Subjective risk perceptions and risk compensation *

Yunrong Zhou†

Aug 2025

Abstract

Understanding how agents perceive risks and demand compensation is of first-order importance, but has received little empirical examination. I construct a novel dataset linking bond analysts' recommendations to examine how they perceive risk and its associated compensation. I document two novel facts. First, I find that analysts favor higher-rated bonds, even when fundamentals are comparable across ratings. Second, perceived credit risk extracted from comments predicts subsequent credit deterioration but is not reflected in recommendations for higher-rated bonds. Traditional incentive-based explanations, including client or investor catering, cannot explain these patterns. Rather, I find evidence that these patterns are driven by categorical thinking—whereby analysts mental classify bonds based on rating labels, leading them to overvalue yields and underweight risks for purportedly safe bonds.

*I thank Huseyin Gulen, Kaushik Vasudevan, Sergey Chernenko, Dejanir Silva for regular meetings and feedbacks.

†Purdue University, zhou1181@purdue.edu

A key principle of asset pricing is that investors demand compensation for risk. Accordingly, it is of first-order importance to understand how agents perceive risks and form expectations of the compensation for risk. Despite their importance, risk perceptions have received little empirical examination, in part due to the difficulty of obtaining data. Indeed, as Adam and Nagel (2023) highlight, "*Subjective risk perceptions may be as important as the perceived first moments of returns and pay-offs... We need more work that explores how investors form beliefs about asset risks.*"

In this paper, I provide some of the first direct evidence on agent perceptions of risk and compensation by constructing and examining a novel dataset of bond analyst reports on corporate bonds spanning more than a decade. This setting is uniquely well-suited to study risk perceptions for several reasons. First, because the cash flows of corporate bonds are known in advance, the core driver of corporate bond pricing is perceptions of the risk that issuers may fail to make contractual payments, and the associated compensation demanded by investors. Second, analyst reports provide detailed comments on analysts' perceptions of bond risks, alongside explicit Overweight-Neutral-Underweight recommendations. These comments provide a rich source of information on the relevant risks and their pricing, as perceived by market participants. Finally, perceptions of risk can have significant real consequences. For example, overoptimistic beliefs about credit risk can drive excessive credit growth and increase the economy's vulnerability to financial crises (Minsky, 1977; Gennaioli, Shleifer and Vishny, 2012, 2015).

Analyzing the analyst reports, I document two novel facts. First, analysts tend to believe higher-yielding, riskier bonds provide higher compensation for the level of risk they perceive, but only within safer rating categories—even when fundamentals are comparable to lower-rated bonds. Second, analysts' comments predict future credit risk deterioration. However, analysts overlook their acknowledged risks when providing recommendations for higher-rated bonds, suggesting a gap between risk perceptions and compensation for risk. Traditional explanations, such as catering to clients or investors, cannot explain these patterns. Instead, I propose and provide evidence for categorical thinking—in which analysts mentally classify bonds using rating categories, causing them to overweight yields and underweight risks in purportedly safe bonds. My findings suggest that credit rating labels can significantly shape risk perceptions and associated compensation, and underscore the importance of incorporating category-based framing when modeling subjective risk perceptions and compensation in asset pricing.

I begin the empirical analysis by examining how analysts form recommendations. Their comments indicate that recommendations hinge on whether current yield spreads adequately compensate credit risk. Consistent with this interpretation, I test how yield spreads map

into recommendations and find a clear, robust asymmetry: analysts issue more favorable recommendations for bonds with higher yields, but only within the higher-rated segment (A-rated group). Quantitatively, a one-standard-deviation widening of the spread (about 86 basis points) increases the unconditional probability that an A-rated bond receives the most favorable recommendation by roughly seven percentage points in the cross-section, but has no effect on BBB-rated bonds. Moreover, this asymmetry persists even when comparing bonds simultaneously covered by the same analyst within the same industry and after controlling for comprehensive bond- and issuer-level characteristics, so it is unlikely to be driven by analyst preferences over these characteristics. Additionally, the tendency to recommend higher-yielding bonds is particularly acute when the yield curve is flat and default spreads are low, especially among A-rated bonds. In other words, analysts perceive the compensation of higher-yielding bonds as more attractive when overall market compensation for risk declines.

Next, I examine how analysts incorporate objective credit risk into their recommendations. The credit risk is measured using firm-level Expected Default Frequency (EDF), a forward-looking default probability derived from the Merton (1974) model framework. Analysts prefer bonds with higher credit risk only within A-rated bonds, indicating that their yield-seeking behavior reflects greater tolerance for credit risk. However, recommendations for BBB-rated bonds do not reward incremental default risk. That is, analysts perceive higher-yielding, riskier bonds as offering more attractive compensation relative to their perceived risk, but only among higher-rated bonds. Although regulatory capital requirements often differ sharply between A- and BBB-rated bonds for institutional investors, the dominant holders of corporate bonds, these differences alone appear insufficient to explain why analysts treat yield spreads and objective risk so differently across investment-grade bonds.

To better understand how yields and perceptions of risk affect analysts' recommendations, I study the text of analysts' comments. Specifically, I investigate how analysts differ across rating categories along two key dimensions: (1) assessment of credit risk, and (2) interpretation of yield information. To investigate how analysts assess risk, I construct two novel quantitative measures of forward-looking risk perceptions inferred from comments.[1] One measure involves prompting a large language model (LLM) to parse each comment carefully, evaluate both the frequency and intensity of risk-related language, and assign a risk score. The other measure follows a traditional textual analysis approach (Loughran and McDonald, 2011), counting occurrences of uncertain, weak modal, and negative words within analyst comments.

[1]The analysts' comments encompass the assessment of recent conditions and the projection of those perceptions into the future, consistent with the model framework of forward-looking risk perceptions in (Pflueger, Siriwardane and Sunderam, 2020).

By relating analysts' perceived risk scores to EDF, I examine how analysts translate public information into risk perceptions. Intuitively, credit ratings serve as reference points, and a moderate rise in EDF may be viewed as "normal" for BBB-rated issuers but alarming for A-rated issuers. Analysts may therefore exhibit greater sensitivity to EDF changes among A-rated bonds. However, the data reveals a significantly flatter slope for higher-rated bonds, probably suggesting attenuated recognition of credit risk.

Such attenuation may not matter economically. I find that analysts' risk assessments still contain forward-looking credit signals that forecast subsequent credit risk realizations. However, analysts' recommendations fail to forecast them in the higher-rated segment. This suggests a gap between analysts' informative risk comments and their not informative recommendations for A-rated bonds to predict future credit events.

Motivated by this gap, I examine how recommendations load on perceived-risk scores. Controlling for yield spreads, higher perceived risk leads analysts to issue less favorable recommendations for BBB-rated bonds, but has no significant effect for A-rated bonds. This selective response explains the divergence between analysts' stated risk concerns and their actual recommendations: analysts explicitly acknowledge downside risks but systematically fail to reflect them in recommendations for purportedly safe bonds. Such behavior aligns with the "local thinking" theory of risk neglect (Gennaioli, Shleifer and Vishny, 2012), wherein investors underestimate adverse outcomes deemed improbable. Given the lower default risk and regulatory capital charges associated with A-rated bonds, analysts may perceive adverse scenarios as unlikely, discounting their own stated concerns. Such rating-related risk neglect also aligns closely with psychological explanations based on representativeness, the cognitive bias where people overestimate the probability of outcomes perceived as more likely (Kahneman and Tversky, 1972; Gennaioli, Shleifer and Vishny, 2015). Moreover, this finding provides the first empirical support for the mechanism outlined by Becker and Ivashina (2015), who suggest that reaching-for-yield behaviors can reflect underlying risk neglect.[2]

Next, by investigating how analysts evaluate yield information, I find that the yield asymmetry arises when analysts translate yield spreads into qualitative descriptions. Specifically, analysts systematically describe equivalent increases in yield spreads as more attractive for A-rated bonds after controlling for their perceived risks. Once analysts characterize yields as attractive, they issue more favorable recommendations across both rating categories.

Evaluating analysts' ability to generate α, I find that the portfolios that go long on Overweight and short on Underweight produce insignificant monthly alpha for both A-rated

[2]From Becker and Ivashina (2015), *"We expect reaching for yield over 'post-crisis' period to become weaker due to an increase in downgrade probabilities, scrutiny, and awareness of previously neglected risks, or a shift in risk preferences."*

bonds (-2.1 bps, t-stat = -0.42) and BBB-rated bonds (5 bps, t-stat = 1.41) after controlling for the market risk factor. When using other common factors, the more persistent positive alphas are observed in BBB-rated bonds, suggesting that analysts more carefully incorporate risk considerations into their recommendations in lower-rated bonds.

The above empirical results raise two puzzles. First, why do analysts reward higher yield and risk only for A-rated bonds? Second, why do they neglect risks that they already explicitly recognized in A-rated bonds but not in those rated BBB? I consider three potential mechanisms: (1) catering to clients, (2) catering to investors, and (3) behavioral explanations.

Traditional catering explanations appear insufficient to explain these asymmetric patterns. Specifically, agency conflicts may incentivize analysts to issue strategically optimistic recommendations for bonds issued by their banks' clients, potentially generating the A–BBB recommendation asymmetry. To test this catering-to-client mechanism, I link explicitly disclosed client relationships from analyst reports to bond recommendations, and find that the A–BBB asymmetry persists within client and non-client subsamples, across current, past, or prospective client ties. In addition, analysts may strategically tailor their recommendations to institutional investors' preferences. Specifically, they might cater to insurance companies in the A-rated segment, as insurers face regulatory constraints, primarily hold A-rated bonds, and systematically reach for yield (Becker and Ivashina, 2015), and cater to mutual funds in the BBB-rated segment, as these funds face fewer constraints and exhibit limited yield-seeking behavior (Choi and Kronlund, 2018). I test this mechanism by linking analyst recommendations to institutional ownership data and find no supporting evidence.

I propose and provide evidence of a mechanism related to psychology: categorical thinking. According to Mullainathan (2002), individuals simplify inference by relying on coarse categories rather than continuously updating beliefs as rational Bayesians. Credit ratings provide precisely such natural groupings—salient, widely recognized, and reinforced by regulations—with the most prominent distinction within investment grade between A and BBB. Under this categorization, analysts may adopt different stereotypes: they treat A-rated bonds as inherently safe, leading them to overweight yield and discount negative risk signals, while viewing BBB-rated bonds as inherently risky, causing them to regard incremental yield increases as insufficient compensation for additional risk. This cognitive framing can simultaneously generate risk-underweighting and yield-seeking dynamics.

Since ratings reflect firm fundamentals, rating labels alone carry no incremental information for predicting future bond pricing. Rational analysts, conditional on these fundamentals, should respond smoothly and consistently to credit risk and spreads across ratings. Any discontinuity at the A/BBB boundary would thus indicate that analysts' inference depends on

rating labels themselves, contradicting rational updating. To test this, I match A-rated and BBB-rated firms on industry and key financial characteristics. Within this matched sample, analyst recommendations still load positively with credit spreads and credit risk for A-rated bonds but not for BBB-rated bonds. This difference in slopes across rating categories is statistically significant, providing direct evidence of categorical thinking.

I provide additional supportive evidence through textual analysis on analyst comments. Analysts employ coarser reasoning by using fewer information categories, frequently omit factual details, and only provide subjective predictions when describing A-rated bonds compared to BBB-rated bonds. Moreover, favorable recommendations predict downgrades for BBB+ bonds but not for A- bonds, while predicting upgrades for A- bonds but not for BBB+ bonds. These patterns are consistent with categorical thinking: analysts neglect downside risk in higher-rated bonds and discount upside potential in riskier-rated bonds.

My work is related to a nascent literature that studies subjective risk perceptions. Existing literature primarily infers market-level risk perceptions using surveys (Giglio, Maggiori, Stroebel and Utkus, 2021; Couts, Gonçalves and Loudis, 2023; Jo, Lin and You, 2024) or observed financial outcomes (Pflueger, Siriwardane and Sunderam, 2020; Lochstoer and Muir, 2022; Nagel and Xu, 2023). My paper advances this literature by constructing a new dataset where market participants directly discuss their risk perceptions at the firm/bond level and explicitly linking these perceptions to required risk compensation.

My paper also contributes to recent work on risk neglect in credit markets. Minsky (1977) argues that overly optimistic investor beliefs can drive credit expansion and trigger financial crises. Theoretically, Gennaioli, Shleifer and Vishny (2012) argues that such overoptimism often takes the form of risk neglect. Empirical studies have largely assumed or indirectly inferred risk neglect from the finding that investors do not receive higher yield compensation for taking on higher risk (Chernenko, Hanson and Sunderam, 2016; Sørensen, 2022; Zhang, Zhao and Zhao, 2024). My paper provides direct empirical evidence of risk neglect: professional bond analysts neglect their acknowledged risks for safe rated bonds. Such rating-induced neglect aligns closely with psychological explanations based on representativeness, the cognitive bias where people overestimate the probability of outcomes perceived as more likely (Kahneman and Tversky, 1972; Gennaioli, Shleifer and Vishny, 2015).

Additionally, my work relates to research on reaching for yield (RFY)—the preference for higher yielding bonds within a given rating category. This behavior has been documented among institutional investors, especially within higher rated segments (Becker and Ivashina, 2015; Choi and Kronlund, 2018; Chen and Choi, 2024). My paper provides the first empirical support for the potential mechanism proposed by Becker and Ivashina (2015), showing that RFY behavior in higher-rated bonds reflects risk neglect induced by categorical thinking.

More broadly, my paper relates to empirical work studying expectation formation. A rich literature has explored it in various asset classes such as equities, real estate, and interest rates. Prior studies typically infer expectations in bond markets from related markets (e.g., equity or interest rate markets) and emphasize aggregate-level analysis (Piazzesi, Schneider et al., 2009; Buraschi, Piatti and Whelan, 2022; Gulen, Ion, Jens and Rossi, 2024). My study contributes to this literature by offering the first direct empirical evidence on expectations in the corporate bond market, with a particular focus on the cross-section.

1 Data

This section describes the construction of my main datasets, starting with corporate bond analyst reports. These reports are merged with issuer-level characteristics from CRSP and Compustat, and with security-level information using the CRSP Corporate Bond Panel. I also use other datasets such as manually constructed analyst characteristics, client information, and eMAXX holdings. Finally, I address data-related issues before turning to results.

1.1 Corporate bond analyst reports

Analysts evaluate issuers' credit fundamentals to assess their ability to meet debt obligations and provide issuer- or bond-level recommendations in three categories: Overweight, Neutral, or Underweight. The recommendations are based on expected performance relative to the rating index (A and BBB) or sectors over the next three months.[3] To form recommendations, analysts mainly consider credit ratings from credit rating agencies, current yield spreads (emphasized in bold in the report), and some other firm and bond information. Additionally, research quality, accuracy, and client feedback can affect analysts' compensation.

I focus on Investment Grade (IG) industry summary reports from September 2013 to March 2023. The reports were issued monthly until 2017, then quarterly from 2017 to 2020, and semiannually. My analysis focuses on U.S. public non-financial firms and the final sample includes over 20 industries. Table 1 shows the number of times covered for each industry. Most industries are covered repeatedly. I apply issuer-level recommendations to individual bonds following disclosures in the IG reports and also to better correct for bond characteristics, such as duration and coupon rate.[4]

[3]From the disclosures of IG reports: Overweight, Neutral, and Underweight indicate expectations that, over the next three months, the recommended risk position will outperform, perform in line with, or underperform the relevant benchmark, respectively. The A-rated category primarily encompasses A+, A, and A– bonds, with AA and AAA-rated bonds accounting for approximately 6% of the sample. The BBB-rated category comprises BBB+, BBB, and BBB– bonds.

[4]According to the reports, *"an issuer recommendation applies to all bonds at the same level of the issuer's capital structure unless I specify a different recommendation for the individual security."* Analysts

The final sample comprises 327 unique non-financial firms with 104,534 bond-month recommendations, contributed by over 40 analysts over a decade. Panels A, B, and C of Table 2 provide a detailed breakdown of issuers, bonds, and recommendations. Issuers in my sample tend to be larger firms and issue more bonds compared to all public issuers in the WRDS Corporate Bond Dataset, as shown in Panel A of Table A2. In the IG market, the median size of issuers in my sample is $20,286.37 million in 2018 dollars, compared to $14,357.75 million in the full sample. Additionally, my bond-month panel closely represents the broader bond universe along key dimensions. During the 2010–2023 period, it covers approximately 80% of all investment-grade bond-month observations in the CRSP/TRACE universe (Panel B of Table 2 and Table A2), with similar distributions of credit ratings, durations, yield spreads, and monthly returns.

Regarding recommendation characteristics, Figure 1 shows an asymmetric distribution of analyst recommendations across rating groups: analysts issue less favorable recommendations (underweight) for higher-rated bonds and more favorable recommendations (overweight) for lower-rated bonds.[5] Alongside their recommendation, analysts also provide detailed comments to give a rationale for their recommendations and focus on their perceptions of bond risks.

1.1.1 Constructing variables from comments by LLM

I generate a number of empirical measures of interest by analyzing comments using Large Language Models (LLMs). This approach leverages recent advances in natural language processing and addresses limitations inherent in traditional textual methods. Traditional methods, such as keyword counts or unsupervised topic models, rely heavily on predefined dictionaries or statistical groupings. Such reliance restricts their ability to interpret nuanced financial contexts or answer analytical queries. In contrast, LLMs employ transformer-based neural networks trained extensively on large-scale textual data. These models can capture subtle economic meanings and respond to targeted questions. For instance, analyst comments usually combine factual information (e.g., reported earnings) with subjective interpretations (e.g., earnings forecasts). Traditional methods cannot reliably separate objective financial metrics from forward-looking judgments.

generally issue uniform recommendations across a firm's bonds. However, in fewer than 5% of the firm-month observations, they assign different recommendations to individual bonds based on specific factors, such as the bond's maturity date or its association with the firm's primary business. In such cases, I review their comments to identify the rationale and apply it consistently across the firm's bonds.

[5]This pattern is consistent with the finding in De Franco, Vasvari and Wittenberg-Moerman (2009), which shows that the positive skewness in bond analysts' recommendations is greater for low than high credit quality bonds, largely due to the lack of upside return potential and bond investors' more asymmetric demand for negative information in high-quality bonds.

I use Llama-3.3-70b, the largest publicly available variant of Meta's third-generation Llama model family.[6] Unlike proprietary models, Llama-3.3-70B provides full access to its weights, and such openness enables deterministic replication, in-domain fine-tuning, and local deployment. These features ensure methodological transparency, making the model particularly suitable for extracting structured financial information (Meta AI, 2024b,a).

In section 3, I discuss empirical measures and prompts and present the full text of the prompts in section IA.A To ensure the reliability and reproducibility of the LLM's outputs, I implement several steps. First, I adopt a "chain-of-thought" prompting approach, requiring the LLM to articulate intermediate reasoning steps before providing a final answer. Prior research has found that chain-of-thought prompting significantly enhances model accuracy, especially in complex tasks (Wei et al., 2022). Second, I explicitly instruct the LLM to support its reasoning by directly quoting relevant passages from analyst comments, firmly anchoring its conclusions in the source text. Finally, wherever possible, I validate the LLM-generated variables against observable numerical data, ensuring empirical consistency and accuracy throughout my analysis.

1.2 CRSP and Compustat

Bond-month level information, such as yields, returns, maturity, and ratings, are from the WRDS Corporate Bond Panel. This panel merges TRACE trade data with FISD security descriptors and provides an end-of-month quote for every CUSIP of publicly traded U.S. corporate bonds starting in 2021. I clean the raw data following Dick-Nielsen, Feldhütter, Pedersen and Stolborg (2023). The yield spread is measured as the bond's yield minus the interpolated Treasury yield of the same maturity following Dickerson, Mueller and Robotti (2023b). U.S. Treasury yield curve is obtained from The Federal Reserve Board.[7] I include bonds that have available information on credit rating, yield spread, and time to maturity, with maturities exceeding one year. Also, I exclude all convertible and putable bonds along

[6] On standard NLP benchmarks, Llama-3.3-70B was released April 18, 2024 and achieves 82% accuracy on MMLU and 64% on GSM-8K. Its performance is comparable to GPT-3.5 Turbo and within six percentage points of GPT-4. However, it incurs less than one-tenth the inference cost of GPT-4. The model includes 70 billion parameters and supports an 8,000-token context length. It was trained on approximately 15 trillion tokens sourced from news articles, code repositories, and specialized domain texts.

[7] For a given target maturity τ^*, I construct the synthetic Treasury yield y_{τ^*} by linear interpolation:

$$y_{\tau^*} = y_{\tau_L} + \frac{y_{\tau_U} - y_{\tau_L}}{\tau_U - \tau_L}(\tau^* - \tau_L),$$

where

$$\tau_L = \max\{\tau \le \tau^* : \text{Treasury yield at } \tau \text{ is observed}\}, \quad \tau_U = \min\{\tau \ge \tau^* : \text{Treasury yield at } \tau \text{ is observed}\}.$$

with bonds with sinking fund provisions in order to better identify the relationship between credit risk and yield spreads.

Firm-level fundamentals come from the CRSP and Compustat database, from which I construct financial variables used in the matching procedure described in Table 2. Bond characteristics such as returns, yields, and firm characteristics are all winsorized at the 1% level to mitigate concerns that outliers might be driving the results.

1.2.1 No break in firm characteristics at A-BBB cutoff

Our identification strategy and falsification tests require differences in firm characteristics at the A-BBB cutoff to be similar to those across other rating cutoffs. Panel D of Table 2 reports the means of firm characteristics by credit rating from AAA down to BBB-. There are few AAA and AA firms, so I don't use more granular rating notches within these two rating groups.

As expected, as ratings fall, firms become smaller, more leveraged, and less solvent, i.e., Altman's Z-score falls in near-linear steps from 1.57 (AAA) to 0.65 (BBB-). Profitability measures, such as ROA, operating margin, and cash flow to assets, also decrease notch by notch, whereas asset tangibility (PPE/Assets) increases, consistent with collateral-backed borrowing among riskier issuers.

What matters for identification is to see whether the A/BBB cutoff shows a structural break in fundamentals. Moving from A to BBB+ reduces log assets by roughly 0.5 points, raises market leverage by five percentage points, and lowers the Z-score by 0.17. These shifts are indistinguishable from the changes observed between other rating notches, and a few variables even look slightly stronger for BBB+ than for A issuers. Such patterns support the core assumption of the matching design. Once matched on observables, A and BBB-rated issuers are comparable counterfactual pairs for testing how rating label affects analysts' process of credit information.

