How Employee Voice Informs Firm Valuation: Evidence from Glassdoor

Preliminary Version. Please do not circulate.

Dennis Bams[1,2], Peiran Jiao[1], and Yanxiu Piao[1*]

[1] Maastricht University, The Netherlands

[2] Open University, The Netherlands

Abstract

Studies have identified employee satisfaction as an important predictor of firm outcomes. However, the information we can extract from employees goes beyond simple satisfaction. In this paper, we construct multidimensional measures of employee-generated content on Glassdoor, ranging from the platform's numeric ratings to the tone (sentiment) and content (risk-related buzz) in written reviews. We embed our measures in a Credit Default Swap (CDS) valuation framework. We find that even after orthogonalization relative to firm fundamentals and the S pillar score in ESG, our Glassdoor measures boost the weekly cross-sectional R^2 of our CDS model by 2.44 percentage points. Both sentiment and risk-related buzz channels add significant contributions, with the following distinction. The influence of sentiment exhibits a cyclical pattern, whereas that of risk buzz is steadier but becomes more important during high credit risk periods. Our additional mechanism checks and robustness tests provide converging evidence. Overall, our findings reveal the novel and robust value of employee-generated content for credit risk assessment.

JEL Classification: D83, G11, M12

Keywords: Employee-generated Information,Online Reviews, Textual Analysis, Credit Risk

*Corresponding author: `yanxiu.piao@maastrichtuniversity.nl`

1 Introduction

Extant literature in finance acknowledges that employees matter for firm value. However, much of the evidence relies on broad aggregates to indirectly proxy employees' influence through, for instance, payroll expenses, satisfaction indices, or the social aspect in ESG scores, which could mask more granular and dynamic aspects of information from employees. Indeed, employees nowadays often possess key information related to firm risks.[1] In this study, we demonstrate that Glassdoor, as an independent, bottom-up information source, contains non-redundant signals overlooked in conventional disclosures and standard financial fundamentals. To do so, we develop multidimensional measures based on Glassdoor, including numeric ratings and our novel measures of written reviews, and embed them in a credit-risk assessment model of Credit Default Swap (CDS) spreads. Within this framework, we isolate two channels for the effects of written reviews: the tone or the behavioral channel, reflecting employees' sentiment and how they feel about their workplace, and the content or the informational channel, capturing substantive risk-related buzz words in employees' narratives. Both channels exert significant influences and can independently explain cross-sectional differences in CDS spreads.

In particular, we develop multidimensional measures of employee voice using data from Glassdoor, an online platform where employees anonymously evaluate their firms through both structured numeric ratings and unstructured written reviews (see Figure 1 as an example). The numeric ratings include an overall rating, and structured assessments on career opportunities, forward-looking items like business outlook, and so on. However, the written reviews (pros, cons, advice to management) offer richer, unconstrained narratives. We leverage this textual data to distill two distinct facets of employee voice, which form the core of our analysis: the sentiment channel based on the tone of the reviews and the risk-related buzz channel based on the content of the reviews. This distinction aligns with two established pathways through which employee

[1]For example, in 2024 the U.S. Securities and Exchange Commission reported about 24,980 whistleblower tips. Although submissions came from both insiders and outsiders, insiders accounted for roughly 62% of award recipients that year; see U.S. Securities and Exchange Commission (2024).

perceptions can influence firm value. The sentiment channel corresponds to a behavioral channel, and it captures the affective tone, such as optimism, pessimism or overall morale expressed in employee narratives. This can influence firm outcomes by affecting productivity and coordination (Barsade & Gibson, 2007; Edmans, 2011; Green et al., 2019; Shan & Tang, 2023). We measure it using a BERT-based sentiment classifier, which has been widely used in literature (Devlin et al., 2019; Gentzkow et al., 2019; A. H. Huang et al., 2023). In contrast, the risk-related buzz channel corresponds to an informational channel, and it captures the informational content, where employees directly report observations of internal frictions, operational issues, or other risk factors that may foreshadow financial distress. This is particularly relevant in a credit context. We estimate this channel using multinomial inverse regression (MNIR) to isolate phrases predictive of future firm risks (Campbell & Shang, 2022; Taddy, 2013, 2015). By integrating numeric ratings with the two written-review-based measures, our framework offers a concise yet comprehensive way to separate the sentimental and informational dimensions of employee voice.

We study the contribution of these Glassdoor measures to credit risk assessment in the CDS market over and beyond traditional financial fundamentals and ESG scores. To achieve this, building on the methodology of Bai and Wu (2016), we model CDS spreads in a stepwise framework that starts with the Merton-model factors, then incorporates financial fundamentals and ESG scores (especially the most relevant S score), and ultimately introduces our Glassdoor-based measures. At each layer, the newly introduced variables are orthogonalized with respect to all variables from the preceding layers, ensuring that each layer contributes distinct rather than redundant information to the valuation. We introduce the structured and unstructured Glassdoor measures into two layers: first the structured numeric ratings, and then the unstructured written reviews with the sentiment and risk-related buzz channels. This complete layered design allows us to test how structured ratings, employee sentiment, and risk-related buzz are distinctly incorporated into market-based valuations of CDS spreads.

