Using corporate bond market evidence, we show that sophisticated investor attention allocation explains large variation in price underreaction to information. We hypothesize that investors have information-processing constraints so prices underreact to information, but investors optimally allocate more attention to more payoff-relevant news, so prices underreact less to such news. Consistent with this “sophisticated inattention” hypothesis, corporate bonds with higher credit risk underreact less to default risk news and bonds with longer duration underreact less to interest rate news. As the credit quality and duration of a bond changes over time, the price response speed to different news also changes accordingly, consistent with investors adjusting attention allocation to changes in circumstances. Many investors do appear to face binding attention constraints, as larger shocks from one risk consume more attention and result in slower price response to the other risk. Differences in bond liquidity cannot explain our findings, because if a bond is illiquid, it should be equally slow to respond to all news, but we find large differences in response speeds at the risk level. Finally, the amount of “money left on the table” is small enough to be plausibly explained by information-processing costs.

**Keywords:** Investor Inattention, Investor Sophistication, Market Efficiency

**JEL classification:** G14, G41

---

*Stanford Graduate School of Business ([jiacui@stanford.edu](mailto:jiacui@stanford.edu)). I am grateful for comments from Yu An, Jonathan Berk, Svetlana Bryzgalova, Darrell Duffie, Matthew Gentzkow, Steven Grenadier, Benjamin Hebert, David Hirshleifer, Robert Hodrick, Peter Koudijs, Arvind Krishnamurthy, Charles Lee, Ken Li, Yucheng Liang, Hanno Lustig, Ulrike Malmendier, Filip Matejka, Xuelang Mu, Paul Pfeiderer, Yang Song, David Sraer, Jeremy Stein, Xinyuan Tao, Laura Veldkamp, Edward Michael Watts, Joanna Yan, and David Yang, as well as seminar participants at the Colorado Finance Summit, Fudan Fanhai, Peking Guanghua, Stanford GSB, and Yale SOM.
1 Introduction

Contrary to the efficient market hypothesis, asset prices often respond slowly to information. After decades of work, researchers still do not agree on the explanation. Some researchers believe that investors can process information immediately, therefore any price underreaction necessarily reflects illiquidity. At the same time, other researchers argue that prices underreact because investors are naive and neglect useful information. In this paper, using corporate bond market data, we show evidence for a “sophisticated inattention” view that is different from both existing explanations. We show that investors do have information-processing constraints, but they are not naive, and they purposefully allocate their scarce cognitive budget according to marginal costs and benefits.

The key empirical challenge is to circumvent liquidity-based explanations. Illiquid securities are bound to underreact more to information, and liquidity is hard to measure, so any security-level variation in price underreaction may always be explainable by omitted liquidity variables. The key innovation of our empirical strategy is to investigate differences in price underreaction to different risks with different payoff-relevance. If investors are sophisticated in allocating attention, they will pay more attention to risks that are more payoff-relevant, and thus prices would respond more quickly to those risks but less quickly to less payoff-relevant risks. We find strong evidence for this prediction in the corporate bond market. To our best knowledge, this is the first paper to empirically demonstrate the impact of purposeful attention allocation in explaining price underreaction to information.

To derive testable predictions, we first develop a simple trading model in which investors allocate scarce attention across multiple risks. Following Van Nieuwerburgh and Veldkamp (2010), we model attention allocation as investors choosing the precision of the signals they

---

1 As examples of slow price response to information, prior works document that stock prices underreact to earnings news (Bernard and Thomas (1990)), new security analyst recommendations (Womack (1996)), revenue news (Jegadeesh and Livnat (2006)), and news about economically related firms (Cohen and Frazzini (2008), Menzly and Ozbas (2010)).

2 As an example of the first perspective, Hou and Moskowitz (2005) discuss slow stock price reaction to market returns as a measure of “market frictions.” As examples for the investor neglect view, most empirical papers on investor inattention assume that investors are making mistakes. Hirshleifer, Lim, and Teoh (2009) document slower stock price response to earnings announcements on days with more news releases, interpreting this as “extraneous” information distracting investors. DellaVigna and Pollet (2009) document slower price response to earnings announced on Fridays and call this finding the “market response to a bias.”

3 We focus on exploiting the payoff-relevance of risks because this is a sharper test of conscious, purposeful attention allocation. A number of existing papers measure impact of exogenous shocks to attention costs which are harder to interpret as tests of purposeful attention allocation. For instance, Fedyk (2018) shows that Bloomberg news on the front page are viewed more and trigger faster price adjustment. Because investors see front-page news first before scrolling to the next page, even if they are not sophisticated and do not purposefully allocate attention between front-page and non-front-page news, we would expect the findings to be true.
receive on each risk. The model generates two main predictions. First, investors will allocate more attention to risks with higher payoff-relevance, trade on them more aggressively, and cause prices to respond faster to these more important risks. Second, if investors face attention constraints and cannot easily expand attention capacity, then there will be a “distraction effect”: if one risk (e.g., credit risk) is more important, then that consumes investor attention and slows down price reaction to the other risk (e.g., interest rate risk).

We test these predictions in the corporate bond market. To avoid stale quotes or lack of trading, we only use actual transaction prices institutional size trades to compute weekly returns, and drop all observations without trading in the current or the previous week. For each bond, we use the returns of duration-matched Treasury bonds to proxy for interest rate shocks, and we use company stock returns to proxy for fundamental risk shocks. While price response to these shocks are sometimes slow, eight weeks appears sufficient for the price adjustment process to converge, so we use the eight-week price response to measure full price reaction. Following the model, we measure underreaction as the fraction of price reaction that does not happen immediately when shocks realize.

Both model predictions are strongly supported in the data. Consistent with the first prediction, bonds with higher credit risk are faster to respond to default risk shocks, and bonds with longer duration are faster to respond to interest rate shocks. Consistent with the second “distraction effect” prediction, if interest rate risk (credit risk) is more relevant, then the price response to default shocks (interest rate shocks) is slower. As an example, Figure 1 plots the response of the prices for bonds with opposite payoff-relevance profiles. Bonds with lowest quintile interest rate relevance and highest quintile default risk relevance underreact to interest rate shocks by 75% and default risk shocks by 48% (blue lines). In contrast, bonds with the opposite risk exposure show the opposite pattern of price adjustments, underreacting to interest rate shocks only by 19% but to default risk shocks by 74% (black lines). As explained earlier, this risk-level variation cannot be attributed to differences in bond-level liquidity differences.

These risk-level variation of price response speeds also hold up when examining within-bond variation. That is, when the credit quality or interest rate sensitivity of a bond changes, the price underreaction patterns also change accordingly, consistent with investors adjusting attention allocation according to changes in circumstances. To test this, we split the sample into three equal-length periods and measure underreactions separately for each bond and each period. To exploit within-bond changes, we regress risk-level underreactions on payoff-relevance with bond fixed effects. We find that when bond exposure to a risk increases (or decreases), the price underreaction to that risk decreases (or increases), as our theory predicts.
Two additional model predictions are also supported. First, after controlling for payoff relevance, the model predicts that risks with lower attention cost should be incorporated into prices more quickly. In corporate bonds, the cost of learning about interest rate shocks is arguably lower. Consistent with this, price reaction to interest rate shocks is 15 to 20% faster than to default risk shocks. Second, we use trading volume as a proxy for investor attention and conduct an event study around earnings announcements. Consistent with investors understanding which bonds are more sensitive to default-relevant information, bonds with higher credit risk experience much larger trading volume spikes around earnings announcements.

We consider a number of alternative hypotheses. First, illiquidity can also cause slow price reaction to information. However, as explained earlier, because our tests focus on risk-level variation, the facts we document cannot be explained by liquidity differences at the bond level. Without selective investor attention allocation, illiquid bonds will be slow to respond to all information, but the degree of underreaction cannot differ by risk. Not only that liquidity cannot explain our risk-level results, but we also find that the risk-level
underreaction variation explained by our mechanism is much larger than the bond-level
variation explained by bond-level liquidity proxies.

We then show that our findings cannot be explained by transaction costs. This alternative
hypothesis goes as follows. If investors face transaction costs, such as bid-ask spreads, then
they may only trade when a bond experiences large shocks. Because shocks in more payoff-
relevant risks tend to be larger, this may explain the first prediction of our model: faster
price response to more important risks. However, this alternative explanation predicts the
opposite of our second prediction (the “distraction effect”) and is empirically rejected. We
also rerun our tests in periods with large ex-post shock sizes and find our results to be robust,
while under this alternative hypothesis, our results should be explained away.

In the final part of the paper, we quantify the implied attention cost by measuring
how much “money is left on the table.” If there is too much money left on the table to
be reasonably explained by information-processing costs, this will entail a rejection of the
sophisticated inattention view. A back-of-envelope calculation shows that the manager of
an average corporate bond mutual fund (around $1 billion in assets, 130% annual turnover)
can at most increase her annual compensation by 2.7% ($74,000) if she pays full attention.
We err on the side of overestimation, so the true gain may be significantly lower. Given this,
we argue that the foregone profit is small enough to plausibly reflect information-processing
costs.

This paper contributes to the existing literature in two ways. First, it is the first paper
to demonstrate that purposeful allocation of attention can explain price underreaction to
information. Second, researchers frequently assume that institutional investors, to a first
order approximation, have zero information-processing costs. By studying the impact of
information-processing costs in an institution-dominated market, we challenge this existing
view. Section 1.1 explains the relationship of this paper to the related literature in more
detail.

We explain our findings in more detail in the remainder of this paper. Section 2 de-
scribes our endogenous investor inattention model and derives testable implications. Section
3 explains how we measure relevant variables, and section 4 shows evidence that price un-
derreaction depends on the payoff-relevance of risks. Section 5 tests our explanation against
alternative hypotheses. Section 6 estimates the implied cost of investor inattention and
section 7 concludes.

4Under this alternative hypothesis, price response speed to a shock depends upon the ex-post shock size.
Therefore, if the bond experiences a large default risk shock, more investors will trade on the bond in that
period, speeding up price response to interest rate shock. The opposite is true if investors face attention
constraints, as more of their attention will be consumed in processing default risk information, leaving less
energy to respond to interest rate information.
1.1 Literature review

Our paper is most closely related to the burgeoning literature linking investor inattention to asset prices. Making use of exogenous distractions to investors, Hirshleifer et al. (2009) and Drake, Gee, and Thornock (2016) show that inattention causes price underreaction. DellaVigna and Pollet (2009) argue that investors are less attentive to earnings released on Fridays. Da, Engelberg, and Gao (2011) and Ben-Rephael, Da, and Israelsen (2017) develop internet and Bloomberg search-based proxies of investor inattention and show that they are related to stock returns. Using Google search volumes, Liu and Peng (2015) show that investor attention to stocks increases upon earnings and macroeconomic news releases.

While existing papers focus on showing that inattention impacts asset pricing, we differ by focusing on showing that inattention is endogenously determined.

More broadly, our work is related to using rational inattention models to explain economic behavior. Most of the relevant papers are in macroeconomics, starting with Sims (2003), and are surveyed in Mackowiak, Matejka, and Wiederholt (2018). There is also a more behavioral-oriented literature on endogenous attention allocation (Gabaix (2014), Gabaix (2017)). In finance, there are a number of theory papers that link attention allocation to asset prices (Peng (2005), Hirshleifer, Lim, and Teoh (2011)) and category learning (Peng and Xiong (2006)). In terms of modeling techniques, we borrow most heavily from Van Nieuwerburgh and Veldkamp (2010).

A number of recent papers apply rational inattention to asset pricing. The closest one to our paper is Katz, Lustig, and Nielsen (2016) who show that stock prices are slow to respond to local inflation shocks and interpret their finding using rational inattention. Huang, Huang, and Lin (2018) show that when Taiwan investors are distracted by large jackpot lotteries, stock return comovement increases, consistent with investors allocating more attention to market-wide information and less attention to firm-specific information. Peng, Xiong, and Bollerslev (2007) show that stock returns comove more with the market in periods of more macroeconomic uncertainty, and interpret this using endogenous investor attention allocation. Finally, Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) show that mutual funds make more profits by timing the market in recessions and picking stocks in other periods, consistent with fund managers dynamically allocating their attention to more profitable tasks across the business cycle.
2 Conceptual framework

We develop a model to derive testable predictions of sophisticated investor attention allocation. It is a standard static trading model, with the additional feature that investors choose how to allocate their attention across risks. The modeling of attention choice builds upon Van Nieuwerburgh and Veldkamp (2010). The model timeline is shown in Figure 2.

2.1 Model

Asset There is a single risky asset with zero net supply. The fundamental asset value is the sum of $N$ independent risk components,

$$V = \sum_{i=1}^{N} X_i, \quad X_i \sim \mathcal{N}(0, \sigma_i^2) \quad \forall i$$

where each risk $X_i$ is normally distributed with mean zero and variance $\sigma_i^2$. We assume the risks are independent for analytical tractability. We think of the different variances $\{\sigma_i^2\}_{i=1}^{N}$ as the payoff-relevance of risks. For example, take the asset to be a corporate bond with $X_1$ representing the interest rate risk component, $X_2$ representing the default risk component, and $X_3, \ldots, X_N$ representing others such as the liquidity risk components, and so forth. Bonds with longer duration will have higher $\sigma_1$ and those with higher credit risk will have higher $\sigma_2$.

Trading At time 1, there is a continuous unit mass of competitive, identical risk-neutral investors who submit market orders. Each investor $j \in [0, 1]$ incurs a quadratic trading cost.
\( \frac{\psi}{2} q_j^2 \) when submitting her order of \( q_j \) shares, where parameter \( \psi > 0 \) governs how costly trading is. The aggregate order flow \( Y = \int_0^1 q_j dq_j \) is then absorbed by market makers who set price equal to \( p_1 = \lambda \cdot Y \).

**Investor attention allocation** Before trading at time 1, each investor \( j \) receives noisy signals on the \( N \) risks:

\[
s_{ij} = X_i + \sigma_i \cdot u_{ij}, \quad u_{ij} \sim N(0, K^{-1}_{ij})
\]

where the signal errors \( u_{ij} \) are independent across risks and investors. The precisions of the normalized signals \( \{s_{ij}/\sigma_i\}_{i=1}^N \) are \( \{K_{1j}, ..., K_{Nj}\} \), and they are investor choice variables.

We model investor attention using these signal precisions. If the investor pays zero attention to risk \( i \) (\( K_{ij} = 0 \)), then her signal has infinite variance and she knows nothing about \( X_i \). If she chooses a very large \( K_{ij} \), then she knows \( X_i \) very precisely. We think of this precision choice as capturing investor effort to processing information. For corporate bond investors, if they very closely follow news about the company and work diligently to infer the value impact of those news, that will show up as them choosing a high precision on the default-risk component.

To capture the idea that attention is costly, we assume that investors can reduce the amount of noise in signals by paying attention resources. We closely follow the modeling technique in Van Nieuwerburgh and Veldkamp (2010). Specifically, at time 0, each investor \( j \) choose signal precisions \( \{K_{ij} \geq 0\}_{i=1}^N \) subject to a cost of

\[
\sum_{i=1}^N c_i \cdot K_{ij}
\]

where \( c_i > 0 \) are risk-specific attention costs. We allow for different costs across risks because some risks are arguably less costly to process. In the corporate bond market, it is arguably the case that interest rate relevant news – macroeconomics news, central bank announcements, etc – are easier to process than company-level default-relevant news. For

---

5This quadratic cost can be thought of as a reduced-form way to capture risk aversion of agents with constant absolute risk aversion. We can think of \( \psi \) as equal to \( A \cdot Var(R|\{s_{ij}\}_{i=1}^N) \) where \( A \) is the absolute risk aversion coefficient and \( R \) is the return on the asset from time 1 to time 2. Our assumption of a constant \( \psi \) thus amounts to ignoring changes in the conditional payoff variance in \( Var(R|\{s_{ij}\}_{i=1}^N) \) when investors choose different signal structures (to be explained later). This assumption should be innocuous if the amount of variance reduction from learning is small relative to the overall payoff risk, which is arguably the case in our empirical setting of corporate bonds.

6While we take the price-impact parameter \( \lambda \) to be fixed in this section, we show in Appendix A.2 that it can be endogenized using the Kyle (1985) set-up. That is, \( \lambda \) can arise from competitive and uninformed market makers learning from order flow and setting break-even price \( p_1 = E(V|Y) \).
a typical corporate bond portfolio manager, she only needs to learn about interest rate risk once and can apply it to all her bond holdings. Learning about default news requires reading news about all companies in her portfolio. Therefore, after dividing by the number of bonds held, the cost of learning about interest rates should be considerably lower.

2.2 Equilibrium

We characterize the equilibrium and provide intuition about the underlying mechanisms. Our solution concept is Perfect Bayesian Equilibria in which each investor \( j \) chooses optimal attention allocation \((K_{1j}, K_{2j}, ..., K_{Nj})\) and optimal order submission rules \(q_j(s_{1j}, s_{2j}, ..., s_{Nj})\). This baseline version of our model can be solved analytically. Proofs are in Appendix A.

**PROPOSITION 1.** (Equilibrium) There is a unique equilibrium where time 1 price is

\[
p_1 = \sum_{i=1}^{N} (1 - D_i) \cdot X_i
\]

(4)

For each risk \( i \), if payoff-relevance is high enough \((\sigma_i \geq \sqrt{2\psi c_i})\), then the price underreaction and all investors’ attention choices are:

\[
D_i = \frac{\psi + \lambda \sqrt{2}\psi c_i}{\psi + \lambda - \sigma_i \sqrt{2}\psi c_i}
\]

(5)

\[
K_i = \frac{\lambda + \sigma_i \sqrt{2}\psi c_i}{\lambda + \psi}
\]

(6)

Otherwise, payoff-relevance isn’t high enough, then all investors pay no attention \((K_i = 0)\) and price completely underreacts \((D_i = 1)\).