1.3 Others

1.3.1 Analyst characteristics

I manually collect demographic and career details for each analyst in my sample from LinkedIn and MarketScreener profiles. The data include gender, birth year, undergraduate major, highest degree, prior work experience, year joined J.P. Morgan, and LinkedIn followers. Approximately 40 percent of analysts are female. Two-thirds hold economics degrees, with the remainder in STEM fields. Forty percent hold master's degrees.

1.3.2 Client information

I manually collect and analyze disclosures from analyst reports. Each disclosure explicitly identifies the bank's specific relationship with issuers. I consolidate the client relationships into three categories: (i) investment banking, which primarily includes lead manager or co-manager underwriting roles and broader advisory services such as mergers, restructuring, etc; (ii) broker and market maker, which are related to liquidity providing and secondary-market trading; and (iii) other significant financial interests, primarily reflecting proprietary positions of analysts' employer in issuing debt securities exceeding one million USD. Among the matched relationship-linked observations, underwriting mandates represent the largest category.

1.3.3 eMAXX holding

For each bond in our sample, we extract its institutional holdings data from Thomson Reuters eMAXX, covering the most sample period from 2013Q4 to 2022Q4. eMAXX contains detailed fixed-income holdings for insurance companies, mutual funds, pension funds, and other investors at the quarterly level. eMAXX compiles holding disclosures from mandatory regulatory filings with the National Association of Insurance Commissioners (NAIC) for insurers, and with the Securities and Exchange Commission (SEC) for mutual funds, asset managers, and public pension funds, along with voluntary disclosures from major private pension plans. However, coverage of banks and hedge funds is limited (Becker and Ivashina, 2015).

2 Recommendation formation

2.1 Recommendation formation: yield spread and risk

Analysts evaluate issuers' credit fundamentals to assess their ability to meet debt obligations and provide recommendations based on expected performance relative to the index or sectors over the next three months. To form these recommendations (e.g., Overweight, Neutral, or Underweight), analysts assess whether the bond's yield spread offers sufficient compensation for its perceived credit risk. They incorporate credit ratings from agencies, market-implied indicators such as yield spreads (highlighted in bold in reports), and other relevant information.

To formalize the decision framework on how corporate bond analysts issue recommendations, I first express the bond's excess return explicitly following Elton, Gruber, Agrawal and Mann (2001); Brooks, Katz and Richardson (2016). Over a horizon Δ, the realized excess

return of a bond i can be approximated as:

$$\text{ExcessRet}_{i,t\to t+\Delta} = \underbrace{S_{i,t} \times \Delta}_{\text{Carry}} - \underbrace{Dur_{i,t}\left(S_{i,t+\Delta} - S_{i,t}\right)}_{\text{Price change}}, \tag{1}$$

where $\text{ExcessRet}_{i,t\to t+\Delta}$ is the bond's excess holding-period return relative to a maturity-matched Treasury over a horizon of length Δ (e.g., $\Delta = 0.25$ for holding three months). This excess return can be decomposed into two components. The first component, carry, is the compensation investors earn from holding the bond based on its initial yield spread, $S_{i,t}$. The second component, $-Dur_{i,t}(S_{i,t+\Delta} - S_{i,t})$, captures the price changes due to spread movements. Here, $Dur_{i,t}$ is the bond's duration.

Then, consider yield spread $S_{i,t}$, it's a sum of expected default loss, risk premium and others (Becker and Ivashina, 2015):

$$S_{i,t} \approx \underbrace{\text{PD}_{i,t} \times \text{LGD}_{i,t}}_{\text{Expected default loss}} + \underbrace{\phi\big(\text{PD}_{i,t},\, \text{LGD}_{i,t},\, \delta_t,\, M_t\big)}_{\text{Default risk premium}} + \underbrace{\ell_{i,t}}_{\text{Liquidity and other risk premium}} + \underbrace{\alpha_{i,t}}_{\text{Mispricing}}, \tag{2}$$

where PD_i is the market-assessed probability of default. LGD_i is the market-assessed loss given default. δ is the market overall risk preference and M is the market price of bearing each unit of risk, which I assume are common to all bonds at a given date.

Combining Eqs. (1) and (2), the expected excess return on bond i over a horizon Δ is

$$\begin{aligned}
\mathbb{E}_t\big[\text{ExcessRet}_{i,t\to t+\Delta}\big] &= S_{i,t}\,\Delta \;-\; \text{Dur}_{i,t}\left(\mathbb{E}_t[S_{i,t+\Delta}] - S_{i,t}\right) \\
&= S_{i,t}\big(\Delta + \text{Dur}_{i,t}\big) \;-\; \text{Dur}_{i,t}\,\mathbb{E}_t[S_{i,t+\Delta}] \\
&= \underbrace{S_{i,t}\big(\Delta + \text{Dur}_{i,t}\big)}_{\substack{\text{Carry plus}\\ \text{duration-adjusted}\\ \text{initial spread}}} \;-\; \underbrace{\text{Dur}_{i,t}\,\mathbb{E}_t\Big[\text{PD}_{i,t+\Delta}\,\text{LGD}_{i,t+\Delta} + \phi_{i,t+\Delta} + \ell_{i,t+\Delta} + \alpha_{i,t+\Delta}\Big]}_{\substack{\text{Duration-adjusted}\\ \text{expected yield spread}}}.
\end{aligned} \tag{3}$$

Across the cross-section at a given point in time, assuming homogeneous market risk preferences (δ) and a common market price of risk (M), a bond's expected excess return over a fixed forecast horizon should be positively associated with its current yield spread ($S_{i,t}$) and negatively associated with its future default risk, after properly accounting for loss given default, liquidity risk, and other bond-specific factors affecting future yields. In the columns (5) and (11) of **??**, I confirm that future three-month excess returns are positively related to the current spread and negatively related to default risk empirically.

I use EDF as the proxy for credit risk. $EDF_{i,t}$ is the firm's expected default frequency

and reflects the probability that a firm's asset value will fall below its debt obligations over a fixed horizon. It is derived from the structural model of Merton (1974) and calculated following the implementation in Bharath and Shumway (2008). Specifically, EDF is computed as the standard normal cumulative density function evaluated at a standardized distance to default:

$$EDF_{it} = \mathcal{N}\left(\frac{-\log(V_{it}/B_{it}) - \left(\mu_{V_{it}} - \frac{1}{2}\sigma^2_{V_{it}}\right)}{\sigma_{V_{it}}}\right), \tag{4}$$

where V_{it} is the market value of firm i's assets, B_{it} is the book value of its debt, $\mu_{V_{it}}$ is the expected asset return, and $\sigma_{V_{it}}$ is the asset return volatility. Conceptually, EDF captures the distance between the firm's asset value and its debt obligation, scaled by the standard deviation of the firm's asset value.

I use EDF as my primary measure of firm-level credit risk for several reasons. First, it is a theory-based forward-looking measure, which incorporates market-based inputs that adjust continuously with changes in firm value and volatility. Second, prior research shows that EDFs strongly predict actual default events and credit rating downgrades. I also confirm this using my sample. Finally, it's widely used in both academia (e.g., Gilchrist and Zakrajšek 2012, Greenwood and Hanson 2013, Gomes et al. 2019) and industry, making EDF a natural proxy for capturing objective credit risk in my analysis.[8]

2.2 Recommendation and yield spreads

Building on the above conceptual framework, I first look at how yield spread affects recommendations. The empirical strategy is delineated as follows:

$$\text{Rec}_{i,t} = \alpha + \sum_{g \in G} \beta_{1,g}\, \text{YieldSpread}_{i,t} \times G_{g,i,t} + \beta_2\, \mathbf{1}_{\text{g=A},i,t} + \beta_3' \mathbf{X}_{i,t} + \epsilon_{i,t}, \tag{5}$$

where $Rec_{i,t}$ is the analyst recommendation for bond i at time t. $YieldSpread_{i,t}$ is the bond's yield spread over the yield of treasuries in the same maturity. $G_{g,i,t}$ indicates the bond's rating group, where $g = A$ or BBB, and the interaction terms allow the sensitivity of recommendations to yield spreads to vary across rating groups. Table 3 reports the results. In column (1), the interaction term for yield spreads is positive and highly significant for A-rated bonds, whereas it is much smaller and statistically insignificant for BBB-rated bonds. A one-standard-deviation widening in the spread (about 86 basis points) raises the unconditional probability that an A-rated bond receives an Overweight recommendation by roughly seven percentage points in the cross section, while leaving the probability for BBB-rated bonds

[8] See Eisenbach, Kovner and Lee (2017); Moody's Analytics (2024)

unchanged.[9]

The industry-specific comparison matters if analysts rank bonds relative to sector peers rather than the whole cross-section. Hence, in column (2), I use industry-fixed effects, and the A-rated group slope falls from 21.76 to 12.884 and remains highly significant. In contrast, the BBB slope remains economically and statistically negligible. In the next two columns, I focus on the cases where the same analyst covers both A-rated and BBB-rated bonds in the same time and industry. Column (3) introduces analyst fixed effects to absorb persistent analyst styles. Column (4) further adds industry fixed effects and the results keep consistent, indicating that the asymmetric pattern is unlikely to be explained by analysts' recommendation style.

In the remaining columns, I test whether the credit-spread slope for A bonds differs significantly from the slope for BBB bonds and run the following regression:

$$\text{Rec}_{i,t} = \alpha + \beta_1 \, \text{YieldSpread}_{i,t} \times G_{g=A,i,t} + \beta_2 \, \mathbf{1}_{\text{g=A},i,t} + \beta_3' \mathbf{X}_{i,t} + \epsilon_{i,t}, \tag{6}$$

where the coefficient of $BBB \times yieldspread$ is equal to the slope gap relative to A. The baseline difference in column (5) is 19.794 and remains large and significant when the comparison is confined to the same industry, the same analyst, or both. Using ordered-logit regression reported in Table A3, this gap means that a one-standard-deviation widening in the spread significantly increases the likelihood of an Overweight by roughly seven percentage points for A-bucket bonds relative to BBB-bucket bonds.

Finally, I examine whether bond and firm characteristics explain the pattern. Following Gilchrist and Zakrajšek (2012), I control for bond-level characteristics that could influence bond yields through term or liquidity premiums, including bond's duration, issue size, coupon rate, issue age, and an indicator for callability. I also control for key firm characteristics reflecting financial health, including size, leverage, cash holdings, PPE to assets, cash flow to PPE, and ROA. Analyst reports disclose that they use these characteristics to formulate credit recommendations. Columns (9)–(12) incorporate these bond and firm characteristics, and the A-minus-BBB yield spread coefficients remain statistically significant. Thus, bond and firm fundamentals cannot explain the differential analyst response. Notably, even comparing bonds from the same industry covered by the same analyst and controlling for firm

[9]Marginal effects are calculated using predictions from the ordered logit estimates in Table A3. Specifically, I first evaluate the predicted probability of an Overweight recommendation at the mean credit spread level observed in the sample, and then recalculate the predicted probability after increasing the spread by one standard deviation. The difference between these two probabilities, computed separately for A-rated and BBB-rated bonds, yields the marginal effect. For A-rated bonds, the probability of an Overweight recommendation increases from approximately 10 percent to 17 percent, while for BBB-rated bonds, the corresponding change is negligible (from 26.52 percent to 27.58 percent.

and bond characteristics, analysts still issue more favorable recommendations to A-rated bonds as yield spreads widen, while leaving BBB-rated bonds unchanged or downgraded.

2.3 Recommendation and credit risk

Next, I assess how analysts incorporate credit risk into their recommendations by replacing yield spreads with EDF in Equation 5 as the measure of default risk (Panel A, Table 4). Column (1) shows that EDF coefficients are significantly positive for A rated bonds, but small and statistically insignificant for BBB rated bonds. Thus, analysts systematically upgrade A rated bonds as model-implied default risk increases, but remain unresponsive to similar changes among BBB rated bonds. Adding industry fixed effects in column (2) modestly reduces the A-rated coefficient, without affecting its significance. Columns (3) and (4) isolate within-analyst comparisons. The coefficient for A rated bonds remains positive and significant, while the BBB coefficient remains insignificant. These findings suggest that analysts' asymmetric responses are not driven by differences in analyst risk preferences or coverage style.

Columns (5) through (8) directly measure the difference in EDF slopes between A and BBB bonds by interacting EDF with a BBB indicator. The resulting treatment gap remains consistently large and negative, even when restricting comparisons to bonds within the same industry, covered by the same analyst, or both. The ordered logit estimates in Table A3 further illustrate this asymmetry. Raising EDF from its mean (0.565) to one standard deviation above (0.685) increases the likelihood of an Overweight recommendation for A rated bonds by roughly 4.6 percentage points (from 8.6% to 13.2%), yet leaves recommendations for BBB rated bonds effectively unchanged. Columns (9)–(12) add bond-level and firm-level controls, and the treatment gap remains virtually identical, confirming it is not driven by observable bond or firm characteristics. Overall, analysts interpret the same default-risk signals differently across the A and BBB rating groups.

Panel B of Table 4 brings yield spreads and EDF together in the same regression, so that I can test whether one signal subsumes the other and to see whether the asymmetry survives once analysts can react to both yield spread and credit risk at the same time. By including both variables, the estimate captures the incremental weight analysts assign to each signal while holding the other signal constant.

The first four columns show the joint model with both signals in the equation. Compared with the single-signal specifications in Panels A and B, the pattern is remarkably stable. The interaction between EDF and the A rating remains positive, while the corresponding BBB interaction remains negative. The spread interaction is still strongly positive for A-rated bonds and economically negligible for BBB-rated bonds. The slight reduction in the credit-

spread coefficient indicates that the spread and EDF share some information, but each still contributes distinctly to explaining recommendations once both are included.

Columns (5)–(12) show that, with both EDF and the spread in the equation, the BBB-interaction terms for each signal remain negative and highly significant, only marginally smaller than in the single-signal panels. Adding industry and analyst fixed effects or bond controls leaves these gaps essentially intact. Including both EDF and the spread raises the within-pair R square by only a few percentage points, because each signal already captures most of the default-risk variation that matters once pair-month effects are removed.

Across every specification the standalone BBB indicator carries a positive and significant coefficient. One interpretation is that, once fundamentals are matched, a BBB label signals misclassification: the bond looks A-like on the fundamentals and analysts anticipate an eventual upgrade, so they lean more favorably toward it. A second possibility is that analysts may infer that BBB bonds, all else equal, offer a better payoff than their A-rated twins and therefore deserve a stronger recommendation.

2.4 Yield-seeking behavior over time

Analysts' recommendations may respond to shifts in market conditions. Institutional investors, driven by agency conflicts and client demands, often seek higher-yielding securities during periods of low interest rates to boost portfolio performance (Becker and Ivashina, 2015; Choi and Kronlund, 2018; Acharya and Naqvi, 2019). As information intermediaries, analysts may cater to their clients or achieve their own earning goals by assigning greater weights to yield spreads when safe yields are low or compensation for bearing risk declines.

To examine this possibility, I track how their sensitivity to yield spreads evolves over time. Figure 2 plots the time series of yield spread coefficients alongside key measures of aggregate market compensation for risk. First, the one-year Treasury rate shows no negative relation with the spread coefficient, likely because policy rates were persistently low—between zero and two percent—throughout most of the sample. Second, the default-risk premium, measured as the BBB–AAA corporate spread, is strongly negatively related to the yield spread coefficient, with correlations of –0.28 for A-rated bonds and –0.26 for BBB-rated bonds: as the market's compensation for credit risk narrows, analysts become more responsive to yield spreads, especially for A bonds. Finally, I evaluate the term spread, the difference between 30-year and 1-year Treasury yields, which measures compensation for bearing duration risk. Similarly, when the yield curve flattens and duration compensation decreases, analysts again exhibit greater sensitivity to yield spreads for A-rated bonds (correlations of -0.52 for A-rated bonds and -0.5 for BBB-rated bonds).

In sum, analysts' recommendations assign greater weight to yield spreads when market

compensation for bearing risk declines, particularly within A-rated securities. Although analysts do not directly invest, their recommendations can shape investor portfolio allocations. Thus, analyst behavior can amplify investors' incentives to reach for yield, especially in low-risk compensation environments, consistent with well-documented patterns of investor risk-taking behaviors (Becker and Ivashina, 2015; Choi and Kronlund, 2018; Lian, Ma and Wang, 2019).

2.5 Cross-sectional variation of yield-seeking behavior

Prior literature shows that individual characteristics such as gender, experience, and education significantly influence how market participants interpret information and formulate forecasts (Bradley, Gökkaya and Liu, 2017; Cole, Paulson and Shastry, 2014; Peng, Teoh, Wang and Yan, 2022; Bacher, 2024). Motivated by these findings, I explore whether these analyst-specific traits amplify or mitigate the observed asymmetry in responses to yield spreads.

To test for such heterogeneity, I augment the baseline specification by interacting analyst characteristics directly with the difference in yield-spread sensitivity between A-rated and BBB-rated bonds. Specifically, I estimate the following model:

$$\begin{aligned}\text{Rec}_{it} &= \alpha_0 + \beta_1 A_{i,t} + \beta_2 \,\text{Yield Spread}_{it} + \beta_3\big(A_{i,t} \times \text{Yield Spread}_{it}\big) + \beta_4 M_{i,t} + \beta_5\big(A_{i,t} \times M_{i,t}\big) \\ &\quad + \beta_6\big(M_{i,t} \times \text{Yield Spread}_{it}\big) + \beta_7\big(A_{i,t} \times M_{i,t} \times \text{Yield Spread}_{it}\big) + \gamma_{yrmon} + \gamma_{industry} + u_{it}.\end{aligned} \tag{7}$$

In this specification, A_i equals one for A-rated bonds and zero for BBB-rated bonds. M_j denotes one of analyst-specific characteristics, including gender, major, highest degree obtained, LinkedIn follower count, and tenure at JP Morgan. The key coefficient of interest is the triple interaction term $A \times YieldSpread \times Trait$. This coefficient measures how analyst characteristics influence the difference in yield-spread sensitivity between A-rated and BBB-rated bonds. Table 5 presents the results of β_7. Columns (1)–(2) link recommendation asymmetry to gender and show that male analysts exhibit a significantly smaller yield-seeking gap between A-rated and BBB-rated bonds compared to female analysts.

It's possible that an educational background and training shape analysts' frameworks for forming expectations. To investigate this dimension, I use STEM versus economics undergraduate majors and master's versus bachelor's degrees as proxies for analytical rigor and training intensity. Columns (3) to (6) show that analysts with STEM degrees or master's degrees display a significantly reduced yield-seeking gap relative to those with economics majors or bachelor's degrees, respectively.

In addition, analysts' professional incentives, shaped by career tenure and public visibility,

can further influence their recommendation patterns. I proxy public visibility with two top-half indicators: high public visibility, defined as having more than 750 LinkedIn followers (the median in my sample), and long firm tenure, defined as more than ten years at J.P. Morgan (again, the sample median). Columns (7)–(10) reveal that analysts with higher public visibility or longer professional tenure exhibit a significantly larger yield-seeking gap between A-rated and BBB-rated bonds. These findings imply that career incentives and reputational concerns amplify analysts' selective responsiveness to yield spreads.

Taken together, these results demonstrate that analyst characteristics—particularly gender, education, public visibility, and professional tenure—shape the extent to which analysts differentially respond to yield spreads across bond-rating categories.

3 Thinking process in comments

Having documented that analysts' recommendations tend to favor higher-yielding, riskier bonds only when those bonds carry a safer rating, this section examines the underlying belief formation process through analysts' own written comments. These comments explicitly reveal the information analysts consider and their interpretative frameworks. Analyzing the textual evidence allows me to document how analysts' reasoning differs across rating categories and sheds light on potential explanations for the observed recommendation asymmetry.

I structure the analysis around two dimensions: (1) the evaluation of credit risk, and (2) the interpretation of yield spreads. Credit risk evaluation comes first because, unlike yield spreads, which are directly observable market data, default risk requires subjective inference from firm fundamentals, industry context, and other relevant factors. Analysts first form risk assessments, then evaluate whether the yield adequately compensates for the perceived risk, and issue recommendations. Additionally, while analysts frequently omit explicit references to yield, they consistently highlight factors related to risk improvement or deterioration, which underscores the foundational role of risk perception in their decision-making process. For each dimension, I first describe how I transform analysts' free-form commentary into structured data using a large language model (LLM) pipeline, and then present empirical comparisons across rating groups.

3.1 Credit risk evaluation

Risk evaluation can be viewed as a two-step process. First, analysts translate publicly available information into an internal perceived credit risk. Second, they convert this perceived risk, together with yield and other considerations, into specific investment recommendations.

The EDF-related recommendation asymmetry may originate at either stage. Analysts may understate risk when interpreting available information, or they may perceive sufficient risk but fail to incorporate it into their recommendations fully. Both channels may also operate jointly. Investigating these channels requires a quantitative measure of analysts' perceived risk. Ideally, analysts would directly provide such numerical scores of their perceived risk; however, such scores don't exist. Consequently, I construct two complementary text-based measures.

3.1.1 Perceived risk measure

My first measure is to use a large language model (LLM) to transform analysts' comments into a single perceived risk score. Recent studies demonstrate that large language models (LLMs) can reliably capture nuanced textual contexts, interpret specialized financial terminology, and assign scores guided by prompts (Laarits, Matthies, Vasudevan and Yang, 2025). Specifically, I prompt the LLM to emulate an experienced credit analyst. For each analyst comment, the LLM identifies the signals that either heighten (downside) or reduce (upside) credit risk and carefully assesses both the frequency and intensity of such expressions. It then aggregates the information into a perceived-risk score, ranging from -4 (extreme safety) to $+4$ (extreme risk). The sign of the score captures the overall stance embedded in the analyst's language: positive for a net perceived risk, negative for a net perceived safety, and zero for neutrality. The magnitude captures the strength of the perceived risk or safety. The LLM prompt used for this step is provided in the Figure A1.

Table A1 presents examples of comments, corresponding recommendations, and the perceived-risk score assigned by the LLM. The first two examples, DuPont (January 2015) and GE (August 2018), highlight explicitly negative language, such as "revenue headwinds," "lukewarm profitability," and "elevated leverage," earning higher risk scores (+3 and +2). By contrast, the DTE comment (November 2014) emphasizes improvements in the balance sheet and management's commitment to strong BBB ratings, receiving a negative risk score (–2) and an Overweight recommendation. The final excerpt, Coca-Cola (May 2015), features highly favorable language and yields an even lower risk score (–3). These examples confirm that the LLM-derived scale effectively captures both the direction and intensity of risk perceptions. Also, they demonstrate the measure's flexibility in accurately capturing a diverse range of issuer characteristics, market conditions, and credit events. For instance, the perceived risk for GE notably declines from August 2018 to December 2020, as management's deleveraging strategy gains credibility.