Using the above method, we document a significant linkage between Glassdoor-based

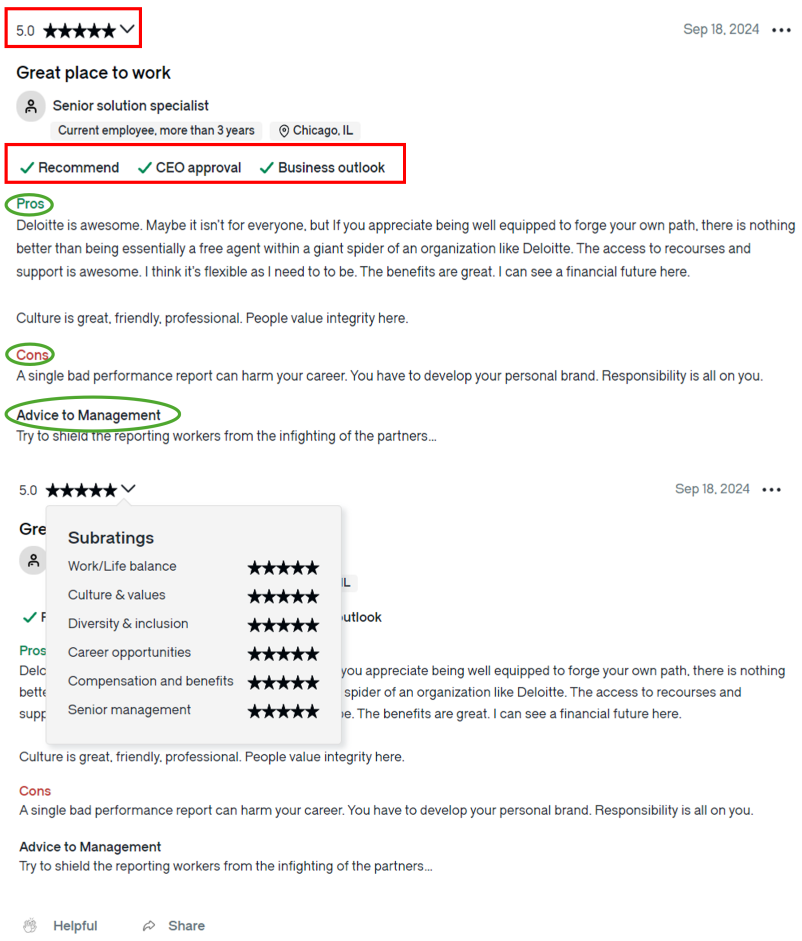


Figure 1: **An example of an employee review on Glassdoor**

Notes: This screenshot illustrates a typical layout of an employee review on Glassdoor, highlighting key information such as overall rating and review comments.

measures and CDS spreads in a weekly cross-sectional analysis of U.S. firms from 2012 to 2023. The estimation is conducted at the firm-week level with a sample of 417 publicly listed U.S. firms.[2] In our stepwise cross-sectional regressions, the inclusion of structured ratings raises the model's explanatory power by 1.4 percentage points in terms of cross-sectional R-squared, relative to a baseline that includes standard financial fundamentals and the Social ('S') pillar of the ESG score, which is the most closely tied to workforce outcomes in ESG.

We next respectively examine the sentiment and risk-related buzz channels in greater

[2]We selected the sample firms with complete fundamentals, ESG coverage, and Glassdoor information over 2012–2023. See Section 2 for the sample construction details.

detail. Conditional on financial fundamentals, the ESG Social (S) pillar, and the numeric ratings, the two written-review-based measures raise the weekly cross-sectional R^2 by 1 percentage point. On average, a 1-standard-deviation (SD) change in Sentiment is associated with a 0.002-SD change in ln(CDS), while a 1-SD change in Risk-Related Buzz is associated with a 0.014-SD change in ln(CDS). We also find that the explanatory power of the two channels varies over time: the sentiment channel is more cyclical, while the risk-related buzz channel is more persistent. The importance of each channel varies over time, implying that markets price different facets of employee voice depending on different conditions. Accordingly, when modeling the influence of these channels, one should allow their weights to be time-varying rather than constant.

We further seek identification and validation for the two channels. For the sentiment channel, we address the endogeneity concern using an IV method based on plausibly exogenous affective shocks from blockbuster releases and aviation fatalities, which should shift employees' mood without altering fundamentals or aggregate conditions. The IV estimates reveal a direct link from employee sentiment to contemporaneous CDS spreads. Moreover, we conduct heterogeneity analyses as an additional validation of both channels. We first split the sample firms by labor intensity, and we find that the sentiment effect is stronger in capital-intensive firms while the risk-related buzz effect is similar between high and low labor-intensity firms. A 1-SD increase in Sentiment narrows CDS spreads deviation by about 15 bps in capital-intensive firms, and by about 11 bps in labor-intensive firms, a statistically significant difference of about 31%. By contrast, a 1-SD increase in Risk-Related Buzz narrows CDS spreads deviation by a similar amount in both groups, about 15 bps increase in both groups with no statistical difference. This is consistent with the buzz channel capturing concrete, event-type risk information that is less sensitive to labor intensity. We next split by baseline credit risk, and we find larger effects in the high-risk group in both channels. For the sentiment channel, a 1-SD increase in Sentiment associates with a 16 bps narrowing of CDS spreads deviation in low-risk firms, and about 18 bps narrowing in high-risk firms, roughly 12% larger. For the risk-related buzz channel, a 1-SD increase in Risk-Related Buzz associates with a 16 bps narrowing of CDS spreads deviation in the low-risk group and 21 bps narrowing in the high-risk group, about one-third larger. The differences are all significant at the 5% level.

Our study contributes to the literature in several ways. First and foremost, we innovate a measure of multidimensional employee-generated information based on Glassdoor, and demonstrate its value in a model of credit risk assessment based on CDS spreads. Glassdoor information is produced by employees, who are typically considered as human capital of a firm. Much of the human-capital literature operationalizes the 'employee' through input-based proxies. These measures, such as payroll aggregates, education stocks, and accounting disclosures, speak to what firms invest in employees rather than how employees experience the firm from within. This approach is exemplified by the widespread use of surrogate measures, cost-based human-capital accounts (Abraham & Mallatt, 2022; Folloni & Vittadini, 2010; Friedman & Lev, 1974), and disclosure-based indicators (Lajili & Zeghal, 2006; Widener, 2006).