Finally, each investor \( j \) submits the following market order:

\[
q_j(s_{1j}, ..., s_{Nj}) = \frac{1}{\psi} \cdot \sum_{i=1}^{N} D_i \cdot \frac{K_i}{K_i + 1} \cdot s_{ij}.
\]

(7)

The price path is illustrated in Figure 3. The time 1 price will partially react to risks, and \( D_i \) is the fraction of price underreaction to risk \( i \). We make a few remarks about the model:

1. Investor attention determines price underreaction. If investors pay more attention to a risk by choosing higher \( K_i \), then they will trade more heavily on that risk (equation (7)), speeding up price incorporation of information.
Figure 3. Illustration of price response to risk \( i \) in the model. Price incorporates \( 1 - D_i \) fraction of the realization \( X_i \) at time 1, so \( D_i \) (red) is the fraction of underreaction.

2. Investors profit from more precise signals by trading on price drifts. For each risk \( X_i \), the time 1 to time 2 price drift is \( D_i \cdot X_i \).
3. Investor attention allocation to each risk \( i \) is decreasing in the cost of attention \( c_i \) and increasing in the payoff-relevance \( \sigma_i \) (equation (6)).
4. The equilibrium uniqueness comes from two forces. First, investors are competitive and take \( D_i \) as exogenously given. Second, the attention allocation of different investors are strategic substitutes. That is, if other investors pay more attention to risk \( i \), \( D_i \) will decrease, then some investors will react by decreasing attention to risk \( i \). In contrast, strategic complementarity tends to lead to multiple equilibria.

Note that \( \lambda \) is assumed fixed in this baseline model. In Appendix A.2 we show that the equilibrium structure is the same if we endogenize \( \lambda \) as in Kyle (1985).

2.3 Two testable predictions

The key feature of our model is that price underreaction will be risk-specific. We derive two predictions to bring to data.

1. **Main effect:** for each risk \( i \), price underreaction \( D_i \) is weakly decreasing in payoff-relevance \( \sigma_i \) and weakly increasing in attention cost \( c_i \).

   This is a direct corollary from the equilibrium solution in Proposition 1. The “weak” part in the statement is only for extreme cases where investors pay zero attention. Otherwise, the comparative statics is strict.

   This prediction can be best understood from a first order condition intuition. Because investors trade on drifts \( D_i \cdot X_i \), the degree of profitability from learning risk \( X_i \) is related to \( D_i \cdot \sigma_i \). Obviously, higher attention cost \( c_i \) means investors will optimally
pay less attention and thus price underreaction increases. Fixing attention cost $c_i$, the higher the payoff-relevance $\sigma_i$, the smaller $D_i$ has to be for agents to be optimizing.

2. “Distraction effect”: If investors face attention constraints, i.e. they have a finite attention budget ($\sum_{i=1}^{N} c_iK_{ij} \leq \bar{K}_j$), then price underreaction to each risk $i$ ($D_i$) is weakly increasing in the payoff-relevance $\sigma_r$ and weakly decreasing in the attention cost $c_r$ of other risks ($r \neq i$).

This is a test of whether investors have attention constraints. If they don’t, as in the baseline model, then the payoff-relevance and attention costs of other risks are irrelevant for $D_i$.

Mapping to the empirical setting, we would expect investors to not face constraints if they can flexibly add attention capacity by working harder or hiring more analysts. This may be a reasonable assumption for institutional investors. At the same time, if there are frictions in hiring and delegating tasks to others, then investment managers may indeed face constraints.

In our empirical tests, the main variation will come from payoff-relevance. Corporate bond value depends on interest rate risks and default risks, but depending on the duration and credit quality of the bond, there is significant variation in payoff-relevance. In terms of the attention cost prediction, the only testable variation is comparing between the two type of risks. We think attention cost to interest rate should be lower, so after controlling for payoff-relevance, underreaction to interest rate should be smaller.

3 Data and measures

We choose to test our theory in the corporate bond market because there is substantial variation in the payoff-relevance of interest rate risk and default risk. Also, we have timely proxies of shocks. Specifically, we will use Treasury returns and stock returns as proxies for interest rate and default risk shocks. This section summarizes the data and explains how we measure the two key variables: payoff-relevance and price underreaction.

Our test contingents on the assumption that it is not costless to process these two types of information. Costly default risk information should be uncontroversial. However, one may wonder why processing interest rate information is not costless. Don’t bond investor
just need to look at yield movements of duration-matched treasuries, and adjust their bond valuation one-for-one?

This is not the case. Corporate bonds do not move one-to-one with treasuries. Treasury market movements is frequently associated with other macroeconomic news (e.g. employment reports, Jones, Lamont, and Lumsdaine (1998)) that necessitate adjustments in credit spreads, and the impact is also heterogeneous across bonds (Collin-Dufresn, Goldstein, and Martin (2001)). We contend that inferred the credit impact of macroeconomics news is not costless.

Appendix B shows evidence. We use the contemporaneous movement of investment grade CDX index spread to proxy for the amount of credit-relevant macro news. The bond price response to treasuries is heterogeneous and depends on the information being released. When the Treasury movement is not accompanied by much CDX movement, the bond return sensitivity to treasuries is closer, although still not equal, to the duration measure. The more the macro news release, the smaller the bond sensitivity to treasuries.

### 3.1 Data

#### 3.1.1 Corporate bond data

We use corporate bond price data from the enhanced Trade Reporting and Compliance Engine (TRACE). This dataset contains the records of all dealer-intermediated corporate bond transactions in the U.S. from July 2002 to December 2014. Because most corporate bonds do not trade every day, we first convert data into weekly prices. Following Bessembinder, Kahle, Maxwell, and Xu (2009), we first compute daily volume-weighted average prices for each bond and then use the latest day of the week on which trading occurred. We also follow Bessembinder et al. (2009) to only use transactions with value higher than $100,000 in order to exclude retail trading which may happen at noisy prices. We then merge the data with FISD Mergent to get bond characteristics such as coupon rate, issuance size, optionality, and so forth.

We apply a number of cleaning procedures to enhance data quality. At the raw transactions level, we use the algorithm in Dick-Nielsen (2014) to delete canceled, reversed, and double-counted trades. Because testing our theory requires non-negligible interest rate risk exposure, we do not use bonds with floating-rate coupons and require at least one year of maturity remaining. To reduce confounding effects from currency risks or optionality, we

---

Because we use the latest observation in each week, one may be concerned that bonds that trade more frequently will thus have an “advantage” and appear to respond faster to information. This is not an issue for two reasons. First, the underreactions we document later are multi-week delays, not within-week delays. Second, many of our tests are based on differences in response speed of different risks for the same bond.
restrict attention to USD-denominated bonds that are not callable, puttable, or convertible. Finally, we require the bond issuance size to be greater than $10 million.

For some of the exercises, we use weekly excess returns. For each bond $i$ in week $t$, we compute returns using

$$\text{BondRet}_{i,t} = \frac{\text{Price}_{i,t} + (7/365) \times \text{CouponRate}_i}{\text{Price}_{i,t-1}} - (1 + \text{RiskFree}_t)$$ \hspace{2cm} (8)$$

where $\text{Price}_{i,t}$ is the weekly bond price. The risk-free rate is the one-month Treasury bill rate. To reduce the impact of data errors, we eliminate cases with $\text{Price}_{i,t-1}$ above 150% of par value and winsorize bond returns at (0.5%, 99.5%) levels.

We take care not to identify slow price reactions due to mechanical reasons. First, we only use actual transactions so stale quotes do not affect our inference. Second, if a bond does not trade for either of week $t - 1$ or week $t$, the week $t$ return is designated as missing, so we will not mechanically infer “slow reaction” from lack of trading.$^8$

3.1.2 Proxies for default risk and interest rate risk shocks

We use returns of duration-matched U.S. Treasury portfolios as proxies for interest rate shocks for each bond. Matching on duration accommodates changes in bond duration over time and the fact that the movements of different parts of the Treasury yield curve can be different. The treasuries data come from the Center for Research in Security Prices (CRSP) fixed-term Treasury indices.$^9$

Similarly, we use company stock returns as proxies for shocks to fundamental risk. According to the Merton model, when the asset value of the company changes, the default probability of debt changes accordingly. The response of bond returns to stock returns is nonlinear and varies across bonds, but we will empirically estimate and adjust for that later. We use the bond-CRSP link file from Wharton Research Data Services (WRDS) to match our data to excess stock returns in CRSP. We winsorize stock returns at the (0.5%, 99.5%) levels to reduce the impact of outliers.

---

$^8$We also verify that corporate bond-based repo transactions are not contained in the TRACE dataset. Otherwise, this would impact the subsequent analysis because the reversing repo trade will happen at predetermined prices, and thus mechanically be “slow to respond to information.” See FINRA, “Trade Reporting Frequently Asked Questions” Question 106.6, accessible on the FINRA website.

$^9$We construct duration-matched Treasury portfolios as follows. For each week $t$, let $\text{Dur}_{i,t-1}$ denote the modified duration of bond $i$ at the end of the previous week. We first find the two benchmark treasuries with nearest duration $\text{TrsyDur}_{i,t-1}^L$ and $\text{TrsyDur}_{i,t-1}^H$ such that $\text{TrsyDur}_{i,t-1}^L \leq \text{Dur}_{i,t-1} \leq \text{TrsyDur}_{i,t-1}^H$. We then form a Treasury portfolio with weights of $\left( w_{i,t-1}^{L} = \frac{\text{TrsyDur}_{i,t-1}^H - \text{Dur}_{i,t-1}}{\text{TrsyDur}_{i,t-1}^H - \text{TrsyDur}_{i,t-1}^L}, w_{i,t-1}^{H} = \frac{\text{Dur}_{i,t-1} - \text{TrsyDur}_{i,t-1}^L}{\text{TrsyDur}_{i,t-1}^H - \text{TrsyDur}_{i,t-1}^L} \right)$ so that the value-weighted duration of the Treasury portfolio matches the bond. We use benchmark Treasury maturities of 1, 2, 5, 7, 10, 20, and 30 years.
Using these returns as shock proxies relies on two facts. First, weekly stock and Treasury returns are essentially unpredictable so they are close to pure innovations. Second, stock and Treasury prices are much faster than corporate bonds to respond to information. For treasuries, Fleming and Remolona (1999) document that U.S. treasuries react immediately to macroeconomic announcements. For stocks, existing literature generally finds that equity prices reflect information faster than corporate bonds (e.g., Kwan (1996), Gebhardt, Hvidkjaer, and Swaminathan (2005), and Downing, Underwood, and Xing (2009)).

Our later analysis also confirms that Treasury and stock returns lead corporate bond returns. Admittedly, these shock proxies also incorporate other risks that can impact corporate bond pricing. In addition to pure interest rate movements, U.S. treasuries also respond to macroeconomic news, such as economy-wide changes in funding liquidity, which can impact the default risk of corporations. Similarly, stock return movements may reflect changes in investor risk preferences that also impact bond pricing. However, we believe interest rate movements and default risk movements are still the main drivers of our shock variables.

3.1.3 Summary statistics

After requiring stock and duration-matched Treasury returns, our sample contains 1,432,062 bond-week observations with 13,655 unique bonds issued by 1,736 companies. Summary statistics are in Table 1. Because we only calculate weekly returns when a bond trades in two successive weeks, more liquid bonds are more likely to enter into the sample. Our sample captures 74% of trading volume of the U.S. corporate bond market in the corresponding period.

3.2 Measuring payoff-relevance of risks

Our model predicts that price underreaction depends upon the payoff-relevance of risks. We use the model to guide our measure of payoff-relevance.

In our model, payoff-relevance of a risk is defined by the volatility of bond returns explained by that risk. To measure this empirically, for each bond \(i\) with at least 26 weeks (half year) of data, we use time-series regressions to decompose bond returns using stock

\[\text{Volatility of bond returns explained by risk} = \text{Regression coefficient} \times \text{Risk factor} \]

One exception in the literature is Even-Tov (2017) who finds that, for non-investment grade companies, bond price response to earnings events have predictive power over post-earnings announcement stock drifts. In robustness checks, our results are unaffected if we subtract the market-wide component from stock return shocks. This implies that our results are driven by the idiosyncratic, company-level risk in stock returns.
<table>
<thead>
<tr>
<th>Year</th>
<th>Bond-weeks</th>
<th>Bonds</th>
<th>Companies</th>
<th>Credit spread (%)</th>
<th>Duration (years)</th>
<th>Issuance size ($mil)</th>
<th>Weekly volume ($mil)</th>
<th>Trades per week</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>41,094</td>
<td>3,586</td>
<td>733</td>
<td>6.25</td>
<td>4.89</td>
<td>541</td>
<td>30.5</td>
<td>15.4</td>
</tr>
<tr>
<td>2003</td>
<td>92,994</td>
<td>4,282</td>
<td>814</td>
<td>5.06</td>
<td>5.25</td>
<td>562</td>
<td>29.2</td>
<td>14.9</td>
</tr>
<tr>
<td>2004</td>
<td>96,607</td>
<td>4,286</td>
<td>862</td>
<td>4.83</td>
<td>5.31</td>
<td>586</td>
<td>28.6</td>
<td>12.6</td>
</tr>
<tr>
<td>2005</td>
<td>90,847</td>
<td>4,172</td>
<td>880</td>
<td>5.38</td>
<td>5.32</td>
<td>614</td>
<td>26.8</td>
<td>12.5</td>
</tr>
<tr>
<td>2006</td>
<td>91,365</td>
<td>4,197</td>
<td>902</td>
<td>6.09</td>
<td>5.34</td>
<td>632</td>
<td>23.7</td>
<td>11.0</td>
</tr>
<tr>
<td>2007</td>
<td>82,631</td>
<td>4,054</td>
<td>876</td>
<td>6.16</td>
<td>5.46</td>
<td>700</td>
<td>23.7</td>
<td>10.9</td>
</tr>
<tr>
<td>2008</td>
<td>82,119</td>
<td>4,047</td>
<td>805</td>
<td>7.09</td>
<td>5.34</td>
<td>789</td>
<td>22.5</td>
<td>13.4</td>
</tr>
<tr>
<td>2009</td>
<td>103,750</td>
<td>4,376</td>
<td>853</td>
<td>6.46</td>
<td>5.44</td>
<td>802</td>
<td>26.0</td>
<td>17.1</td>
</tr>
<tr>
<td>2010</td>
<td>124,783</td>
<td>4,802</td>
<td>956</td>
<td>4.44</td>
<td>5.71</td>
<td>816</td>
<td>24.0</td>
<td>15.2</td>
</tr>
<tr>
<td>2011</td>
<td>135,926</td>
<td>5,113</td>
<td>1,025</td>
<td>4.17</td>
<td>5.91</td>
<td>804</td>
<td>20.8</td>
<td>14.1</td>
</tr>
<tr>
<td>2012</td>
<td>145,270</td>
<td>5,605</td>
<td>1,067</td>
<td>3.40</td>
<td>6.13</td>
<td>811</td>
<td>26.0</td>
<td>16.2</td>
</tr>
<tr>
<td>2013</td>
<td>167,568</td>
<td>6,068</td>
<td>1,091</td>
<td>3.37</td>
<td>6.22</td>
<td>812</td>
<td>20.2</td>
<td>14.8</td>
</tr>
<tr>
<td>2014</td>
<td>177,108</td>
<td>6,422</td>
<td>1,124</td>
<td>3.32</td>
<td>6.36</td>
<td>840</td>
<td>19.6</td>
<td>13.8</td>
</tr>
</tbody>
</table>

Table 1. Summary statistics of the corporate bond sample. Corporate bond data comes from enhanced TRACE.

and Treasury returns:

\[
\text{BondRet}_{i,t} = \alpha_i + \beta_{i,0}^{\text{stock}} \cdot \text{StockRet}_{i,t} + \beta_{i,1}^{\text{stock}} \cdot \text{StockRet}_{i,t-1} + \ldots + \beta_{i,8}^{\text{stock}} \cdot \text{StockRet}_{i,t-8} \\
+ \beta_{i,0}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t} + \beta_{i,1}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t-1} + \ldots + \beta_{i,8}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t-8} + \mu_{i,t} \quad (9)
\]

We include lags to accommodate slow price response to information. Our results are not sensitive to using 2, 4, or 12 lags instead of 8. We then compute the sample volatility of each estimated component and use the results as payoff-relevance proxies:

\[
\sigma_i^{\text{stock}} = \text{StandardDeviation} \left( \hat{\beta}_{i,0}^{\text{stock}} \cdot \text{StockRet}_{i,t} + \ldots + \hat{\beta}_{i,8}^{\text{stock}} \cdot \text{StockRet}_{i,t-8} \right) \quad (10)
\]

\[
\sigma_i^{\text{trsy}} = \text{StandardDeviation} \left( \hat{\beta}_{i,0}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t} + \ldots + \hat{\beta}_{i,8}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t-8} \right) \quad (11)
\]

where coefficients \( \{ \hat{\beta}_{i,0}^{\text{stock}}, \ldots, \hat{\beta}_{i,8}^{\text{stock}}, \hat{\beta}_{i,0}^{\text{trsy}}, \ldots, \hat{\beta}_{i,8}^{\text{trsy}} \} \) are the point estimates in regression (9).

We use \( \sigma_i^{\text{stock}} \) and \( \sigma_i^{\text{trsy}} \) as proxies for payoff-relevance of default and interest rate risks for bond \( i \), respectively. As expected, \( \sigma_i^{\text{trsy}} \) varies with bond duration and \( \sigma_i^{\text{stock}} \) varies with credit quality. To see this, we sort bonds by these two payoff-relevance measures using common breakpoints into 5 \( \times \) 5 portfolios.\(^{12}\) Panel B in Table 2 shows the average duration and credit spreads for those portfolios. Bonds in the bottom \( \sigma_i^{\text{trsy}} \) quintile have an average duration in the 2.28 to 4.46 year range, and that rises monotonically to 7.42 - 13.42 years in the top quintile. Similarly, bonds with higher \( \sigma_i^{\text{stock}} \) have higher credit spreads.

\(^{12}\)That is, the break-points are the 20\%, 40\%, 60\%, and 80\% percentiles of all the payoff-relevance measures \( \{ \sigma_1^{\text{stock}}, \ldots, \sigma_n^{\text{stock}}, \sigma_1^{\text{trsy}}, \ldots, \sigma_n^{\text{trsy}} \} \) where \( n \) is the number of bonds.
We chose our payoff-relevance measures to most closely align with the model. Another benefit of using these measures is that they can be interpreted quantitatively, so we can later control for payoff-relevance levels in some tests. However, in subsequent exercises, we also verify that our results are robust to using duration as the proxy for interest rate relevance and using credit ratings or credit spreads as the proxy for default risk relevance.