To further ensure reliability, I construct a second perceived-risk measure using the dictionary-based method of Loughran and McDonald (2011); Zhang, Zhao and Zhao (2024). Specifi-

cally, I parse each analyst comment and count occurrences of uncertain, weak-modal, and negative words from predefined word lists. Frequently appearing uncertain words in analyst commentary include terms such as "risk(s)," "could," "exposure," "negative," and "weaker." Because analysts naturally employ more uncertain and negative language when discussing serious risk factors, higher frequencies of these words indicate a greater extent of risk factors and higher perceived credit risk.

The dictionary-based measure is well-established, transparent, and perfectly reproducible. However, the cost of transparency is rigidity, so this measure has several limitations in a few cases. For instance, it can overlook contextual nuances and misinterpret negations (e.g., "not a liquidity concern"). Additionally, the dictionary approach fails to adapt to new, specialized financial terminology, such as "covenant-lite structure," that emerges after the word lists were created in 2011. Lastly, it only captures downside sentiment, neglecting potential upside or risk improvements. In contrast, the LLM-based measure effectively addresses these limitations by interpreting entire sentences rather than isolated words, accurately capturing negations, contextual subtleties, and new financial concepts. Additionally, it provides a symmetric and directional assessment of risk, distinguishing clearly between increasing and decreasing perceived risk. However, the LLM approach lacks the complete transparency of a dictionary-based measure and may potentially introduce model-based subjectivity over time (Ye et al., 2024; Lin et al., 2024).

Each measure brings distinct strengths and weaknesses, so I use both measures together in my main analysis. In my sample, the two perceived-risk measures are positively and significantly correlated (correlation = 0.27), suggesting they capture related but distinct dimensions of analysts' credit-risk perceptions.

3.1.2 From public information to risk assessment

With measures of perceived risk established, I now examine how analysts translate objective default risk (EDF) into their textual assessments of credit risk. Table 6 presents regression results of the textual risk measures on EDF interacted with bond-rating categories and plus a separate dummy variable for A-rated bonds.

In column (1), I regress the LLM-based perceived risk score on EDF, controlling for time fixed effects. The EDF slope is positive and insignificant for A-rated bonds, and positive and significant for BBB-rated bonds. In column (2), I add industry fixed effects, and the EDF slope for A-rated bonds becomes significant, with other results remaining largely unchanged. These results indicate that analysts perceive higher risk as default risk increases. The intercept for A-rated bonds is positive. Given EDF values range from zero to one, analysts have a lower perceived risk for A-rated bonds relative to BBB-rated bonds in general.

Next, I test whether the two objective risk sensitivities differ significantly. I re-estimate the model with BBB as the reference group, so the coefficient on the interaction term now captures the incremental loading for A relative to BBB. The estimate is negative and highly significant. In other words, an extra SD increase in EDF translates into a lower boost in the perceived risk when the bond is A-rated than when it is BBB-rated.

In columns (5) to (8), I re-estimate the model using the number of uncertain texts as the dependent variable, and findings remain largely unchanged. A higher intercept and a flatter slope for A bonds together imply that analysts start more cautiously but accelerate their concern more slowly as EDF rises. This finding is interesting. Intuitively, credit ratings serve as reference points for analyst expectations, and a moderate increase in default risk (EDF) might be acceptable for a BBB issuer, but would be alarming for a safer A-rated issuer. If analysts react more strongly when default risk deviates from rating-based expectations, their comments should reflect greater EDF slope for A-rated bonds than BBB bonds.

Predicting Future Realization of Risk

Based on analyst comments and prior literature, CRA downgrades are salient and can carry reputational consequences for analysts. Especially when the downgrade risk is predictable using public information, such as EDF, a rational analyst is expected to predict future downsides and defaults in the correct direction.[10] I test this prediction by relating analyst recommendations and perceived risks to subsequent downgrades and defaults. Specifically, I estimate a series of logit models, where the dependent variable equals one if the issuer experiences a downgrade (DG) within three, six, or twelve months, or a default (DF) within three years, and zero otherwise. Table 7 reports these regression results.

Columns (1)–(4) examine the predictive power of analyst recommendations for downgrades and defaults. Column (1) shows that for BBB-rated bonds, moving from Underweight to Overweight in the cross section reduces the predicted three-month downgrade probability from 7.6% to 2.7%. However, recommendations have no significant predictive content for A-rated bonds. Extending the prediction horizon to six and twelve months (columns 2–3), the results remain robust. Column (4) focuses on defaults over a three-year horizon and the sign of predictability reverses for A-rated bonds. A more favorable recommendation (from Underweight to Overweight) increases the default probability from 0.7% to 3.3%. For BBB-rated bonds, the corresponding effect remains statistically insignificant. These findings suggest that recommendations can predict short-term downgrades only within the BBB rating category. For safer A-rated bonds, favorable recommendations fail to predict short-term deterioration and instead appear associated with increased longer-term default risk.

[10] Appendix shows that EDF is able to predict future downgrades and defaults in the granular rating group in my sample.

Defaults rarely occur without warning and typically follow a sequence of credit deterioration and rating downgrades.

Columns(5)–(12) exploit the LLM–derived perceived risk score and uncertain word count as key explanatory variables. For BBB bonds, the rise in the risk score (from 0 to +3) lifts the predicted downgrade risk from 3.7% to 9.5%. The corresponding increase for A bonds is similar. Word-count uncertainty yields a similar pattern. For default, the predictive strength of the LLM score is more pronounced in A-rated bonds; however, the effect becomes statistically less precise for BBB-rated bonds at the three-year horizon. Overall, the comments-based perceived risk measure flags credit deterioration well ahead of rating agency actions.

These results are puzzling, as analysts' recommendations fail to predict subsequent risk realizations in the higher-rated segment. To address this puzzle, I next explore how analysts incorporate their perceived risks into final recommendations.

3.1.3 From risk assessment to recommendations

Table 8 reports regressions of analysts' recommendations on textual risk measures. The key explanatory variables are the perceived risk proxies, the LLM-generated risk scores (columns 1 to 4) or uncertainty word counts (columns 5 to 8), each interacted with bond rating indicators (A and BBB). All regressions control for a rating group dummy and time fixed effects. Alternate columns further include yield spreads interacted with rating groups and industry fixed effects.

In columns (1)–(4), the interaction term between perceived risk and the BBB rating is negative and statistically significant, indicating that the bonds with higher perceived risk will receive less favorable recommendations in this rating category. Conversely, the same perceived risk score shows a positive but insignificant relationship with recommendations for A-rated bonds, suggesting analysts do not adjust recommendations downward in response to higher perceived textual risk in higher-rated bonds. Columns (5)–(8) repeat the analysis using the uncertain word-count measure. Consistent with the LLM-based results, the loadings of uncertain text in the A group are positive and in the BBB group are negative. The weaker statistical significance likely reflects the reduced nuance and inability to fully capture sentence-level context of the uncertain word-count measure relative to the LLM-generated measure.

Taken together, these results help resolve the gap between what analysts say and what they recommend. Although analysts explicitly acknowledge downside risks in their comments, they systematically fail to incorporate these risks into recommendations for bonds perceived as safer. This selective response aligns closely with the "local thinking" frame-

work of risk neglect proposed by Gennaioli, Shleifer and Vishny (2012), wherein investors underestimate the probability of adverse outcomes deemed unlikely to happen. Because an A rating carries lower default risk and lower regulatory capital charges, analysts may dismiss adverse scenarios as remote or irrelevant. Consequently, they discount or disregard even their own articulated risk concerns when issuing investment recommendations. In short, analysts neglect risk in purportedly safe bonds, and this is a plausible driver of the recommendation asymmetry. In addition, this finding provides the first empirical evidence for the mechanism suggested by Becker and Ivashina (2015), who argue that reaching-for-yield behavior can reflect risk neglect.

3.2 Yields evaluation

This section examines analysts' evaluations of yield. Unlike credit risk, which must be inferred from firm fundamentals, yield spreads are directly observable market data. Thus, analysts do not estimate the numerical level of yield spreads; instead, they qualitatively assess whether a given spread is sufficiently "wide" or "tight," conditional on their underlying risk perceptions. This framework implies two potential mechanisms behind the recommendation asymmetry. First, analysts might perceive identical increases in yield spreads as more attractive for A-rated bonds than for BBB-rated bonds. Second, analysts might place differing weights on identical qualitative yield assessments when making final recommendations across rating categories.

To disentangle these possibilities, I proceed in three steps. First, I construct a numerical yield-tone measure by categorizing analysts' qualitative assessments (e.g., "wide," "fair," or "tight"). Second, I test whether a given numeric spread increase is more frequently described as "wide" for A-rated bonds than for BBB-rated bonds. Third, I examine whether analysts translate identical qualitative yield assessments differently into final recommendations across ratings.

3.2.1 Yields tone measure

I quantify analysts' qualitative evaluations of yield through a structured yield tone score, derived directly from their comments. Specifically, the LLM scans each paragraph to identify explicit references to yields, and comments without valuation references are excluded from subsequent tone analyses. Then, the LLM classifies language by directional tone: positive terms (e.g., "wide," "attractive," "rich," etc.) receive a score of +1, neutral expressions (e.g., "fair," "range-bound," etc.) receive 0, and negative phrases (e.g., "tight," "compressed,"

"limited room," etc.) receive –1.[11] The exact LLM prompt is provided in Figure A2. The procedure results in approximately 38,000 yield-tone scores. Among these, 42% are classified as "tight," 27% as "neutral," and 31% as "wide."

3.2.2 From true yield to yield tone

Panel A of Table 9 examines how numeric yield spreads shape analysts' qualitative language. Columns (1)–(4) regress the LLM-derived yield tone on yield spreads, rating indicators, and their interactions. Column (1) controls for time fixed effects only. It shows that yield spread increases are characterized as more attractive for A-rated than BBB-rated bonds. Column (2) adds perceived risk scores. The risk score coefficient is negative, consistent with higher perceived risk making analysts less likely to describe yields as attractive or wide, as the spread no longer adequately compensates for elevated risk. After adding industry fixed effects in columns (3) and (4), the key findings remain robust.

Notably, the sensitivity of qualitative yield tone to yield spreads is approximately twice as large for A-rated bonds as for BBB-rated bonds, highlighting substantial differences in how analysts interpret identical numeric information across rating categories. Columns (5)–(8) further confirm this asymmetry: analysts systematically describe equivalent increases in numeric spreads as more attractive when the issuer carries an A rating.

3.2.3 From yield tone to recommendation

Having established the asymmetry in how analysts qualitatively interpret identical yield spreads, I now examine whether differences also appear in how they translate these qualitative assessments into recommendations. Panel B of Table 9 regresses analysts' numeric recommendations on the yield-tone measure, rating-category indicators, and their interactions. Columns (1) and (2) show a consistent and strong relationship between yield tone and analyst recommendations. Columns (3) and (4) show that the interaction term between yield tone and bond rating is negligible and not statistically significant, with and without industry fixed effects. Thus, the influence of yield-tone language on recommendations remains essentially uniform across ratings.

This symmetry implies that the observed recommendation asymmetry does not stem from how analysts incorporate qualitative yield assessments into recommendations. Rather, the divergence occurs earlier, when analysts convert numeric spreads into qualitative judgments. Specifically, identical yield spreads are more frequently described as "wide" for A-rated bonds.

[11] Analysts explicitly discuss yield spreads at roughly equal frequencies (around 40 percent of notes) across both rating groups. This rules out the possibility that analysts rarely mention yields for BBB-rated bonds.

4 Future Performance of Recommendation

The central hypothesis is that if analyst recommendations for A-rated bonds reflect yield-seeking and risk neglect behaviors, then bonds receiving favorable recommendations should not deliver positive risk-adjusted returns (alphas). I test this hypothesis directly by evaluating the alphas of portfolios formed from analyst recommendations.

Specifically, I form two-way sorted trading portfolios. At the end of each month, I separately sort A-rated and BBB-rated bonds into three analyst recommendation categories: Underweight (UW), Neutral (N), and Overweight (OW), and then form both value-weighted (VW) and equal-weighted (EW) portfolios for each recommendation group. I hold these portfolios for three months and measure return performance following the rolling-portfolio methodology of Jegadeesh and Titman (1993).

Panel A of Table 10 reports summary statistics for the bonds in these two-by-three portfolios, which hold on average 550 bonds each month. Within the A-rated group, Overweight (OW) portfolios consistently exhibit wider yield spreads compared to Underweight (UW) portfolios, confirming that analysts reward higher yields. Additionally, OW portfolios have longer maturities, indicating that higher-yielding bonds in this segment generally involve greater duration. By contrast, yield and maturity differences within the BBB-rated group are modest, aligning with prior evidence that analysts emphasize safety rather than yield in this category. Further, higher yields or yield spreads are not necessarily associated with poor recent performance, a finding consistent with Chen and Choi (2024). UW portfolios exhibit both lower yields and lower past returns compared to OW portfolios. Age and liquidity (proxied by trading volume and offering amount) are similar across recommendation categories, suggesting these dimensions are unlikely to drive subsequent return differences.

In Panel B of Table 10, I report average excess returns and alphas for the portfolios sorted on ratings and recommendations. To gauge risk-adjusted performance, I report alphas under three benchmark models. I start with the Fama and French (1993) term-structure and investment-grade default factors, the most widely used two-factor system for capturing macro risks that move bond prices. I then include the single corporate-bond market factor advocated by Dickerson, Mueller and Robotti (2023b), who show that, previously proposed bond risk factors do not have any incremental explanatory power over the corporate bond market factor.

Focusing on the excess returns, the average excess returns are higher for higher recommendations in both the A and BBB buckets. The long OW and short UW portfolio (OW-UW) earns 10.2 basis points per month in the A bucket and 9.9 basis points in the BBB bucket for the VW portfolios. The EW portfolios show the same pattern. However,

after risk-adjusting returns, the picture diverges across rating buckets. In the A portfolio, the alpha shrinks to zero under the Fama–French two-factor model and even turns negative once the single corporate-bond market factor is used. In the BBB portfolio, the alpha remains significantly positive relative to the two-factor benchmarks, but it is not significant when measured against the market factor alone, with a t-stat equal to 1.41. Such patterns exist if one only looks at OW portfolios.

Next, I examine the individual factors in terms of alphas and factor loadings using the OWUW value-weighted portfolios. Figure 3 reports the betas and alphas. Besides the ones used earlier, I add stock-momentum, bond-momentum, reversal, and downside-risk.[12] For A-rated bonds, the alpha vanishes whenever the regression includes any of the bond factors. And the factors, including bond-market, duration, momentum, downside, or term-structure, carry significant positive loadings, implying that the apparent excess return is simply compensation for conventional risks. By contrast, BBB portfolios continue to deliver economically meaningful and statistically significant alphas in all specifications except the aggregate bond-market factor, with momentum providing only marginal explanatory power. Loadings on other standard factors are economically small and statistically indistinguishable from zero.

Overall, these results closely align with earlier findings. Analysts' yield-seeking recommendations for A-rated bonds primarily reflect compensation for conventional risk factors rather than superior insight, and thus fail to generate genuine alpha. In contrast, the more persistent alpha observed in BBB-rated bonds indicates that analysts more carefully incorporate risk considerations into their recommendations for this segment.

Until now, the empirical findings from recommendations and comments raise several fundamental questions. Why do analysts reward higher yield and risk only when a bond carries an A rating, but not when rated BBB? Why do they neglect risks explicitly recognized in A-rated bonds but not in those rated BBB? Does this asymmetric behavior reflect fundamental differences associated with each rating—such as distinct investor bases or client relationships—or does it instead stem from the rating label itself? I explore these questions in the next section.

[12] I use factors from Dickerson, Julliard and Mueller (2023a). Specifically, the TERM factor is defined as the return spread between long-term government bonds and the one-month Treasury bill. The default risk factor (DEF) is the difference in monthly returns between a market portfolio of long-term corporate bonds and long-term government bonds. The liquidity factor (LIQ) follows Pástor and Stambaugh (2003), calculated as the average liquidity measure estimated from daily individual bond returns after removing market-wide effects. The bond momentum factor (MOMB), following Gebhardt, Hvidkjaer and Swaminathan (2005), is constructed by independently sorting bonds into five rating groups and five groups based on their past 12-to-2-month cumulative returns (MOM). Within each rating quintile, the long-short momentum return is computed by subtracting the return of the lowest MOM portfolio from the highest. MOMB is the average of these long-short returns across rating quintiles.

5 Conventional incentive mechanisms

5.1 Catering to clients

Client relationships may explain the observed recommendation asymmetry between A-rated and BBB-rated bonds, as analysts face incentives to issue strategically biased recommendations in the presence of agency conflicts (Lin and McNichols, 1998; Michaely and Womack, 1999; Mehran and Stulz, 2007). An extensively studied type is investment banking (IB), where analysts issue more optimistic research to attract or retain underwriting and advisory mandates. Also, analysts can face brokerage incentives, as optimistic recommendations boost trading volume and commission revenues (Ajinkya, Atiase and Gift, 1991; Agrawal and Chen, 2012).

To evaluate these client-driven incentives, I manually collect and analyze disclosures from analyst reports. Each disclosure explicitly identifies the bank's specific relationship with issuers. I consolidate the client relationships into three categories: (i) investment banking, which primarily includes lead manager or co-manager underwriting roles and broader advisory services such as mergers, restructuring, etc; (ii) broker and market maker, which are related to liquidity providing and secondary-market trading; and (iii) other significant financial interests, primarily reflecting proprietary positions of analysts' employer in issuing debt securities exceeding one million USD.

I match these unstructured relationships by company name to my sample dataset. The resulting matched sample contains 14,315 observations across 170 unique firms with at least one disclosed relationship. Among these relationship-linked observations, underwriting mandates represent the largest category (95%), followed by brokerage roles (5%) and proprietary interests (3%).[13]

To test whether client relationships drive the observed A-versus-BBB asymmetry, I first look at recommendation distributions. If analysts cater to issuers that are clients, these issuers should receive more OW recommendations. However, the data do not support this prediction. Among 14,315 observations with client relationships, the distribution—OW (22%), N (52%), and UW (26%)—is aligned with the unconditional sample. Also, within A-rated bonds, OW recommendations account for just 9.8%, compared to 38.3% for UW. Conversely, BBB-rated bonds have a higher OW proportion of 35% and a much lower UW proportion of 11.4%. Thus, the distribution of recommendations for client-affiliated issuers is consistent with the overall pattern and shows no systematic optimism toward such clients.

While summary statistics provide initial evidence, regression analysis explicitly tests

[13]Percentages sum to over 100%, as the broker house can provide multiple services simultaneously to the same issuer.

whether client relationships drive the rating-specific asymmetry, controlling for firm and bond characteristics, as well as time and industry fixed effects. Columns (1)–(4) of Table 11 report OLS estimates of the baseline coefficient β_1 from equation (6), separately for issuers with and without disclosed client relationships. The significant positive coefficients on the interaction between the A-rating indicator and yield spreads confirm that the core recommendation asymmetry holds in both subsamples.

Columns (5)–(10) test explicitly whether client relationships amplify this asymmetry by estimating the specification of Equation 7 and focusing on β_7, which is the coefficient of a triple interaction term ($A \times \text{Yield Spread} \times \text{M}$), where M is the relationship. Columns (5)–(6) separately examine relationship types: underwriting (manager), broker, and other financial interests. Among these, only the residual "other" category shows a weakly significant positive effect. Because brokerage and other relationships constitute only a small fraction of disclosed client ties and all client categories should have similar directional effects if analysts cater to clients, I combine the three categories into a single client dummy. Columns (7) and (8) report the results and show that the triple interaction remains statistically indistinguishable from zero.

Lastly, even if an analyst's employer has no active relationship with a *currently* covered firm, it may anticipate *future* business opportunities or retain the *previous* business. Therefore, columns (9) and (10) broaden the client relationship definition to include ties established within 12 months before the official start date or maintained within 12 months after the official end date of a disclosed relationship. Results provide no evidence that analysts cater to issuers with current or potential client ties.

The limited impact of client relationships is unsurprising given heightened regulatory scrutiny following the landmark 2003 analyst settlement, which, combined with analysts' reputational concerns, can constrain strategic distortions and keep analysts honest with their decision-making.[14]

6 Behavioral mechanism: Rating-Induced Categorical Thinking

6.1 Institutional background: the A–BBB rating boundary

This section outlines two institutional features of credit ratings that motivate the empirical approach. After reviewing the institutional background, I describe the empirical design that

[14] In April 2003, ten major Wall Street firms paid $1.4 billion in penalties to resolve charges that their analysts routinely issued overly optimistic research to attract investment banking business. The settlement also imposed structural reforms, prohibiting analysts from participating directly in sales activities and mandating separation between research and investment banking operations (U.S. Securities and Exchange Commission, 2003).

exploits these features.

Institutional regulations on A-BBB

Regulations and market conventions have long drawn distinctions between A- and BBB-rated bonds. These differences date back over a century to the origins of formal credit ratings and continue to shape market practices today. When Moody's initially introduced its credit rating system in 1909, it grouped AAA, AA, and A bonds into a "first-grade," while BBB and BB bonds into a "second-grade" (Harold, 1938; Fons, 2004). Early regulatory interpretations reinforced this distinction. For instance, *American Banker* concluded that early regulations practically limited investments to A-rated bonds. It was not until 1938 that Moody's persuaded regulators that BBB-rated bonds were neither "distinctly nor predominantly speculative," thereby classifying BBB bonds as investment-grade (Harold, 1938). Today, although both A-rated and BBB-rated bonds are classified as investment-grade under most regulatory frameworks, many institutions—including banks, insurance companies, and market indices—still distinguish significantly between these ratings.[15] Over time, the regulatory and institutional reinforcement has led market participants to naturally believe that A-rated bonds are inherently safer than BBB-rated bonds.

Limitations of credit ratings

Credit ratings are designed to give investors an informed assessment of a firm's creditworthiness (S&P Global Ratings, 2023). However, extensive research highlights structural limitations of credit ratings that reduce their accuracy and timeliness. A notable limitation is rating inertia, the deliberate emphasis on stability that discourages frequent rating adjustments (S&P Global Ratings, 2010; Cantor and Mann, 2006). Credit rating agencies (CRAs) prioritize stability over accuracy and often require sustained changes in firm fundamentals before updating ratings. This conservatism can delay both downgrades and upgrades.