However, a significant strand of research has recently shifted toward directly measuring employee perceptions and satisfaction, recognizing that workforce experience captures time-varying organizational effectiveness beyond static labor inputs. Firms with better management and HR practices exhibit higher productivity and total factor productivity even holding capital and headcount constant, indicating that 'inside-the-firm' conditions shift the production residual rather than the input bundle itself (Bloom & Van Reenen, 2007; Ichinowski et al., 1997). In response, contemporary research and practice now leverage a suite of tools, including structured surveys, HR metrics, and technology-enabled analytics, to capture employee perceptions and experience.[3] Within this landscape, ESG frameworks have emerged as a codified, policy-centric benchmark, and its Social (S) pillar is supposed to capture firm-disclosed indicators related to human-

[3]The literature reflects this multi-faceted approach. In the domain of structured surveys and certifications, Edmans (2011) use an employer ranking based on employee surveys to proxy satisfaction, while Guiso et al. (2015) utilize survey data on perceived managerial integrity. Research employing core HR metrics includes Aldatmaz et al. (2018), who analyze employee turnover data, and Boone et al. (2024), who study employee reactions to CEO pay disparity disclosures. Furthermore, the use of technology-enabled analytics is exemplified by Li et al. (2021), who apply natural language processing to managerial speech to quantify cultural dimensions.

capital management and stakeholder outcomes.[4]

While informative about policy intent and compliance, the S pillar in ESG shares many of the limitations inherent in mainstream employee measurement approaches. For instance, they are infrequently updated (typically quarterly to annually), predominantly top-down, and reflect what firms choose to disclose rather than how employees actually experience (Beer & Eisenstat, 2004; Coff, 1997). Their highly aggregated, single-score nature compresses heterogeneous and evolving workplace dynamics, obscuring granular signals and early warnings that may be critical for assessing operational risks (Abraham & Mallatt, 2022). Moreover, these firm-reported metrics are susceptible to presentation bias and may rely on non-validated internal surveys that produce inflated measures not predictive of actual performance (Ed & Andrew, 2022).[5] Most critically, the temporal misalignment between these periodic updates and the speed at which workplace conditions and markets move leaves a substantial gap in capturing timely, authentic employee-side information.

Our approach addresses these limitations by leveraging the high-frequency, granular, and employee-generated nature of Glassdoor reviews. Unlike prior studies that analyze numeric ratings or textual information in isolation (Campbell & Shang, 2022; Green et al., 2019; K. Huang et al., 2020), we integrate these elements within a unified credit-risk assessment framework that delivers four specific advances. First, we overcome the temporal misalignment of traditional metrics by utilizing a data source that is updated continuously, capturing the rapid evolution of workplace sentiment and risk. Second, we preserve the rich, multidimensional nature of employee voice by analyzing structured ratings and unstructured narratives side by side rather than compressing them into a single aggregated score, thereby revealing granular signals and early warnings that

[4]The Social pillar covers labor management and human-capital practices, supply-chain labor standards and human rights, community relations, and product/customer responsibility. Scores are constructed via rule- and weight-based frameworks and are refreshed periodically. Provider-specific definitions, sources, and our frequency-alignment procedure are detailed in Section 2.

[5]According to Ed and Andrew (2022), a key methodological critique is that many firms utilize non-validated employee engagement surveys which produce inflated metrics ('junk' metrics) that are not predictive of performance. To address this, they recommend employing psychometrically validated survey items and transparently reporting the full distribution of survey responses to ensure metric integrity, particularly for ESG disclosures.

aggregate measures obscure. Third, we alleviate top-down and presentation biases by grounding our analysis in the direct, unfiltered discourse of employees, moving beyond firm-curated disclosures and potentially inflated internal surveys. Fourth, and most fundamentally, we provide a cohesive modeling framework that places the various facets of employee-generated information within a layered valuation structure, allowing us to quantify their incremental contributions to credit risk assessment relative to conventional fundamentals and ESG benchmarks, thereby offering a comprehensive evaluation of how different dimensions of employee voice inform firm valuation.

Another advantage of our method is that we clearly distinguish two important aspects of employee voice, namely the sentiment channel and the risk-related buzz channel. Employee sentiment is economically important because workplace morale directly impacts productivity, coordination, discretionary effort, retention, and innovation. It affects expected cash flows and risk, thereby influencing firm valuations (Edmans, 2011; Green et al., 2019; Guiso et al., 2015; K. Huang et al., 2020). This link is further reinforced by the broader asset-pricing literature, which shows that public sentiment and mood can move market prices (Edmans et al., 2007; Hirshleifer & Shumway, 2003; Jiao et al., 2020). Distinct from sentiment, employee voice can potentially convey valuable, forward-looking information that reflect insiders' view about operational risks in the firm, before they materialize in observed firm characteristics. The critical role of employees in surfacing corporate malfeasance is demonstrated by real-world evidence. For instance, Dyck et al. (2010) show, in a study of major U.S. fraud cases, that employees are among the most effective detectors, with monetary incentives playing a key role in motivating whistleblowing. This aligns with the premise that front-line employees are often the first to observe internal frictions, safety issues, customer attrition, or supply-chain disruptions, long before such information surfaces in formal reports (Coff, 1997). Moving from specific fraud events to systematic risk prediction, several research confirms that narratives from employee reviews can be quantified to anticipate future misconduct and distress events (Campbell & Shang, 2022; Graham et al., 2023). Despite the distinct natures of employee voice, extant studies often conflate these affective and informational signals within aggregate measures.

Our contribution in this regard is to explicitly disentangle and evaluate the role of each channel within a unified framework. This approach offers several key advantages for identification and interpretation. First, it prevents conflating the two distinct mechanisms. This can avoid attribution errors, when a significant overall effect might be mistakenly assigned to a single channel, as well as double-counting of information when these measures are used alongside other controls like ESG scores. Second, our unified specification allows for a clean test of each channel's marginal informativeness, enabling us to quantify their respective explanatory power for credit spreads and determine which factor dominates under different conditions. Finally, and most critically for credit risk assessment, separating sentiment from risk-related buzz provides a clearer theoretical link to firm value. This becomes important when ostensibly positive employee sentiment does not always translate into narrower credit spreads, while negative risk signals show a more direct and potent association with credit risk premiums.