### Panel A: Payoff-relevance proxies

| Relevance of interest rate risk ($\sigma^{\text{trsy}}$) | Relevance of default risk ($\sigma^{\text{stock}}$) |
|---|---|---|---|
| quintile | Low | 2 | 3 | 4 | High | quintile | Low | 2 | 3 | 4 | High |
| Low | 1.24% | 2.17% | 3.23% | 4.27% | 7.23% | Low | 0.83% | 1.03% | 1.25% | 1.28% | 1.39% |
| 2 | 1.40% | 2.23% | 3.26% | 4.42% | 6.88% | 2 | 2.19% | 2.19% | 2.20% | 2.22% | 2.27% |
| 3 | 1.41% | 2.26% | 3.27% | 4.58% | 7.02% | 3 | 3.24% | 3.21% | 3.20% | 3.26% | 3.24% |
| 4 | 1.44% | 2.28% | 3.25% | 4.64% | 6.94% | 4 | 4.70% | 4.56% | 4.64% | 4.64% | 4.71% |
| High | 1.36% | 2.34% | 3.29% | 4.65% | 8.40% | High | 7.52% | 7.63% | 8.30% | 8.29% | 10.17% |

### Panel B: Bond characteristics

<table>
<thead>
<tr>
<th>Duration (years)</th>
<th>Yield spread to treasuries</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^{\text{stock}}$</td>
<td>$\sigma^{\text{trsy}}$</td>
</tr>
<tr>
<td>quintile</td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>2.28</td>
</tr>
<tr>
<td>2</td>
<td>3.67</td>
</tr>
<tr>
<td>3</td>
<td>4.06</td>
</tr>
<tr>
<td>4</td>
<td>4.46</td>
</tr>
<tr>
<td>High</td>
<td>3.53</td>
</tr>
</tbody>
</table>

Table 2. Characteristics of bonds by payoff-relevance of risks. Bonds are sorted into quintiles by the payoff-relevance of default risk ($\sigma^{\text{stock}}$) and interest rate risk ($\sigma^{\text{trsy}}$). These payoff-relevance measures, shown in Panel A, are computed as the (annualized) volatility of bond returns explained by stock and Treasury returns, respectively. Panel B shows that bonds with higher $\sigma^{\text{trsy}}$ have longer duration and those with higher $\sigma^{\text{stock}}$ have larger credit spreads.

### 3.3 Measuring price underreaction to different risks

We now measure price underreaction, defined as the fraction of price response that do not happen immediately in our model. To do this, we need to first define full price reaction, which we measure using the response of long horizon bond returns. To determine the appropriate horizon, we run the following panel regression of weekly bond returns on current

---

13 We do not use structural models to estimate full price reaction because they do not always generate reliable bond price sensitivity to risks. For instance, Schaefer and Strebulaev (2008) find that structural models of corporate bonds are bad at explaining how bond prices respond to interest rate shocks in practice.
and lagged shocks on the whole sample:

\[
\text{BondRet}_{i,t} = \alpha + \beta_0^{stock} \cdot \text{StockRet}_{i,t} + \ldots + \beta_7^{stock} \cdot \text{StockRet}_{i,t-12} + \beta_0^{trsy} \cdot \text{TrsyRet}_{i,t} + \ldots + \beta_7^{trsy} \cdot \text{TrsyRet}_{i,t-12} + \sum_i \gamma_i \cdot 1_{\text{bond } i} + u_{i,t}
\]  

(12)

where we use bond dummies BondFE to absorb fixed average return differences across bonds (riskier bonds have higher average returns). We cluster standard errors by week and bond.

Figure 4 plots the rolling sums of regression coefficients to trace out the cumulative price response to shocks. For instance, the cumulative three-week price response to stock returns is measured by $\beta_0^{stock} + \beta_1^{stock} + \beta_2^{stock}$. The immediate reaction coefficients $\beta_0^{stock}$ and $\beta_0^{trsy}$ are marked by triangles. There is clear evidence of slow price reaction to both types of shocks. In response to one unit of stock return, we observe cumulative bond return responses of 0.052 in one week, 0.089 by two weeks, until converging to around 0.123. In response to one unit of duration-matched Treasury return, bond return responds by 0.469 in one week and converges to around 0.65 a few weeks later.

After inspecting Figure 4, we choose to use eight weeks to measure full responses, because responses to both shocks appear to have converged by eight weeks. Note that we face a bias-variance trade-off here: using a longer horizon is better for capturing complete responses but increases estimation error.

In subsequent work, we will use one-week and eight-week bond returns to estimate immediate and full price responses, and then calculate price underreaction as one minus the ratio between them. For instance, using coefficients from the full sample panel regression (12), price underreaction to risks will be defined as:

\[
\text{Underreaction}^{stock} = 1 - \frac{\beta_0^{stock}}{\beta_0^{stock} + \ldots + \beta_7^{stock}},
\]

(13)

\[
\text{Underreaction}^{trsy} = 1 - \frac{\beta_0^{trsy}}{\beta_0^{trsy} + \ldots + \beta_7^{trsy}}.
\]

(14)

The point estimates are Underreaction$^{stock} = 57\%$ and Underreaction$^{trsy} = 27\%$, and standard errors calculated from delta method are 2.7% and 7.1%, respectively.

While we can use alternative proxies such as duration and credit spreads for payoff-relevance, because they are slow-moving, we cannot use them to measure the speed of price responses.

---

14 The fact that bond prices do not respond one-to-one with duration-matched Treasury returns is consistent with the long-standing finding that conventional bond duration measures overestimate interest rate sensitivity (Longstaff and Schwartz (1995), Schaefer and Strebulaev (2008)).
Figure 4. Cumulative response of corporate bond prices to shocks. The immediate response to shocks are marked as triangles. Because these graphs show that eight weeks is approximately enough for shocks to be fully incorporated into prices (marked by squares), we use eight weeks to measure full price responses in subsequent work. The responses are obtained by regressing weekly corporate bond returns on current and lagged stock and Treasury returns (equation (12)), and then adding up cumulative coefficients. For example, the $h$-week response to stock returns is obtained as $\beta_0^{\text{stock}} + ... + \beta_h^{\text{stock}}$. The immediate reaction coefficients $\beta_0^{\text{stock}}, \beta_0^{\text{trsy}}$ are marked using triangles. Dotted lines are two standard error bounds and standard errors are clustered by week and bond.

4 Testing theory predictions

We now test the two theory predictions in section 2. As a reminder, if one risk is more payoff-relevant, than the bond will:

1. respond to that risk more quickly, \textit{("main effect")}
2. respond to the other risk less quickly. \textit{("distraction effect")}

We first conduct tests at the portfolio level and then at the bond level. Each test has its own benefits. The portfolio-level test requires weaker parametric assumptions. The bond-level test imposes stronger parametric assumptions but has more power. The latter test also allows us to control for bond liquidity and test how underreaction changes over time.

4.1 Portfolio-level tests

Similar to the exercise in Table 2, we first measure payoff-relevance $\sigma_i^{\text{stock}}$ and $\sigma_i^{\text{trsy}}$ for each bond. We then double sort bonds into $5 \times 5$ portfolios by $\sigma_i^{\text{stock}}$ and $\sigma_i^{\text{trsy}}$ using common
quintile break-points. For bonds in each of the 25 portfolios, we run a panel regression of bond return response to shocks:

\[
\text{BondRet}_{i,t} = \alpha + \beta_{0}^{\text{stock}} \cdot \text{StockRet}_{i,t} + \ldots + \beta_{7}^{\text{stock}} \cdot \text{StockRet}_{i,t-7} \\
+ \beta_{0}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t} + \ldots + \beta_{7}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t-7} + \sum_{i} \eta_{i} \cdot 1_{\text{bond } i} + u_{i,t},
\]

and then calculate price underreaction using equations (13) and (14). We use bond fixed effects to absorb average return differences across bonds. The average underreaction estimates for the 25 portfolios are shown in Table 3.

<table>
<thead>
<tr>
<th>Panel A. Underreaction to stock returns</th>
<th>Panel B. Underreaction to Treasury returns</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>(\sigma^{\text{stock}}) quintile</th>
<th>(\sigma^{\text{trsy}}) quintile</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
<th>High minus low</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td>70%</td>
<td>71%</td>
<td>68%</td>
<td>66%</td>
<td>74%</td>
<td>4%</td>
<td>(0.48)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>56%</td>
<td>68%</td>
<td>68%</td>
<td>66%</td>
<td>65%</td>
<td>8%</td>
<td>(0.88)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>57%</td>
<td>60%</td>
<td>63%</td>
<td>65%</td>
<td>68%</td>
<td>11%</td>
<td>(1.34)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>58%</td>
<td>61%</td>
<td>64%</td>
<td>60%</td>
<td>64%</td>
<td>6%</td>
<td>(0.81)</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>47%</td>
<td>53%</td>
<td>49%</td>
<td>55%</td>
<td>56%</td>
<td>8%</td>
<td>(1.50)</td>
</tr>
<tr>
<td>High minus low</td>
<td></td>
<td>-22%***</td>
<td>-19%***</td>
<td>-19%**</td>
<td>-11%</td>
<td>-18%**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td></td>
<td>(-3.00)</td>
<td>(-2.62)</td>
<td>(-2.34)</td>
<td>(-1.39)</td>
<td>(-2.51)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\sigma^{\text{stock}}) quintile</th>
<th>(\sigma^{\text{trsy}}) quintile</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
<th>High minus low</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td>35%</td>
<td>31%</td>
<td>27%</td>
<td>24%</td>
<td>18%</td>
<td>-16%</td>
<td>(-1.56)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>56%</td>
<td>42%</td>
<td>29%</td>
<td>28%</td>
<td>18%</td>
<td>-38%***</td>
<td>(-3.25)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>58%</td>
<td>38%</td>
<td>35%</td>
<td>31%</td>
<td>24%</td>
<td>-34%**</td>
<td>(-2.57)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>52%</td>
<td>44%</td>
<td>37%</td>
<td>29%</td>
<td>22%</td>
<td>-30%*</td>
<td>(-1.78)</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>76%</td>
<td>70%</td>
<td>42%</td>
<td>34%</td>
<td>20%</td>
<td>-55%***</td>
<td>(-3.64)</td>
</tr>
<tr>
<td>High minus low</td>
<td></td>
<td>41%***</td>
<td>39%***</td>
<td>15%</td>
<td>10%</td>
<td>2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td></td>
<td>(3.02)</td>
<td>(4.13)</td>
<td>(1.32)</td>
<td>(0.99)</td>
<td>(0.15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Price underreaction of bond portfolios with different risk exposures. Bonds are double sorted into 5 \(\times\) 5 portfolios by \(\sigma^{\text{stock}}\) and \(\sigma^{\text{trsy}}\), the payoff-relevance proxies, using common breakpoints. For each of the 25 portfolios, we then estimate the bond price underreaction to stock returns (Panel A) and Treasury returns (Panel B). Price underreaction is defined as the fraction of price response that happens in the delay window, i.e., from week \(t + 1\) to \(t + 7\), in response to information in week \(t\). Stars *, **, and *** represent statistical significance at 10%, 5%, and 1% levels, respectively.

There is strong evidence for the main effect. In Panel A of Table 3 for each \(\sigma^{\text{trsy}}\) quintile,

\(^{15}\)Figure 14 in Appendix C shows that the distribution of \(\sigma^{\text{stock}}\) and \(\sigma^{\text{trsy}}\) is similar, so using common break-points results in fairly balanced quintiles for both.
bonds with higher $\sigma^{stock}$ have monotonically decreasing price underreaction to stock returns. The differences between the top and bottom $\sigma^{stock}$ quintiles range from 11% to 22% and are statistically significant at the 5% level in four out of five cases. Similarly, in Panel B, for each $\sigma^{stock}$ quintile, bonds with higher $\sigma^{trsy}$ are faster to respond to Treasury returns. The difference between the top and bottom $\sigma^{trsy}$ quintiles range from 16% to 55%, and the difference is statistically significant in four out of five cases.

There is also evidence for the distraction effect. For each $\sigma^{stock}$ quintile in Panel A, higher $\sigma^{trsy}$ is associated with higher underreaction to stock returns. Although the differences (“high minus low” column) are not statistically significant, they all point in the predicted direction. Panel B shows that bonds with higher $\sigma^{stock}$ have larger underreaction to Treasury returns, and the effect is statistically significant for the bottom two quintiles of $\sigma^{trsy}$.

The magnitudes of distraction effect appear weaker than the main effect, and this suggests that only a fraction of investors face attention constraints. In Appendix A.3, we show that if all investors face binding constraints, then the two effects would be equally large$^{16}$ Because we use common breakpoints in sorting the two dimensions, the difference between top and bottom quintiles of $\sigma^{stock}$ is similar to the difference between top and bottom quintiles of $\sigma^{trsy}$. Therefore it is meaningful to quantitatively compare how price underreaction varies with the quintile number of $\sigma^{stock}$ and $\sigma^{trsy}$.

Bond-level liquidity differences cannot explain away our findings on explaining price underreaction variation at the risk level. For instance, Figure 1 in the introduction plots the price response paths for the two corner cases in the 25 portfolios. The black-line portfolio contains bonds with top-quintile $\sigma^{trsy}$ and bottom-quintile $\sigma^{stock}$, while the blue-line portfolio has the exact opposite risk profile (bottom-quintile $\sigma^{trsy}$ and top-quintile $\sigma^{stock}$). Relative to the other portfolio, the black-line portfolio underreacts less to Treasury returns but more to stock returns. This cannot be explained by liquidity differences: if the black-line portfolio is more liquid, it should respond faster to both types of risks, not faster to one risk and slower to the other.

4.2 Bond-level tests

We now measure price underreactions at a more granular level and conduct more powerful tests. Because firm credit quality and bond duration change over time, the payoff-relevance of

---

16 Precisely speaking, if all investors face attention constraints, then $\frac{d\text{Underreaction}^{stock}}{d\sigma^{trsy}} = -\frac{d\text{Underreaction}^{stock}}{d\sigma^{stock}}$ when $\sigma^{stock} = \sigma^{trsy}$. Therefore, the appropriate comparison requires looking at cases where the two risks have similar payoff-relevance. As an example, in Panel A of Table 3 we should compare each $n^{th}$ ($n = 1,...,5$) column in the “high minus low” row (main effect) with the $n^{th}$ row in the “high minus low” column (distraction effect). The proof is given in a two-risk case of our model in Appendix A.3.

---
risks also changes over time. If investors are sophisticated, they should also adjust attention allocation accordingly. To test this, we separate the data into roughly equal-length periods \( p \in \{2002 - 2006, 2007 - 2010, 2011 - 2014\} \) and test model predictions at the bond-period level.

We measure payoff-relevance of risks, \( \sigma_{i,p}^{stock} \) and \( \sigma_{i,p}^{trsy} \), for each bond \( i \) in each period \( p \). To measure price underreaction, we cannot use the method in the portfolio-tests because there is not enough data for each bond-period. We thus need to impose stronger parametric structure. For each bond-period, we use generalized method of moments (GMM) to estimate two equations:

\[
\text{BondRet}_{i,p,t} = (1 - \text{Underreaction}_{i,p}^{stock}) \cdot b_{i,p}^{stock} \cdot \text{StockRet}_{i,p,t} \\
+ (1 - \text{Underreaction}_{i,p}^{trsy}) \cdot b_{i,p}^{trsy} \cdot \text{TrsyRet}_{i,p,t} + u_{i,p,t} \tag{16}
\]

\[
\text{BondRet}_{i,p,t\rightarrow t+7} = b_{i,p}^{stock} \cdot \text{StockRet}_{i,p,t} + b_{i,p}^{trsy} \cdot \text{TrsyRet}_{i,p,t} + u_{i,p,t\rightarrow t+7} \tag{17}
\]

where we jointly estimate four parameters: \( \text{Underreaction}_{i,p}^{stock}, \text{Underreaction}_{i,p}^{trsy}, b_{i,p}^{stock}, \) and \( b_{i,p}^{trsy} \). The additional restriction we impose is that the underreaction parameters are constrained to be in \([0, 1]\). The other two parameters, \( b_{i,p}^{stock} \) and \( b_{i,p}^{trsy} \), are estimates of the full price reaction to shocks after eight weeks. In both equations, all left- and right-hand-side variables are demeaned so we do not need intercepts. As there are two different right-hand-side variables, there are two moments for each equation so the GMM is exactly identified.

There are two reasons why we chose to estimate (16) and (17) jointly using GMM, rather than separately with two OLS regressions and then taking the ratio of regression coefficients. First, GMM makes it convenient to impose the restriction that underreaction needs to be between 0 and 1. Second, GMM estimation allows for direction estimation of the standard error of the underreaction parameter. We will use that standard error as an estimation quality filter, to be explained below.

We conduct the GMM estimation for each bond-period having at least 26 weeks of observations. Because of the small sample size per bond-period, the underreaction parameters are sometimes estimated with sizable errors. To screen out bad estimates, we apply two mild data filters. First, we require the estimated full price responses (\( b_{i,p}^{stock} \) or \( b_{i,p}^{trsy} \)) to be positive.\(^{17}\) Second, we require the standard error of underreaction estimates to be below 1.\(^{18}\)

\(^{17}\)Structural models of corporate bonds indicate positive relationships, so negative estimates likely arise due to small sample randomness.

\(^{18}\)The reader may wonder how the standard error can be above 1. This is because, as is explained in more detail in Appendix C.3, we restrict these underreaction coefficients to be between 0 and 1 by a change of variable to \( \eta_{i,p}^{risk} \) defined so that \( \text{Underreaction}_{i,p}^{risk} = 1/(1 + e^{-\eta_{i,p}^{risk}}) \) is a logistic transform. We then estimate standard errors for \( \eta_{i,p}^{risk} \) and derive standard errors for Underreaction\(^{risk}\) through the delta method, which can sometimes be above 1.
After applying these filters, we are left with 7,480 bonds with Underreaction\textsuperscript{stock}\textsubscript{i,p} estimates and 9,014 bonds with valid Underreaction\textsuperscript{trsy}\textsubscript{i,p} estimates. In Appendix C.1, we explain our GMM estimation and data filters in more detail. As a check on estimation precision, we also confirm that the GMM-based estimates are similar to the earlier regression-based estimates.

