t(\alpha) > 1.96)$	7	2	5	8	2	6	30
Panel C. Net CAPM α After Borrowing and Delay Costs							
C1. Small Trades							
Avg. α	−0.129	−0.044	0.000	0.202	−0.005	−0.022	0.021
Avg. $t(\alpha)$	(−1.77)	(−0.27)	(−0.02)	(0.89)	(−0.21)	(−1.26)	(−0.35)
Avg. BidAsk Cost	0.332	0.328	0.105	0.153	0.123	0.202	0.188
Avg. Delay Cost	0.049	0.108	0.005	0.050	0.009	0.029	0.035
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	0	0
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	0	0	0	0	0	0	0
C2. Large Trades							
Avg. α	−0.044	−0.009	0.036	0.212	0.034	0.030	0.062
Avg. $t(\alpha)$	(−0.77)	(−0.05)	(0.62)	(0.98)	(0.32)	(−0.45)	(0.21)
Avg. BidAsk Cost	0.205	0.203	0.064	0.093	0.074	0.125	0.115
Avg. Delay Cost	0.090	0.199	0.011	0.100	0.019	0.054	0.067
$\#(t(\alpha) > 1.96)$	0	0	0	0	0	1	1
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	1	0	2	0	1	6

The table reports the average bond CAPM α, the associated average t-statistic, and the number of factors with a t-statistic greater than 1.96 before and after accounting for transaction costs, including bid-ask spread costs, borrowing costs for short sales, and delay costs across factors associated with the six categories from Section 5.2. Of the 341 factors that we generate, we consider the 58 factors with significant average excess returns. We regress each factor's returns on the corporate bond market factor and estimate the intercept of the regression, α. We group the 58 factors into six categories and average the estimates within each group. The values in parentheses are the average of the t-statistics (Newey-West adjusted with 12 lags). In Panel B, we subtract the borrowing fee per month from the returns on the short positions to compute the CAPM α. In Panel C, we employ our own estimates of trade intensity to calculate the cost of delays and present the net-of-cost alphas and the number of factors that remain statistically significant. All factors are formed with the `PyBondLab` Python package.

Table A.11: CAPM Alphas, Long-Only Factors

Category	Equity Momentum	Equity Reversal	Investment	Spreads	Value Profit	Volatility Risk	All
Count	9	4	13	14	7	11	58
Panel A. Gross CAPM α							
Avg. α	0.226	0.339	0.157	0.402	0.155	0.237	0.254
Avg. $t(\alpha)$	(3.44)	(2.62)	(2.68)	(2.33)	(2.08)	(2.69)	(2.64)
Avg. Turnover	142.58	150.29	30.34	45.46	32.10	82.43	69.77
$\#(t(\alpha) > 1.96)$	9	4	13	11	4	9	50
Panel B. Net CAPM α, Own Estimates of Delay							
B1. Small Trades							
Avg. α	−0.052	0.041	0.082	0.294	0.076	0.077	0.108
Avg. $t(\alpha)$	(−1.15)	(0.22)	(1.38)	(1.58)	(0.88)	(0.21)	(0.67)
Avg. BidAsk Cost	0.167	0.177	0.054	0.074	0.056	0.099	0.094
Avg. Delay Cost	0.112	0.120	0.021	0.034	0.022	0.061	0.053
$\#(t(\alpha) > 1.96)$	1	0	1	2	0	2	6
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	0	2	2	0	2	8
B2. Large Trades							
Avg. α	−0.082	0.009	0.081	0.289	0.075	0.063	0.097
Avg. $t(\alpha)$	(−1.65)	(−0.04)	(1.36)	(1.54)	(0.87)	(−0.03)	(0.52)
Avg. BidAsk Cost	0.103	0.110	0.033	0.045	0.034	0.061	0.057
Avg. Delay Cost	0.205	0.220	0.043	0.069	0.046	0.113	0.100
$\#(t(\alpha) > 1.96)$	0	0	1	2	0	2	5
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	5	2	8	8	3	5	31
Panel C. Net CAPM α, Kargar et al. (2023) Estimates of Delay							
C1. Small Trades							
Avg. α	−0.075	0.016	0.077	0.286	0.071	0.064	0.097
Avg. $t(\alpha)$	(−1.54)	(0.02)	(1.30)	(1.52)	(0.81)	(0.00)	(0.52)
Avg. BidAsk Cost	0.167	0.177	0.054	0.073	0.056	0.099	0.093
Avg. Delay Cost	0.135	0.145	0.026	0.042	0.027	0.074	0.064
$\#(t(\alpha) > 1.96)$	0	0	1	2	0	2	5
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	2	0	2	2	0	3	9
C2. Large Trades							
Avg. α	−0.062	0.030	0.086	0.296	0.081	0.074	0.107
Avg. $t(\alpha)$	(−1.33)	(0.13)	(1.45)	(1.59)	(0.95)	(0.14)	(0.65)
Avg. BidAsk Cost	0.103	0.110	0.033	0.045	0.035	0.061	0.058
Avg. Delay Cost	0.185	0.199	0.038	0.060	0.040	0.102	0.090
$\#(t(\alpha) > 1.96)$	1	0	1	2	0	2	6
$\#(t(\alpha^{BidAskOnly}) > 1.96)$	5	2	8	8	3	5	31