In terms of the outcome variable, we extend the literature from its dominant focus on equity markets to credit markets, showing that employee-generated information also plays a significant role in debt valuation. Much of the existing evidence has been developed in equity markets, where proxies such as employee satisfaction and workplace sentiment have been shown to influence stock returns and volatility, attract institutional participation, and exhibit better post-SEO operating performance (Chemmanur et al., 2021; Edmans, 2011; Green et al., 2019). Equity prices provide a broad perspective on firm value, but equity markets are usually subject to the influence of complicated factors, such as noise trading, heterogeneous beliefs, transitory sentiments, and macroeconomic shocks.[6] Additionally, the equity investor base is tilted toward retail investors

[6]This view is supported by a substantial body of literature. The seminal work of Shleifer and Vishny (1997) explains how noise trader risk and limits to arbitrage can cause prices to deviate from fundamentals. Theoretical and empirical studies further show how investor sentiment can systematically affect cross-sectional returns (Baker & Wurgler, 2006; Barberis et al., 1998). Moreover, stock prices can reflect a biased view of value under short-sale constraints and heterogeneous beliefs (Diether et al., 2002; Miller, 1977). Finally, equity valuations are significantly influenced by macroeconomic shocks, such as unanticipated monetary policy changes (Bernanke & Kuttner, 2005; Fisher et al., 2022).

and mutual funds.[7] This composition can amplify sentiment-driven trading and short-term performance pressures, potentially leading to pricing dynamics where employee voice is temporarily obscured or mispriced.

Complementing the equity-based evidence, the credit default swap (CDS) market offers a distinct and valuable environment in which to study the value relevance of employee reviews. At its core, the credit market's defining feature is its direct link to default risk. CDS spreads are explicit pricing mechanisms for default probability and loss given default, offering a clean channel through which employee-generated information affecting firm solvency can be impounded into value. This intrinsic property leads to a second key advantage: a pronounced sensitivity to downside and tail risks. CDS spreads load heavily on deteriorating fundamentals and vary more monotonically with negative shocks, making them exceptionally well-suited to capture the early-warning signals of operational and compliance risks embedded in employee narratives (Acharya & Johnson, 2007; Hand et al., 1992; Jorion & Zhang, 2007; Longstaff et al., 2005). Beyond these fundamental attributes, the market's participant structure enhances its informational efficiency. Credit markets are populated primarily by sophisticated institutions, such as hedge funds, banks, and insurers, which have the expertise and incentive to gather and process firm-specific soft information for hedging and risk management (Acharya & Johnson, 2007; Arora et al., 2012).[8] This sophistication supports the CDS market's documented efficacy in price discovery, especially in reacting to and anticipating negative firm-specific news ahead of other markets (Augustin et al., 2020; Blanco et al., 2005; Lee et al., 2018). Furthermore, the differing arbitrage limits in credit markets, characterized by higher trading costs and lower liquidity, provide a solid ground for ex-

[7]Household participation on the equity side is broad and often mediated through registered funds. In 2024, about 54-56% of U.S. households owned shares of U.S.-registered investment companies (mutual funds, ETFs, closed-end funds, or UITs), according to the Investment Company Institute. See Investment Company Institute (2024a) and Investment Company Institute (2024b). In addition, households held a substantial share of U.S. corporate equities by market value in 2024. See Board of Governors of the Federal Reserve System (2024), and Federal Reserve Bank of St. Louis (2024).

[8]U.S. corporate bond ownership and trading are predominantly institutional. Historically, corporate bonds have been held mainly by institutions, especially insurance companies, rather than directly by households (Koijen & Yogo, 2023). The corporate bond market is described as institution-dominated in recent Federal Reserve research (Brunetti et al., 2024)

amining whether subtle signals like employee voice can have a persistent impact (Bao et al., 2011; Schestag et al., 2016). Finally, studying credit markets offers significant external relevance. It directly illuminates the firm financing channel, as credit mispricing immediately impacts corporate borrowing costs and investment decisions, unlike equity mispricing which may remain in secondary markets (Ashcraft & Santos, 2009; Baker & Wurgler, 2002; Chava & Roberts, 2008; Saretto & Tookes, 2013). These features make the credit market particularly well suited to examine whether employee voice is impounded into firm valuations.

Methodologically, we apply the Bayesian shrinkage framework of Bai and Wu (2016) to the setting of employee-generated information. Whereas Bai and Wu (2016) use shrinkage priors to anchor CDS spreads to firm fundamentals, we add Glassdoor-based measures in a layered information structure that builds sequentially on fundamentals and the Social (S) dimension of ESG scores. At each step, newly introduced variables are residualized with respect to all variables from the preceding layers. This sequential orthogonalization ensures that fundamentals, the S score, and our Glassdoor measures each contribute distinct, non-overlapping signals to valuation. By doing so, our methodology directly engages with the modern asset-pricing literature's emphasis on parsimony and avoiding redundant factors (Barillas & Shanken, 2017, 2018).

We address these issues with a Bayesian shrinkage design that disciplines noisy inputs and improves out-of-sample stability relative to unregularized OLS/PCA mixtures (Freyberger et al., 2020; Kozak et al., 2020). Within the Glassdoor layer, we treat sentiment and risk-related content as separate inputs and let shrinkage allocate time-varying weights across the structured ratings and written-review-based measures, thereby limiting overfitting and stabilizing estimation while preserving economically interpretable channels (Freyberger et al., 2020; Harvey et al., 2016; Kozak et al., 2020). This framework provides a more robust foundation for incorporating employee-generated information into asset pricing models and offers new insights into how markets dynamically respond to different dimensions of workplace conditions.