We now test model predictions by regressing price underreaction to each risk on the payoff-relevance of the risk and also the other risk. We pool together estimates of underreactions to stock and Treasury returns in panel regression:

$$\text{Underreaction}_{i,p}^{\text{risk}} = \alpha + \sum_{q=1}^{5} \gamma_q^{\text{risk}} \cdot 1_{\sigma_{i,p}^{\text{risk}} \text{ in quintile } q} + \sum_{q=1}^{5} \gamma_q^{\text{other risk}} \cdot 1_{\sigma_{i,p}^{\text{other risk}} \text{ in quintile } q} + \beta \cdot 1_{\text{risk=Treasury}} + \text{LiquidityControls}_{i,p} + \delta_p \cdot 1_{\text{period } p} + \epsilon_{i,p}^{\text{risk}}$$

(18)

where when “risk” is stock, the “other risk” refers to Treasury, and vice versa. Because the relationship can be nonlinear, similar to the portfolio-test, we sort payoff-relevance measures into quintiles using common break-points and regress on the quintile indicators.

To test our theory, we are interested in the three sets of coefficients marked with underbraces in the regression. We add a number of bond liquidity controls: average bid-ask spread, log weekly trading volume, log weekly trading frequency, and log offering size. To facilitate interpreting coefficients, the liquidity proxies are standardized to unit variance. We also add period fixed effects. Standard errors are double clustered by bond and period. The results are shown in Table 4, where column 1 is the main specification and columns 2 to 4 are robustness checks.

There are four take-aways. First, consistent with the portfolio-test results, we find clear evidence for both the main effect and distraction effect, and the former is slightly stronger than the latter. Underreaction to a risk decreases by 31.89% from the bottom to top quintile of $\sigma_{i,p}^{\text{risk}}$ (main effect) and by 30.75% from the bottom to top quintile of $\sigma_{i,p}^{\text{other risk}}$ (distraction effect). The size of these two effects are very similar, consistent with most investors facing attention constraints. These estimated effect magnitudes are broadly similar to the estimates in the portfolio-test.

Second, we also find evidence consistent with investors changing attention allocation over time in response to bond risk exposure changes. Column 5 adds bond fixed effects to exploit cases where a bond changes payoff-relevance quintile assignment across periods.\footnote{The drop in $R^2$ for this specification is because we use marginal $R^2$ which adjusts for the number of regressors. There are many bond dummies being added.} In total,
### Table 4. Explaining price underreaction using payoff-relevance of risks.

Price underreaction to stock and Treasury returns are estimated for each bond in each period using GMM. The regression pools underreaction to both risks together. In the variables, when “risk” refers to stock, “other risk” refers to Treasury, and vice versa. The first regressor is a fixed effect that estimates the average difference between underreaction to Treasury and underreaction to stock. The next eight regressors are payoff-relevance measures for “risk” and “other risk,” turned into quintile indicator variables to capture possible nonlinearity. The next four regressors are liquidity proxies, and they are standardized to facilitate interpreting coefficients. The regression specification in column (4) weights data by bond issuance sizes and specification (5) includes bond fixed effects. Standard errors are clustered by bond and period. Coefficients significant at the 10%, 5%, and 1% level are noted with *, **, and ***, respectively.
there are 1,231 cases of bonds changing $\sigma_{stock}$ quintile assignment and 1,142 cases for $\sigma^{trsy}$

Both main and distraction effects are robust to adding bond fixed effects.

Third, across all specifications, we find that underreaction to Treasury is lower than underreaction to stock by 14 - 20%. This is consistent with lower attention cost in learning about interest rate risk.

Finally, while bond liquidity proxies do have explanatory power, the risk-level variables explain price underreaction much better. Although the $R^2$ measures are significantly downward biased because of estimation errors in the left-hand-side variables, we can still compare the marginal $R^2$ of different variables. The four bond level liquidity proxies jointly explain marginal $R^2$ of 1.98%, while the payoff-relevance indicator variables $1_{\sigma_{risk}}$, $1_{\sigma_{other\ risk}}$ explain 6.47%, which is more than 3 times as much. If we add the $1_{\sigma_{risk}} = Treasury$ indicator, the total explanatory power of risk-level variables increases to 15.31%.

4.3 Trading volume around earnings events

To further test the attention mechanism, we use trading activity as a proxy for investor attention and test whether investors pay more attention to a risk when it is more payoff-relevant. Trading volume has often been used in prior literature as a proxy for inattention (e.g., Hou, Xiong, and Peng (2009)). To investigate investor attention allocation to default risk shocks, we study corporate bond trading activity around company quarterly earning events because they are concentrated releases of default-relevant information.

For each bond, we calculate the weekly trading turnover (trading volume/ issuance size) and trading frequency using TRACE. Because earnings releases are company-level events, we aggregate both trading activity measures and payoff-relevance proxies at the company level, weighting bonds issued by the same company using issuance sizes. We sort companies into quintiles by $\sigma_{i,p}^{stock}$, our proxy for default risk relevance. We obtain quarterly earnings dates from Compustat. The merged data has a total of 35,335 earnings events from 1,411 companies.

We use an event window of six weeks prior to six weeks after each announcement, as six weeks is roughly midway between two successive quarterly announcements. For companies

---

20 Table 8 in Appendix C.2 shows the assignment migration patterns. As expected, most of these cases see decreasing $\sigma^{trsy}$ as bond duration decreases over time. There is no clear upward or downward trend for $\sigma^{stock}$ changes.
in each default exposure quintile, we run panel regressions:

\[
\log(Y_{k,t}) = \alpha + \sum_{\tau=-6}^{6} \beta_{\tau} \cdot 1_{\text{event week } \tau} + \sum_{k} \gamma_{k}^{\text{company}} \cdot 1_{\text{company } k} + \sum_{t} \gamma_{t}^{\text{week}} \cdot 1_{\text{week } t} + \epsilon_{k,t}
\]  

(19)

where the left-hand side variable \( Y \) is either turnover or trading frequency. Company and week are denoted by \( k \) and \( t \). We include company and week fixed effects to absorb fixed trading activity differences across companies and over time. Standard errors are clustered by calendar week and company. Figure 5 plots the event-week fixed effects.

**Figure 5. Event study of bond trading activity around earnings announcements.** Companies are sorted into quintiles by \( \sigma^{\text{stock}} \), our proxy for the payoff-relevance of default risk. We then plot the average log trading turnover (left) and log trading frequency (right) for selected quintiles, after controlling for company and calendar week fixed effects. The standard errors are clustered by company and calendar week. To avoid cluttering up the graph, two standard error bands are only plotted for the first quintile.

Consistent with our theory, companies with large default risk exposure experience more significant spikes in trading activity in the event week. For companies with the top quintile default risk exposure, log turnover spikes by 17% on earnings weeks relative to the baseline (six weeks prior to the event), while the bottom quintile only spikes by 7.4%. The difference is statistically significant with a p-value of 0.005. When using log trading frequency as the activity measure, the top quintile spikes by 13.7% relative to 1.5% for the bottom quintile, a difference that is statistically significant with a p-value of \( 5.8 \times 10^{-7} \).

Unfortunately, we do not have pure interest rate information release events to test the differential trading activity response when sorting on \( \sigma^{\text{trsy}} \). The closest candidate events are probably the open market operation announcements during which the Federal Reserve
communicates interest rate policies. However, a number of papers have shown that these
announcements also move the stock markets significantly (Savor and Wilson (2013), Lucca
and Moench (2015)), indicating that they likely contain information relevant for economy-
wide credit risk as well.

5 Alternative explanations

In this section, we show that transaction costs cannot explain our findings. We also
consider other alternative explanations.

5.1 The transaction cost explanation

One may wonder if our results can be due to transaction costs. Suppose that investors
face trading costs due to, for instance, bid-ask spreads. They will then only trade when the
bond value has changed sufficiently to justify the transaction cost. Because trading activity
speeds up price responses, this could generate our “main effect”: more payoff-relevant risks
tend to have larger shocks, so the average price response would be faster.

Figure 6 explains this for a hypothetical bond with high default risk exposure and low
interest rate risk exposure. The shaded areas represent regions where bond value movement
is larger than the transaction cost, and therefore the bond will experience fast price response.
Notice that when default shock realizations are large, the total bond value change also tends
to be large, so the price response tends to be fast. Due to this correlation between default
risk shock size and response speed regime, we would expect this bond to, on average, respond
to default shocks more quickly.

We explicitly model this alternative explanation in Appendix A.4, and use that to derive
two tests to test it against our inattention explanation. It is worth clarifying that we are not
testing whether transaction cost exists. We are only testing whether transaction costs can
explain our findings of risk-level variation of underreaction.

5.2 Test 1: Controlling for shock sizes

Under the transaction cost-based explanation, price underreaction is determined by the
size of \textit{ex post} realization of bond value shock. We show this formally in Appendix A.4. It is
worth noting that it is the total shock that matters, not the individual shocks from each risk.
Intuitively, if the bond value increased by $10, it does not matter whether the composition
is $7 from interest rate risk and $3 from credit risk, or the reverse. In contrast, under our
Figure 6. Illustration of how transaction costs impact price underreaction. The concentric ellipses are contour lines of the joint distribution of interest rate and default shocks. When the bond value movement – the sum of the two shocks – is larger than a transaction cost (e.g. bid-ask spread), there is more trading activity and thus bond price adjusts to the new value quickly (shaded region). Otherwise, the response is slow (unshaded region).

In the inattention model, the determinant of price reaction speed is the \textit{ex ante} payoff-relevance of risks.

Guided by this, we estimate ex post bond value changes, control for them, and test whether ex ante payoff-relevance still matters. For each bond $i$ in each week $t$, we estimate the value change using

$$ValueChange_{i,t} \triangleq \hat{b}_{i}^{stock} \cdot StockRet_{i,t} + \hat{b}_{i}^{trsy} \cdot TrsyRet_{i,t}$$

(20)

where coefficients $\hat{b}_{i}^{stock}$ and $\hat{b}_{i}^{trsy}$ are GMM estimates of bond-specific full price reaction to shocks. Because we are not accounting for other drivers of bond returns, this is an
underestimation of the total bond value movement.\footnote{Directly using realizations of BondRet_{i,t} or BondRet_{i,t-\tau+7} as proxies for bond value change – the explanatory variable in the subsequent exercise – is invalid. This is because these variables also affect measurement of price underreaction (based on 1 – BondRet_{i,t}/BondRet_{i,t-\tau+7}), the independent variable, thus creating a spurious correlation. Concretely, in periods where BondRet_{i,t} is small due to chance, or when BondRet_{i,t-\tau+7} is large due to chance, the underreaction estimate will be mechanically high.}

We split the sample by the shock size (|ValueChange_{i,t}|) into three bins: \{0 – 25bp, 25 – 75bp, > 75bp\}. This divides the sample into three roughly equal-sized parts. Then, for weekly observations of each bond \(i\) in each shock size bin \(s = \{1, 2, 3\}\), we use the GMM methodology in section 4.2 to estimate price underreactions. As before, we require at least 26 weeks of observations per (bond, shock size), and impose the data filters that 1) estimated price response coefficients are positive and 2) the standard error of underreactions is below one. After applying these filters, we end up with 4,863, 4,839, and 3,040 bonds in each of the three shock size bins, respectively.

If the transaction cost mechanism explains our findings, then we would expect payoff-relevance to lose explanatory power after controlling for shock sizes. To test this, for estimates in each shock size bin \(s = 1, 2, 3\), we run the regression

\[
\text{Underreaction}_{i,s} = \alpha_s + \sum_{q=1}^{5} \gamma_{s,q}^{\text{risk}} \cdot 1_{\sigma_{i}^{\text{risk}} \text{ in quintile } q} + \sum_{q=1}^{5} \gamma_{s,q}^{\text{other risk}} \cdot 1_{\sigma_{i}^{\text{other risk}} \text{ in quintile } q} + \beta \cdot 1_{\text{risk=Treasury}} + \text{LiquidityControls}_i + \epsilon_{i,s} \tag{21}
\]

which is the same as the main regression (18) except for one difference: because we need to ensure there is enough data in each (bond, shock size bin) to estimate underreaction, we no longer separate by period. We measure payoff-relevance \(\sigma_i^{\text{stock}}, \sigma_i^{\text{tray}}\) at the bond level and cluster standard errors by bond.

Columns 2 to 4 in Table 5 report the regression results for \(s = 1, 2, 3\), respectively. For comparison, column 1 reports the result when we pool together estimates from all three bins. Our findings show up in all three bins: price underreaction decreases in \(\gamma_{\text{risk}}\) and increases in \(\gamma_{\text{other risk}}\). To visualize the estimates, the left and right panels of Figure 7 plot the average price underreaction as a function of \(\gamma_{\text{risk}}\) and \(\gamma_{\text{other risk}}\), respectively. While there is some evidence that overall price underreaction is weaker when shock sizes are larger, the impact of payoff-relevance is roughly the same within each shock size bin.

Admittedly, our ex post shock measure ValueChange_{i,t} contains noise, so it is still possible...
<table>
<thead>
<tr>
<th></th>
<th>Full sample</th>
<th>&lt; 25bp</th>
<th>25 – 75bp</th>
<th>&gt; 75bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>72.29***</td>
<td>71.68***</td>
<td>77.30***</td>
<td>69.09***</td>
</tr>
<tr>
<td></td>
<td>(1.34)</td>
<td>(2.06)</td>
<td>(2.19)</td>
<td>(3.02)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{risk in quintile 2}}}$</td>
<td>-8.09***</td>
<td>-6.96***</td>
<td>-8.40***</td>
<td>-5.14***</td>
</tr>
<tr>
<td></td>
<td>(1.35)</td>
<td>(1.91)</td>
<td>(2.26)</td>
<td>(3.05)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{risk in quintile 3}}}$</td>
<td>-12.61***</td>
<td>-9.09***</td>
<td>-15.08***</td>
<td>-10.88***</td>
</tr>
<tr>
<td></td>
<td>(1.25)</td>
<td>(1.91)</td>
<td>(2.07)</td>
<td>(2.78)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{risk in quintile 4}}}$</td>
<td>-18.84***</td>
<td>-13.37***</td>
<td>-22.68***</td>
<td>-18.43***</td>
</tr>
<tr>
<td></td>
<td>(1.23)</td>
<td>(1.96)</td>
<td>(2.03)</td>
<td>(2.71)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{risk in quintile 5}}}$</td>
<td>-24.54***</td>
<td>-9.51***</td>
<td>-23.07***</td>
<td>-29.95***</td>
</tr>
<tr>
<td></td>
<td>(1.23)</td>
<td>(2.27)</td>
<td>(2.12)</td>
<td>(2.70)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{other risk in quintile 2}}}$</td>
<td>5.16***</td>
<td>4.55***</td>
<td>3.45**</td>
<td>5.32***</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
<td>(1.70)</td>
<td>(1.50)</td>
<td>(1.52)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{other risk in quintile 3}}}$</td>
<td>7.35***</td>
<td>9.47***</td>
<td>3.15**</td>
<td>6.70***</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(1.77)</td>
<td>(1.66)</td>
<td>(1.53)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{other risk in quintile 4}}}$</td>
<td>11.54***</td>
<td>16.16***</td>
<td>7.33***</td>
<td>9.02***</td>
</tr>
<tr>
<td></td>
<td>(1.02)</td>
<td>(1.79)</td>
<td>(1.69)</td>
<td>(1.57)</td>
</tr>
<tr>
<td>$1_{\sigma_{\text{other risk in quintile 5}}}$</td>
<td>19.05***</td>
<td>22.75***</td>
<td>21.59***</td>
<td>14.18***</td>
</tr>
<tr>
<td></td>
<td>(1.16)</td>
<td>(2.14)</td>
<td>(1.91)</td>
<td>(1.77)</td>
</tr>
<tr>
<td>$1_{\text{risk = Treasury}}$</td>
<td>-19.72***</td>
<td>-26.30***</td>
<td>-22.36***</td>
<td>-11.56***</td>
</tr>
<tr>
<td></td>
<td>(0.69)</td>
<td>(1.26)</td>
<td>(1.17)</td>
<td>(1.00)</td>
</tr>
<tr>
<td>Liquidity Controls</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Obs</td>
<td>9,076</td>
<td>3,041</td>
<td>3,146</td>
<td>2,889</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.240</td>
<td>0.264</td>
<td>0.292</td>
<td>0.238</td>
</tr>
</tbody>
</table>

Table 5. Explaining price underreaction within each shock size bin. For each bond in each week, we first estimate how much its value has changed based on concurrent stock and Treasury movements. We then sort data using the absolute bond value change ("shock sizes") into three bins: {< 25 bp, 25 - 75bp, > 75bp}. For each bond in each shock size bin, we estimate price underreaction to stock and Treasury returns using GMM. The table reports the output from regressing price underreactions to risks on the payoff-relevance proxies and liquidity controls (regression [21]). Regression specification (1) uses underreaction in all three shock size bins, while specifications (2) to (4) only use estimates from one shock size bin. Coefficients significant at the 10%, 5%, and 1% level are noted with *, **, and ***, respectively.

that payoff-relevance appears to matter because it provides additional information about ex post shocks. In fact, the reverse is also true: shock sizes may also matter because they
Figure 7. Price underreaction for each shock size bin. We plot the average price underreaction when varying $\sigma^{\text{risk}}$ (left) or $(\sigma^{\text{other risk}})$. For example, when “risk” is stock, “other risk” is Treasury, and vice versa. The left graph plots how underreaction to a risk varies with the payoff-relevance of that risk (“main effect”), while the right graph plots the variation with the payoff-relevance of the other risk (“distraction effect”). Details: the plotted estimates come from regression (21), with the left graph plotting $\{\alpha_s, \alpha_s + \gamma_s^{\text{risk}}, ..., \alpha_s + \gamma_s^{\text{risk}}\}$, and the right graph plotting $\{\alpha_s, \alpha_s + \gamma_s^{\text{other risk}}, ..., \alpha_s + \gamma_s^{\text{other risk}}\}$. The dotted lines are 95% confidence intervals.

provides additional information about the variation of payoff-relevance over time. From this perspective, it is ultimately hard to differentiate the two explanatory variables in a “horse race” because each may proxy for the other.