Conflicts of interest within CRAs can also compromise rating accuracy and credibility (Opp, Opp and Harris, 2013; Mathis, McAndrews and Rochet, 2009; Bolton, Freixas and Shapiro, 2012; Kronlund, 2020). Theoretical and empirical evidence suggests reputational concerns fail to prevent CRAs from catering to issuer interests. Conflicts of interest can arise from agencies deliberately understating risk to secure issuer business, issuers selectively pur-

[15] Under Basel III regulations, commercial banks apply credit rating-based risk weights to corporate bond holdings, directly influencing their regulatory capital requirements. Specifically, bonds rated A receive a 20% risk weight, while bonds rated BBB receive a 50% risk weight (Basel Committee, 2019). Similarly, insurance companies must hold more capital when investing in lower-rated bonds. Specifically, under the NAIC framework, the risk-based capital charges increase from 0.4% for A-rated bonds to 1.3% for BBB-rated (National Association of Insurance Commissioners, 2021). Additionally, prominent market indices, such as Bloomberg and ICE corporate bond indices, provide separate performance metrics for A-rated and BBB-rated bonds (Bloomberg, 2024; ICE, 2024). Although the SEC's net capital rule does not explicitly assign rating-specific risk weights, it implicitly relies on credit ratings when broker-dealers calculate net capital requirements based on asset risk categories (U.S. Securities and Exchange Commission, 2022).

chasing the most favorable ratings available, etc. Recent literature further documents credit rating inflation that can also compromise the accuracy of credit ratings (Griffin, Nickerson and Tang, 2013; He, Qian and Strahan, 2012; Efing and Hau, 2015).[16]

Given these distortions within CRAs, there is reason to believe that ratings tend to lag behind market-based indicators of credit risk (Cantor and Mann, 2006). As Bongaerts, Cremers and Goetzmann (2012) highlights, credit ratings are assigned primarily for regulatory purposes and do not provide significant incremental information about a firm's credit quality. The disconnect between ratings and underlying fundamentals creates a natural setting where firms with similar fundamental characteristics can receive different ratings within the investment-grade category.

6.2 Empirical design: matching

The previous section suggests two stylized facts that motivate my empirical approach. First, institutional regulations and investment mandates treat A- and BBB-rated bonds differently, reinforcing the natural market perception that these categories reflect meaningful economic differences. As a result, investors and analysts may aggregate information at the coarse rating group level and form assessments based on rating group stereotypes. A-rated bonds are commonly viewed as "safe and low return" assets, while BBB-rated bonds are viewed as "risky and high return."

Second, structural limitations in credit ratings, such as rating inertia and conflicts of interest, imply that some A-rated firms are very similar to BBB in terms of financial health characteristics . That is, there are firms rated in the BBB group that should be in the A group and vice versa. By leveraging such rating-label frictions, I use matching methodology to isolate the effect of the rating category itself. Specifically, I match A and BBB firms based on industry and firm characteristics and compare analysts' interpretations of the same credit signal of the matched firms.

My benchmark matching procedure uses industry, size, leverage, Altman's z-score, EDF and cash holdings. These variables have the most explanatory power in regressions of the A versus BBB group on firm characteristics. Each month, I take a firm rated A and find a BBB firm that is in the same industry and is the closest in terms of my matching variables. I use a caliper-based nearest-neighbor matching adapted to a panel setting, following Asker, Farre-Mensa and Ljungqvist (2015). Although it would be possible to mechanically find the closest BBB firm for every A firm, I would like to ensure high-quality matches in which the

[16]Specifically, CRAs inflated ratings beyond their models when facing competition, and they tend to assign overly optimistic ratings to asset-backed securities structured by large issuers who generate significant fee revenue. Becker and Milbourn (2011) demonstrate that increased competition among rating agencies has reduced the quality and informativeness of credit ratings from established agencies like S&P and Moody's.

A firm and its BBB match are very similar. To do so, I impose a tight 0.35 caliper value to ensure quality matches.

The matching sample provides a clean setting to directly test rating-driven categorical thinking: whether analysts interpret the same credit signals differently purely depending on the firm's rating label. Next, I introduce the matching procedure to construct pairs of otherwise similar A rated and BBB rated firms.

6.2.1 Characteristics of matched A and BBB firms

Each month, an A (A−, A, A+) non-utility firm is matched to the BBB (BBB−, BBB, BBB+) firm that is closest in terms of log assets, book leverage, market leverage, EDF, cash to assets, Z−score, and PPE to assets. AAA and AA issuers are excluded because they are more likely to differ markedly in credit quality from BBB firms. I measure distance as the absolute difference between the log odds of the propensity scores obtained from the logit model, and a BBB candidate is accepted as the nearest neighbor only when the distance lies within a caliper of 0.35.[17] Utility firms are matched similarly, except that the Z-score is omitted.

Table 12 reports the mean characteristics for each set of firms and the difference in means. The full sample contains 3,265 A and 5,287 BBB firm-month observations. The matched sample consists of 1,556 A-rated firm-month observations matched to 846 unique BBB-rated firm-month observations. The share of outlying notches (A+ and BBB−) falls from 25% in the full sample to 10% in the matched sample. Most matched pairs link issuers right on the A/BBB boundary. This concentration suggests that the matching algorithm lines up bonds with comparable credits.

Our matching procedure successfully picks A and BBB firms that have similar size, leverage, solvency, tangibility, and profitability. In the raw data, BBB issuers carry 14.5 percent more leverage than their A-rated counterparts. After matching, the gap shrinks to 2.7 percent and is no longer statistically significant. Furthermore, none of the other differences in characteristics between A firms and matched BBB firms are statistically or economically significant. Overall, my matching procedure selects a sample of A and BBB firms that are very similar along observable dimensions. Next, I explore what credit signals analysts should respond to when forming recommendations.

[17] Prior simulation and empirical work (Cochran Rubin 1973; Austin 2011) shows that keeping matches within roughly 0.20–0.50 standard deviations of the logit of the propensity score removes most observable covariate bias while preserving sample size. In my data, the pooled standard deviation of the log-odds score is ≈ 1.4, so I choose caliper $\approx 0.25 \times 1.4 \approx 0.35$.

6.3 Rational benchmark: The same credit signal shouldn't be treated differently in the matched A and BBB firms

My identifying assumption is that, for A-rated and BBB-rated firms with identical observable fundamentals at a given time, the expected response of future bond performance to underlying credit signals is the same. Under this assumption, the rating label itself carries no additional predictive content once fundamentals are controlled. In other words, within fundamentally matched firms, the predictive power of yield spreads or EDF for future excess returns should not differ significantly across rating groups.

Specifically, I test the specification using Equation 5 and Equation 6, and replace the dependent variable with the next 3-month excess returns. Columns (1) to (6) report estimates for the full sample of U.S. public non-financial firms from 2013 to 2023.[18] The yield spread predicts future returns much more strongly for BBB bonds than for A bonds. A 10 basis points widening in the spread raises next three-month excess returns on BBB bonds by about 11 basis points, whereas the corresponding increase in A bonds is only 0.8 basis points and is statistically indifferent from zero. Column (2) confirms that the 10.2-basis-point slope gap is highly significant. Column (3) shows that a higher EDF helps forecast positive excess returns for BBB bonds but slightly negative ones for A bonds, and column (4) shows that the differential response on EDF is statistically significant. Columns (5) and (6) enter both signals jointly, and the EDF coefficients become statistically insignificant, indicating that the risk information is already embedded in the yield spread. The difference of yield spread effect remains positive and highly significant.

Columns (7)–(12) repeat the above analysis in the matched sample. After matching, the spread's and EDF's forecasting slope converges across rating buckets. Column (7) shows that a ten-basis-point widening in yield spreads predicts only a 2.3-basis-point higher excess return for BBB bonds relative to their A-rated matches, and this difference is statistically indistinguishable from zero (t = 0.74), as shown in column (8). Column (10) similarly shows that the corresponding difference based on EDF is insignificant. Columns (11) and (12) combine yield spreads and EDF, the results keep consistent. Overall, the evidence supports the assumption that matched A- and BBB-rated issuers exhibit similar return dynamics, with yield spreads and EDF forecasting future returns indistinguishably across these A and BBB rating groups.

Overall, if analysts are rational, then I should expect $\beta_{1,A}$ is statistically indifferent from $\beta_{1,BBB}$ and $\beta_{2,A}$ is statistically indifferent from $\beta_{2,BBB}$. Any systematic divergence in these coefficients would indicate rating-driven inference behavior consistent with categorical

[18] I also exclude utility firms from this test because the Z-score, used in matching, is not applicable to utilities.

thinking.

6.4 Categorical thinking evidence: The same credit signal is treated differently in the matched A and BBB firms

Table 13 presents estimates from Equation 5 and Equation 6. I construct the sample in two steps: Each month and industry, an A-rated issuer is matched to a BBB-rated issuer with nearly identical fundamentals, and every bond those matched firms have outstanding that month is then added to the panel. All specifications include pair-by-month-by-industry fixed effects, so any remaining difference is attributable to the rating label itself.

In column (1) the spread interaction is positive and highly significant for A-rated bonds but negative and insignificant for BBB-rated bonds, so the same price signal is interpreted as good news on the A side and as noise or worse on the BBB side. Column (2) shows that $\beta_{1,A}$ is statistically indifferent from $\beta_{1,BBB}$. Column (3) controls for bond characteristics, including Macaulay duration, the log of issue size, coupon, and a callability flag, and the A-minus-BBB spread coefficients become slightly larger and their t-statistic strengthens, confirming that differences in bond characteristics do not drive the pattern. Column (4) adds analysts FE, and the results show that even when two bonds come from the same industry, share comparable issuer and bond fundamentals, and are covered by the same analyst, the analyst still reacts to an identical spread change in markedly different ways. Analysts upgrade A-rated bonds with higher yield spreads but leave their BBB-rated counterparts largely unchanged or mark them down.

Columns (5)-(8) replace the yield spread with the issuer's expected default frequency (EDF). I can find that $\beta_{2,A}$ is statistically different from $\beta_{2,BBB}$ and the treatment gap remains large and negative even when the comparison is restricted to the same industry, the same analyst, or both. Across every specification the standalone BBB indicator carries a positive and significant coefficient. One interpretation is that, once fundamentals are matched, a BBB label signals misclassification: the bond looks A-like on the fundamentals and analysts anticipate an eventual upgrade, so they lean more favorably toward it. A second possibility is that analysts may infer that BBB bonds, all else equal, offer a better payoff than their A-rated twins and therefore deserve a stronger recommendation.

6.5 Other supportive evidence 1: Coarser categories

Because analysts mention the things they deem relevant, systematic differences in the types and counts of cited concepts across rating buckets would reveal whether their information sets shift with the rating. To measure both what analysts mention and how much they mention, I

convert each comment into structured variables through a three-stage large-language-model (LLM) pipeline.

6.5.1 Information Extraction

To fully digest and summarize the textual information in analyst reports, one needs a way to comprehend the content in the reports and to extract relevant information. The traditional approach usually involves some degree of manual reading by humans, but given my sample of more than 1.2 million analyst reports, manual reading comprehension and information extraction is prohibitively inefficient. Instead, I leverage the recent development of LLMs and extract information.

Large-language models (LLMs) offer the most effective tool to extract financial information embedded in analysts' comments. An LLM is a transformer-based neural network trained on extensive textual data. It learns to sequentially predict tokens, enabling near-human performance on tasks like zero-shot classification, question-answering, and summarization.

Traditional text-analysis methods, such as keyword-frequency counts or unsupervised topic models, cannot reliably capture context-specific details or precisely answer targeted questions. For instance, traditional methods cannot accurately identify which factual metrics analysts mention, or which forward-looking risks they emphasize. In contrast, LLMs interpret textual context, enabling precise extraction of structured, economically relevant information. Specifically, I use an LLM to separate analysts' factual observations from subjective judgments and systematically catalog the financial concepts they discuss. This method provides detailed, interpretable measures of analysts' information sets, a precision unattainable with conventional textual analysis.

In this paper, I use Llama-3.3-70b as the generator, which is the largest publicly available variant of Meta's third-generation Llama model family. The model contains 70 billion parameters and supports a context length of 8k tokens, with weights trained on a carefully curated corpus of approximately 15 trillion tokens from sources including news articles, code repositories, and domain-specific documents. In standard NLP benchmarks, Llama-3-70B achieves accuracy of 82 percent on MMLU and 64 percent on GSM-8K, comparable to GPT-3.5 Turbo and within approximately six percentage points of GPT-4, but at less than one-tenth the inference cost. Moreover, unlike proprietary models, Llama-3-70B provides full access to model weights and tokenizers, enabling deterministic replication, in-domain fine-tuning, and local deployment that meets stringent data protection requirements. These advantages make Llama-3-70B an ideal large-scale text generator for extracting structured financial information while ensuring methodological transparency.

To guarantee the quality of information extraction, I implement a multi-step prompting process for better extraction and generation quality. The steps are described below.

Step one: Decompose comments to factual and subjective content

Analysts' comments generally mix observable facts about current corporate conditions with subjective forecasts of future outcomes. Facts reflect analysts' information sets, while subjective statements provide insight into their forward-looking beliefs. Since the determinants of analysts' subjective expectations may differ substantially from those of their factual observations, clearly separating these two types of content is essential to systematically understand the expectation-formation process.

I thus begin by decomposing each analyst comment into two distinct parts using LLM: (i) a factual description, which explicitly references realized events or publicly observable firm conditions, and (ii) a subjective prediction, which expresses the analyst's personal judgments, forecasts, or expectations about future developments. The LLM prompt used for decomposition is provided in the Figure A3.

Figure A4 illustrates the decomposition approach using an analyst's commentary on General Electric from August 14, 2018. The factual description summarizes objectively observable conditions such as GE's wider bond spreads relative to other single-A industrial peers, its strategic plan announced in June (including the spin-off of the Healthcare business and exit from Baker Hughes JV), elevated leverage, and its stated financial targets. In contrast, the subjective prediction captures the analyst's judgment-driven views: specifically, that uncertainty, execution risk, and the weak outlook for key end markets justify the Neutral recommendation. By explicitly separating the factual from subjective content, this method yields structured variables that distinctly represent analysts' stated information sets apart from their forward-looking beliefs and opinions.

Step two: Extract financial concepts

Next, the LLM scans both factual descriptions and subjective predictions to identify key financial concepts explicitly mentioned by analysts. This step serves two key purposes. First, it transforms unstructured, high-dimensional analyst commentary into discrete, economically interpretable terms (e.g., "bond spreads," "leverage," "execution risk"), enabling manageable analysis of information sets. Second, by strictly limiting extraction to explicitly stated phrases, it ensures transparency and reproducibility, avoiding inference. The LLM prompt used for this step is provided in the **??**.

Figure A4 illustrates the financial concept extraction. The LLM identifies explicitly mentioned financial concepts separately from the factual and subjective components. From the factual content, the model extracts concepts such as "underlying business trends," "credit metrics," "execution risk," etc. From the subjective component, it identifies concepts reflect-

ing the analyst's forward-looking judgments, including "execution risk," "weak outlook," and "changes in business," etc. This example demonstrates the LLM's capability to reliably capture financial concepts from both factual observations and subjective predictions, thereby laying the groundwork for structured analysis.

Step three: Label concepts

To further distill extracted financial concepts into economically interpretable variables, I classify each concept into predefined categories at three levels: firm-specific, sector-specific, and macroeconomic. At the firm level, I follow the standard factor structure commonly used by credit rating agencies (CRAs), including Business Profile, Leverage, Liquidity, Profitability, and Firm-specific Events (**?**). To ensure greater granularity, I supplement these categories with basic balance-sheet components, such as Assets, Cash, Debt, and Equity. The sector-specific categories include Regulation/Policy and Supply–Demand dynamics, while macro-level concepts comprise Monetary & Fiscal Policy and Country/Global Growth & Shocks. The complete taxonomy is detailed in Appendix. To guarantee consistency, the LLM assigns each concept to exactly one category using a structured and rule-based prompt.

For example, in the GE commentary, the concept "bond spreads" is categorized under Bond (firm-level), "elevated leverage" under Leverage (firm-level), "end-market weakness" under Supply–Demand dynamics (sector-level), and "execution risk" under Firm-specific Events (firm-level). This labeling procedure converts high-dimensional textual information into structured data suitable for further analysis.

6.5.2 What info - Types of financial concepts

The pie chart of Figure 4 shows the proportion of different types of financial concepts that analysts consider. My taxonomy captures more than 70 percent of all extracted terms, leaving only 27 percent in the residual "Other" bucket (grey). Firm-level concepts (purple) represent approximately 60% of the total. Sector-level considerations (green) account for about 10%, and macroeconomic factors (blue) represent only around 4%.

This distribution is stable across bond ratings, which is unsurprising given that both ratings fall within the investment-grade category. Analysts consistently rely on the core set of credit fundamentals required by clients, investors, and regulators. Additionally, my analysis focuses exclusively on public nonfinancial firms, which are subject to similar disclosure requirements, so analysts can have access to similar types of information across these ratings.

6.5.3 How much info - Quantities of financial concepts

Having identified what types of financial concepts analysts mention, I now explore how extensively they discuss each category. Specifically, I examine whether analysts consistently provide the same level of detail across bond ratings, or trim the discussion when a bond already wears a safer label.

The bar chart of Figure 4 plots the number of financial concepts mentioned per comment. The blue bars are A-rated bonds, and the purple bars are BBB-rated bonds. Analysts systematically mention fewer concepts for A-rated bonds relative to BBB-rated bonds across nearly every category, as indicated by the lower blue bars compared to purple. The numbers above each bar indicate that most differences are statistically significant. This pattern implies that when a bond carries a safer A rating, analysts streamline their comments, possibly assuming certain risks or financial details require less explicit elaboration. Conversely, the more extensive discussion observed for BBB-rated bonds likely reflects analysts' greater attention to marginal credits, necessitating more detailed justifications of their assessments.

Next, I shift from what topics are covered to where the detail is less inside each note. This slide presents a similar idea in a regression analysis to ensure the result is not driven by time, sector, or issuer mix. Table 14 reports concept counts for the whole note, then splits them into the factual section, the forward-looking section, and finally the probability that no factual block is given at all.

In columns (1)-(3), I use the number of concepts in the whole note, and show that analyst comments on A-rated bonds contain roughly two fewer financial concepts per note on average. This difference remains significant even after controlling for credit risk and industry-time fixed effects. In columns 4-9, I break down the comments into factual descriptions and subjective predictions. And I find that the gap is particularly strong in subjective, forward-looking comments.

Another interesting observation is that analysts sometimes omit factual descriptions altogether and directly provide their predictions, especially for bonds that appear safe. To better understand this behavior, in columns (10)-(12), I specifically examine the likelihood that analysts skip factual information entirely. I find that analysts are indeed more likely to omit factual details when discussing A-rated bonds. However, this effect becomes smaller and statistically insignificant after controlling for issuer-specific risk factors.

Together, these findings strongly suggest that analysts use coarser analysis and may expend less cognitive effort when discussing bonds perceived as safer. Such lower mental effort—fewer financial concepts, omitted factual sections, and reduced detail—may lead analysts to underestimate underlying risks in higher-rated bonds.

6.6 Other supportive evidence 2: A- and BBB+ upgrades and downgrades

I next focus on the boundary between A- and BBB+, where even a single-notch move carries significant implications. A downgrade from A- to BBB+ can raise funding costs, prompt index exclusion, and draw closer investor scrutiny, and is followed by negative return performance, whereas an upgrade from BBB+ to A- can have the opposite effects. If analysts paid enough attention to risk on both sides of this line, their calls should flag downgrades in A– and upgrades in BBB+. Panel C shows the opposite. A more favorable recommendation has no predictive content for downgrades in A–, but it anticipates downgrades in BBB+. Conversely, the recommendation predicts upgrades only in A– and not in BBB+. The pattern is consistent with analysts discounting downside risk once a bond is rated A while overlooking upside potential when it sits in the lower investment grade bucket.

The fact that strong analyst recommendations precede defaults in A-rated bonds, despite the observable signals, suggests that they may underweight downside risk in bonds perceived as categorically safe. In contrast, recommendations in the BBB segment are more aligned with future credit outcomes, showing more accurate and timely responsiveness to deteriorating fundamentals. This asymmetry in predictive power supports the original hypothesis that analysts apply different inference rules across rating groups: they emphasize upside and discount risk in A-rated bonds while treating BBB-rated bonds with greater caution. These patterns are consistent with categorical thinking.

7 Discussion and conclusion

Overall, I studied how key economic agents form expectations with respect to risk in the credit market. I start with documenting three stylized facts. First, analysts favor higher-yielding, riskier bonds only when these bonds carry safer credit ratings. Second, perceived credit risk extracted from comments predicts subsequent credit deterioration but is not reflected in recommendations for higher-rated bonds. Third, analysts characterize identical yield increases as more attractive when bonds appear safer. Traditional incentive-based explanations cannot explain these patterns. I then propose and provide evidence of a mechanism related to psychology—categorical thinking—whereby analysts mentally classify bonds based on rating labels, leading them to overvalue yields and underweight risks for purportedly safe bonds. Investors following these recommendations inadvertently reach for yield.

Figure 1. Distribution of recommendations across rating

This figure shows the distribution of analyst recommendations (Underweight, Neutral, and Overweight) across rating groups from A+ to BBB-. The sample at the bond-month level and from 2013 to 2023. Blue bars represent bonds rated A+ to A-, and purple bars represent bonds rated BBB+ to BBB-. As there are few AA and AAA firms, I combine these firms into A+ categories.
The figure illustrates a clear pattern: analysts issue more cautious recommendations (underweight) for higher-rated bonds and relatively more favorable recommendations (overweight) for lower-rated investment-grade bonds.