Taken together, our findings have important implications for research, investment practice, and policy. For academic work, our findings demonstrate the feasibility and value of systematically incorporating employee-generated, high-frequency qualitative data into firm valuation frameworks, opening new avenues for studying the role of employee-generated information in credit markets. By investigating the sentiment and risk-buzz channels in a CDS model, this study bridges behavioral and informational perspectives on how employee-generated information is priced in credit markets, providing a scalable measurement approach that can be adopted across asset classes. For practitioners, the results highlight that employee-generated information, properly decomposed into the tone and content components, offers added predictive power beyond conventional financial and ESG disclosures, enabling more responsive and accurate credit risk assessments. For policymakers and standard setters, our evidence suggests that the current 'S' metrics in ESG scores underweight forward-looking, employee-generated information on operational and compliance risk, indicating scope to improve existing measures.

The remainder of this paper is structured as follows: Section 2 describes data and sample construction, Section 3 outlines our methodology, Section 4 presents empirical results, and Section 5 concludes.

2 Data Collection and Sample Construction

We collect data on U.S. publicly traded corporations from multiple sources. Our dataset begins with the universe of firms with available credit default swap (CDS) data from S&P Global (formally known as Markit), which we then match with financial statement information from Compustat, stock option implied volatilities from Ivy DB OptionMetrics, and stock market data from the Center for Research in Security Prices (CRSP). Then, we incorporate ESG scores from MSCI ESG Ratings and employee review data from Glassdoor.

We obtain a sample of 417 publicly traded U.S. firms with all necessary information for our analyses from 2012 to 2023. A firm is included in our final sample if it satisfies the following criteria: (i) it has a valid five-year CDS spread quote, (ii) its financial statements contain book value of debt and total assets, (iii) it has at least one year of daily stock return history to compute realized volatility and market capitalization, and (iv) it has non-missing observations in both the MSCI ESG and Glassdoor databases. Following Bai and Wu (2016), we construct a weekly panel anchored on CDS quotes: the weekly observation uses the Wednesday five-year CDS spread (or the nearest trading day if Wednesday is missing). Meanwhile, due to the structural limitations of the MSCI ESG database, the company coverage expanded significantly in 2012. Additionally, the OptionMetrics IvyDB database, which undergoes periodic updates, only provides option data up to August 31, 2023. Therefore, to ensure consistency in data availability and to meet the requirements for our empirical analyses, we define our sample period from August 2012 to August 2023.

2.1 CDS and Firm Fundamentals

Credit default swaps (CDS) are over-the-counter contracts that provide protection against credit events of a reference entity. The buyer of protection makes periodic premium payments to the seller until either the contract reaches maturity or a credit event occurs, triggering a settlement. Unlike credit ratings, which distinguish between temporary shocks and permanent shocks to a company's value and will only change in the event of a permanent shock (Gredil et al., 2022), CDS spreads incorporate real-time market perceptions of a firm's ability to meet its financial obligations, and consider the temporary shocks which may trigger contractual terms affecting a firm's ability to purchase raw materials from suppliers and its production, making them a forward-looking indicator of firm value.

Our CDS data are sourced from S&P Global. The dataset provides CDS spreads across various contract terms, currencies, and documentation types. Consistent with prior literature (Bai & Wu, 2016), we focus on five-year CDS contracts denominated in U.S. dollars with modified restructuring (MR) clauses, as this contract type is the most liquid. To ensure reliability, we exclude observations with CDS spreads exceeding 10,000 basis points, as such extreme values often indicate illiquid or distressed trading conditions.

The database contains CDS spreads for 1,259 unique U.S. company names from 2012 to 2023. To integrate CDS data with firm fundamentals, we match each entity to its financial statement information retrieved from Compustat. Through this matching process, we identify a subset of 466 publicly traded U.S. firms with complete financial fundamental information. Given the limited ESG coverage in earlier years and our requirement that each firm have complete ESG information, our final dataset includes 417 publicly traded firms after integrating ESG ratings into the broader dataset.

Following prior literature (Bai & Wu, 2016), we sample CDS data on a weekly basis, selecting Wednesday as the reference date each week, covering a total of 706 active weeks. For consistency, we match financial statement information with market-based variables using a 45-day rule: for each accounting quarter, the associated CDS and stock variables are taken from the subsequent window that begins 45 days after the quarter end. Specifically, we match CDS spreads and stock market data between:

- **May 15 to August 14** with Q1 balance sheet data,
- **August 15 to November 14** with Q2 balance sheet data,
- **November 15 to February 14** with Q3 balance sheet data, and
- **February 15 to May 14** with Q4 balance sheet data.

This 45-day rule ensures that financial statement information from the most recent quarter is available before it is linked to CDS spreads and stock market variables. By contrast, Glassdoor reviews are aligned to the accounting quarter in which they are posted. Each review is time-stamped by its posting date and assigned to that quarter. This alignment allows us to avoid the look-ahead bias by matching employee-generated information from Glassdoor contemporaneously to the firm's fundamental information, and to test whether Glassdoor information contains incremental information beyond firm fundamentals.

2.2 ESG Data from MSCI

To incorporate ESG considerations, we use MSCI ESG Ratings, a standard proxy for firm-level sustainability practices with monthly updates that capture evolving risks. We focus on the Social (S) pillar insofar as it codifies firms' labor and human-capital practices (training and career development, compensation and benefits, health & safety), supply-chain labor standards and human rights, community relations and engagement, and product/customer responsibility. In line with industry practice, S scores are produced by rule- and weight-based methodologies that combine firm-disclosed policies and management systems with ongoing monitoring of incidents and controversies; inputs are drawn from company filings and regulatory reports, supplemented by continuous news surveillance, and scores are refreshed on periodic (quarterly-to-annual) cycles with ad hoc updates for major events.[9]

Integrating MSCI ESG data presents a sample construction challenge due to a structural expansion in 2012. Before 2012, ESG coverage focused mainly on large-capitalization publicly listed firms, which limits the intersection with our credit dataset. After 2012, MSCI broadened coverage to include many mid-capitalization and some small-capitalization publicly listed firms. However, a substantial proportion of these newly covered firms do not have CDS observations, which reduces the overlap between the ESG universe and the CDS universe and constrains the size and representativeness of the joint sample. Accordingly, we begin our sample in 2012 to ensure stable coverage and to minimize missing observations.