To provide a sharper test, we now focus on one aspect where the transaction cost-based explanation and our inattention explanation give opposite predictions.

5.3 Test 2: How one shock impact price response to the other risk

The transaction cost-based explanation generates the opposite prediction to the “distract effect” in our inattention model. If transaction cost is the only friction, then if default risk receives a larger shock, this increases trading and thus speeds up the incorporation of contemporaneous interest rate information. The opposite is true in our inattention model which predicts a “distraction effect”: more value-relevant information from one risk consumes

---

23The ex ante payoff-relevance is likely to be time-varying, and investors can also adjust attention accordingly. For example, investors can allocate more attention to default risk shocks on company earnings days and to interest rate risk on CPI announcement days.
investor attention, *slowing down* price response to the other.

To test this prediction, for each bond $i$ in each week $t$, we separately measure the shocks from two sources:

$$\text{shock}_{i,t}^{\text{stock}} = \hat{b}_{i}^{\text{stock}} \cdot \text{StockRet}_{i,t}$$  \hspace{1cm} (22)$$

$$\text{shock}_{i,t}^{\text{trsy}} = \hat{b}_{i}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t}$$  \hspace{1cm} (23)$$

where $\hat{b}_{i}^{\text{stock}}$, $\hat{b}_{i}^{\text{trsy}}$ are bond-specific price response coefficients estimated using GMM. We then sort the sample by the sizes of the two shocks, $|\text{shock}_{i,t}^{\text{stock}}|$ and $|\text{shock}_{i,t}^{\text{trsy}}|$, into $3 \times 3$ bins. We use the same breakpoints to form the bins: $\{< 25\text{bp}, 25 - 75\text{bp}, > 75\text{bp}\}$.

After dividing observations into the $3 \times 3$ bins, each individual bond has too few observations. Therefore, we group together every 4 bonds with similar price coefficients into a “superbond.” We use GMM to estimate price underreactions for each superbond in each of the $3 \times 3$ bins, and then regress the underreaction estimates on shock size bin indicator:

$$\text{Underreaction}_{I,\Delta_s,\Delta_t}^{\text{stock}} = \sum_{\Delta_s=1}^{3} \sum_{\Delta_t=1}^{3} \gamma_{I,\Delta_s,\Delta_t}^{\text{stock}} \cdot \left(1_{|\text{shock}_{i,t}^{\text{stock}}| \text{ in bin } \Delta_s} \times 1_{|\text{shock}_{i,t}^{\text{trsy}}| \text{ in bin } \Delta_t}\right)$$

$$+ \sum_{I} \eta_{I}^{\text{stock}} \cdot 1_{\text{superbond } I} + \epsilon_{I}^{\text{stock}}$$  \hspace{1cm} (24)$$

$$\text{Underreaction}_{I,\Delta_s,\Delta_t}^{\text{trsy}} = \sum_{\Delta_s=1}^{3} \sum_{\Delta_t=1}^{3} \gamma_{I,\Delta_s,\Delta_t}^{\text{trsy}} \cdot \left(1_{|\text{shock}_{i,t}^{\text{stock}}| \text{ in bin } \Delta_s} \times 1_{|\text{shock}_{i,t}^{\text{trsy}}| \text{ in bin } \Delta_t}\right)$$

$$+ \sum_{I} \eta_{I}^{\text{trsy}} \cdot 1_{\text{superbond } I} + \epsilon_{I}^{\text{trsy}}$$  \hspace{1cm} (25)$$

where $I$ denotes superbonds and $\Delta_s, \Delta_t \in \{1, 2, 3\}$ denote the stock- and Treasury-shock size bins. We cluster standard errors at the superbond level. The main coefficients of interest, $\gamma_{I,\Delta_s,\Delta_t}^{\text{stock}}$ and $\gamma_{I,\Delta_s,\Delta_t}^{\text{trsy}}$, are shown in the two panels of Table 6. The evidence supports the inattention explanation and rejects the transaction cost explanation. When $|\text{shock}_{i,t}^{\text{trsy}}|$ is larger, the price underreaction to Treasury is lower rather than higher. The same is true with how underreaction to stock returns varies with $|\text{shock}_{i,t}^{\text{trsy}}|$.

---

24Specifically, let $n$ denote the number of bonds. We double sort bonds by $\hat{b}_{i}^{\text{stock}}$ and $\hat{b}_{i}^{\text{trsy}}$ into $m \times m$ “superbond” portfolios where $m = \lfloor \sqrt{n/4} \rfloor$ so that there are roughly 4 bonds per “superbond.”

25We apply the same data quality filters to the GMM estimates: the price reaction coefficient needs to be positive and the standard error of underreaction needs to be below 1.
### Panel A. Underreaction to stock

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>shock\textsuperscript{stock}</th>
<th></th>
<th>shock\textsuperscript{trsy}</th>
<th>bin</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>3 minus 1</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64%</td>
<td>65%</td>
<td>62%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-0.95)</td>
</tr>
<tr>
<td>2</td>
<td>58%</td>
<td>62%</td>
<td>63%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2.67)</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>52%</td>
<td>57%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.33)</td>
</tr>
</tbody>
</table>

| 3 minus 1 | | -14%\textsuperscript{**} | -12%\textsuperscript{**} | -5%\textsuperscript{**} | |
| t-stat | | (-11.46) | (-7.55) | (-2.61) | |

### Panel B. Underreaction to Treasury

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>shock\textsuperscript{stock}</th>
<th></th>
<th>shock\textsuperscript{trsy}</th>
<th>bin</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>3 minus 1</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65%</td>
<td>54%</td>
<td>55%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-9.13)</td>
</tr>
<tr>
<td>2</td>
<td>69%</td>
<td>56%</td>
<td>56%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-7.19)</td>
</tr>
<tr>
<td>3</td>
<td>80%</td>
<td>66%</td>
<td>63%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-6.16)</td>
</tr>
</tbody>
</table>

| 3 minus 1 | | 15%\textsuperscript{***} | 13%\textsuperscript{***} | 8%\textsuperscript{***} | |
| t-stat | | (6.53) | (6.15) | (4.18) | |

Table 6. The effects of stock and Treasury shock sizes on price underreaction. For each bond in each week, we estimate shock\textsuperscript{stock} and shock\textsuperscript{trsy} which are the bond value changes due to stock and Treasury returns, respectively. We then double sort weekly observations into 3 x 3 bins by |shock\textsuperscript{stock}| and |shock\textsuperscript{trsy}|. The table reports average price underreaction for these 3 x 3 bins. Details are explained in regressions (24) and (25); Panels A and B show the estimated coefficients $\gamma_{\Delta s, \Delta t}^{\text{stock}}$ and $\gamma_{\Delta s, \Delta t}^{\text{trsy}}$, respectively.

### 5.4 Time-varying risk premium

One may wonder if the documented price underreaction patterns can be explained by some frictionless asset pricing models with time-varying risk premium.\textsuperscript{26} However, Appendix D.1 shows that a simple strategy that trades on slow corporate bond price response to stock returns generates an annual Sharpe ratio of 2.61 before transaction costs. Given that a Sharpe ratio of 0.4 is already considered a “puzzle” in the equity risk premium literature (Mehra and Prescott (1985), Hansen and Jagannathan (1991)), we view this as strong evidence against frictionless equilibrium pricing explanations.

\textsuperscript{26}For instance, take the fact that corporate bond prices tend to continue drifting up (or down) after periods of high (or low) stock returns. This can potentially be explained if corporate bond risk premium increases (decreases) after good (bad) news.
6 Quantifying the implied cost of attention

In this section, we quantify the degree of pricing inefficiency and estimate the implied cost of investor attention. In section 6.1, we first show that the documented pricing inefficiency is large, in the sense that simple trading strategies can achieve very high Sharpe ratios. However, while the equilibrium pricing inefficiency is large, due to large transaction costs in the corporate bond market, arbitrageurs cannot make profits after transaction costs.

While the return predictability is not large enough to induce arbitrageurs to initiate new trades, regular market participants are still leaving some money on the table due to inattention. This is because if they pay more attention, they can profit by better timing their existing trades. In section 6.2, we estimate the amount of foregone profits for an average corporate bond mutual fund and find it to be rather small. Therefore, the documented pricing inefficiency can be plausibly justified by attention costs.

Finally, we investigate how return predictability varies with risk source and bond. All profits come from using lagged stock returns. When making out of sample forecasts, lagged Treasury return is no longer useful. This is consistent with the fact that it is much less costly to process interest rate information than default risk information. We also find that predictability is larger for bonds with lower credit quality, suggesting that cost of attention is higher for companies closer to distress.

6.1 Out-of-sample return predictability

![Figure 8. Illustration of the walk-forward return prediction strategy.](image)

Can the documented price underreactions help predict returns out of sample? As illustrated in Figure 8, in each year \( T = 2004, \ldots, 2014 \), we use data up to year \( T - 1 \) to estimate

---

33
forecasting regressions\textsuperscript{27}

\begin{align*}
\text{BondRet}_{i,t} &= \alpha^{\text{stock}} + \beta_{1}^{\text{stock}} \cdot \text{StockRet}_{i,t-1} + \ldots + \beta_{8}^{\text{stock}} \cdot \text{StockRet}_{i,t-8} + \epsilon_{i,t} \\
\text{BondRet}_{i,t} &= \alpha^{\text{trsy}} + \beta_{1}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t-1} + \ldots + \beta_{8}^{\text{trsy}} \cdot \text{TrsyRet}_{i,t-8} + \epsilon_{i,t}.
\end{align*}

Because bonds with different risk exposures can have different lag patterns, we estimate the two regressions separately for each of the $5 \times 5$ portfolios sorted by $\sigma^{\text{stock}}$ and $\sigma^{\text{trsy}}$. We then apply the estimated regressions to form two sets of return forecasts in year $T$, one based on lagged stock returns and another based on lagged Treasury returns.

For each of the 25 portfolios, we sort bonds weekly by return predictions, and calculate returns to a strategy that goes long on the top 50% and short on the bottom 50%. The portfolio weights are adjusted so that the gross exposure in each calendar week is always 100%.\textsuperscript{28} The hypothetical trading profits are shown in Panel A of Table 7.

The results show that there is only out-of-sample return predictability when using lagged stock information. The predictability is higher for higher $\sigma^{\text{stock}}$ quintiles, and is statistically significant for the second to fifth $\sigma^{\text{stock}}$ quintiles (Panel B). Strikingly, there is no out of sample return predictability using past Treasury information. In fact, trading on this signal produces negative returns, although this is barely statistically significant. In Appendix E.2, we show that the lack of Treasury-based predictability is robust to using fewer lagged Treasury returns as predictors (1, 2, or 4 instead of 8), or fitting the model using elastic net regularization instead of ordinary least squares (Zou and Hastie (2005)). Appendix E.3 demonstrates that the lack of predictability is explained by instability of the forecasting relationship over time.\textsuperscript{29}

Over all, these findings are consistent with lower costs in learning interest rate news than company-level fundamental news. The finding of higher stock-related predictability for bonds with higher $\sigma^{\text{stock}}$ is consistent with greater difficulty in analyzing default risk of companies that are closer to distress.\textsuperscript{30}

\textsuperscript{27}To ensure there is no time overlap between the left- and right-hand-side variables, we adjust the time interval for measuring weekly returns $\text{StockRet}_{i,t-1}$ and $\text{TrsyRet}_{i,t-1}$ so that they cannot overlap with the time interval for measuring $\text{BondRet}_{i,t}$. Specifically, recall that we measure $\text{BondRet}_{i,t}$ using price data from the latest trading day in week $t-1$ and week $t$. If the last bond trade in week $t-1$ happens on day $s$, then we measure $\text{StockRet}_{i,t-1}$ and $\text{TrsyRet}_{i,t-1}$ from the end of week $t-2$ to day $s-1$.

\textsuperscript{28}For example, if there are $N$ longs and $N$ shorts, then weight $\frac{1}{2N}$ is applied to each position.

\textsuperscript{29}Welch and Goyal (2007) make a similar point about a number of aggregate equity market return predictors. Although there is in-sample predictability, the relationship changes so much over time such that an investor cannot make money if only using historical data to estimate the predictive slopes.

\textsuperscript{30}There are other interpretations. For instance, it is possible that there are fewer experts capable of studying the default risk of bonds closer to distress. In our model, this amounts to a smaller attention budget $\bar{K}$ for distressed companies.
Table 7. Average predictable weekly corporate bond returns. Bonds are double sorted into $5 \times 5$ portfolios using $\sigma_{\text{stock}}$ and $\sigma_{\text{trsy}}$, the payoff-relevance proxies for default risk and interest rate risk. For each of the 25 portfolios, we predict out of sample bond returns using eight lags of stock returns (left panels) or Treasury returns (right panels). We then estimate the weekly profits of a strategy that, in each week, goes long on bonds with top 50% predicted returns and short on the bottom 50%. Panel A reports the weekly calendar profits of these strategies and Panel B reports the t-statistics.

6.2 Quantifying the dollar value of attention

We now quantify the implied dollar value of attention costs. The key question is, who is the marginal determinant of the return predictability patterns we document?

Arbitrageurs, such as hedge funds, are likely not the marginal investors. They are not leaving any money on the table, because the documented return predictability is not large enough to be profitable after accounting for transaction costs. To see this, Table 9 compares the return predictability to the average trade-weighted bid-ask spreads estimated using the Hong and Warga (2000) methodology.

This comparison is appropriate because an arbitrageur incurs a whole bid-ask spread in each trade, half when initiating the position and half when closing out the position. Because corporate bond transaction costs decrease with trad-

---

31 The TRACE dataset reports whether a trade is a customer selling to a dealer at the bid price, a dealer selling to a customer at the ask price, or an interdealer trade. For each bond-day, we calculate trade-weighted bid and ask prices, and take their differences to get a bid-ask spread estimate when both sides are available. This is a crude estimator but Schestag, Schuster, and Uhrig-Homburg (2016) show that it is comparable to more sophisticated measures of transaction costs.
ing sizes (Edwards et al. (2007)), we estimate spreads for trades above $100,000, $1 million, and $5 million separately. Even for trades above $5 million, the bid-ask spreads are still much larger than return predictability. Trades above $5 million are already relatively rare, with an average bond in our sample only having such trades in 27% of all traded weeks. Thus, we conclude that the documented underreactions are not profitable for arbitrageurs. Appendix E.2 shows that our conclusion remains robust when only trading on extreme predictions or using alternative statistical models to predict returns.

<table>
<thead>
<tr>
<th>Weekly predictable returns (basis points)</th>
<th>Transaction cost (size &gt; $100k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{\text{stock}} ) quintile</td>
<td>( \sigma_{\text{trsy}} ) quintile</td>
</tr>
<tr>
<td>Low</td>
<td>2</td>
</tr>
<tr>
<td>Low</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>2.90</td>
</tr>
<tr>
<td>3</td>
<td>5.93</td>
</tr>
<tr>
<td>4</td>
<td>6.44</td>
</tr>
<tr>
<td>High</td>
<td>6.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transaction cost (size &gt; $1 million)</th>
<th>Transaction cost (size &gt; $5 million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{\text{stock}} ) quintile</td>
<td>( \sigma_{\text{trsy}} ) quintile</td>
</tr>
<tr>
<td>Low</td>
<td>2</td>
</tr>
<tr>
<td>Low</td>
<td>10.41</td>
</tr>
<tr>
<td>2</td>
<td>17.25</td>
</tr>
<tr>
<td>3</td>
<td>19.95</td>
</tr>
<tr>
<td>4</td>
<td>22.19</td>
</tr>
<tr>
<td>High</td>
<td>25.08</td>
</tr>
</tbody>
</table>

Figure 9. Comparing return predictability and transaction costs. Bonds are double sorted into 5 \( \times \) 5 portfolios using our proxies of default risk and interest rate risk payoff-relevance (\( \sigma_{\text{stock}} \) and \( \sigma_{\text{trsy}} \)). For each portfolio, the top left panel is an estimate of the weekly profit when trading on lagged stock information (also reported in top left panel of Table E.7). The other three panels are average transaction costs of the bonds for different trading sizes, estimated using the Hong and Warga (2000) approach. All numbers are reported in basis points.

Even though the return predictability is not profitable for arbitrageurs, it is still profitable for regular market participants such as mutual funds. As these investors trade regularly as part of their business, they are already paying transaction costs. If they are more attentive to information, they can time their trades better by delaying certain trades to their benefit. This is illustrated in Figure 10 from the perspective of a buyer who wants to buy a bond in week \( t \). If the bond experienced a negative shock before week \( t \), the buyer knows that the price will continue to decline and can delay the trade to week \( t+1 \). If the shock is positive, the buyer cannot profit by delaying. Because positive and negative shocks happen roughly half of the time, and if we assume no correlation between shocks and intended trading directions, then attentive investors can capture approximately half of the predictable drifts.
Figure 10. Illustration of making profits by timing trades. Consider an investor who, for exogenous reasons, plans to buy a bond in week $t$. She can profit from paying attention to learn the shock in week $t - 1$. If the shock is negative, she knows that the price will continue to decline further, so she can profit by waiting until week $t + 1$ to buy at a lower price. She cannot profit if the shock is positive. The case for selling is symmetric.

Note that the scope of profits is much more limited than the typical arbitrageur perspective. Arbitrageurs can initiate new positions, while in this case, the amount of additional profit is limited by the amount of trading that the investor originally planned to do.

Under our inattention hypothesis, the amount of “money left on the table” has to reflect investor attention costs. We now estimate foregone profits for an average corporate bond mutual fund. Such a fund has assets under management of $1,002 million and has an annual turnover of 130%, so it trades a total of $1,002 \times 130\% \approx 1,303$ million per year. Because trading on treasuries is not profitable, we suppose the fund trades on lagged stock information, which gives an average predictable return of 5.25 basis points per trade. Therefore, a back-of-envelope estimate of the annual foregone profit due to inattention is:

$$\frac{1,002 \text{ million AUM}}{1,303 \text{ million trading volume per year}} \times \frac{5.7 \text{ bp return predictability}}{2} \approx 371,000$$

32 The equilibrium price underreaction should reflect the wealth-weighted average inattention of all participating investors. Inattention is likely heterogeneous across investors, with larger investors being less inattentive. In Appendix E.1, we show that, under some parametric assumptions, the equilibrium price underreaction roughly reflects the behavior of the average-sized investor.