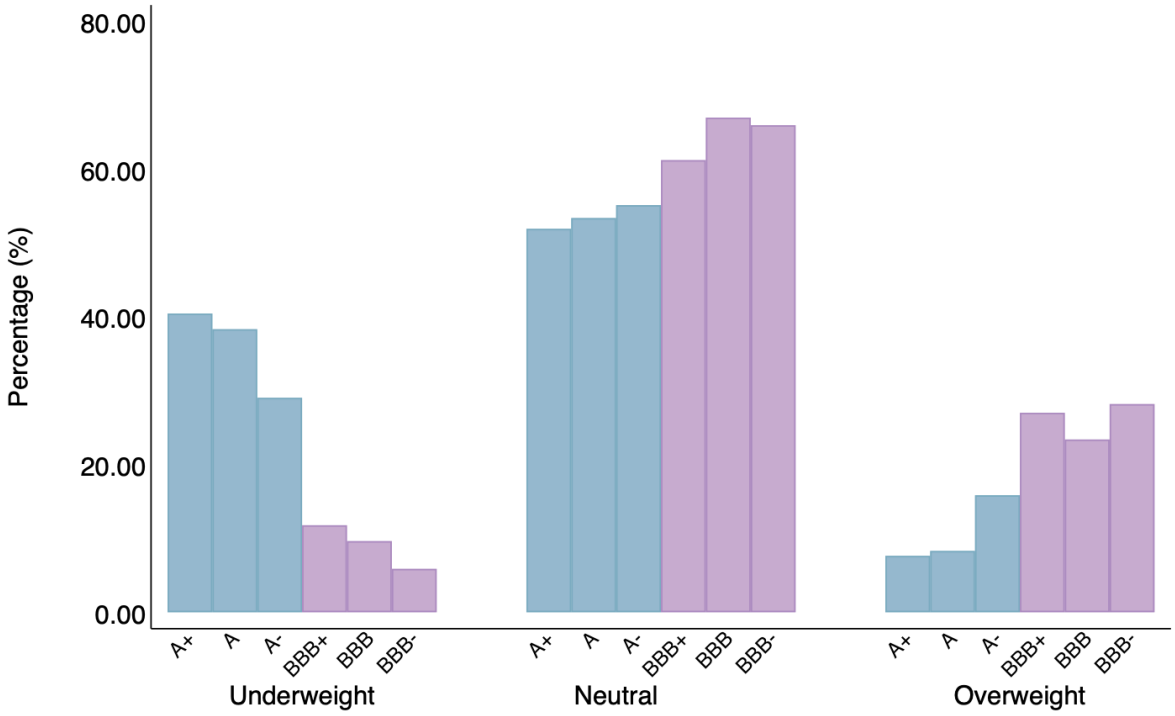
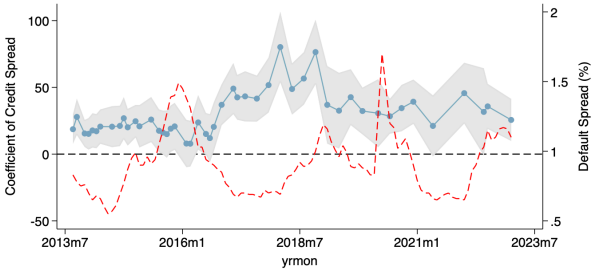


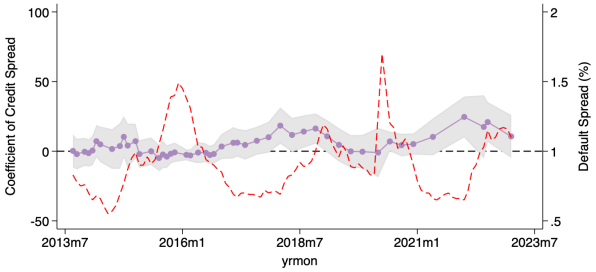
Figure 2. Time series of yield-seeking behavior

This figure shows time series of RFY. The coefficients of yield spread is negatively correlated with default spreads and slopes of treasury yields.

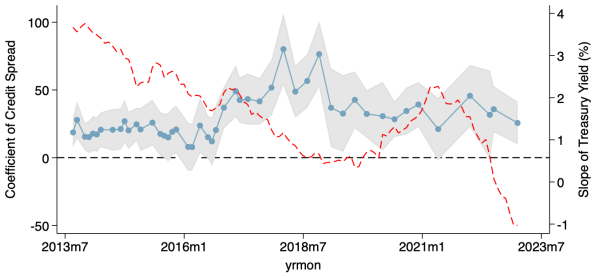
(a) A Bonds



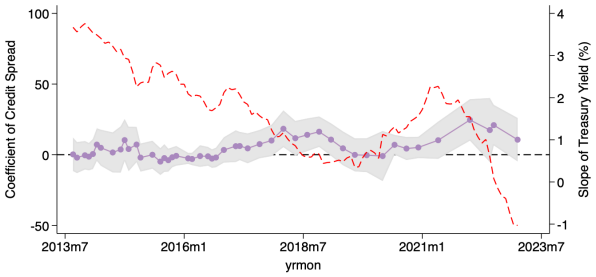
(b) BBB Bonds



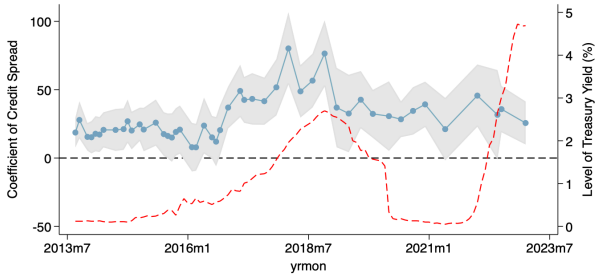
(c) A Bonds



(d) BBB Bonds



(e) A Bonds



(f) BBB Bonds

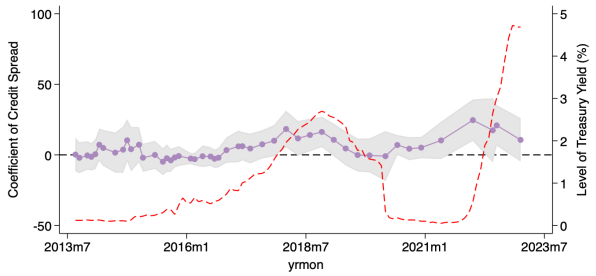
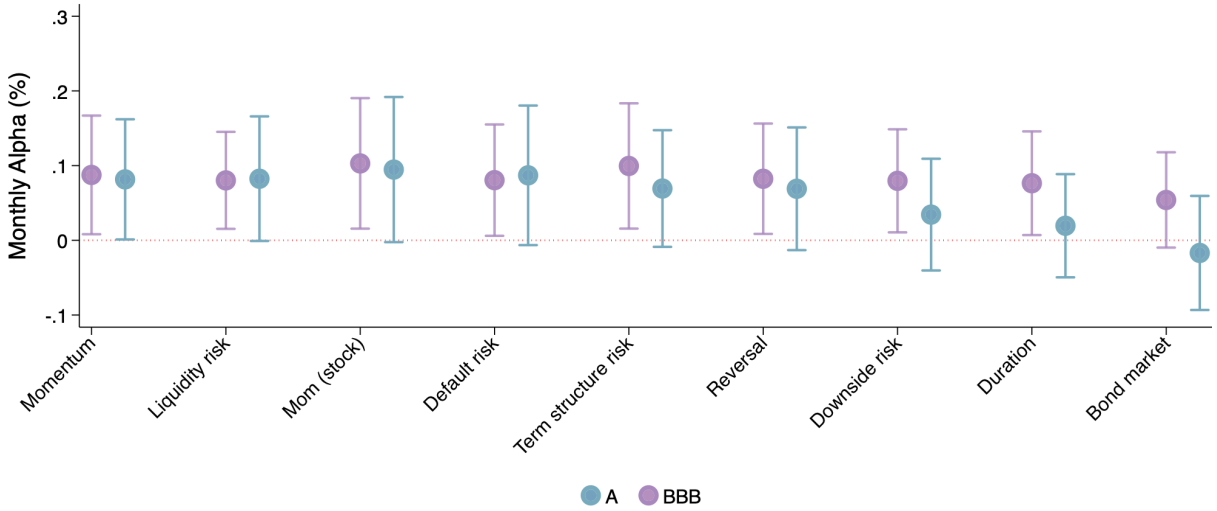


Figure 3. Alphas and Betas

This figure shows the distribution of analyst recommendations (Underweight, Neutral, and Overweight) across rating groups from A+ to BBB-. The sample at the bond-month level and from 2013 to 2023. Blue bars represent bonds rated A+ to A-, and purple bars represent bonds rated BBB+ to BBB-. As there are few AA and AAA firms, I combine these firms into A+ categories.



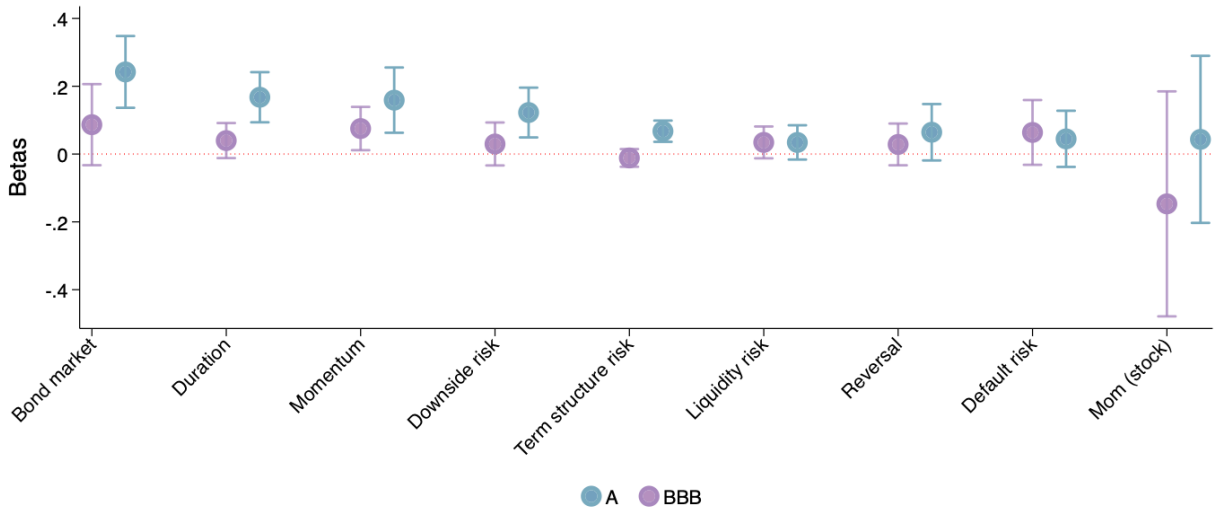
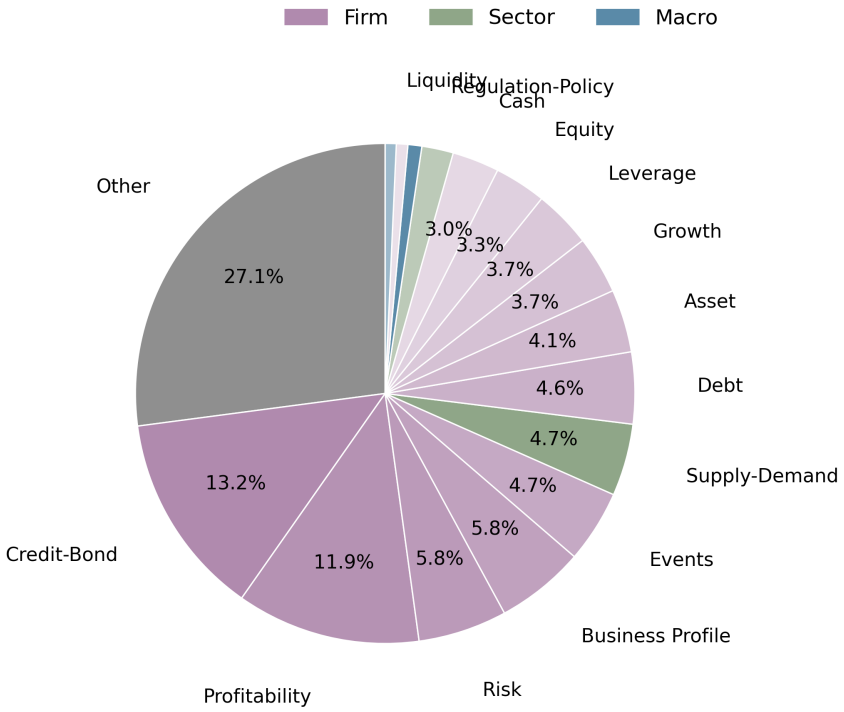
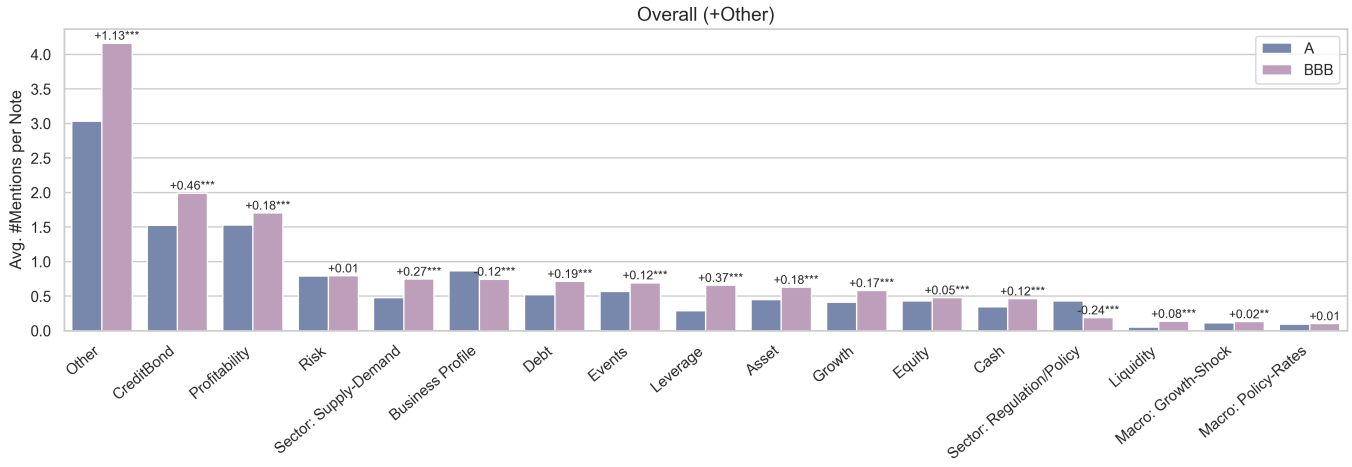


Figure 4. Analysts considerations

The pie chart shows the proportion of different types of financial concepts that analysts consider. The colors represent different concept types: purple for firm, green for sector, blue for macro, and grey for concepts that do not fit into these categories. The bars plot the number of financial concepts mentioned per comment. The blue bars are A-rated bonds, and the purple bars are BBB-rated bonds.



(a) Types of financial concepts



(b) Quantities of financial concepts

Table 1. Summary statistics of industry

This table presents the number of months each industry is covered in my sample, across Investment Grade (IG) reports. IG reports are available from September 2013 to March 2023, issued monthly until 2017, then quarterly from 2017 to 2020, and semi-annually after that. Most industries receive repeated coverage over time. The 'Others' row under IG includes additional sectors, such as Lodging and Housing (Homebuilders and Building Materials).

Sector	Months	Bond–month obs.	Share(%)
Utilities	48	22 560	22.5
Energy/MLPs/Midstream	43	13 012	13.0
Food/Beverage/Tobacco	48	11 875	11.8
Healthcare	48	9 598	9.6
Technology	48	6 057	6.0
Non-Food Retail	48	5 347	5.3
Manufacturing	47	4 597	4.6
Transportation	45	4 637	4.6
Telecom	47	4 491	4.5
Aerospace/Defense	48	4 018	4.0
Chemicals	47	3 085	3.1
Consumer Products	48	2 796	2.8
Automotive	37	2 571	2.6
Paper & Packaging	48	1 496	1.5
Cable/Satellite	47	1 376	1.4
Diversified Media	47	1 160	1.2
Metals/Mining	48	914	0.9
Others	48	850	0.8
Total		100 408	100.0

Table 2. Summary statistics of issuers, bonds, and recommendations in the sample

The table provides descriptive statistics for issuers, bonds, and recommendations in my sample. Panel A presents firm-month-level financial characteristics, including the number of bonds issued per firm and firm size. Panel B summarizes bond-month-level characteristics, such as credit rating, duration (in years), time to maturity (in years), yield spread, and monthly bond returns. For credit ratings, I use the lower rating if rated by two CRAs and the middle rating if rated by three CRAs (Becker and Ivashina, 2015). Yield spreads are calculated based on the methodology in Dickerson, Mueller and Robotti (2023b). Panel C presents the distribution of recommendations at both the firm and bond levels. I assign firm-level recommendations to bonds based on the disclosure. When analysts issue different recommendations for individual bonds (less than 5% cases), I calculate firm-level recommendations as the average of bond-level recommendations. A firm-level recommendation is classified as OW if the average falls within [0.5, 1], N if within (-0.5, 0.5), and UW if within [-1, -0.5]. Panel D reports the means of firm characteristics by credit rating for my sample of nonfinancial firms.

Panel A: Firm-month level variables

	N	Mean	SD	1st	25th	50th	75th	99th
Number of Bonds per Firm	8,323	12.06	10.86	1	5	9	16	53
Size (in million)	8,323	41,966.26	77,533.09	1,536.92	8,940.99	20,286.37	46,060.24	287,491.6

Panel B: Bond-month level variables

Credit rating	100,408			BBB−	BBB	BBB+	A	AAA
Duration (years)	100,135	7.73	4.94	1.16	3.55	6.44	12.26	17.85
Time to maturity (years)	100,408	11.41	9.28	1.17	3.83	7.50	20.00	29.67
Credit spread (%)	100,408	1.30	0.86	0.13	0.71	1.13	1.66	4.25
EDF	100,408	0.56	0.11	0.19	0.51	0.58	0.64	0.73
Monthly bond return (%)	99,309	0.58	2.50	−5.27	−0.39	0.27	1.39	8.50
Offering amount (log)	100,408	13.27	0.68	11.79	12.77	13.12	13.81	14.96
Coupon (%)	99,309	4.46	1.76	1.05	3.20	4.25	5.70	8.88
Trading volume (log)	100,408	16.23	1.91	10.84	15.13	16.49	17.57	19.82
Number of uncertain words	100,408	3.26	2.99	0	1	3	5	14

Panel C: Recommendations in the firm-month or bond-month level

	Firm		Bond	
	(N)	(%)	(N)	(%)
UW (Rec = -1)	1,987	23.87	21,955	21.87
N (Rec = 0)	4,848	58.25	59,803	59.56
OW (Rec = 1)	1,488	17.88	18,650	18.57
Total	8,323	100.00	100,408	100.00

Panel D: Means of firm characteristics by credit ratings

	AAA	AA	A+	A	A-	BBB+	BBB	BBB-
Assets	11.998	11.618	11.101	10.516	10.628	10.098	10.045	9.746
Book leverage	0.229	0.448	0.487	0.516	0.478	0.489	0.527	0.487
Market leverage	0.069	0.229	0.254	0.251	0.301	0.255	0.302	0.339
Cash/Asset	0.182	0.073	0.181	0.079	0.065	0.080	0.068	0.058
Z-Score	1.574	1.433	1.052	1.172	1.001	1.002	0.757	0.654
PPE/Asset	0.224	0.352	0.322	0.403	0.420	0.404	0.375	0.407
ROA	0.031	0.025	0.019	0.018	0.014	0.016	0.013	0.007
PM	0.186	0.100	0.133	0.107	0.101	0.102	0.073	0.024
Sales growth	0.021	-0.007	0.010	0.011	0.031	0.025	0.033	0.091
Earnings Growth	-0.006	0.042	0.095	0.134	0.124	0.077	0.096	-0.003
CF/PPE	0.422	0.267	0.423	0.228	0.313	0.291	0.281	0.193

Table 3. Recommendation and yield spreads

This table presents two complementary OLS regression specifications. The first quantifies how yield spreads shape analyst recommendations across matched A- and BBB-rated firms. The second tests whether the sensitivity of recommendations to yield spreads differs significantly between the two rating buckets.

$$\text{Rec}_{i,t} = \alpha + \sum_{g \in G} \beta_{1,g} \, \text{YieldSpread}_{i,t} \times G_{g,i,t} + \beta_2 \, \mathbf{1}_{\text{g=A},i,t} + \beta_3' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

$$\text{Rec}_{i,t} = \alpha + \beta_1 \, \text{YieldSpread}_{i,t} \times G_{g=A,i,t} + \beta_2 \, \mathbf{1}_{\text{g=A},i,t} + \beta_3' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

where the dependent variable $Rec_{i,t}$ is the analyst recommendation for bond i at time t, coded -1 = Underweight, 0 = Neutral, +1 = Overweight. $YieldSpread_{i,t}$ is the bond's yield spread over the yield of treasuries in the same maturity and expressed in decimal yield units. $G_{g,i,t}$ indicates the bond's rating group, where $g = A$ or BBB, and the interaction terms allow the sensitivity of recommendations to yield spreads to vary across rating groups. Controls $\mathbf{X}_{i,t}$ include bond characteristics, including duration, issue size, coupon rate, issue age, and an indicator for callability, and firm characteristics, including size, leverage, cash holdings, PPE to assets, cash flow to PPE, and ROA. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	Specification 1				Specification 2				Specification 2 with controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A × Yield Spread	21.760***	12.884**	17.034***	12.951**	19.794***	11.695**	16.096**	12.301**	17.197***	12.576**	16.702***	12.018*
	[3.82]	[2.66]	[3.26]	[2.56]	[3.28]	[2.32]	[2.74]	[2.27]	[2.86]	[2.27]	[2.92]	[2.05]
BBB × Yield Spread	1.966	1.189	0.938	0.650	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[1.05]	[1.02]	[0.56]	[0.52]	[.]	[.]	[.]	[.]	[.]	[.]	[.]	[.]
A	-0.601***	-0.547***	-0.625***	-0.568***	-0.601***	-0.547***	-0.625***	-0.568***	-0.616***	-0.570***	-0.660***	-0.571***
	[-4.34]	[-4.57]	[-4.53]	[-4.69]	[-4.34]	[-4.57]	[-4.53]	[-4.69]	[-3.94]	[-3.80]	[-4.02]	[-3.70]
Yield Spread					1.966	1.189	0.938	0.650	-2.543	-0.405	-1.432	-2.382
					[1.05]	[1.02]	[0.56]	[0.52]	[-0.78]	[-0.19]	[-0.68]	[-1.60]
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Analyst FE	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Controls	N	N	N	N	N	N	N	N	Y	Y	Y	Y
Observations	104,061	104,061	95,617	95,617	104,061	104,061	95,617	95,617	96,059	96,059	89,042	89,042
R Square	0.132	0.203	0.179	0.223	0.132	0.203	0.179	0.223	0.190	0.238	0.226	0.264