To ensure proper integration of ESG data with firm fundamentals and CDS spreads, we match the datasets at the firm level using Compustat firm identifiers, thereby maintaining consistency in firm coverage. Given that MSCI ESG scores are updated monthly, we then align them with our quarterly financial and CDS data by taking the most recent ESG rating available before the end of the financial quarter for each firm-quarter observation. This procedure ensures that the ESG information used in our analysis reflects

[9]Methodological details are broadly consistent across leading providers; see, for example, FTSE Russell (2023), LSEG (2024), MSCI Inc. (2024), and S&P Global (2025).

the most relevant data available to investors.

2.3 Employee Reviews from Glassdoor

Glassdoor is an influential online platform that provides employees with the opportunity to anonymously review their employers. Since its inception in 2008, Glassdoor has emerged as a pivotal source of a real-time, crowd-sourced channel to reflect workplace conditions, employee satisfaction, corporate culture, and managerial effectiveness. Employees offer assessments through both structured numeric ratings and unstructured textual comments, thus capturing nuanced aspects of firms' internal dynamics that traditional corporate disclosures frequently miss.

Our analysis utilizes employee reviews on Glassdoor as a source of a multidimensional measure of employees' assessments of the workplace. From the platform, we extract both structured numeric ratings and unstructured written reviews. The numeric ratings capture discrete aspects of the workplace. Specifically, employees rate their firms across several dimensions on a scale from 1 to 5, with ratings available for 'Overall Rating' and six detailed categories reflecting workplace conditions: Work-Life Balance, Career Opportunities, Compensation & Benefits, Senior Management, Culture & Values, and Diversity & Inclusion. Employees further provide categorical ratings on broader dimensions including Business Outlook, CEO Approval, and the likelihood of recommending the company to peers. Business Outlook and CEO Approval ratings use a ternary coding scheme (positive = 1, neutral = 0, negative = –1), while Recommendation ratings adopt a binary measure (yes = 1, no = –1). However, it is important to note that 'Diversity & Inclusion' was only introduced on the Glassdoor platform in 2020. As a result, this dimension contains substantial missing data throughout most of the sample period. To maintain consistency in our empirical analysis and ensure comparability across firms and time, we exclude this variable from our baseline specifications and primary inferences.

Additionally, we process the unstructured written reviews using machine-learning techniques to derive two distinct components: the tone (or sentiment) and the content (or risk-related buzz). The sentiment is the overall positive/negative valence of employee assessments, and the risk-related buzz is a topic-based summary of employee discussions, where we examine, in a CDS setting, whether the content includes expressions plausibly associated with potential credit risks. We explain further details in Section 3.

3 Methodology

3.1 Credit Risk Modeling

Our empirical strategy builds upon the structural framework introduced by Merton (1974), which links a firm's credit risk to its asset volatility and leverage through the well-known distance-to-default metric. While Merton's model provides an intuitive and theoretically grounded starting point for analyzing credit spreads, it is constrained by restrictive assumptions regarding debt structure and default thresholds, and omits non-financial dimensions such as ESG considerations and employee-related information.

To retain the foundational logic of structural credit risk modeling while addressing these limitations, we adopt the Bayesian shrinkage methodology developed by Bai and Wu (2016). This approach integrates traditional financial fundamentals with non-financial information, including the S score and employee-generated information from Glassdoor, in a sequential and disciplined manner. Each layer of information is orthogonalized with respect to the preceding one, ensuring that only incremental explanatory power is captured and avoiding redundancy and multicollinearity.

We begin by estimating a baseline CDS valuation model using the Merton model's distance-to-default measure (MCDS). Recognizing the systematic biases in structural models documented by prior literature (Eom et al., 2004), we calibrate the MCDS to market-observed CDS spreads using a nonparametric local quadratic regression. This calibration allows flexible correction of model misspecifications without imposing additional parametric restrictions, aligning theoretical valuations more closely with observed market conditions.

The baseline model is then augmented with firm-specific financial fundamentals to produce a fundamental-adjusted CDS valuation (FCDS). Each fundamental characteristic is orthogonalized relative to the MCDS predictions, isolating the portion of variation not explained by the structural model. These orthogonalized signals are combined using Bayesian shrinkage estimation, in which the weights are dynamically updated over time based on historical predictive performance. This Bayesian updating smooths temporal fluctuations in coefficients, mitigating instability due to transient noise in the data. Furthermore, the methodology incorporates a reliability-based imputation procedure for handling missing values, whereby missing observations are replaced with weighted averages of other characteristics, with the weights reflecting each characteristic's historical explanatory power. Operational details and implementation choices are provided in Appendix A.

The framework is next extended to incorporate non-financial information, focusing specifically on the 'Social' (S) dimension of MSCI ESG ratings. This targeted integration ensures thematic consistency with the employee-focused nature of our Glassdoor measures, as the S dimension is supposed to capture aspects of workforce management, employee well-being, and labor relations. The Social score is orthogonalized relative to the fundamental-adjusted valuation and then incorporated using the same Bayesian weighting process, producing an S-adjusted CDS valuation (SCDS).