33 We choose mutual funds for data availability. We obtain CRSP corporate bond mutual fund data over our sample period and restrict attention to funds with assets under management above $10$ million.
This is the total profit for the fund, but the relevant decision makers are the portfolio managers who bear the attention costs. We thus estimate how much profit accrues to managers. Because we lack data on the strength of manager incentive, we estimate an upper bound by taking the “two plus twenty” hedge fund contract where managers get 20% of additional fund returns. With this assumption, we estimate profits to managers for paying more attention to be $371,000 \times 20% \approx $74,200. This is approximately 2.7% of the annual compensation of an average mutual fund manager ($2.7 million). \[34\]

We need to note that this is an upper bound. We have assumed not only a strong managerial incentive, but also that funds can delay trades with no cost. This is a strong assumption in the corporate bond market because many bonds do not trade frequently. Thus, the manager may be concerned that once she passes up a trading opportunity, she may not be able to trade the same bond again later. With this in mind, the actual benefit of attention is likely much smaller, so we argue that the documented price underreaction can be plausibly explained by attention costs.

7 Conclusion

Investor inattention can cause slow price reaction to information, but existing work does not distinguish between behavioral neglect and sophisticated attention allocation. This paper develops tests to differentiate these two. We apply the tests to the corporate bond market and find evidence for the latter: investors are attention constrained, but they are sophisticated and allocate attention where it matters the most.

We find strong support for a unique prediction of sophisticated attention allocation: prices are faster to respond to more payoff-relevant risks. Concretely, for corporate bonds with higher credit risk, prices respond to default risk innovations more quickly. For bonds with higher interest rate risk, prices respond to interest rate innovations more quickly. When the risk exposure of bonds changes over time, the pattern of price underreaction also changes accordingly, indicating that investors reallocate attention based on changes in circumstances. All these findings are consistent with sophisticated attention allocation and inconsistent with behavioral neglect. We also show that our findings cannot be explained by differences in bond liquidity or transaction costs.

Finally, we estimate that an average market participant is only leaving a small amount

\[34\] To estimate mutual fund manager compensation, we use Table IA.VIII in the online Appendix of Ma, Tang, and Gómez (2018) which estimates that the average advisory fee per fund is around $8 million. Using corporate bond fund data from Morningstar, we find that funds with above $10 million in assets under management (AUM) on average have 2.96 portfolio managers. Thus we estimate that an average mutual fund manager makes $8 million / 2.96 \approx $2.7 million annually.
of money “on the table” due to their inattention, indicating that the documented price underreaction can truly explained by attention costs. In summary, our paper shows that corporate bond investors understand their own cognitive constraints and allocate attention in a sensible fashion.
References


42


A  Proofs and additional theoretical results

A.1  Proof for Proposition 1 (equilibrium)

We start with conjecturing the pricing functional form in (4):

\[ p_1 = \sum_{i=1}^{N} (1 - D_i) \cdot X_i + \lambda \cdot Z. \quad (28) \]

Under this conjecture, asset return from time 1 to time 2 reflects continued drifts in the \( N \) components plus reversion of the noise-trading component:

\[ R = V - p_1 = \sum_{i=1}^{N} D_i \cdot X_i - \lambda \cdot Z. \quad (29) \]

**Time 1 trading with fixed attention allocation**  Investors with more accurate signals can make more profits from trading on the drifts. We first derive investor portfolio choice with fixed attention choices. After receiving signals \( \{s_{ij}\}_{i=1}^{N} \), investor \( j \in [0, 1] \) updates belief about return:

\[ E(R|\{s_{ij}\}_{i=1}^{N}) = \sum_{i=1}^{N} D_i \cdot E(X_i|s_{ij}) = \sum_{i=1}^{N} D_i \cdot \theta_{ij} \cdot s_{ij} \]

where the projection coefficient \( \theta_{ij} \) is an alternative way to parametrize the attention choices:

\[ \theta_{ij} = \frac{Cov(X_i, s_{ij})}{Var(s_{ij})} = \frac{Cov(X_i, X_i + \sigma_i \cdot u_{ij})}{Var(X_i + \sigma_i \cdot u_{ij})} = \frac{\sigma_i^2}{\sigma_i^2 (1 + K_{ij}^{-1})} = \frac{K_{ij}}{K_{ij} + 1}. \]

Investor submits optimal market order \( q_j \), given by

\[ q_j = \arg \max_q q \cdot E(R|\{s_{ij}\}_{i=1}^{N}) - \frac{\psi}{2} q^2 = \frac{E(R|\{s_{ij}\}_{i=1}^{N})}{\psi} = \frac{\sum_{i=1}^{N} D_i \cdot \theta_{ij} \cdot s_{ij}}{\psi}. \]
The aggregate order flow is:

\[ Y = \int_0^1 q_j \, dj + \underbrace{Z}_{\text{noise traders}} \]

\[ = \frac{1}{\psi} \sum_{i=1}^N D_i \cdot \left( \int_0^1 \theta_{ij} \cdot s_{ij} \, dj \right) + Z \]

\[ = \frac{1}{\psi} \sum_{i=1}^N D_i \cdot \bar{\theta}_i \cdot X_i + Z \]  \hspace{1cm} (30)

where \( \bar{\theta}_i = \int_0^1 \theta_{ij} \, dj \) and the last step uses the exact law of large numbers (Duffie and Sun (2012)). Thus, the aggregate order flow will more heavily trade on risks with more underreaction (higher \( D_i \)) and if investors paid more attention to that risk (higher \( \bar{\theta}_i \)). Now, under the exogenous \( \lambda \) set up, we have verified the pricing conjecture (28):

\[ p_1 = \lambda \cdot Y \]

\[ = \sum_{i=1}^N \frac{\lambda}{\psi} \cdot D_i \cdot \bar{\theta}_i \cdot X_i + \lambda Z \]

**Solve for attention allocation at time 0** We now solve for the investor attention choice at time 0. After receiving signals \( \{s_{ij}\}_{i=1}^N \), investor \( j \)'s expected utility is:

\[ U_j(\{s_{ij}\}_{i=1}^N) = q_j(\{s_{ij}\}_{i=1}^N) \cdot \underbrace{E(R|\{s_{ij}\}_{i=1}^N)}_{\text{expected profit}} - \frac{\psi^2}{2} q_j^2(\{s_{ij}\}_{i=1}^N) \cdot \underbrace{E(R|\{s_{ij}\}_{i=1}^N)^2}_{\text{trading cost}} \]

optimal \( q_j(\{s_{ij}\}_{i=1}^N) = \frac{E(R|\{s_{ij}\}_{i=1}^N)}{\psi} \cdot \frac{E(R|\{s_{ij}\}_{i=1}^N)^2}{2\psi} \cdot \]

To get ex ante expected utility, we simply integrate over signal realizations\(^{35}\)

\[ U_j = E(U_j(\{s_{ij}\}_{i=1}^N)) \]

\[ = \frac{1}{2\psi} \cdot Var(R|\{s_{ij}\}_{i=1}^N) \]  \hspace{1cm} (31)

where the last step uses the law of total variance and the fact that \( Var(R) = 0 \). Equation (31) is intuitive: investor utility only depends on how informative her signals are in predicting

\(^{35}\)Due to the normal-normal updating structure, \( Var(R|\{s_{ij}\}_{i=1}^N) \) does not depend on signal realizations.
returns. We further expand out (31) as a function of signal precisions:

\[ U_j = \frac{1}{2\psi} \cdot \text{Var} \left( \sum_{i=1}^{N} D_i \cdot X_i \mid \{s_{ij}\}_{i=1}^{N} \right) \]

\[ = \frac{1}{2\psi} \cdot \sum_{i=1}^{N} D_i^2 \cdot \text{Var}(X_i \mid s_{ij}) \]

\[ = \frac{1}{2\psi} \cdot \sum_{i=1}^{N} \frac{K_{ij}}{K_{ij} + 1} \cdot \frac{D_i^2 \sigma_i^2}{\sigma_i^2} \]  

Equation (32) has a natural interpretation: an investor who knows \( X_i \) completely (taking signal precision \( K_{ij} \) to infinity) will increase her utility by a factor proportional to \( D_i^2 \sigma_i^2 \). The more payoff-relevant is risk \( i \) (higher \( \sigma_i \)), the higher the degree of underreaction \( (D_i) \), the more profitable it is to learn about and trade on. The \( \frac{K_{ij}}{K_{ij} + 1} \) factor is increasing in the signal precision chosen, and it is concave in \( K_{ij} \) due to decreasing returns to scale in precision acquisition: the more precise the signals, the less valuable is another unit of precision in reducing variance.

We now solve for optimal precision choices for the two attention specifications. Because investors face exactly the same optimization problem, they will make identical attention choices.

1. If we use the attention cost specification (without an upper bound), then, notice that the objective (32) is separable and concave in precisions while the cost \( \sum_{i=1}^{N} c_i K_{ij} \) is linear. Thus, as long as attention costs are not too large, the optimal solution comes from taking first order conditions separately for each risk:

\[ \frac{dU_j}{dK_{ij}} = \frac{D_i^2 \sigma_i^2}{2\psi} \cdot \frac{1}{(K_{ij} + 1)^2} \leq c_i \]

\[ \Rightarrow K_{ij}^* = \max \left( 0, \frac{D_i \sigma_i}{\sqrt{2\psi c_i}} - 1 \right) \]  

2. If the total investor attention has an upper bound, then we take first order conditions of the Lagrangian:

\[ \mathcal{L} = \frac{1}{2\psi} \sum_{i=1}^{N} \frac{K_{ij}}{K_{ij} + 1} \cdot D_i^2 \sigma_i^2 + \Phi \cdot \left( K - \sum_{i=1}^{N} c_i K_{ij} \right) \]

\[ \Rightarrow K_{ij}^* = \max \left( 0, \frac{D_i \sigma_i}{\sqrt{2\psi \Phi c_i}} - 1 \right) \]
where $\Phi$ is the Lagrange multiplier.

**Solving for underreaction parameters**  So far we have solved the model taking $\{D_1, ..., D_N\}$ as given. Now we solve for them. Note that price is given by:

$$p_1 = \lambda \cdot Y$$

$$= \lambda \cdot \left( \frac{1}{\psi} \sum_{i=1}^{N} D_i \cdot \frac{K_i^*}{K_i^* + 1} \cdot X_i + Z \right)$$

(36)

where $K_i^*$ are the optimal attention choices in (35) and we omit the agent index $j$ for simplicity. Notice that these attention choices are functions of $\{D_1, ..., D_N\}$. Equating coefficients in (36) and (28) gives equations for $D$'s:

$$1 - D_i = \frac{\lambda}{\psi} \cdot D_i \cdot \frac{K_i^*}{K_i^* + 1}$$

(37)

1. In the attention cost specification, if $\sigma_i \leq \sqrt{2\psi c_i}$, then $K_i^* = 0$ and $D_i = 1$ (full underreaction). Otherwise, substitute in $K_i^* = \frac{D_i \sigma_i}{\sqrt{2\psi c_i}} - 1$:

$$1 - D_i = \frac{\lambda}{\psi} D_i \cdot \frac{D_i \sigma_i - \sqrt{2\psi c_i}}{D_i \sigma_i}$$

$$\Rightarrow D_i = \frac{\psi + \frac{\lambda \sqrt{2\psi c_i}}{\sigma_i}}{\psi + \lambda} \quad \text{as long as } \sigma_i \geq \sqrt{2\psi c_i}$$

Thus, underreaction $D_i$ increases in attention cost $c_i$ and decreases in payoff-relevance $\sigma_i$. Note that it does not depend on the parameters of the other risks.

2. In the attention constraint specification, the expression becomes:

$$D_i = \frac{\psi + \frac{\lambda \sqrt{2\psi \Phi c_i}}{\sigma_i}}{\psi + \lambda} \quad \text{as long as } \sigma_i \geq \sqrt{2\psi \Phi c_i}$$

where $\Phi$ is the Lagrange multiplier that makes the attention constraint bind.

**A.2 Proving equilibrium under endogenous $\lambda$**

We now show that the equilibrium structure and uniqueness do not depend on an exogenous $\lambda$. Following [Kyle (1985)](Kyle1985), we assume market makers are competitive and uninformed, and they learn from aggregate order flow and set prices equal to $p_1 = E(V|Y)$ where $Y$ is the aggregate order flow.
We again start by conjecturing the pricing equation (28) and solving through to get aggregate order flow expression (30). Because $Y$ is linear in $\{X_1, ..., X_N, Z\}$, $p_1$ will also be linear in these variables through standard Bayesian updating. Competitive market makers will set time 1 price equal to expected value conditional on their information:

$$p_1 = E(V|Y) = \sum_{i=1}^{N} E(X_i|Y)$$

$$= \left( \sum_{i=1}^{N} \frac{Cov(X_i, Y)}{Var(Y)} \right) \cdot Y$$

$$= \left( \sum_{i=1}^{N} \frac{Cov(X_i, \frac{1}{\psi} \cdot D_i \bar{\theta}_i X_i)}{Var(\frac{1}{\psi} \sum_{i=1}^{N} D_i \bar{\theta}_i X_i + Z)} \right) \cdot Y$$

$$= \frac{1}{\psi^2} \cdot \sum_{i=1}^{N} D_i^2 \bar{\theta}_i \sigma_i^2 \cdot \left( \frac{1}{\psi} \sum_{i=1}^{N} D_i \bar{\theta}_i X_i + Z \right)$$

(38)

Therefore the pricing conjecture is verified. Later we will show that investors choose unique identical attention choices $\{K_i^*\}_{i=1}^{N}$ in equilibrium, therefore $\bar{\theta}_i = \frac{K_i^*}{K_i^* + 1}$, so the $\lambda$ parameter is given by:

$$\lambda = \frac{\psi \sum_{i=1}^{N} D_i^2 \frac{K_i^*}{K_i^* + 1} \sigma_i^2}{\sum_{i=1}^{N} D_i^2 \left( \frac{K_i^*}{K_i^* + 1} \right)^2 \sigma_i^2 + \psi^2 \sigma_Z^2}$$

(39)

Solving for underreaction parameters $(D_1, ..., D_N)$ We have solved the model taking $\{D_1, ..., D_N\}$ as given. We now solve for them explicitly. For notational convenience, express attention choices in the $\theta$-space as functions of underreaction parameters:

$$\theta_{ij}^*(D_i) = \frac{K_{ij}^*}{K_{ij}^* + 1}$$

$$= \max \left( 0, 1 - \frac{\sqrt{2\psi c_i}}{D_i \sigma_i} \right)$$

(40)

Note that all investors will make identical choices so $\bar{\theta}_i(D_i) = \theta_{ij}^*(D_i)$ for any $j$. Plugging (40) into (38) and matching coefficients with (28), we get a system of $N$ equations in $\{D_i\}_{i=1}^{N}$:

$$1 - D_i = \frac{\lambda(\{D_k\}_{k=1}^{N})}{\psi} \cdot \bar{\theta}_i(D_i) \cdot D_i$$

(41)
where $\lambda(\{D_i\})_{i=1}^N$ is defined in (38). We can have corner solutions: if $c_i$ is very large or $\sigma_i$ is very small, then we should expect $\bar{\theta}_i = 0$ and $D_i = 1$.

**Existence**

We now treat the case with interior solutions as it is easy to adapt the proof to treat corner solutions. Rewrite (41) as:

$$D_i = \frac{1}{1 + \frac{\lambda(D_1,...,D_N)}{\psi}\bar{\theta}_i(D_i)}$$

(42)

Define the mapping $f : [0,1]^N \rightarrow [0,1]^N$ using the right-hand side of (42):

$$f_i(D_1,...,D_N) = \frac{1}{1 + \frac{\lambda(D_1,...,D_N)}{\psi}\bar{\theta}_i(D_i)}$$

Clearly, for the interior case, $f$ is a continuous function so there exists at least one fixed point in $D_1,...,D_N$ by Brouwer’s fixed point theorem.

**Uniqueness**

Now we just need to show that the solution is unique. We note that it is sufficient to show that $\lambda$ cannot take on two values. For a given $\lambda$, we can solve for $\theta_i$ and $D_i$ from (40) and (42):

$$D_i(\lambda) = \frac{1 + \frac{\lambda}{\psi}\sqrt{2\psi c_i}}{1 + \frac{\lambda}{\psi}}$$

(43)

$$\theta_i(\lambda) = \frac{\sigma_i - \sqrt{2\psi c_i}}{\sigma_i + \frac{\lambda}{\psi}\sqrt{2\psi c_i}}$$

(44)

So a unique value of $\lambda$ will yield unique solutions of $D_1,...,D_N$. We now plug (43) and (44) into the expression of $\lambda$ in (38) to get a mapping from $\lambda$ to itself:

$$g(\lambda) = \frac{1}{\psi^2} \cdot \frac{\sum_{i=1}^N D_i(\lambda)\theta_i(\lambda)\sigma_i^2}{\sum_{i=1}^N D_i^2(\lambda)\theta_i^2(\lambda)\sigma_i^2 + \sigma_Z^2}$$

plug in (43) and (44), define $z_i \equiv \frac{\sqrt{2\psi c_i}}{\sigma_i}$

$$= \frac{1}{\psi^2} \cdot \frac{\sum_{i=1}^N (1-x_i)\sigma_i^2}{\sum_{i=1}^N (1-x_i)^2\sigma_i^2 + \sigma_Z^2}$$

(45)

Let $u(\lambda)$ and $v(\lambda)$ denote the numerator and denominator in (45). By the ranking lemma in [Milgrom (2004)], to prove uniqueness it suffices to show that, for all fixed point $\bar{\lambda}$, we have
$g'(\lambda) < 1$. We can check this via brute force differentiation:

$$
g'(\lambda) = \frac{u'(\lambda)}{v(\lambda)} - \frac{u(\lambda) \cdot v'(\lambda)}{v^2(\lambda)}$$

$$= \frac{2\lambda}{\psi^2} \sum_{i=1}^{N} \frac{(1-x_i)^2 \sigma_i^2}{(1+\lambda/\psi)^2} - \frac{1}{\psi^2} \sum_{i=1}^{N} \frac{(1-x_i)^2 \sigma_i^2}{(1+\lambda/\psi)^2} + \frac{1}{\psi^2} \sum_{i=1}^{N} \frac{(1-x_i)^2 \sigma_i^2}{(1+\lambda/\psi)^2} + \sigma_Z^2$$

$$< \frac{2\lambda}{\psi^2} \sum_{i=1}^{N} \frac{(1-x_i)^2 \sigma_i^2}{(1+\lambda/\psi)^2} - \frac{1}{\psi^2} \sum_{i=1}^{N} \frac{(1-x_i)^2 \sigma_i^2}{(1+\lambda/\psi)^2} + \sigma_Z^2$$

$$= \frac{\lambda-\psi}{\lambda+\psi} \cdot \left( \frac{1}{\psi^2} \sum_{i=1}^{N} \frac{(1-x_i)^2 \sigma_i^2}{(1+\lambda/\psi)^2} \right) + \sigma_Z^2$$

Thus the equilibrium has to be unique.