The table reports the average bond CAPM α, the associated average t-statistic, and the number of factors with a t-statistic greater than 1.96 before and after accounting for transaction costs, including bid-ask spread costs and delay costs across factors associated with the six categories from Section 5.2. Of the 341 factors that we generate, we consider the 58 factors with significant average excess returns. We regress each factor's excess returns on the corporate bond market factor and estimate the intercept of the regression, α. We group the 58 factors into six categories and average the estimates within each group. The values in parentheses are the average of the t-statistics (Newey-West adjusted with 12 lags). In Panel B, we employ our own estimates of trade intensity to calculate the cost of delays and present the net-of-cost alphas and the number of factors that remain statistically significant. In Panel C, we use the Kargar et al. (2025) delay estimates for the trade intensity parameter and compute the cost of delays for small and large trades respectively. All factors are formed with the `PyBondLab` Python package.

Table A.12: Forecasting Regressions of Corporate Bond Transactions: Liquid Bonds (Customer Buy)

		Bonds with High Turnover				Large Bonds			
		\$100,000		\$2M		\$100,000		\$2M	
Panel A. Regression Estimates									
$\log InvComp$		0.014	(12.93)	0.022	(11.30)	0.018	(12.32)	0.031	(11.77)
$\Delta Inventory$		0.007	(13.50)	0.040	(34.31)	0.010	(19.65)	0.047	(38.19)
CAP		−0.001	(−0.72)	0.011	(3.88)	−0.010	(−3.43)	−0.005	(−1.68)
$Rating$		0.040	(19.98)	0.077	(34.98)	0.052	(23.47)	0.107	(41.01)
$Maturity$		−0.020	(−9.98)	−0.003	(−1.84)	−0.016	(−7.35)	0.005	(2.61)
$\log FaceValue$		0.092	(50.65)	0.175	(104.61)				
$Coupon$		−0.026	(−13.09)	0.009	(3.23)	−0.054	(−21.08)	−0.048	(−16.92)
Age		−0.028	(−15.25)	−0.059	(−34.82)	−0.034	(−16.54)	−0.074	(−33.36)
$HighTurnover$		0.091	(26.66)	0.209	(67.32)				
$LargeFaceValue$						0.165	(44.81)	0.326	(92.74)
Intercept		0.823	(264.79)	0.472	(143.08)	0.791	(190.05)	0.424	(114.70)
Adj. R^2		0.166		0.284		0.127		0.207	
N		814,145		814,145		814,145		814,145	
Panel B. Model-Based Estimates of Trade Intensity and Delays									
Prob. Trade	$Prob^{Data}$	0.914		0.681		0.956		0.750	
	$Prob^{LB}$	0.972		0.773		-		0.882	
λ	λ^{Data}	0.117		0.054		0.149		0.066	
	λ^{LB}	0.170		0.071		-		0.102	
Exp. Delay	$1/\lambda^{Data}$	8.573		18.373		6.716		15.143	
(days)	$1/\lambda^{LB}$	5.876		14.162		-		9.811	

The table reports the coefficient estimates for the regressions of the bond trading dummy for the customer buy side on characteristics as in Eq. (11). The variable $InvComp$ is the investor composition of Li and Yu (2025), which measures the activeness of bond investors at the end of the previous quarter. The variable $\Delta Inventory$ is the difference between customer buys and customer sells in the preceding 28 days, while CAP is the intermediary capital ratio of He et al. (2017) in the previous month. $HighTurnover$ is a dummy which is one if the bond is above the median in terms of bond turnover rate in month t and zero otherwise. $LargeFaceValue$ is a dummy which is one if the bond is above the median in terms of amount outstanding in month t and zero otherwise. Except for the two dummy variables, the right-hand-side variables are standardized for ease of interpretation. Values in parentheses are t-statistics clustered at the bond and month levels. The sample is monthly, spanning August 2002 to November 2022.