In the final step, we add Glassdoor-based measures, including structured ratings, a Sentiment indicator extracted from employee narratives, and a Risk-Related Buzz indicator constructed via multinomial inverse regression, to produce a Glassdoor-adjusted CDS valuation (GCDS). This integration isolates the informational content of employee perceptions and workplace risk cues beyond what is reflected in the Social (S) pillar of ESG and traditional financial fundamentals. Table 1 summarizes the layered construction of our valuation measures. Merton-based CDS (structural leverage–volatility risk), fundamental-adjusted CDS (incremental signals from financial statements), S-adjusted CDS (policy-centric workforce metrics captured by the S pillar of ESG), and GCDS (numeric ratings, as well as Sentiment and Risk-Related Buzz from Glassdoor written reviews). Detailed implementation procedures are provided in Appendix A. Next, we explain the methodologies to extract the Sentiment and the Risk-Related Buzz components from written reviews.

Table 1: **Descriptions of CDS Measures**

Abbreviation	Full term	Definition
MCDS	Merton-based CDS	Structural Anchors: Market-implied pricing of leverage and volatility risk.
FCDS	Fundamental-adjusted CDS	Fundamental Adjustments: Incremental signals from financial statement analysis.
SCDS	Social-adjusted CDS	'S' Pillar Adjustments: Incremental signals from S score.
GCDS	Glassdoor-adjusted CDS	Employee-Generated Information Adjustments: Incremental information from Glassdoor numeric ratings and written-review-based measures.

Notes: This table provides definitions of the CDS measures, detailing the interpretation of each measure based on its data source and adjustments.

3.2 Sentiment Analysis for Written Reviews

We use text mining and sentiment analysis to analyze the tone within employees' written reviews. First, we apply text-mining techniques to clean and normalize the Glassdoor reviews. This involves removing stop words, punctuation, non-text characters, and converting the text to lowercase.

Our sentiment analysis is then based on a machine-learning model, specifically BERT (Bidirectional Encoder Representations from Transformers). Machine learning-based sentiment models have become increasingly prevalent in financial research due to their ability to capture context-sensitive meanings and complex linguistic structures (Gentzkow et al., 2019; A. H. Huang et al., 2023; Tetlock, 2007). These models have demonstrated superior accuracy compared to dictionary-based methods, particularly in settings where context significantly affects the sentiment conveyed (Loughran & McDonald, 2011), which makes them more adaptable to various textual inputs. BERT's bidirectional training enables it to consider both preceding and succeeding words in a sentence, which significantly improves its ability to assess sentiment in employee reviews.

We compute the Sentiment indicator as the difference between the model-estimated probabilities of positive and negative language:

$$\text{Sentiment}_r = \overline{p_r^+} - \overline{p_r^-}, \tag{1}$$

where $\overline{p_r^+}$ and $\overline{p_r^-}$ are the average sentence-level probabilities (across all sentences in review r) that a fine-tuned BERT classifier assigns to the positive and negative classes, respectively. The classifier returns, for each sentence, a probability for each class (positive/neutral/negative) that lies in $[0, 1]$ and sums to one. We then average these probabilities across sentences before taking the difference.

This construction captures both prevalence (more sentences leaning positive or negative) and strength of evidence (higher per-sentence probabilities increase their influence on the average). A higher Sentiment indicates that, on average, sentences are more likely and more confidently classified as positive. A lower Sentiment suggests more confidently negative language. When sentences are ambiguous or the model is unsure, the positive/negative probabilities cluster around 0.5 and are less likely to shift the average away from neutrality, which limits the influence of such weak sentiment.

3.3 Measuring Risk-Related Buzz from Written Reviews

We measure the content dimension of written reviews by constructing a text-based indicator, Risk-Related Buzz, designed to summarize phrases in employee reviews that are statistically associated with subsequent credit conditions. To extract this measure from high-dimensional text, we employ Multinomial Inverse Regression (MNIR), which projects token frequencies onto a one-dimensional score linked to an externally defined credit-condition label (Taddy, 2013). Reviews are processed separately for Glassdoor's Pros, Cons, and Advice to Management text fields, to preserve section-specific semantics; the final indicator averages the three section-level scores.

Risk-Related Buzz is anchored to credit risk events defined at each date t. Following Acharya and Johnson (2007), we assign an event label at t by looking forward from t to the end of the sample and marking t as an event date whenever the subsequent CDS path meets a prespecified criterion. In our main specification, the label equals 1 if there exists at least one trading day $s \geq t$ on which the firm's five-year CDS increases by more than 50 bps relative to $s - 1$; otherwise it is 0. As a robustness check, the label equals 1 if the firm's CDS level remains above 100 bps on all trading days from t through the end of the sample; otherwise it is 0. We use the 50-bps definition in the baseline MNIR estimation and report separate results for the 100-bps definition.

To estimate Risk-Related Buzz, we apply MNIR (Campbell & Shang, 2022), which projects reweighted word frequencies onto a one-dimensional score linked to the forward credit-condition label, while conditional on controls. This projection yields per-term loadings that quantify each word's statistical association with subsequent credit conditions (without imposing any prior assumption on which terms should matter). The MNIR specification is:

$$W_{j,i,t} = \frac{d_{j,i,t}}{N_{i,t}} \times \ln\left(\frac{N_{i,t}}{1 + d_{j,i,t}}\right), \tag{2}$$

$$E(W_{j,i} \mid x_{i,t}, v_{i,t+1}) = \exp\left(\alpha_j + \beta_j x_{i,t} + \phi_j v_{i,t}\right), \tag{3}$$

where, $d_{j,i,t}$ is the count of all Glassdoor reviews for a given firm i at quarter t containing word j; $N_{i,t}$ is the total word count for firm i at quarter t; $W_{j,i,t}$ is the reweighted frequency for word j; $x_{i,t}$ is a vector of controls for firm i in quarter t; $v_{i,t}$ is the forward credit-event indicator defined above.