### A.3 The magnitude of “distraction effects”

Consider a simplified case of our model where the asset is driven by two risks (e.g., interest rate and default risks), and suppose we are in the interior case where both risks receive some attention. We then have the following result.

**PROPOSITION 2.** (Size of “distraction effect”)

If all investors face attention constraints ($c_1 \cdot K_1 + c_2 \cdot K_2 \leq \bar{K}$), then the “distraction effect” is as large as the “main effect” after adjusting for the ratio of payoff-relevance. Precisely, for risks $i \neq k$:

$$\frac{dD_i}{d\sigma_k} = \frac{\sigma_k}{\sigma_i} \cdot \left( \frac{-dD_i}{d\sigma_i} \right)$$

“distraction effect”

“main effect”

**Proof.** We first solve the two-risk model in closed form, and then the result will become obvious. Following derivation in [34], investors solve the following optimization:

$$\max_{K_1, K_2} \frac{K_1}{K_1 + 1} \cdot D_1^2 \sigma_1^2 + \frac{K_2}{K_2 + 1} \cdot D_2^2 \sigma_2^2$$

subject to $c_1 \cdot K_1 + c_2 \cdot K_2 \leq \bar{K}$
We write $K_2 = K_2(K_1)$ using the binding precision constraint and then set the derivative of the objective function to zero:

$$
0 = \frac{D_1^2 \sigma_1^2}{(K_1 + 1)^2} + \frac{K_2'(K_1) \cdot D_2^2 \sigma_2^2}{(K_2(K_1) + 1)^2}
$$

$$
= \frac{D_1^2 \sigma_1^2}{(K_1 + 1)^2} - \frac{c_1}{c_2} \cdot \frac{D_2^2 \sigma_2^2}{(K_2(K_1) + 1)^2}
$$

$$
\Rightarrow \frac{D_1^2 \sigma_1^2}{(K_1 + 1)^2} = \frac{c_1 c_2 D_2^2 \sigma_2^2}{(\tilde{K} - c_1 K_1 + c_2)^2}
$$

$$
\Rightarrow \frac{\tilde{K} - c_1 K_1 + c_2}{K_1 + 1} = \sqrt{\frac{c_1 c_2 \cdot D_2 \sigma_2}{D_1 \sigma_1}}
$$

$$
\Rightarrow K_1 = \frac{\tilde{K} + c_2 - \eta}{c_1 + \eta}
$$

$$
\Rightarrow \frac{K_1}{K_1 + 1} = \frac{\tilde{K} + c_2 - \eta}{\tilde{K} + c_1 + c_2}
$$

Plug this into the fixed point equation (derived in (37)) for $D_1$:

$$
1 - D_1 = \frac{\lambda}{\psi} \cdot \frac{K_1}{K_1 + 1} \cdot D_1
$$

$$
\Rightarrow D_1 = \left(1 + \frac{\lambda}{\psi} \cdot \frac{K_1}{K_1 + 1}\right)^{-1}
$$

$$
= \left(1 + \frac{\lambda}{\psi} \cdot \frac{\tilde{K} + c_2 - \eta}{\tilde{K} + c_1 + c_2}\right)^{-1}
$$

(48)

Similarly, we can write out the fixed point problem for $D_2$:

$$
D_2 = \left(1 + \frac{\lambda}{\psi} \cdot \frac{K_2}{K_2 + 1}\right)^{-1}
$$

$$
= \left(1 + \frac{\lambda}{\psi} \cdot \frac{\tilde{K} + c_1 - c_1 c_2 / \eta}{\tilde{K} + c_1 + c_2}\right)^{-1}
$$

(49)

We now plug in the expression of $\eta = \sqrt{c_1 c_2} \cdot \frac{D_2 \sigma_2}{D_1 \sigma_1}$ (defined in (47)) into (48) and (49). This gives us two linear equations in $D_1, D_2$:

$$
\left(1 + \frac{\lambda}{\psi} \cdot \frac{\tilde{K} + c_2}{\tilde{K} + c_1 + c_2}\right) \cdot D_1 - \frac{\lambda \sqrt{c_1 c_2}}{\psi(K + c_1 + c_2)} \cdot \frac{\sigma_2}{\sigma_1} \cdot D_2 = 1
$$

$$
- \frac{\lambda \sqrt{c_1 c_2}}{\psi(K + c_1 + c_2)} \cdot \frac{\sigma_1}{\sigma_2} \cdot D_1 + \left(1 + \frac{\lambda}{\psi} \cdot \frac{\tilde{K} + c_1}{\tilde{K} + c_1 + c_2}\right) D_2 = 1
$$

52
solving which yields:

\[
D_1 = \Delta \cdot \left( 1 + \frac{\lambda}{\psi} \cdot \frac{\bar{K} + c_1 + \sqrt{c_1 c_2} \cdot \sigma_2}{\bar{K} + c_1 + c_2} \right)
\]

\[
D_2 = \Delta \cdot \left( 1 + \frac{\lambda}{\psi} \cdot \frac{\bar{K} + c_2 + \sqrt{c_1 c_2} \cdot \sigma_1}{\bar{K} + c_1 + c_2} \right)
\]

where \( \Delta = \left( 1 + \frac{\lambda}{\psi} \cdot \frac{2\bar{K} + c_1 + c_2 + \lambda^2}{\bar{K} + c_1 + c_2} \cdot \frac{\bar{K}^2 + (c_1 + c_2)\bar{K}}{(\bar{K} + c_1 + c_2)^2} \right)^{-1} \)

Now we have explicit solutions for the equilibrium. To prove the proposition, we only differentiate \( D_1 \) with respect to \( \sigma_1 \) and \( \sigma_2 \):

\[
\frac{dD_1}{d\sigma_1} = -\frac{\lambda\Delta}{\psi} \cdot \frac{\sqrt{c_1 c_2}}{\bar{K} + c_1 + c_2} \cdot \frac{\sigma_2}{\sigma_1^2}
\]

\[
\frac{dD_1}{d\sigma_2} = \frac{\lambda\Delta}{\psi} \cdot \frac{\sqrt{c_1 c_2}}{\bar{K} + c_1 + c_2} \cdot \frac{1}{\sigma_1} = \frac{\sigma_2}{\sigma_1} \cdot \left( -\frac{dD_1}{d\sigma_1} \right)
\]

Because the model is symmetric, the same is true if we swap risk 1 and 2. \( \square \)

### A.4 Alternative hypothesis: transaction cost

In this section, we show that if investors have no inattention but face a transaction cost, then this can generate the “main effect” in our inattention model but generates the opposite prediction to the “distraction effect”.

Consider an alternative model where investors face no attention costs so they observe value \( V \) perfectly, but needs to pay transaction cost \( c > 0 \) to submit an order. We think of this cost as capturing explicit bid-ask spreads and also other back-office expenses, such as recording and reporting expenses, associated with trading. Recall that investors submit market orders, and market makers set price to \( p = \lambda \cdot Q \) where \( Q \) is the aggregate order submitted.

This alternative model produces an intuitive equilibrium where the degree of price delays depends on \( |V| \), illustrated in Figure 11. When \( |V| \) is very small, investors choose not to trade due to the transaction cost. When \( |V| \) is large, all investors trade. For intermediate values of \( |V| \), a fraction of investors trade, and the degree of price inefficiency is such that investors are indifferent between trading or not.

**PROPOSITION 3.** *(Equilibrium) Define price underreaction \( D(V) \) such that price is \( p(V) = (1 - D(V)) \cdot V \). Let \( k(V) \) denote the fraction of investors submitting orders. Then, the unique equilibrium in the transaction model is given by:
Figure 11. Illustration of the model with transaction cost. Left panel plots the fraction of delayed price response and right panel plots the fraction of investors submitting orders. When value change $|V|$ is too small to overcome fixed cost of transactions, investors do not trade and there is full price underreaction $D = 100\%$. When $|V|$ is very large, all investors trade and price underreaction reaches the lowest level. For intermediate values of $|V|$, investors choose mixed strategies and only a fraction of investors submit orders. Parameters: bond value volatility $\sigma_V = 1$, fixed transaction cost $c = 0.5$, variable transaction cost $\eta = 0.25$, and price impact coefficient $\lambda = 1$.

1. When $|V| \leq \sqrt{2\psi c}$, $k(V) = 0$, and $D(V) = 1$.
2. When $|V| \in \left(\sqrt{2\psi c}, \frac{\psi + \lambda}{\psi} \cdot \sqrt{2\psi c}\right]$, $k(V) = \frac{\psi}{\lambda} \cdot \frac{V - \sqrt{2\psi c}}{\sqrt{2\psi c}}$, and $D(V) = \frac{\sqrt{2\psi c}}{V}$.
3. When $|V| > \frac{\psi + \lambda}{\psi} \cdot \sqrt{2\psi c}$, $k(V) = 1$, and $D(V) = \frac{\psi}{\psi + \lambda}$.

Proof. We discuss each of the three regions. Note that, if an investor submits an order, he will trade $q = \frac{V - p}{\psi}$ units and obtain expected utility $\frac{D^2(V) \cdot V^2}{2\psi} - \frac{c}{\psi + \lambda}$.

1. When $|V| \leq \sqrt{2\psi c}$, trading gains does not justify transaction cost $c$, so investors choose not to trade.
2. When $|V| \in \left(\sqrt{2\psi c}, \frac{\psi + \lambda}{\psi} \cdot \sqrt{2\psi c}\right]$, $R = D(V) \cdot V = \sqrt{2\psi c}$, so trading gains exactly offset transaction cost, so investors are indifferent so they can adopt a mixed strategy.
3. When $|V| > \frac{\psi + \lambda}{\psi} \cdot \sqrt{2\psi c}$, trading gains are higher than transaction cost, so all investors submit orders $q = \frac{D(V) \cdot V}{\psi}$. Solving for $D(V)$ using $p = \lambda \cdot q$ gives $D(V) = \frac{\psi}{\psi + \lambda}$.
To compare predictions from this transaction cost model with our inattention model, suppose $V = X_1 + X_2$, with $X_1$ representing interest rate shock and $X_2$ representing credit risk shock. It is true that this transaction cost model can generate the “main effect” in our inattention model, because price delay $D(V)$ is a function of $|V|$. The larger is $\sigma(X_1)$, the more correlated is $|X_1|$ with $|V|$, so larger values of $|X_1|$ will more likely coincide with periods with lower price delays. This, if we simply run a regression of $R = D(V) \cdot V$ on $X_1$, we will find lower average delays associated with risks with higher payoff relevance.

However, this transaction cost model generates the opposite prediction for the “distraction effect”. The intuition is as follows. Under the transaction cost model, larger $\sigma(X_1)$ will on average increase $|V|$, and thus reduce price underreaction to all shocks, resulting in lower price underreaction to $X_1$. We show this numerically in Table 8. However, under our model with attention constraints, larger $\sigma(X_1)$ will consume more investor attention, resulting in slower price response to $X_2$.

<table>
<thead>
<tr>
<th>Underreaction to shock $X_1$</th>
<th>Underreaction to shock $X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma(X_1)$</td>
<td>$\sigma(X_2)$</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>93%</td>
<td>74%</td>
</tr>
<tr>
<td>74%</td>
<td>62%</td>
</tr>
<tr>
<td>57%</td>
<td>51%</td>
</tr>
<tr>
<td>46%</td>
<td>42%</td>
</tr>
<tr>
<td>38%</td>
<td>37%</td>
</tr>
</tbody>
</table>

Table 8. Average underreaction to shocks when varying payoff-relevance. We consider a model where value is given by $V = X_1 + X_2$, and we vary payoff relevance of the two risks ($\sigma(X_1)$ and $\sigma(X_2)$). Investors observe value $V$ perfectly but face a transaction cost $c$. Average price underreaction to risk 1 and 2 are plotted in the left and right panels, respectively. Due to the symmetric model structure, the two panels are simply transposes of each other. Apart from $\sigma(V) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2)}$, model parameters are the same as those in Figure 11.
B Why isn’t it costless to learn interest rates?

In the paper, we argue that corporate bond investors do not adjust bond valuation one-to-one with duration-matched treasuries, and the exact adjustment sensitivity depends on contemporaneous shocks to credit risk. As empirical evidence for this view, we first show that corporate bond sensitivity to treasuries varies with the market level credit risk shock. To proxy for the amount of contemporaneous credit-related information, we use the size of the weekly change of five-year Markit CDX investment grade (IG) index spread. Because our Markit data starts in 2006, we only use data during 2006 - 2014. We then sort the sample period into quintiles by CDX spread change sizes, and regress bond returns on stock and duration-matched Treasury returns for each quintile:

\[
\text{BondRet}_{i,t} = a + \beta^{stock} \cdot \text{StockRet}_{i,t} + \beta^{trsy} \cdot \text{TrsyRet}_{i,t} + \sum_{i} \gamma_i \cdot 1_{\text{bond } i} + u_{i,t}. \tag{50}
\]

We then plot the variation of \(\beta^{trsy}\) across quintiles in the right panel of Figure 12. Standard errors are clustered by week and bond.

**Figure 12.** Explaining why bond returns do not move one-to-one with treasuries. **Left:** weekly changes in five-year CDX IG index spreads plotted against weekly changes in five-year Treasury yields. **Right:** bond return sensitivity to duration-matched treasuries as a function of CDX movements. We sort the sample into quintiles using the absolute value of the changes in five-year CDX IG index spreads. For each quintile, we regress weekly corporate bond returns on duration-matched Treasury returns and stock returns, and report the regression coefficients on treasuries. Dotted lines are two standard deviation bands and standard errors in the regression are clustered by week and bond.
As is clearly shown in the right panel of Figure 12, bond price sensitivity to treasuries is lower in periods when credit risk changes are larger. The difference between the first and last quintile is statistically significant with a t-statistic of 3.09. The fact that $\beta^{trsy}$ is lower than 1 is consistent with the negative correlation between Treasury yields and credit spreads, shown in left panel of Figure 12. When Treasury yields rise (fall), credit spread narrows (widens), so bond yield moves less than one-for-one with Treasury yields.

While this is slightly outside the scope of our paper, we also use an approximation to show that this mechanism generates roughly the right magnitude of $\beta^{trsy}$ deviation from unity. Consider a bond whose yield decomposes into $r_t$, the risk-free interest rate, plus $c_t$, the credit spread. Then, a regression of bond return on duration-matched Treasury return gives a regression coefficient of:

$$\frac{\text{Cov}(\Delta r_t + \Delta c_t, \Delta r_t)}{\text{Var}(\Delta r_t)} = 1 - \frac{\text{Cov}(\Delta c_t, \Delta r_t)}{\text{Var}(\Delta r_t)} \quad (51)$$

Using five-year Treasury yield to proxy for $r_t$ and five-year CDX IG index spread to proxy for $c_t$, we estimate the deviation term in (51) to be 0.23. The total deviation is 0.39 over the full sample. The remaining $0.39 - 0.23 = 0.16$ is explained away by slow price reaction to Treasury movements. $^{36}$

The negative correlation between Treasury yields and credit spreads also helps explain the variation of deviation across the five quintiles sorted by CDX spread changes. To see this, in Figure 13, we plot the actual deviation of $\beta^{trsy}$ from unity against the estimated deviation using equation (51) across deciles.

$^{36}$In section 3.3 we estimate that 27% of Treasury-induced bond movements are delayed. This means that underreaction can explain approximate $0.61 \cdot \frac{27\%}{1 - 27\%} = 0.226$ of deviation.
Figure 13. Explaining deviation of bond-Treasury sensitivity from unity. Data is sorted into quintiles using the size of five-year CDX IG spread changes. Black line plots one minus the estimated bond price sensitivity to duration-matched treasuries for each quintile, and the dotted lines are two standard error bands. The blue line plots the amount of deviation explained by the negative correlation between Treasury yields and credit spreads.
C Additional details on testing model predictions

Figure 14. Sorting payoff-relevance measures into quintiles. The payoff-relevance measures, $\sigma^{\text{stock}}$ and $\sigma^{\text{trsy}}$, are defined as the annualized volatility of bond returns explained by stock and Treasury returns, respectively. The dotted green lines are common breakpoints for the five quintiles.
C.1 Details on GMM-based underreaction estimation

We give more details about the GMM-based estimation here. To restrict the underreaction parameters to be between 0 and 1, we apply logistic transforms:

\[
\text{Underreaction}_{i,p}^{\text{stock}} = \frac{1}{1 + e^{-\eta_{i,p}^{\text{stock}}}},
\]
\[
\text{Underreaction}_{i,p}^{\text{trsy}} = \frac{1}{1 + e^{-\eta_{i,p}^{\text{trsy}}}},
\]

where \(\eta_{i,p}^{\text{stock}}, \eta_{i,p}^{\text{trsy}}\) are unconstrained real parameters. For each bond \(i\) in each period \(p\), we estimate \((\eta_{i,p}^{\text{stock}}, \eta_{i,p}^{\text{trsy}}, b_{i,p}^{\text{stock}}, b_{i,p}^{\text{trsy}})\) jointly. Because the estimation uses overlapping windows, we compute autocorrelation-consistent standard errors with truncated kernel of eight lags (Zeileis (2004)). We then use the delta method to derive standard errors for Underreaction_{i,p}^{\text{stock}} and Underreaction_{i,p}^{\text{trsy}}.