Table 4. Recommendation and EDF

This table presents two complementary OLS regression specifications. The first quantifies how credit signals shape analyst recommendations across matched A- and BBB-rated firms. The second tests whether the sensitivity of recommendations to each credit signal differs significantly between the two rating buckets.

$$\mathrm{Rec}_{i,t} = \alpha + \sum_{g \in G} \beta_{1,g}\, \mathrm{EDF}_{i,t} \times G_{g,i,t} + \beta_2\, \mathbf{1}_{\mathrm{g=A},i,t} + \beta_3' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

$$\mathrm{Rec}_{i,t} = \alpha + \beta_1\, \mathrm{EDF}_{i,t} \times G_{g=A,i,t} + \beta_2\, \mathbf{1}_{\mathrm{g=A},i,t} + \beta_3' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

where the dependent variable $Rec_{i,t}$ is the analyst recommendation for bond i at time t, coded –1 = Underweight, 0 = Neutral, +1 = Overweight. $CreditSpread_{i,t}$ is the bond's yield spread over the yield of treasuries in the same maturity and expressed in decimal yield units. $G_{g,i,t}$ indicates the bond's rating group, where $g = A$ or BBB, and the interaction terms allow the sensitivity of recommendations to yield spreads to vary across rating groups. $EDF_{i,t}$ is the firm's expected default frequency. Controls $\mathbf{X}_{i,t}$ include Macaulay duration, issue size, coupon rate, and an indicator for callability. Panel A reports estimates that include only the credit-spread interactions, Panel B repeats the exercise for EDF, and Panel C includes both signals jointly. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

Panel A: EDF

	Specification 1				Specification 2				Specification 2 with controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A × EDF	1.139**	0.871**	0.733*	0.819**	0.964**	0.659**	0.688*	0.671*	1.478***	1.088***	1.188***	1.226***
	[2.27]	[2.34]	[1.86]	[2.14]	[2.71]	[2.11]	[1.91]	[1.95]	[3.55]	[2.91]	[2.91]	[3.29]
BBB × EDF	0.174	0.212	0.045	0.148	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[0.49]	[0.70]	[0.12]	[0.47]	[.]	[.]	[.]	[.]	[.]	[.]	[.]	[.]
A	-0.940***	-0.779***	-0.848***	-0.802***	-0.940***	-0.779***	-0.848***	-0.802***	-1.264***	-1.030***	-1.151***	-1.111***
	[-6.79]	[-5.03]	[-4.80]	[-5.00]	[-6.79]	[-5.03]	[-4.80]	[-5.00]	[-6.07]	[-5.14]	[-6.20]	[-6.35]
EDF					0.174	0.212	0.045	0.148	-0.238	0.386	-0.056	0.133
					[0.49]	[0.70]	[0.12]	[0.47]	[-0.79]	[1.62]	[-0.16]	[0.42]
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Analyst FE	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Controls	N	N	N	N	N	N	N	N	Y	Y	Y	Y
Observations	104,061	104,061	95,617	95,617	104,061	104,061	95,617	95,617	96,059	96,059	89,042	89,042
R Square	0.139	0.208	0.179	0.227	0.139	0.208	0.179	0.227	0.201	0.252	0.233	0.278

Panel B: yield spreads and EDF

	Specification 1				Specification 2				Specification 2 with controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A × Yield Spread	18.115**	10.570*	15.282**	10.673*	16.470**	9.421*	14.163**	9.895	12.354*	7.634	12.356*	6.718
	[2.57]	[1.87]	[2.41]	[1.83]	[2.36]	[1.70]	[2.15]	[1.65]	[1.83]	[1.33]	[1.93]	[1.11]
BBB × Yield Spread	1.644	1.150	1.119	0.779	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[1.02]	[1.16]	[0.86]	[0.73]	[.]	[.]	[.]	[.]	[.]	[.]	[.]	[.]
Yield Spread					1.644	1.150	1.119	0.779	-1.046	0.416	-0.302	-1.209
					[1.02]	[1.16]	[0.86]	[0.73]	[-0.35]	[0.22]	[-0.15]	[-0.88]
A × EDF	1.018**	0.806**	0.654*	0.752*	0.881**	0.617*	0.621*	0.621*	1.338***	1.014**	1.067**	1.141***
	[2.09]	[2.15]	[1.72]	[1.98]	[2.65]	[1.98]	[1.77]	[1.80]	[3.38]	[2.73]	[2.64]	[3.01]
BBB × EDF	0.137	0.188	0.033	0.131	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[0.39]	[0.62]	[0.09]	[0.41]	[.]	[.]	[.]	[.]	[.]	[.]	[.]	[.]
EDF					0.137	0.188	0.033	0.131	-0.181	0.399	-0.027	0.161
					[0.39]	[0.62]	[0.09]	[0.41]	[-0.65]	[1.67]	[-0.08]	[0.50]
A	-1.037***	-0.839***	-0.934***	-0.862***	-1.037***	-0.839***	-0.934***	-0.862***	-1.306***	-1.061***	-1.199***	-1.133***
	[-7.46]	[-5.25]	[-5.07]	[-5.22]	[-7.46]	[-5.25]	[-5.07]	[-5.22]	[-6.73]	[-5.18]	[-6.74]	[-6.33]
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Analyst FE	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Controls	N	N	N	N	N	N	N	N	Y	Y	Y	Y
Observations	104,061	104,061	95,617	95,617	104,061	104,061	95,617	95,617	96,059	96,059	89,042	89,042
R Square	0.148	0.211	0.185	0.230	0.148	0.211	0.185	0.230	0.204	0.254	0.236	0.278

Table 5. Recommendation and yield spread: Heterogeneity by analysts

The table reports OLS estimates of β_7 from the following specification:

$$\begin{aligned}\text{Rec}_{it} &= \alpha_0 + \beta_1 A_{i,t} + \beta_2 \,\text{Yield Spread}_{it} + \beta_3\big(A_{i,t} \times \text{Yield Spread}_{it}\big) + \beta_4 M_{i,t} + \beta_5\big(A_{i,t} \times M_{i,t}\big) + \beta_6\big(M_{i,t} \times \text{Yield Spread}_{it}\big) \\ &\quad + \beta_7\big(A_{i,t} \times M_{i,t} \times \text{Yield Spread}_{it}\big) + \gamma_{yrmon} + \gamma_{industry} + u_{it},\end{aligned}$$

where Rec_{it} is analyst j recommendation for bond i in month t, A_i equals one for A-rated bonds and zero for BBB-rated bonds, $Spread_{it}$ is the bond's yield spread, and M_j is a dummy for the analyst characteristic shown at the top of each pair of columns. Analyst traits are coded as follows: gender (male = 1, female = 0); undergraduate major (economics = 0, STEM = 1); highest degree (bachelor = 0, master = 1); LinkedIn following (fewer than one thousand = 0, at least one thousand = 1); and JPMorgan tenure (ten years or less = 0, more than ten years = 1). The reported coefficient is β_7, the triple interaction $A \times Spread \times Trait$, which indicates how the A-minus-BBB spread slope varies with the analyst attribute. All regressions include the full vector of bond and firm controls, time fixed effects γ_t, and, where indicated, industry fixed effects $\gamma_{industry}$. Standard errors are two-way clustered by issuer and analyst. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	Gender		Edu-Major		Edu-Degree		Work-followers		Work-years	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A × Yield Spread × Male	-16.373***	-16.994***								
	[-4.02]	[-3.75]								
A × Yield Spread × Stem			-16.571*	-15.069						
			[-1.80]	[-1.45]						
A × Yield Spread × Master					-18.023**	-13.826*				
					[-2.19]	[-1.84]				
A × Yield Spread × Followers>750							15.996*	14.656*		
							[2.01]	[1.82]		
A × Yield Spread × working years>10									9.993**	13.563**
									[2.07]	[2.50]
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	N	Y	N	Y	N	Y	N	Y	N	Y
Observations	96,059	96,059	82,609	82,609	82,609	82,609	96,059	96,059	96,059	96,059
R Square	0.204	0.246	0.212	0.246	0.194	0.234	0.194	0.241	0.203	0.251

Table 6. Textual risk and objective risk

This table presents OLS regression results examining how analysts' textual expressions of risk relate to firms' expected default frequency (EDF). Columns (1)-(4) use the LLM-generated risk score derived from analyst comments as the dependent variable, while columns (5)-(8) utilize the frequency of uncertain words (following Loughran and McDonald, 2011) as an alternative measure of perceived risk. The main independent variables include EDF interacted with rating-group indicators (A-rated and BBB-rated bonds) and an A-rating dummy. Columns (1), (3), (5), and (7) control for time fixed effects only, whereas columns (2), (4), (6), and (8) include both time and industry fixed effects. Robust standard errors clustered at the issuer level are reported in brackets. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	Risk score				Uncertain text			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A × EDF	0.663	0.907**	-0.997*	-0.700*	2.115**	2.186	-4.685***	-3.009**
	[1.19]	[2.18]	[-1.99]	[-1.71]	[2.77]	[1.12]	[-2.88]	[-2.43]
BBB × EDF	1.660***	1.607***	0.000	0.000	6.800***	5.195***	0.000	0.000
	[3.16]	[4.68]	[.]	[.]	[4.24]	[3.66]	[.]	[.]
EDF			1.660***	1.607***			6.800***	5.195***
			[3.16]	[4.68]			[4.24]	[3.66]
A	0.737**	0.677**	0.737**	0.677**	2.067**	1.655**	2.067**	1.655**
	[2.44]	[2.56]	[2.44]	[2.56]	[2.56]	[2.61]	[2.56]	[2.61]
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
Observations	99,942	99,942	99,942	99,942	99,942	99,942	99,942	99,942
R Square	0.049	0.091	0.049	0.091	0.287	0.434	0.287	0.434

Table 7. Recommendation, perceived risk and future risk materialization

This table reports logit regression results examining whether analysts' textual expressions of risk and uncertainty help predict future credit events. The dependent variable is an indicator equal to one if the issuer experiences a credit downgrade within three, six, or twelve months (DG), or defaults within three years (DF), and zero otherwise. Columns (1)-(4) use recommendation as explanatory variables. Columns (5)-(8) use the LLM-generated risk score from analyst comments as the main explanatory variable. Columns (9)-(12) utilize the frequency of uncertain words (following Loughran and McDonald, 2011) as an alternative measure of perceived risk. Both measures are interacted with rating-group indicators (A-rated and BBB-rated bonds). All regressions control for time fixed effects. Robust standard errors clustered at the issuer level are reported in brackets. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	DG(3m)	DG(6m)	DG(12m)	DF(3y)	DG(3m)	DG(6m)	DG(12m)	DF(3y)	DG(3m)	DG(6m)	DG(12m)	DF(3y)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A × Rec	-0.175	-0.154	-0.181	0.754***								
	[-0.84]	[-0.76]	[-0.88]	[3.32]								
BBB × Rec	-0.566***	-0.515***	-0.365**	-0.254								
	[-3.48]	[-3.30]	[-2.43]	[-0.84]								
A × Risk score					0.336***	0.251***	0.221**	0.814***				
					[3.17]	[2.68]	[2.53]	[4.19]				
BBB × Risk score					0.347***	0.254***	0.216***	0.161				
					[3.92]	[3.14]	[3.02]	[0.61]				
A × Uncertain text									0.180***	0.139***	0.135***	0.254***
									[4.02]	[3.71]	[4.08]	[3.50]
BBB × Uncertain text									0.063**	0.063*	0.046	0.039
									[2.25]	[1.95]	[1.35]	[0.35]
A	-0.088	-0.078	-0.005	2.071***	-0.021	-0.011	0.063	3.401***	-0.301	-0.150	-0.126	1.231*
	[-0.49]	[-0.44]	[-0.03]	[5.02]	[-0.11]	[-0.06]	[0.34]	[6.68]	[-1.21]	[-0.62]	[-0.51]	[1.77]
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	100,129	100,214	100,235	58,619	100,129	100,214	100,235	58,619	100,129	100,214	100,235	58,619
Pseudo R^2	0.062	0.059	0.045	0.114	0.075	0.066	0.052	0.115	0.069	0.062	0.049	0.139

Table 8. Recommendations and textual risk

This table reports results from OLS regressions examining the relationship between analysts' textual expressions of risk and their bond recommendations. The dependent variable is the numerical recommendation score assigned by analysts, where higher values correspond to more favorable recommendations. Columns (1)-(4) measure perceived risk using the LLM-generated risk score, while columns (5)-(8) use the frequency of uncertain words (following Loughran and McDonald, 2011). These risk measures are interacted with rating-group indicators (A-rated and BBB-rated bonds). yield spreads are also interacted similarly and included as controls. All regressions control for time fixed effects, and columns (2), (4), (6), and (8) further include industry fixed effects. Robust standard errors clustered at the issuer level are shown in brackets. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	Dep varlable: Recommendation							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A × Risk score	0.049	0.041	0.044	0.038				
	[0.91]	[0.83]	[0.84]	[0.78]				
BBB × Risk score	-0.144**	-0.127*	-0.151**	-0.131*				
	[-2.30]	[-1.80]	[-2.41]	[-1.86]				
A × Uncertain text					0.031	0.029	0.026	0.027
					[1.43]	[1.53]	[1.29]	[1.43]
BBB × Uncertain text					-0.012	-0.002	-0.013	-0.003
					[-1.16]	[-0.16]	[-1.26]	[-0.28]
A × Credit spread			21.903***	13.384**			20.665***	12.236**
			[3.70]	[2.61]			[3.76]	[2.53]
BBB × Credit spread			2.653*	1.711*			1.788	1.045
			[1.86]	[1.88]			[0.99]	[0.93]
A	-0.003	-0.070	-0.157	-0.164	-0.562***	-0.543***	-0.714***	-0.637***
	[-0.02]	[-0.53]	[-1.21]	[-1.15]	[-3.66]	[-4.29]	[-4.03]	[-4.02]
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
Observations	99,942	99,942	99,942	99,942	99,942	99,942	99,942	99,942
R Square	0.130	0.206	0.144	0.211	0.127	0.203	0.140	0.207

Table 9. Yield spreads, yield tone, and recommendations

This table reports results of OLS regressions. Panel A regresses yield tone on option-adjusted yield spreads, an A-rating indicator, and their interactions. Yield tone is an LLM-derived measure of analysts' qualitative assessments of bond yield spreads, coded from –1 (tight) to 0 (neutral) to +1 (wide). Panel B regresses analyst recommendations (–1 = Underweight, 0 = Neutral, +1 = Overweight) on yield tone. All specifications include month fixed effects, and columns (2), (4), (6), and (8) add industry fixed effects. Standard errors are clustered by issuer and analysts. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

Panel A: Yield spreads and yield tone

	Dep variable: Yield tone							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A × Yield Spread	28.689***	29.856***	22.949***	24.493***	14.094**	15.517**	8.005	9.736*
	[4.33]	[4.39]	[3.37]	[3.93]	[2.49]	[2.75]	[1.44]	[1.86]
BBB × Yield Spread	14.595***	14.339***	14.944***	14.757***	0.000	0.000	0.000	0.000
	[5.19]	[5.32]	[5.41]	[5.72]	[.]	[.]	[.]	[.]
Yield Spread					14.595***	14.339***	14.944***	14.757***
					[5.19]	[5.32]	[5.41]	[5.72]
Risk score		-0.166***		-0.163***		-0.166***		-0.163***
		[-5.31]		[-4.71]		[-5.31]		[-4.71]
A	-0.468***	-0.445***	-0.438***	-0.405***	-0.468***	-0.445***	-0.438***	-0.405***
	[-4.32]	[-3.89]	[-3.37]	[-2.97]	[-4.32]	[-3.89]	[-3.37]	[-2.97]
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	N	Y	N	Y	N	Y	N	Y
Observations	38,311	38,311	38,310	38,310	38,311	38,311	38,310	38,310
R Square	0.117	0.171	0.164	0.214	0.117	0.171	0.164	0.214

Panel B: Yield tone and recommendation

	Dep variable: Recommendation			
	(1)	(2)	(3)	(4)
A × Yield Tone	0.323***	0.285***	0.046	0.043
	[5.00]	[4.66]	[0.66]	[0.68]
BBB × Yield Tone	0.276***	0.241***	0.000	0.000
	[6.88]	[6.02]	[.]	[.]
A	-0.343***	-0.343***	-0.343***	-0.343***
	[-5.08]	[-5.53]	[-5.08]	[-5.53]
Time FE	Y	Y	Y	Y
Industry FE	N	Y	N	Y
Observations	38,311	38,310	38,311	38,310
R Square	0.295	0.378	0.295	0.378

Table 10. Performance of Portfolios Sorted on Rating and Recommendation

This table reports summary statistics, average excess returns, and alphas for 3-by-2 portfolios sorted by analyst recommendations and bond ratings. At the end of each month t, bonds are sorted based on their current-month analyst recommendations and held for three months from $t+1$ to $t+3$. Summary statistics are bond-month characteristics observed at month t. Alphas are estimated by regressing portfolio excess returns on the CAPM and the Fama–French bond factors. Newey and West (1987) t-statistics with two lags are shown in bracket.

Panel A: Bond-month level variables

		Underweight	Neutral	Overweight
A	# of bonds	376.75	549.78	113.55
	Rating	A+	A	A
	Time to maturity (yrs)	10.79	12.51	14.24
	Yield spread (%)	0.8	0.96	1.07
	Yield (%)	2.72	2.98	3.15
	Ret (%)	0.36	0.36	0.38
	Trading volume_log (%)	16.11	16.11	15.74
	Offering amount_log (%)	13.25	13.27	13.14
BBB	# of bonds	123.88	836.01	316.44
	Rating	BBB	BBB	BBB
	Time to maturity (yrs)	10.12	10.76	10.91
	Yield spread (%)	1.5	1.63	1.61
	Yield (%)	3.37	3.54	3.51
	Ret (%)	0.3	0.49	0.52
	Trading volume_log (%)	16.26	16.37	16.47
	Offering amount_log (%)	13.19	13.29	13.33

Panel B: Excess return and alpha

		Equal Weighted				Value Weighted			
		UW	N	OW	OWUW	UW	N	OW	OWUW
A	Excess return	0.333***	0.375**	0.433**	0.100	0.332***	0.363**	0.433**	0.102*
		[2.74]	[2.51]	[2.57]	[1.63]	[2.69]	[2.49]	[2.60]	[1.67]
	α_{FF}	0.050	0.021	0.058	0.008	0.037	0.006	0.056	0.019
		[1.38]	[0.55]	[1.16]	[0.21]	[1.11]	[0.15]	[1.00]	[0.40]
	α_{CAPMB}	-0.055	-0.107*	-0.092	-0.036	-0.062	-0.109*	-0.084	-0.021
		[-1.07]	[-1.78]	[-1.30]	[-0.86]	[-1.17]	[-1.82]	[-1.09]	[-0.42]
BBB	Excess return	0.451***	0.474***	0.536***	0.085	0.450***	0.484***	0.549***	0.099**
		[3.31]	[3.30]	[3.56]	[1.65]	[3.09]	[3.20]	[3.46]	[2.17]
	α_{FF}	0.125*	0.129*	0.183**	0.058	0.094	0.113	0.167**	0.073*
		[1.85]	[1.75]	[2.43]	[1.42]	[1.50]	[1.59]	[2.29]	[1.78]
	α_{CAPMB}	0.015	0.0035	0.039	0.023	-0.026	-0.010	0.025	0.051
		[0.41]	[0.09]	[1.08]	[0.66]	[-0.78]	[-0.26]	[0.69]	[1.41]

Table 11. Recommendation and yield spread: client relationship

The first four columns report OLS estimates of β_1 from the following specification:

$$\text{Rec}_{i,t} = \alpha + \beta_1 \,\text{YieldSpread}_{i,t} \times G_{g=A,i,t} + \beta_2 \,\mathbf{1}_{\text{g=A},i,t} + \beta_3' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

where columns (1) and (2) focus on the sample without any client relationship and columns (3) and (4) focus on the sample with client relationship.
Columns (5) to (10) report OLS estimates of β_7 from the following specification:

$$\begin{aligned}\text{Rec}_{it} \;&= \alpha_0 + \beta_1 A_{i,t} + \beta_2 \,\text{Yield Spread}_{it} + \beta_3\big(A_{i,t} \times \text{Yield Spread}_{it}\big) + \beta_4 M_{i,t} + \beta_5\big(A_{i,t} \times M_{i,t}\big) + \beta_6\big(M_{i,t} \times \text{Yield Spread}_{it}\big) \\ &\quad + \beta_7\big(A_{i,t} \times M_{i,t} \times \text{Yield Spread}_{it}\big) + \gamma_{yrmon} + \gamma_{industry} + u_{it},\end{aligned}$$

where Rec_{it} is analyst j recommendation for bond i in month t, A_i equals one for A-rated bonds and zero for BBB-rated bonds, $Spread_{it}$ is the bond's yield spread, and M_j is a dummy for the client relationship. It can be manager, broker, or others. All represents all these three types of client relationships. The reported coefficient is β_7, the triple interaction $A \times Spread \times Relationship$, which indicates how the A-minus-BBB spread slope varies with the client relationship. All regressions include the full vector of bond and firm controls, time fixed effects γ_t, and, where indicated, industry fixed effects $\gamma_{industry}$. Standard errors are two-way clustered by issuer and analyst. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	No relationship		With relationship		Whole sample1		Whole sample2		Whole sample3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A × Yield Spread	16.650**	12.464*	11.673**	1.411						
	[2.47]	[2.03]	[2.58]	[0.39]						
A × Yield Spread × Relationship(all)							-2.468	-4.178	-1.244	-2.422
							[-0.32]	[-0.60]	[-0.16]	[-0.34]
A × Yield Spread × Relationship(manager)					-2.861	-4.958				
					[-0.35]	[-0.67]				
A × Yield Spread × Relationship(broker)					-3.203	4.612				
					[-0.49]	[0.70]				
A × Yield Spread × Relationship(others)					27.462	35.547*				
					[1.25]	[1.80]				
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	N	Y	N	Y	N	Y	N	Y	N	Y
Observations	82,776	82,776	13,283	13,283	96,059	96,059	96,059	96,059	96,059	96,059
Adj. R^2	0.182	0.232	0.266	0.395	0.193	0.240	0.192	0.240	0.196	0.243