Table A.13: Short-Selling Summary Statistics – Markit Indicative Fees

Category	Mean	Std.	Percentiles						
			p1	p10	p25	p50	p75	p90	p99
All	3.63	5.27	1.32	2.06	2.37	3.12	3.32	4.15	20.29
AAA to AA−	3.14	2.71	1.43	2.10	2.68	3.12	3.17	3.67	7.27
A+ to A	3.15	3.15	1.35	2.08	2.33	3.12	3.15	3.72	9.09
A− to BBB+	3.14	4.04	1.26	1.97	2.29	3.12	3.15	3.74	9.50
BBB to BBB−	3.21	3.79	1.25	1.95	2.29	3.12	3.20	3.85	11.50
BB+ to B−	4.36	5.56	1.49	2.16	3.08	3.27	4.15	4.77	30.00
CCC+ to D	12.55	18.42	2.08	3.12	3.55	9.75	11.50	25.00	85.22

The table presents the pooled summary statistics for the Markit `IndicativeFee` variable. The number of pooled bond-month observations is 1,102,569. The statistics are monthly and presented in basis points. The `IndicativeFee` variable is the net buy side fee paid to borrow the underlying bond. Specifically, it is defined as the interest rate on cash funds minus the rebate rate (that is paid for collateral) and is directly provided by Markit. This fee (cost) is used to adjust the short-leg of the bond factors. The data spans the sample period August 2002 to November 2022. `IndicativeFee` data are unavailable for all bonds prior to September 2006. Therefore, we impute these missing values using the average `IndicativeFee` within each rating category, enabling the sample to begin in August 2002 and align with the sample used for the main results.

Table A.14: Summary Statistics for the Corporate Bond Market Factor.

Panel A: Corporate Bond Market Factor Statistics

	$MKTB_{Net}$	$MKTB_{Gross}$	$MKTB$
Mean	0.316	0.367	0.364
	(2.14)	(2.36)	(2.32)
SD	2.06	1.95	1.91
SR	0.53	0.65	0.66

Panel B: Pairwise Correlations

	$MKTB_{Net}$	$MKTB_{Gross}$	$MKTB$
$MKTB_{Net}$	1		
$MKTB_{Gross}$	0.982	1	
$MKTB$	0.973	0.992	1

Panel A reports the monthly factor means (Mean), the monthly factor standard deviations (SD), and the annualized Sharpe ratios (SR). The $MKTB_{Net}$ factor is constructed as the weighted-average of the BlackRock iShares iBoxx Investment Grade (ticker: LQD) and High Yield (ticker: HYG) ETF net returns from the CRSP Mutual Funds database. The $MKTB_{Gross}$ factor is constructed as the weighted-average of the Bloomberg-Barclays Investment Grade and High Yield index gross returns. The $MKTB$ factor is the value-weighted bond market factor publicly available from openbondassetpricing.com. Panels A and B are based on the sample period 2002:08 to 2022:12 (245 months). t-statistics are in round brackets computed with the Newey-West adjustment with 12-lags.

Table A.15: Hyperparameters across the Machine Learning Models.

Panel A: Linear Models with Penalties: LASSO, RIDGE & ENET		
Parameter	`sklearn` mnemonic	Value
Intercept	`fit_intercept=True`	True
ℓ_1 penalty	`alphas`	Variable
ℓ_2 penalty	`alphas`	$\in [0.0001, \ldots, 1]$
Num. Penalties	`n_alphas`	100
ℓ_1 ratio	`l1_ratio`	$\in$ [0.001, 0.01,0.99,0.999]

Panel B: Tree-Based Ensembles: RF and XT		
Parameter	`sklearn` mnemonic	Value
Num. Trees	`n_estimators`	100
Max depth	`max_depth`	$\in$ [2,4,6]
Split features	`max_features`	$\in$ [5,10,20]
Min leaf samples	`min_samples_leaf`	$\in$ [1,10,50]

Panel C: Feed Forward Neural Network: NN		
Parameter	`tensorflow` mnemonic	Value
Layers	`Dense`	1
Neurons	`Dense`	32
Activation	`activation='relu'`	ReLu
Epochs	`epochs`	100
Batch size	`batch_size`	1024
Batch normalization	`BatchNormalization`	True
Optimizer	`optimizers.Adam`	Adam
Patience	`patience`	5
Learning rate	`learning_rate`	$\in$ [0.001, 0.01]
ℓ_1 penalty	`regularizers.l1`	$\in$ [0.001, 0.01]
Ensemble	-	10
Grand Ensemble	-	10

The table reports the relevant hyperparameters that are chosen via a cross-validation scheme with a 70:30 train-validate split that maintains the temporal ordering of the data. The cross-validation is conducted every 5years commencing on 2002:07 using an expanding window. The set of hyperparameters is chosen such that it yields the smallest mean squared error (MSE) in the validation sample. Panel A reports the hyperparameters for the linear models that include Lasso (LASSO), Ridge (RIDGE) and Elastic Net (ENET), respectively. Panel B reports the hyperparameters for the set of tree-based ensembling nonlinear models that include the random forest (RF) and extremely randomized trees (XT). Panel C reports the hyperparameters for the feed forward neural network (NN). All models except for the NN are estimated with `sklearn`. The NN is estimated with `tensorflow`.