Discarding inaccurate estimates

We discard cases where \(\hat{b}_{i,p}^{\text{risk}} < 0\) because negative bond return sensitivity to interest rate and default risk shocks is hard to justify in theory. We also require the estimated standard error of Underreaction_{i,p}^{\text{risk}} to be below 1. The cases of very large standard errors tend to happen when \(b_{i,p}^{\text{risk}}\) is low, in which case the underreaction parameter is not well identified, as shown in Figure 15. Out of a total of 1,2421 bond-periods with at least 26 observations, 7,480 (60%) stock estimates and 9,014 (73%) Treasury estimates survive this filter. Figure 15 provides details about the distribution of estimates kept versus discarded. Figure 16 plots the histogram of underreaction estimates used.

GMM-based and regression-based estimates are similar

To verify the accuracy of GMM-based bond-level underreaction estimates, we check to see if they are similar to the regression-based, portfolio-level estimates. Because the latter rely on weaker parametric assumptions, this helps us check if our GMM parametric restrictions are valid.

In Figure 17, we plot the average of GMM estimates for the 25 portfolios against the regression-based estimates. The green solid line is the 45% degree line and the dotted green lines are two standard deviation error bands. In almost all cases, the average of GMM estimates are within two standard deviations of the portfolio-level estimates, indicating that estimates from the two methods are similar.
Figure 15. Illustration of the GMM estimates kept versus discarded. The figures plot the fractions of $\hat{\sigma}(\text{Underreaction}) > 1$ as a function of $b$ for stock risk (left) and Treasury risk (right). In the paper, we retain a GMM estimate only if $b > 0$ and $\hat{\sigma}(\text{Underreaction}_i) < 1$. Counts of retained estimates are marked in blue and those discarded are marked in black. For instance, in the left figure, 1,027 bonds have $b^{\text{stock}} < 0$ and $\hat{\sigma}(\text{Underreaction}^{\text{stock}}) < 1$; 1,440 bonds have $b^{\text{stock}} < 0$ and $\hat{\sigma}(\text{Underreaction}^{\text{stock}}) > 1$, etc.

Figure 16. Histogram of GMM-estimated price underreaction.
Figure 17. Comparing GMM-based and regression-based underreaction estimates. Bonds are double sorted into $5 \times 5$ portfolios using the payoff-relevance proxies ($\sigma^\text{stock}$ and $\sigma^\text{treasury}$). For each of the 25 portfolios, we plot the average of bond-level GMM-based underreaction estimates in section 4.2 against the panel regression-based estimates in section 4.1. The green solid line is the 45% degree line and the green dotted lines are two standard deviation bands from the portfolio-level regressions.
C.2 Bond-level tests (section 4.2)

Figure 18. Comparing “main effect” of $\sigma_{\text{risk}}$ and “distraction effect” of $\sigma_{\text{other risk}}$.

For explaining underreaction to stock returns, the main effect is the impact of stock payoff-relevance and the distraction effect is the impact of Treasury payoff-relevance, and the reverse is true when explaining underreaction to Treasury returns. These effects are estimated in regression (18) and coefficients are reported in Table 4. The vertical axis plots the regression coefficients on $1_{\sigma_{\text{risk}} \text{ in quintile } q}$ (black) versus coefficients on $1_{\sigma_{\text{risk}} \text{ in quintile } q}$ (blue). The horizontal axis plots the average $\sigma$ by quintile. The left figure uses the equal-weighted specification (1) in Table 4, while the right figure uses the value-weighted specification (4). Dotted lines are two standard error bands.
Table 9. Tally of payoff-relevance quintile assignment changes of bonds. These tables include all bonds with measures of $\sigma_{i,p}^{stock}$ and $\sigma_{i,p}^{trsy}$ in any of the two consecutive periods $p \in \{2002 - 2006, 2007 - 2010, 2011 - 2014\}$. 

<table>
<thead>
<tr>
<th></th>
<th>$\sigma^{stock}$ quintile migration</th>
<th></th>
<th>$\sigma^{trsy}$ quintile migration</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low 2 3 4 High</td>
<td>Low 2 3 4 High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous</td>
<td>Low 41 35 42 36 39</td>
<td>Low 2 3 2 0 1</td>
<td>Previous 2 64 30 24 9 4</td>
<td></td>
</tr>
<tr>
<td>Period</td>
<td>2 80 38 32 37 38</td>
<td>3 120 112 51 49 22</td>
<td>4 61 180 156 74 57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 85 40 20 41 49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 87 76 77 31 54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High 56 100 102 125 88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D Additional results about alternative hypotheses

D.1 Can time-varying risk premium explain findings?

Time-varying risk premium can also potentially explain return predictability patterns. However, the price underreaction we document imply too high a Sharpe ratio for such explanation to be plausible.

To see this, we consider a very simple trading strategy that uses the return predictions constructed in section 6.1. We only use lagged stock returns to forecast returns because lagged treasuries returns are shown to be not useful. In each week, the strategy goes long on bonds with top 10% return forecasts and short on bonds in bottom 10%. We equal weight all positions and adjust exposure such that the long and short sides each have a gross exposure of 100%.

Table 10 shows that this simple, equal-weighted strategy generates an annualized return of 7.8% with a Sharpe ratio of 2.61. Moreover, the returns are not spanned by well-known asset pricing factors. After adjusting for the Fama-French five factors (Fama and French (2015)) plus the momentum factor, the equal-weighted strategy alpha is still as large as the raw return. The results are only slightly weaker if we construct a value-weighted strategy, in which case the annualized return is 6.5% with a Sharpe ratio of 1.94.

Section 6.2 shows that the return predictability is within bid-ask spreads, so this hypothetical trading strategy cannot be implemented. However, time-varying risk premium explanations are about the variation of equilibrium mid-prices, so the fact that the implied Sharpe ratio is too high is still strong evidence against such explanations.

<table>
<thead>
<tr>
<th>Risk adjustment</th>
<th>Return</th>
<th>Volatility</th>
<th>T-statistic</th>
<th>Sharpe ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Adjustment</td>
<td>7.82%</td>
<td>2.99%</td>
<td>8.68</td>
<td>2.61</td>
</tr>
<tr>
<td>CAPM</td>
<td>8.13%</td>
<td>2.92%</td>
<td>9.26</td>
<td>2.79</td>
</tr>
<tr>
<td>Fama-French 3 Factor</td>
<td>8.15%</td>
<td>2.90%</td>
<td>9.32</td>
<td>2.80</td>
</tr>
<tr>
<td>Fama-French-Cahart</td>
<td>8.20%</td>
<td>2.90%</td>
<td>9.40</td>
<td>2.83</td>
</tr>
<tr>
<td>Fama-French 5 Factor + Momentum</td>
<td>8.05%</td>
<td>2.89%</td>
<td>9.25</td>
<td>2.78</td>
</tr>
</tbody>
</table>

Table 10. Pre-transaction cost profits when trading on lagged stock information.
We use lagged stock returns to forecast bond returns and form a weekly rebalancing strategy that goes long on bonds with top 10% predicted returns and short on the bottom 10%. The first row reports the raw annualized strategy return, and the remaining rows report alphas when adjusting with various factor risk models.
E Additional results on quantifying attention costs

E.1 Who is the marginal investor for price underreaction?

The equilibrium degree of price underreaction reflects the wealth-weighted average degree of inattentiveness of all participating investors. Investors are likely heterogeneous in their inattentiveness, with larger investors being less inattentive. Thus, when we back out attention costs from price underreactions, we need to ask, whose attention cost are we estimating?

Using a simplified analytical framework, we find that the equilibrium underreaction approximately reflects the inattention of the average-sized market participant. For a bond, let \( w \) denote the fraction held by an investor. In the data, \( w \) is heavy-tailed so we assume it follows a Pareto distribution with density \( f(w) = \frac{\alpha w_m^\alpha}{w^{\alpha+1}} \) where \( w_m \) is the minimum holding size and \( \alpha > 1 \) is a tail-shape parameter. We use mutual fund holdings data to estimate \( \alpha \) empirically for each bond, as shown in the left panel of Figure 19, and plot the distribution of estimated \( \alpha \) in the right panel. The average \( \alpha \) is 2.24 in our data.

![Figure 19. Distribution of the bond holding fraction by mutual funds. Left: fitting Pareto distribution to the holdings of a bond (CUSIP 92343VBS). Right: distribution of MLE-estimated Pareto tail-shape parameter \( \alpha \) across all bonds, with one estimate per bond. Data: fractional holdings of corporate bonds by corporate bond mutual funds in Q4 of 2014.](image)

Because the main profits to attention is driven by underreaction to credit risk, we only consider each investors’ attention to credit risk now. Suppose all investors face a cost of \( c \) for each unit of signal precision, and let \( K(w) \) denote the precision chosen by the investor with size \( w \). Let \( D(K(w)) \) denote the investor-specific degree of underreaction, which equals \( D(K(w)) = \frac{1}{K(w)+1} \) in our model. Note that \( K(w) \) is the solution from investor optimization
problem:

\[ K(w) = \arg \max_k - D(k) \cdot w - c \cdot k \]

\[ = \sqrt{\frac{w}{c}} - 1 \]

\[ \Rightarrow D(w) = \frac{1}{K(w) + 1} = \sqrt{\frac{c}{w}} \]

Because each investor’s impact on bond price is proportional to their holding \( w \), the equilibrium underreaction is a wealth-weighted average of individual underreactions:

\[ D_{\text{equilibrium}} = \int_{w_m}^{\infty} D(w) \cdot f(w) dw \]

\[ = \int_{w_m}^{\infty} \frac{c^{1/2}}{w^{1/2}} \cdot \frac{\alpha w^{\alpha}}{w^{\alpha+1}} dw \]

\[ = \frac{c^{1/2}}{w_m^{1/2}} \cdot \frac{\alpha}{\alpha + 1/2} \cdot \int_{w_m}^{\infty} (\alpha + 1/2) w_m^{\alpha+1/2} \cdot \frac{w_m^{\alpha+3/2}}{w^{\alpha+3/2}} dw \]

\[ = \frac{2\alpha}{2\alpha + 1} \cdot \left( \frac{c}{w_m} \right)^{1/2} \]

\[ = D \left( \frac{2\alpha + 1}{2\alpha} \right)^2 \cdot w_m \]

Thus, the equilibrium underreaction reflects the behavior of the investor with holdings \( w_{\text{marginal}} = \left( \frac{2\alpha + 1}{2\alpha} \right)^2 \cdot w_m \). As illustrated in Figure 20, this is always larger than the average investor whose size is \( E(w) = \frac{\alpha}{\alpha - 1} \cdot w_m \), and is rather close to the average investor with the \( \alpha \) estimated from data.
E.2 Alternative trading strategies

In section 6, we conclude that 1) the amount of profits from trading on lagged stock information is not sufficient to overcome transaction costs, and 2) there is no predictability trading on lagged Treasury information. Here we examine whether these conclusions are robust to the trading strategy used.

We first examine whether results change when we only use the most extreme $x\%$ of predictions. For example, if $x = 10$, then our strategy goes long on bonds with top 5% return predictions and short on those with bottom 5% return predictions. We still use 8 lags of past stock and Treasury returns to make predictions. The results are reported in Table 11. The conclusions are qualitatively unchanged: it is not profitable to trade on treasuries. Even though there is some evidence that profits from trading on stock information is higher when $x$ is lower, even when $x = 2\%$, the profits in general do not exceed the bid-ask spreads (reported in Panel C of Table 7).

Because using 8 lags in OLS may result in overfitting, we redo the exercise using 4 or 2 lags of past stock and Treasury returns to predict bond returns. We also try fitting a model using elastic net regularization with $\alpha = 1/2$, which amounts to using half LASSO and half ridge penalty (Zou and Hastie (2005)). When implementing elastic net to make
Table 11. Average weekly profits when trading on extreme return forecasts. This is a robustness check of Panel A in Table 7. Bonds are double sorted into 5 × 5 portfolios using \( \sigma^{stock} \) and \( \sigma^{trsy} \). For each portfolio, we use eight lags of stock (left) or Treasury (right) returns to create out of sample bond return forecasts, and calculate pre-transaction cost profits of a strategy that goes long on bonds with the top \( x/2\% \) return forecasts and short on the bottom \( x/2\% \). The three panels report weekly profits (in basis points) when trading on the most extreme \( x = 50\% \), \( 10\% \), and \( 2\% \) return forecasts.

Predictions for year \( T \), we split the training sample before year \( T \) further into two halves by time, using the first half as the training sample and the second half as the validation sample to tune the penalty parameter. Because Table 11 implies that predictability is best when trading on roughly \( x = 10\% \) extreme predictions, we compute the profits from trading on these alternative prediction models 10% of the time. The results reported in Table 12 show that our conclusions are not sensitive to the prediction model used.
## Average weekly trading profits (basis points)

<table>
<thead>
<tr>
<th>Panel A: OLS, with 4 lags</th>
<th>Panel B: OLS, with 2 lags</th>
<th>Panel C: Elastic net regularization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_{stock} )</td>
<td>( \sigma_{trsy} )</td>
<td>( \sigma_{stock} )</td>
</tr>
<tr>
<td>quintile</td>
<td>quintile</td>
<td>quintile</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |
|                       |                       |                                   |</p>
<table>
<thead>
<tr>
<th>trading on lagged stock information</th>
<th>trading on lagged Treasury information</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{stock} )</td>
<td>( \sigma_{trsy} )</td>
</tr>
<tr>
<td>quintile</td>
<td>quintile</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

|                       |                       |                                   |
|                       |                       |                                   |

### Table 12. Average weekly profits when using different return forecast models.

This is a robustness check of Panel A in Table 7. Bonds are double sorted into 5 x 5 portfolios using \( \sigma_{stock} \) and \( \sigma_{trsy} \). For each portfolio, we form return forecasts using regressions with four lagged stock or Treasury returns (Panel A), regressions with two lagged returns (Panel B), or elastic net regularization with up to eight lagged returns (Panel C). We then report pre-transaction cost profits (in basis points) of a strategy that goes long on bonds with the top 5% return forecasts and short on the bottom 5%.
Figure 21. Distribution of predictive regression coefficients over time. Recall that bonds are sorted into $5 \times 5$ portfolios by the payoff-relevance measures $\sigma_{\text{stock}}$ and $\sigma_{\text{trsy}}$. In each year, for bonds in each of the 25 portfolios, we regress weekly bond returns on two lags of stock (left) or Treasury returns (right). We then plot the median (solid line) and 25, 75 percentiles (dotted lines) of the regression coefficients for each year.

E.3 Parameter instability: Why there is no Treasury-based predictability out of sample

While there is slow bond price reaction to Treasury movements, in the paper, we show that Treasury returns are not useful in predicting bond returns out of sample. Here we show evidence that this is due to time-variation in the forecasting parameter. To investigate parameter stability, for each year and each of the 25 exposure bins, we estimate:

$$\text{BondRet}_{i,t} = a + b^{\text{trsy}}_1 \cdot \text{TrsyRet}_{i,t-1} + b^{\text{trsy}}_2 \cdot \text{TrsyRet}_{i,t-2} + \epsilon_{i,t}. \quad (52)$$

We then plot the distribution (25, 50, and 75 percentiles) of the coefficients by year in Figure 21. For comparison, we also do the same exercise when using stock returns as predictors. There is quite a bit of parameter instability. While the estimated stock coefficients are at least usually above zero, the Treasury coefficients fluctuate wildly and frequently turns negative, even though they are on average positive, giving rise to the positive in-sample Treasury delay patterns in the paper.

We use a simple analytical framework to examine whether the negative predictability can be attributed to parameter instability. Consider the following simple example. Suppose the
true relationship in year $T$ is:

$$\text{BondRet}_{i,t} = a_T + b_T \cdot x_{i,t} + \epsilon_{i,t}$$

where $i$ indexes bonds and $x_{i,t}$ is a mean-zero i.i.d. regressor we are considering adding to our prediction. $\epsilon_i$ is a mean-zero, i.i.d. error term. Let $\hat{b}_{T-1}$ be our estimation of $b_T$, and let the estimation error be independent from everything else. For simplicity, assume $a_T$ is known. Then, using $x_{i,t}$ in the prediction reduces mean-square error if and only if:

$$E(\text{BondRet}_{i,t} - (a_T + \hat{b}_{T-1} \cdot x_{i,t}))^2 < E(\text{BondRet}_{i,t}^2)$$

$$\Leftrightarrow \sqrt{E((\hat{b}_{T-1} - b_T)^2)} < \sqrt{E(b_T^2)}$$

This exercise implies that adding $x_i$ as a predictor reduces prediction error if and only if $\sqrt{E(b_T^2)} - \sqrt{E((\hat{b}_{T-1} - b_T)^2)}$ is positive. Otherwise, it is more efficient to discard $x_i$. We apply this framework to explain negative return predictability when using lagged treasuries information. In each year $T = 2004, ..., 2014$, we estimate coefficients $b_T$ in (52), and compare them with the OLS estimates of $\hat{b}_{T-1}$ using data in year 2002, ..., $T-1$. In Figure 22, we plot the difference between the square root of the sample mean of $(\hat{b}_{T-1} - b_T)^2$ and the square root of $b_T^2$. We also plot the average weekly profit when trading on the lagged Treasury-based return predictor. Consistent with our analytical framework, negative profits happen exactly in years where the parameter estimation is off. We take this as evidence that parameter instability explains why it is not profitable to trade on lagged Treasury returns out of sample.
Figure 22. Using parameter instability to explain the variation in Treasury-based trading profits. The plot suggests that a simple measure of forecast parameter stability (black) can explain much of the time-variation in trading profits based on lagged Treasury returns (blue). Recall that bonds are sorted into $5 \times 5$ portfolios by the payoff-relevance measures $\sigma^{stock}$ and $\sigma^{trsy}$. In each year $T = 2004, \ldots, 2014$ and for each of the 25 portfolios, we use data up to year $T - 1$ to estimate how lagged Treasury returns forecast bond returns (regression (52)). The predictive coefficients are denoted by $b_{T-1}$. We also run the same regression using year $T$ data to get $b_T$. The black line plots estimates of $\sqrt{E(\hat{b}_{T-1} - b_T)^2} - \sqrt{E(b_T^2)}$ which, according to our analytical framework, measures the degree of bond-Treasury forecast stability. The blue lines are average weekly profits when trading bonds using lagged Treasury information.