Table 12. Characteristics of matched A and BBB firms

This table reports the characteristics of matched A and BBB firms. Each month, an A (A-, A, A+) non-utility firm is matched to the BBB (BBB-, BBB, BBB+) firm that is closest in terms of log assets, book leverage, market leverage, EDF, Cash to assets, Z-Score, and PPE to assets. Distance is defined as the absolute difference between the log-odds of the propensity scores obtained from the logit model, and a BBB candidate is accepted as the nearest neighbor only when the distance lies within a caliper of 0.35. Utility firms are matched in the same way except that Z-score is omitted. The full sample contains 3,265 A and 5,287 BBB firm-month observations. The matched sample consists of 1,556 A-rated firm-month observations matched to 846 unique BBB-rated firm-month observations. The sample period is from September 2013 to January 2023. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	Full sample			Matched sample		
	A	BBB	Δ	A	BBB	Δ
Assets	10.760	9.974	0.787***	10.507	10.519	-0.012
Book leverage	0.488	0.505	-0.017***	0.498	0.501	-0.003
Market leverage	0.262	0.300	-0.038***	0.223	0.229	-0.006
EDF	0.510	0.583	-0.073***	0.520	0.520	0.000
Cash/Asset	0.090	0.069	0.022***	0.110	0.105	0.005
Z-Score	1.155	0.786	0.368***	1.056	1.046	0.011
PPE/Asset	0.389	0.392	-0.003***	0.272	0.263	0.009
ROA	0.018	0.012	0.006***	0.019	0.019	0.000
PM	0.110	0.067	0.043***	0.097	0.100	-0.003
Sales growth	0.016	0.047	-0.031***	0.017	0.024	-0.008
Earnings growths	0.114	0.062	0.052*	0.074	0.091	-0.017
CF/PPE	0.290	0.259	0.031***	0.354	0.352	0.001

Table 13. Recommendation formation in matched A and BBB firms (matching by time and industry)

This table presents two complementary regression specifications. The first quantifies how credit signals shape analyst recommendations across matched A- and BBB-rated firms. The second tests whether the sensitivity of recommendations to each credit signal differs significantly between the two rating buckets.

$$\text{Rec}_{i,t} = \alpha + \sum_{g \in G} \beta_{1,g}\, \text{CreditSpread}_{i,t} \times G_{g,i,t} + \sum_{g \in G} \beta_{2,g}\, \text{EDF}_{i,t} \times G_{g,i,t} + \beta_3\, \mathbf{1}_{\text{g=BBB},i,t} + \beta_4' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

$$\text{Rec}_{i,t} = \alpha + \beta_1\, \text{CreditSpread}_{i,t} \times G_{g=BBB,i,t} + \beta_2\, \text{EDF}_{i,t} \times G_{g=A,i,t} + \beta_3\, \mathbf{1}_{\text{g=BBB},i,t} + \beta_4' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

where the dependent variable $Rec_{i,t}$ is the analyst recommendation for bond i at time t, coded –1 = Underweight, 0 = Neutral, +1 = Overweight. $CreditSpread_{i,t}$ is the bond's yield spread over the yield of treasuries in the same maturity and expressed in decimal yield units. $G_{g,i,t}$ indicates the bond's rating group, where $g = A$ or BBB, and the interaction terms allow the sensitivity of recommendations to yield spreads to vary across rating groups. $EDF_{i,t}$ is the firm's expected default frequency. Controls $\mathbf{X}_{i,t}$ include Macaulay duration, issue size, coupon rate, and an indicator for callability. Panel A reports estimates that include only the credit-spread interactions, Panel B repeats the exercise for EDF, and Panel C includes both signals jointly. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	Credit Spread				EDF				EDF and Credit Spread			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A × Credit Spread	9.189***								7.219**			
	[3.08]								[2.33]			
BBB × Credit Spread	-0.364	-9.553**	-10.063***	-10.063***					-0.665	-7.884*	-8.238*	-8.238*
	[-0.12]	[-2.82]	[-3.06]	[-3.06]					[-0.22]	[-1.94]	[-2.05]	[-2.05]
Credit Spread		9.189***	15.045***	15.045***						7.219**	11.741***	11.741***
		[3.08]	[3.65]	[3.65]						[2.33]	[3.00]	[3.00]
A × EDF					2.407***				2.340***			
					[3.94]				[3.78]			
BBB × EDF					1.501*	-0.906	-0.904	-0.904	1.529*	-0.811	-0.808	-0.808
					[2.04]	[-1.10]	[-1.09]	[-1.09]	[2.07]	[-0.97]	[-0.96]	[-0.96]
EDF						2.407***	2.388***	2.388***		2.340***	2.278***	2.278***
						[3.94]	[3.87]	[3.87]		[3.78]	[3.67]	[3.67]
BBB	0.412***	0.412***	0.394***	0.394***	0.835*	0.835*	0.826*	0.826*	0.865*	0.865*	0.847*	0.847*
	[5.42]	[5.42]	[5.22]	[5.22]	[1.91]	[1.91]	[1.86]	[1.86]	[2.03]	[2.03]	[1.93]	[1.93]
Pair FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Analyst FE	N	N	N	Y	N	N	N	Y	N	N	N	Y
Observations	9,692	9,692	9,556	9,556	9,692	9,692	9,556	9,556	9,692	9,692	9,556	9,556
R Square	0.540	0.540	0.546	0.546	0.556	0.556	0.560	0.560	0.557	0.557	0.563	0.563

Table 14. Supportive evidence of categorical thinking: Number of financial concepts

This table reports the regression results examining the level of detail analysts provide in their notes, specifically measuring how many financial concepts analysts mention. Columns (1)-(3) use the total number of financial concepts in each analyst note as the dependent variable. Columns (4)-(6) focus only on the factual description sections, while columns (7)-(9) isolate the subjective, forward-looking parts of the notes. Columns (10)-(12) analyze the likelihood that analysts omit factual descriptions entirely, presenting the probability that notes contain only subjective predictions without any accompanying factual information. *, **, and *** indicate significance at the 10%, 5%, and 10% level, respectively.

Dep var: Num of concepts in	Whole comment			Fact. description			Sub. prediction			Prob of no fact.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
main												
A	-2.011***	-0.684***	-0.331*	-1.024***	-0.205	0.001	-0.986***	-0.479***	-0.333***	0.497**	0.539*	0.375
	[-4.95]	[-3.55]	[-2.04]	[-3.78]	[-1.60]	[0.01]	[-5.82]	[-3.76]	[-3.22]	[2.05]	[1.95]	[1.27]
EDF			5.032**			2.950**			2.082***			-2.234**
			[2.75]			[2.25]			[3.78]			[-2.42]
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FEs	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	89,803	89,803	89,803	89,803	89,803	89,803	89,803	89,803	89,803	88,668	87,380	87,380
R Square	0.276	0.456	0.467	0.145	0.308	0.317	0.321	0.440	0.446			

Table 15. Future downgrade, default, upgrade and Recommendation

This table investigates the predictive power of analyst recommendations for future downgrades or defaults using the following regression specification:

$$\text{Downgrade}_{i,t+3} = \alpha + \beta_1 \cdot \text{Rec}_{i,t} \times \sum_{g \in G} \text{RatingGroup}_g + \beta_2 \cdot \text{RatingGroup Dummy}_g + \gamma_t + \epsilon_{i,t}$$

$Downgrade_{i,t+3}$, is an indicator that equals one if any of the three rating agencies (S&P, Fitch, and Moody's) downgrade the bond i within the next 3 months after the recommendation. Columns (1) to (4) focus on the future downgrades in the next 3 months, and columns (5) to (8) look at the future downgrades in the next six months. In columns (8) to (12), the dependent variable Default equals one if the bond defaults within the next three years, and zero otherwise. Standard errors are clustered at the firm level and reported in brackets below the coefficients. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Logit reg - Downgrade and Default

	DG(3m)			DG(6m)			DF		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A × Rec	-0.206	-0.344	-0.375	-0.179	-0.317	-0.352	0.755***	0.616**	0.585**
	[-0.99]	[-1.41]	[-1.64]	[-0.89]	[-1.31]	[-1.55]	[3.34]	[2.55]	[2.48]
BBB × Rec	-0.559***	-0.592***	-0.582***	-0.517***	-0.545***	-0.548***	-0.253	-0.303	-0.246
	[-3.49]	[-3.62]	[-3.46]	[-3.36]	[-3.51]	[-3.37]	[-0.84]	[-0.95]	[-0.70]
A	-0.065	0.070	0.388**	-0.064	0.069	0.385**	2.072***	2.259***	2.562***
	[-0.36]	[0.41]	[2.19]	[-0.36]	[0.41]	[2.29]	[5.03]	[5.60]	[6.28]
EDF		3.730**	2.570		3.550**	2.494*		4.097***	3.600***
		[2.21]	[1.64]		[2.21]	[1.67]		[3.27]	[2.98]
Credit spread			42.112***			44.186***			30.509***
			[6.24]			[5.86]			[2.60]
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	104,285	104,285	104,285	104,374	104,374	104,374	59,143	59,143	59,143
Pseudo R^2	0.063	0.074	0.090	0.063	0.075	0.093	0.114	0.126	0.130

Panel B: Logit reg - Upgrade

	Upgrade(3m)				Upgrade(6m)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A × Rec	0.712***	0.721***	0.740***	0.741***	0.714***	0.724***	0.741***	0.744***
	[3.47]	[3.42]	[3.63]	[3.56]	[3.49]	[3.44]	[3.64]	[3.58]
BBB × Rec	0.133	0.135	0.135	0.136	0.138	0.140	0.140	0.141
	[0.68]	[0.70]	[0.70]	[0.71]	[0.71]	[0.72]	[0.73]	[0.73]
A	0.065	0.053	-0.063	-0.064	0.079	0.067	-0.045	-0.047
	[0.35]	[0.28]	[-0.32]	[-0.32]	[0.43]	[0.35]	[-0.23]	[-0.23]
EDF		-0.289		-0.056		-0.293		-0.070
		[-0.29]		[-0.05]		[-0.29]		[-0.06]
Credit spread			-21.101***	-20.954***			-20.485***	-20.303***
			[-2.99]	[-2.73]			[-2.87]	[-2.62]
Time FEs	Y	Y	Y	Y	Y	Y	Y	Y
Observations	101,807	101,807	101,807	101,807	101,896	101,896	101,896	101,896
Pseudo R^2	0.114	0.114	0.116	0.116	0.114	0.114	0.115	0.115

Panel C: Logit reg - A- and BBB+: Downgrades and Upgrades

	Downgrade(3m)				Upgrade(3m)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A- × Rec	-0.268	-0.299	-0.316	-0.334	0.687***	0.687***	0.696***	0.697***
	[-0.93]	[-0.98]	[-1.09]	[-1.10]	[2.91]	[2.87]	[2.95]	[2.93]
BBB+ × Rec	-0.532**	-0.541**	-0.586**	-0.589**	0.301	0.301	0.316	0.316
	[-2.17]	[-2.15]	[-2.28]	[-2.26]	[0.76]	[0.76]	[0.81]	[0.80]
BBB+	-0.453*	-0.453*	-0.523**	-0.521**	-0.750**	-0.750**	-0.717**	-0.717**
	[-1.82]	[-1.81]	[-2.08]	[-2.07]	[-2.23]	[-2.23]	[-2.13]	[-2.13]
EDF		1.250		0.899		-0.016		0.060
		[0.77]		[0.60]		[-0.01]		[0.05]
Credit spread			36.074***	34.258***			24.478	24.544

References

Acharya, Viral and Hassan Naqvi, "On reaching for yield and the coexistence of bubbles and negative bubbles," *Journal of Financial Intermediation*, 2019, *38*, 1–10.

Adam, Klaus and Stefan Nagel, "Expectations data in asset pricing," in "Handbook of economic expectations," Elsevier, 2023, pp. 477–506.

Agrawal, Anup and Mark A Chen, "Analyst conflicts and research quality," *The Quarterly Journal of Finance*, 2012, *2* (02), 1250010.

Ajinkya, Bipin B, Rowland K Atiase, and Michael J Gift, "Volume of trading and the dispersion in financial analysts' earnings forecasts," *Accounting Review*, 1991, pp. 389–401.

Asker, John, Joan Farre-Mensa, and Alexander Ljungqvist, "Corporate Investment and Stock Market Listing: A Puzzle?," *Review of Financial Studies*, 2015, *28* (2), 342–390.

Bacher, Annika, "The Gender Investment Gap over the Life Cycle," *Review of Financial Studies*, 2024. Forthcoming.

Basel Committee, "Calculation of RWA for Credit Risk – Standardised Approach: CRE20," https://www.bis.org/basel_framework/chapter/CRE/20.htm?inforce=20191215&published=20191215&tldate=20220113 2019.

Becker, Bo and Todd Milbourn, "How Did Increased Competition Affect Credit Ratings?," *Journal of Financial Economics*, 2011, *101* (3), 493–514.

_ **and Victoria Ivashina**, "Reaching for yield in the bond market," *The Journal of Finance*, 2015, *70* (5), 1863–1902.

Bharath, Sreedhar T. and Tyler Shumway, "Forecasting default with the Merton distance to default model," *The Review of Financial Studies*, 2008, *21* (3), 1339–1369.

Bloomberg, "Bloomberg Fixed Income Index Methodology," https://assets.bbhub.io/professional/sites/10/Bloomberg-Index-Publications-Fixed-Income-Index-Methodology.pdf 2024.

Bolton, Patrick, Xavier Freixas, and Joel Shapiro, "The Credit Ratings Game," *The Journal of Finance*, 2012, *67* (1), 85–111.

Bongaerts, Dion, Martijn Cremers, and William N. Goetzmann, "Tiebreaker: Certification and Multiple Credit Ratings," *The Journal of Finance*, 2012, *67* (1), 113–152.

Bradley, Daniel, Sinan Gökkaya, and Xi Liu, "Before an Analyst Becomes an Analyst: Does Industry Experience Matter?," *Journal of Finance*, 2017, *72* (2), 751–792.

Brooks, Jordan, Michael Katz, and Scott Richardson, "Systematic Credit Investing," White Paper, AQR Capital Management 2016.

Buraschi, Andrea, Ilaria Piatti, and Paul Whelan, "Subjective bond returns and

belief aggregation," *The Review of Financial Studies*, 2022, *35* (8), 3710–3741.

Cantor, Richard and Christopher Mann, "Analyzing the Trade-Off Between Ratings Accuracy and Stability," Technical Report, Moody's Investors Service 2006.

Chen, Qianwen and Jaewon Choi, "Reaching for yield and the cross section of bond returns," *Management Science*, 2024, *70* (8), 5226–5245.

Chernenko, Sergey, Samuel G Hanson, and Adi Sunderam, "Who neglects risk? Investor experience and the credit boom," *Journal of Financial Economics*, 2016, *122* (2), 248–269.

Choi, Jaewon and Mathias Kronlund, "Reaching for yield in corporate bond mutual funds," *The Review of Financial Studies*, 2018, *31* (5), 1930–1965.

Cole, Shawn, Anna Paulson, and Gauri Kartini Shastry, "Smart money? The effect of education on financial outcomes," *The Review of Financial Studies*, 2014, *27* (7), 2022–2051.

Couts, Spencer J, Andrei S Gonçalves, and Johnathan Loudis, "The subjective risk and return expectations of institutional investors," *Fisher College of Business Working Paper*, 2023, *14*.

Dick-Nielsen, Jens, Peter Feldhütter, Lasse Heje Pedersen, and Christian Stolborg, "Corporate Bond Factors: Replication Failures and a New Framework," Available at SSRN December 2023. Working paper, latest version December 9, 2023.

Dickerson, Alexander, Christian Julliard, and Philippe Mueller, "The corporate bond factor zoo," *UNSW Business School Research Paper Forthcoming*, 2023.

_ , Philippe Mueller, and Cesare Robotti, "Priced risk in corporate bonds," *Journal of Financial Economics*, 2023, *150* (2), 103707.

Efing, Matthias and Harald Hau, "Structured Debt Ratings: Evidence on Conflicts of Interest," *Journal of Financial Economics*, 2015, *116* (1), 46–60.

Eisenbach, Thomas M., Anna Kovner, and Michael Junho Lee, "Cyber Risk and the U.S. Financial System: Mapping the Threat," Staff Report 826, Federal Reserve Bank of New York December 2017.

Elton, Edwin J., Martin J. Gruber, Deepak Agrawal, and Christopher Mann, "Explaining the Rate Spread on Corporate Bonds," *The Journal of Finance*, 2001, *56* (1), 247–277.

Fama, Eugene F and Kenneth R French, "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 1993, *33* (1), 3–56.

Fons, Jerome S., "Credit Rating History: The Development of Investment-Grade Standards," 2004.

Franco, Gus De, Florin P Vasvari, and Regina Wittenberg-Moerman, "The informational role of bond analysts," *Journal of Accounting Research*, 2009, *47* (5), 1201–1248.

Gebhardt, William R, Soeren Hvidkjaer, and Bhaskaran Swaminathan, "The cross-section of expected corporate bond returns: Betas or characteristics?," *Journal of financial economics*, 2005, *75* (1), 85–114.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Neglected risks, financial innovation, and financial fragility," *Journal of financial economics*, 2012, *104* (3), 452–468.

_ , _ , and _ , "Neglected risks: The psychology of financial crises," *American Economic Review*, 2015, *105* (5), 310–314.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, "Five facts about beliefs and portfolios," *American Economic Review*, 2021, *111* (5), 1481–1522.

Gilchrist, Simon and Egon Zakrajšek, "Credit spreads and business cycle fluctuations," *American economic review*, 2012, *102* (4), 1692–1720.

Gomes, João F., Marco Grotteria, and Jessica A. Wachter, "Cyclical Dispersion in Expected Defaults," *Review of Financial Studies*, 2019, *32* (4), 1275–1308.

Greenwood, Robin and Samuel G Hanson, "Issuer quality and corporate bond returns," *The Review of Financial Studies*, 2013, *26* (6), 1483–1525.

Griffin, John M., Jordan Nickerson, and Dragon Yongjun Tang, "Rating Shopping or Catering? An Examination of the Response to Competitive Pressure for CDO Credit Ratings," *The Review of Financial Studies*, 2013, *26* (9), 2270–2310.

Gulen, Huseyin, Mihai Ion, Candace E Jens, and Stefano Rossi, "Credit cycles, expectations, and corporate investment," *The Review of Financial Studies*, 2024, *37* (11), 3335–3385.

Harold, Gilbert, "Bond Ratings as an Investment Guide: An Appraisal of Their Effectiveness," *Ronald Press*, 1938.

He, Jie, Jun Qian, and Philip E. Strahan, "Are All Ratings Created Equal? The Impact of Issuer Size on the Pricing of Mortgage-Backed Securities," *The Journal of Finance*, 2012, *67* (6), 2097–2137.

ICE, "ICE Bond Index Methodologies," https://www.ice.com/publicdocs/data/Bond_Index_Methodologies.pdf 2024.

Jegadeesh, Narasimhan and Sheridan Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of finance*, 1993, *48* (1), 65–91.

Jo, Chanik, Chen Lin, and Yang You, "Subjective risk-return trade-off," *Available at SSRN 4096443*, 2024.

Kahneman, Daniel and Amos Tversky, "Subjective probability: A judgment of representativeness," *Cognitive psychology*, 1972, *3* (3), 430–454.

Kronlund, Mathias, "Do Bond Issuers Shop for Favorable Credit Ratings?," *Management Science*, 2020, *66* (12), 5944–5968.

Laarits, Toomas, Ben Matthies, Kaushik Vasudevan, and Will Yang, "Policy by Committee," July 2025. SSRN Working Paper.

Lian, Chen, Yueran Ma, and Carmen Wang, "Low interest rates and risk-taking: Evidence from individual investment decisions," *The Review of Financial Studies*, 2019, *32* (6), 2107–2148.

Lin, Luyang, Lingzhi Wang, Jinsong Guo, and Kam-Fai Wong, "Investigating bias in llm-based bias detection: Disparities between llms and human perception," *arXiv preprint arXiv:2403.14896*, 2024.

Lochstoer, Lars A and Tyler Muir, "Volatility expectations and returns," *The Journal of Finance*, 2022, *77* (2), 1055–1096.

Loughran, Tim and Bill McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of finance*, 2011, *66* (1), 35–65.

Mathis, Jérôme, James McAndrews, and Jean-Charles Rochet, "Rating the Raters: Are Reputation Concerns Powerful Enough to Discipline Rating Agencies?," *Journal of Monetary Economics*, 2009, *56* (5), 657–674.

Mehran, Hamid and René M Stulz, "The economics of conflicts of interest in financial institutions," *Journal of Financial Economics*, 2007, *85* (2), 267–296.

Merton, Robert C., "On the pricing of corporate debt: The risk structure of interest rates," *The Journal of Finance*, 1974, *29* (2), 449–470.

Meta AI, "Building the Future of AI With Llama," 2024.

_ , "Meta Llama 3: The Next Generation of Our Open-Source Large Language Model," 2024.

Michaely, Roni and Kent L Womack, "Conflict of interest and the credibility of underwriter analyst recommendations," *The review of financial studies*, 1999, *12* (4), 653–686.

Minsky, Hyman P, "The financial instability hypothesis: An interpretation of Keynes and an alternative to "standard" theory," *Challenge*, 1977, *20* (1), 20–27.

Moody's Analytics, "EDF-Based Bond Valuation Model," https://www.moodys.com/web/en/us/insights/credit-risk/edf_based_bond_valuation_model.html 2024.

Mullainathan, Sendhil, "Thinking through categories," Technical Report, Working Paper, Harvard University 2002.

Nagel, Stefan and Zhengyang Xu, "Dynamics of subjective risk premia," *Journal of*

Financial Economics, 2023, *150* (2), 103713.

National Association of Insurance Commissioners, "Risk-Based Capital (RBC) Requirements on Fixed Income Assets to Change," https://content.naic.org/sites/default/files/inline-files/topic_risk_based_capital_rbc_requirements_on_fixed_income_assets_to_change.pdf 2021.

Opp, Christian C., Marcus M. Opp, and Milton Harris, "Rating Agencies in the Face of Regulation," *Journal of Financial Economics*, 2013, *108* (1), 46–61.

Pástor, L'uboš and Robert F Stambaugh, "Liquidity risk and expected stock returns," *Journal of Political economy*, 2003, *111* (3), 642–685.

Peng, Lin, Siew Hong Teoh, Yakun Wang, and Jiawen Yan, "Face value: Trait impressions, performance characteristics, and market outcomes for financial analysts," *Journal of Accounting Research*, 2022, *60* (2), 653–705.

Pflueger, Carolin, Emil Siriwardane, and Adi Sunderam, "Financial market risk perceptions and the macroeconomy," *The Quarterly Journal of Economics*, 2020, *135* (3), 1443–1491.

Piazzesi, Monika, Martin Schneider et al., *Trend and cycle in bond premia*, Vol. 424, Federal Reserve Bank of Minneapolis, Research Department, 2009.

Sørensen, Jakob Ahm, *Financial Markets with Frictions and Belief Distortions*, Copenhagen Business School [Phd], 2022.

S&P Global Ratings, "General Criteria: Methodology - Credit Stability Criteria," https://www.spglobal.com/ratings/en/research/articles/100503-general-criteria-methodology-credit-stability-criteria-5961504 2010.

_ , "Understanding Credit Ratings," https://disclosure.spglobal.com/ratings/en/regulatory/article/-/view/sourceId/504352 2023.

U.S. Securities and Exchange Commission, "Fact Sheet: Global Analyst Research Settlements," https://www.sec.gov/news/speech/factsheet.htm April 2003.

_ , "Fact Sheet: Rule Amendments to Remove References to Credit Ratings from Rule 15c3-1," https://www.sec.gov/files/34-102022-fact-sheet.pdf 2022.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou et al., "Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, 2022, *35*, 24824–24837.

wei Lin, Hsiou and Maureen F McNichols, "Underwriting relationships, analysts' earnings forecasts and investment recommendations," *Journal of accounting and economics*, 1998, *25* (1), 101–127.

Ye, Jiayi, Yanbo Wang, Yue Huang, Dongping Chen, Qihui Zhang, Nuno Moniz, Tian Gao, Werner Geyer, Chao Huang, Pin-Yu Chen et al., "Justice or prejudice? quantifying biases in llm-as-a-judge," *arXiv preprint arXiv:2410.02736*, 2024.

Zhang, Harold H, Feng Zhao, and Xiaofei Zhao, "Neglected Risks in the Communication of Residential Mortgage-Backed Securities Offerings," *The Journal of Finance*, 2024, *79* (1), 129–172.

Internet Appendix

Title

Yunrong Zhou

Aug 2025

IA.A LLM Prompts and Examples

Prompt 1: Percived risk score

USER PROMPT:

Forget all previous instructions. You are a financial economics expert reading corporate bond research notes.

TASK: Extract only explicitly stated information from the provided context. Do NOT infer or add information.

CONTEXT:
{comment}

Respond strictly with a valid JSON object matching the schema below:

Questions:

Q1: List **ALL** explicitly stated financial or economic concepts related to risk and performance in the company's securities, operations, business, or financials.

- Don't miss any relevant information from the context.
- Provide a comma-separated list of noun phrases (each 3 words) or nouns.
- Do not infer, generate, or add information beyond what is explicitly stated.
- If none exist, answer "N/A".

Q2: Provide the subset of Q1 concepts that signal **HIGHER risk or WORSE performance**.

- Answer "N/A" if none exist.

Q3: Provide the subset of Q1 concepts that signal **LOWER risk or BETTER performance**.

- Answer "N/A" if none exist.

Q4: Provide the subset of Q1 concepts that appear **RISK-NEUTRAL**.

- Answer "N/A" if none exist.

Note for Q2-Q4:

- Q2, Q3, and Q4 must be **mutually exclusive**; no concept may appear in more than one list.

Q5: Based on the provided context, evaluate the company's overall risk profile using the scale below. Provide only the corresponding number:

+4: Extreme Risk

+3: High Risk

+2: Moderate Risk

+1: Mild Risk

0: Neutral

−1: Mild Safety

−2: Moderate Safety

−3: High Safety

−4: Extreme Safety

Required Output Format:

```
{"Q0": ANSWER, "Q1": ANSWER, "Q2": ANSWER, "Q3": ANSWER, "Q4": ANSWER,
"Q5": ANSWER}
```

Prompt 2: Yield mention and tone

Forget all previous instructions. You are a financial economics expert reading corporate bond research notes.

Return exactly the following JSON format:
`{"Q1": 0 or 1, "Q2": -1, 0, 1, or "N/A"}`

Q1: Does the paragraph explicitly mention yield or spread?

- Answer 1 if yes, 0 if no.
- Include synonyms or related terms: : *yield, yields, spread, spreads, basis points, bp, carry, valuation, rich, cheap, etc.*

Q2: If Q1 = 1, classify the yield sentiment

- **1** if yield is described positively (e.g. attractive, wide, compelling, etc).
- **0** if tone is neutral or fair.
- **−1** if yield is described negatively (e.g. tight, compressed, unattractive, etc).
- If Q1 = 0, output `"N/A"`.

Context
<<<
{comment}
>>>

Figure A2. Prompt 2: Identify yield mention and tone

Prompt 3: Decompose comments into factual and subjective content and identify financial concepts

Instruction: Forget all previous instructions. You are a financial economics expert analyzing corporate bond research reports. You will receive an analyst's commentary on corporate bonds. Your task is to extract either factual information or the analyst's subjective views from the provided context.

Context:
{comment}

Answer the following two questions based solely on the provided context. Respond strictly in JSON format:

Q1: Factual information extraction

- Extract only explicitly stated factual information related to the company's securities (bonds, stocks, characteristics), operations, business, or financials.
- Include only publicly known or already realized information.
- Do not include opinions, forecasts, predictions, or subjective statements.
- Do not infer, assume, omit, or add information beyond explicit statements.
- Provide your response as a concise natural-language paragraph. If no factual information exists, respond "N/A".

Q2: Subjective prediction summarization

- Summarize only subjective statements, forecasts, predictions, or opinions explicitly made by the analyst regarding the company's securities, operations, business, or financials.
- Exclude any factual information already covered in Q1.
- Do not infer, assume, omit, or add information beyond explicit statements.
- Begin your response explicitly with: "The analyst believes...".
- Provide your response as a concise natural-language paragraph. If no subjective statements exist, respond "N/A".

Q3: From Q1 of factual information, list **ALL** explicitly stated financial or economic concepts related to risk and performance in the company's securities, operations, business, or financials.

- Don't miss any relevant information from the context.
- Provide a comma-separated list of noun phrases (each 3 words) or nouns.
- Do not infer, generate, or add information beyond what is explicitly stated.
- If none exist, answer "N/A".

Q4: From Q3 of subjective information, list **ALL** explicitly stated financial or economic concepts related to risk and performance in the company's securities, operations, business, or financials.

- Don't miss any relevant information from the context.
- Provide a comma-separated list of noun phrases (each 3 words) or nouns.
- Do not infer, generate, or add information beyond what is explicitly stated.
- If none exist, answer "N/A".

Required Output Format:


```
{
    "Q1": ANSWER,
    "Q2": ANSWER,
    "Q1": ANSWER,
```

General Electric (GE) on 08/14/2018:

Raw comment:
"We recommend a Neutral on GE. GE bond spreads trade much wider than other single A industrials which we view as appropriate given the significant changes underway at the company and related uncertainty and execution risk, weak outlook for important end markets – power, transportation and energy – and elevated leverage. The company outlined its strategic plan in June which includes the spin-off of its Healthcare business and exit from its Baker Hughes JV over the next 2 to 3 years. The company targets $< 2.5x$ adjusted net debt/EBITDA and long-term ratings at single A. Risks to our recommendation include change in underlying business trends and credit metrics, and if proceeds from asset disposals are mostly returned to shareholders versus a balanced approach."

Factual description summarized by LLM:
"The company's GE bond spreads trade wider than other single A industrials. The company has outlined its strategic plan in June, which includes the spin-off of its Healthcare business and exit from its Baker Hughes JV over the next 2 to 3 years. The company targets $< 2.5x$ adjusted net debt/EBITDA and long-term ratings at single A. The company operates in end markets such as power, transportation, and energy. The company has elevated leverage."

Subjective prediction summarized by LLM:
"The analyst believes that a Neutral recommendation is appropriate for GE. The analyst views the company's current situation as having significant changes underway, related uncertainty, and execution risk. The analyst also views the weak outlook for the company's end markets as a concern. The analyst considers the risks to the recommendation to include changes in underlying business trends and credit metrics, and if proceeds from asset disposals are mostly returned to shareholders versus a balanced approach."

Financial economic concepts extracted by LLM: factual content:
significant changes, underlying business trends, credit metrics, asset disposals, execution risk, end markets, weak outlook, changes in business, balanced approach

Financial economic concepts extracted by LLM: subjective content:
significant changes, execution risk, weak outlook, changes in business

Figure A4. Example of comment summarization from prompt 3

Table A1. Illustrative Analyst Comments and Risk Scores

The table presents examples of comments, corresponding recommendations, and the perceived-risk score assigned by the LLM. The LLM identifies and scores the paragraph's overall stance toward credit risk on a scale from –4 (extreme safety) to +4 (extreme risk). Intermediate scores represent progressively weaker risk or safety assessments. Ellipses indicate omitted text without reordering or altering content.

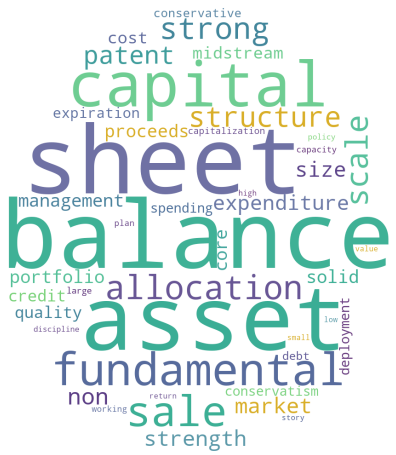
Issuer	Date	Comments	Rec.	Score
DuPont	2015-01	Uncertainty around the spin, revenue headwinds, lukewarm operating profitability, and renewed pressure . . . We believe investors should not rule out . . . updated rating target (BB) . . . downside in owning protection at these levels is largely limited to the roll(s).	Underweight	3
GE	2018-08	. . . bond spreads trade much wider than other A industrials, which we view as appropriate given . . . the related uncertainty and execution risk. The company has . . . a weak outlook . . . and elevated leverage . . . The company targets . . . debt/EBITDA . . . long-term ratings at single A. Risks include change in underlying business trends and credit metrics . . .	Neutral	2
GE	2020-12	. . . relatively wide spreads . . . reflecting GE's elevated leverage and business headwinds . . . GE . . . de-risk and de-lever its balance sheet, and this remains a top management priority. Cash generation has been better than expected . . . Risks . . . include change in underlying business trends and credit metrics as well as downward rating actions.	Neutral	0
DTE	2014-11	. . . DTE . . . will likely remain supportive ..DTE has improved its balance sheet and management has committed to maintaining solid BBB ratings with strong credit metric targets. The company . . . alleviate credit investors concern and . . . adequately managed its . . . business, which has historically been an overhang for some investors.	Overweight	–2
KO	2015-05	. . . Stable, high-quality credit . . . broad global operations with leading brands and market shares. . . .Recent investments . . . add to product diversification . . .	Neutral	–3

IA.B Label Concepts

Figure A5. World cloud of financial concepts

This panel presents the word cloud for each identified financial category. I adopt key categories emphasized by credit rating analysts: Business Profile, Leverage, Liquidity, Profitability, and firm-specific Events. Additionally, I include fundamental balance-sheet items (Asset, Cash, Debt, Equity), and concepts related to firm-level Growth, Credit/Bond Terms, and Risk considerations. Sector-level buckets isolate industry-specific drivers such as Regulation/Policy and Supply-Demand dynamics. Macro-level buckets distinguish top-down influences, including Monetary & Fiscal Policy and broader Country/Global Growth & Shocks. Any residual concepts that do not fit these categories fall under Other (not displayed). For each bucket, we drop stop-words, and the word size is proportional to frequency.

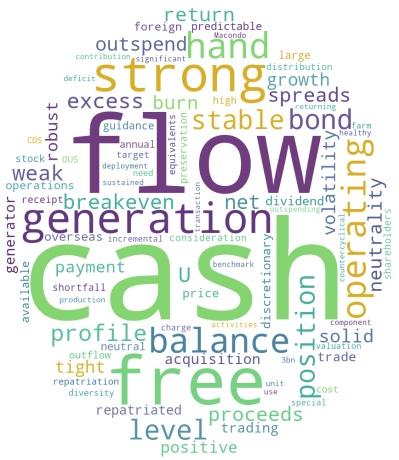
(a) Asset



(b) Business profile



(c) Cash



(d) Bond



(e) Debt



(f) Equity



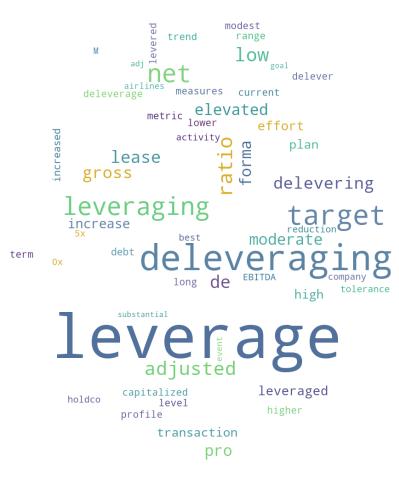
(g) Events

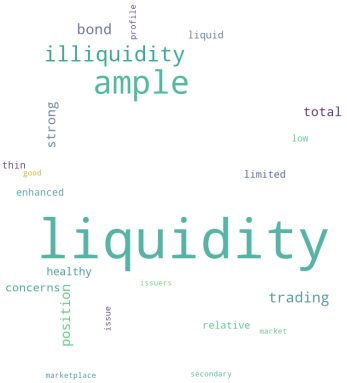


(h) Growth



(i) Leverage





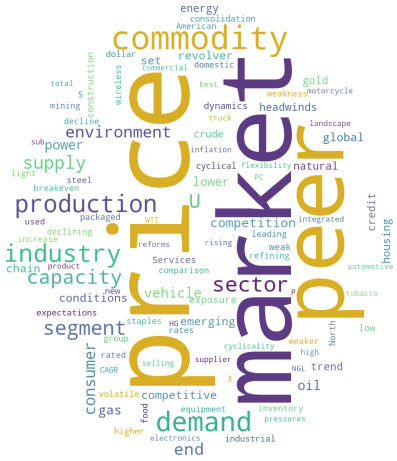
(a) Liquidity



(b) Profitability



(c) Risk



(d) Sector: supply and demand



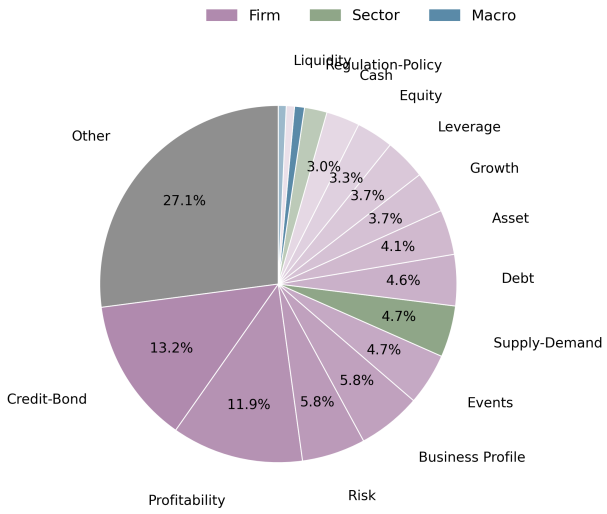
(e) Sector: regulation and policy



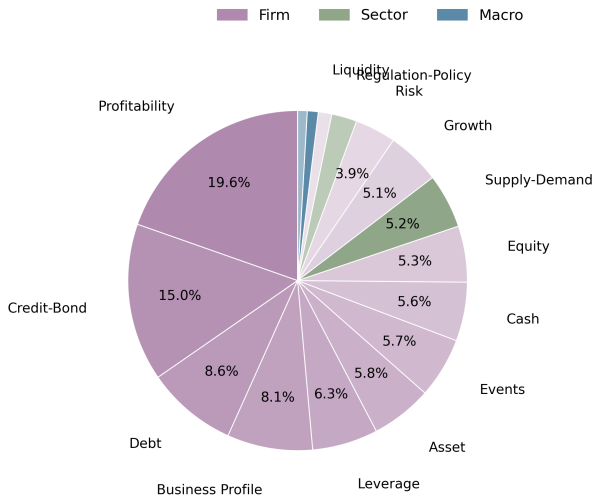
(f) Macro: monetary policy and federal policy



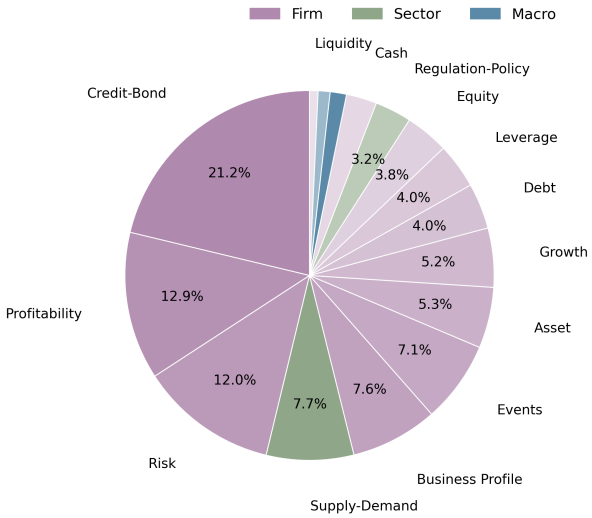
(g) Macro: global Growth & Shocks



(a) All



(b) Factual description



(c) Subjective prediction

IA.C Additional Empirical Analysis

Table A2. Summary Statistics: Firm and bond month level characteristics in the whole sample

The table provides descriptive statistics for issuers and bonds in the full sample. Panel A presents firm-month-level financial characteristics, including the number of bonds issued per firm and firm size. Panel B summarizes bond-month-level characteristics, such as credit rating, duration (in years), time to maturity (in years), credit spread, and monthly bond returns. We construct the full sample from the WRDS Corporate Bond Database for any month with at least one industry report, applying the same filters as the main sample: available credit ratings, credit spreads, and a time to maturity exceeding two years.

Panel A: Firm-month level variables								
	N	Mean	SD	1st	25th	50th	75th	99th
IG report								
Number of Bonds per Firm	28,658	8.09	9.93	1	2	5	10	51
Size (in million)	28,658	37,773.51	89,539.38	518.69	5,979.65	14,357.75	35,687.5	324,827.7
Panel B: Bond-month level variables								
IG report								
Credit Rating	231,723	20.52	1.94	18	19	20	22	28
Credit Rating	231,723			BBB-	BBB	BBB+	A	AA+
Duration (years)	230,384	7.97	4.67	1.91	4.07	6.73	12.06	18.32
Time to Maturity (years)	231,723	12.1	10.26	2.08	4.58	8	19.75	37.08
Credit Spread (%)	231,723	2.73	4.78	-10	0.46	2.62	5.23	13.73
Monthly Bond Return (%)	227,806	0.66	2.98	-6.31	-0.57	0.32	1.65	10.6

Table A3. Recommendation and yield spread and EDF

This table presents *ordered logit regression* specifications with time fixed effects. Columns (1) and (3) quantify how credit signals shape analyst recommendations across matched A- and BBB-rated firms. Columns (2) and (4) test whether the sensitivity of recommendations to each credit signal differs significantly between the two rating buckets. The dependent variable $Rec_{i,t}$ is the analyst recommendation for bond i at time t, coded –1 = Underweight, 0 = Neutral, +1 = Overweight. $YieldSpread_{i,t}$ is the bond's yield spread over the yield of treasuries in the same maturity and expressed in decimal yield units. $EDF_{i,t}$ is the firm's expected default frequency. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	Dep var: recommendation			
	(1)	(2)	(3)	(4)
A × Yield Spread	75.495***	69.222***		
	[4.10]	[3.46]		
BBB × Yield Spread	6.273	0.000		
	[0.86]	[.]		
Yield Spread		6.273		
		[0.86]		
A × EDF			4.316**	3.847*
			[2.04]	[1.93]
BBB × EDF			0.469	0.000
			[0.35]	[.]
EDF				0.469
				[0.35]
A	-2.075***	-2.075***	-3.530***	-3.530***
	[-6.80]	[-6.80]	[-3.23]	[-3.23]
Time FEs	Y	Y	Y	Y
Observations	104,534	104,534	104,534	104,534
Pseudo R^2	0.074	0.074	0.079	0.079

Table A4. Textual risk, objective risk, and future risk materialization

This table reports OLS regression results examining whether analysts' textual expressions of risk and uncertainty help predict future credit events. The dependent variable is an indicator equal to one if the issuer experiences a credit downgrade within three, six, or twelve months (DG), or defaults within three years (DF), and zero otherwise. Panel A use LLM risk and Panel B use uncertain words.
Robust standard errors clustered at the issuer level are reported in brackets. *, **, and *** denote statistical significance at 10%, 5%, and 1%.

	DG(3m)		DG(6m)		DG(12m)		DF	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A × textual risk	0.013***	0.012***	0.018***	0.017***	0.028***	0.027***	0.003	0.004
	[2.97]	[3.03]	[2.99]	[3.11]	[2.91]	[3.29]	[0.96]	[0.99]
A × textual risk	0.014***	0.014***	0.020***	0.019***	0.027**	0.025**	-0.001	0.000
	[3.95]	[3.67]	[3.10]	[2.79]	[2.57]	[2.29]	[-1.24]	[0.96]
A	-0.001	0.003	-0.000	0.011	0.010	0.032	0.007	0.004
	[-0.13]	[0.22]	[-0.00]	[0.53]	[0.38]	[0.92]	[1.13]	[1.07]
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
Observations	89,587	89,587	89,661	89,661	89,677	89,677	89,413	89,405
R Square	0.030	0.036	0.039	0.054	0.045	0.072	0.010	0.045

	DG(3m)		DG(6m)		DG(12m)		DF	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A × no$_u$*ncwords*	0.010**	0.011**	0.013**	0.015**	0.019***	0.020***	0.003	0.004
	[2.24]	[2.42]	[2.27]	[2.40]	[3.05]	[3.22]	[0.99]	[0.99]
A × no$_u$*ncwords*	0.003	0.005**	0.006	0.008*	0.006	0.009	0.000	0.001
	[1.45]	[2.09]	[1.31]	[1.87]	[1.03]	[1.60]	[0.75]	[1.03]
A	-0.015	-0.014	-0.014	-0.004	-0.019	0.003	-0.001	-0.003
	[-1.51]	[-1.16]	[-0.64]	[-0.14]	[-0.63]	[0.08]	[-0.29]	[-0.72]
Time FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
Observations	99,916	99,916	100,001	100,001	100,022	100,022	99,764	99,755
R Square	0.029	0.036	0.040	0.054	0.042	0.068	0.014	0.049