The firm-quarter Risk-Related Buzz indicator then aggregates the estimated per-term loadings using the following normalization:

$$\text{Buzz}_{i,t} = \frac{\phi_1 W_{1,i,t}}{\sum_{j=1} W_{j,i,t}} + \frac{\phi_2 W_{2,i,t}}{\sum_{j=1} W_{j,i,t}} + \cdots + \frac{\phi_j W_{j,i,t}}{\sum_{j=1} W_{j,i,t}}, \tag{4}$$

so that higher values indicate that the review language at (i, t) more closely resembles the vocabulary empirically associated with adverse credit conditions under the chosen label. We compute this score separately for Pros, Cons, and Advice to Management to preserve section-specific semantics, and then take the simple average to obtain a single review-level indicator at the firm-quarter level. We adopt this averaging to keep the construction parsimonious and easily interpretable. As a robustness check, Appendix B reports specifications that include the Pros, Cons and Advice to Management scores separately (without averaging), and the results are quantitatively similar.

To illustrate the type of language captured by our MNIR measure, we list representative vocabulary with high MNIR loadings in Appendix B, along with example excerpts from actual reviews. These terms reveal coherent themes related to organizational risk factors, including operational rigidity, resource constraints, and management quality concerns that empirically precede credit deterioration.

3.4 Comparison of Model Performance

Having developed a sequence of CDS valuation measures through the layered Bayesian shrinkage procedure described above, we next evaluate and compare their ability to explain the cross-sectional variation in observed CDS spreads. This comparison provides a direct empirical link between the model construction process and the central objective of the study, namely, to assess whether incorporating the Social score and Glassdoor-based measures yields statistically and economically meaningful improvements in explanatory power.

For each week, we estimate a set of cross-sectional regressions in which the dependent variable is the logarithm of the observed CDS spread and the explanatory variable is the model-implied CDS valuation from each information layer, following the construction method in Bai and Wu (2016). As a baseline for comparison, we also estimate a bivariate linear regression (BLR) specification that uses the Merton model inputs, the leverage ratio (D/E) and the equity volatility (σ_E), directly, without imposing the Merton model's transformation into a distance-to-default measure. This benchmark enables us to assess the incremental value of adopting the structural framework and its subsequent extensions.

Formally, the BLR specification is given by:

$$\ln(\mathrm{CDS}_{i,t}) = a_t + b_t(D/E)_{i,t} + c_t\sigma_{E,i,t} + e_{i,t}. \tag{5}$$

While the structural and adjusted valuation models are estimated as:

$$\ln(\mathrm{CDS}_{i,t}) = \ln(\widehat{\mathrm{CDS}}_{i,t}^{(k)}) + e_{i,t}, \tag{6}$$

where $\widehat{\text{CDS}}^{(k)}$ denotes the predicted spread from information layer $k \in \{\text{MCDS}, \text{FCDS}, \text{SCDS}, \text{GCDS}\}$. Because these adjusted valuations are designed to be unbiased with respect to market quotes, the intercept is fixed at zero and the slope at one in equation (6), so that the regression residual directly measures the log pricing error.

The sequence of regression results across these models, presented in Section 4, quantifies the marginal contribution of each information set — structural model outputs, firm fundamentals, the S score, and Glassdoor measures — in explaining the cross-sectional dispersion in CDS spreads. This evaluation forms the empirical foundation for our subsequent analysis of the distinct informational channels embedded in employee-generated information.

4 Results

4.1 Incremental Explanatory Power of Glassdoor-Based Measures

We first show the potential link between Glassdoor information and credit risk using descriptive statistics. Table 2 summarizes firm characteristics by CDS quintiles. It shows that well-known risk covariates behave as expected, and the Glassdoor measures vary systematically with credit risk. Moving from Q1 to Q5 of credit risk, firms are smaller, more leveraged, and exhibit higher realized equity volatility and lower interest coverage. ESG overall scores are lower in the high-risk quintile, consistent with standard priors. Glassdoor measures display clear gradients: overall ratings and all subratings decline monotonically with credit risk, and the Sentiment falls from 0.135 in Q1 to 0.083 in Q5. In contrast, the Risk-Related Buzz is essentially flat across quintiles, with values clustered around 0.055. These patterns suggest that broad employee sentiment comoves with overall credit risk, while the Risk-Related Buzz captures concerns not concentrated in any particular risk segment.

We then quantify the incremental contribution to the cross-sectional R^2 obtained by adding the Glassdoor measures. We find that adding the Glassdoor numeric ratings already produces a model that performs much better than adding the Social score, and

Table 2: **Descriptive Statistics across CDS Quintiles**

Panel A: Firm Fundamentals							
Variable	Mean	Std. dev.	Means at CDS Quintiles				
			Q1	Q2	Q3	Q4	Q5
CDS	137.250	247.438	29.143	50.355	74.874	120.197	411.681
Total debt/Market cap	0.566	1.468	0.184	0.253	0.363	0.435	1.594
Realized volatility	0.308	0.153	0.224	0.246	0.290	0.324	0.456
Liability/Market cap	0.716	1.539	0.291	0.390	0.490	0.557	1.855
Total debt/Total assets	0.348	0.190	0.315	0.312	0.332	0.359	0.421
EBIT/Interest expense	13.564	23.952	22.751	16.599	12.365	9.816	6.289
Working capital/Total assets	0.105	0.133	0.088	0.092	0.119	0.122	0.105
EBIT/Total assets	0.031	0.026	0.038	0.034	0.029	0.030	0.021
Retained earnings/Total assets	0.292	0.469	0.457	0.364	0.328	0.298	0.015
ln(Market cap)	9.907	1.454	11.334	10.555	9.866	9.367	8.411
Stock market momentum	0.153	0.389	0.192	0.172	0.137	0.174	0.091
ln(Implied/Realized vol)	0.108	0.211	0.168	0.149	0.101	0.081	0.043
ESG overall score	4.064	1.503	4.422	4.222	4.151	3.965	3.559
ESG Environmental score	4.941	2.504	5.527	4.992	4.888	4.619	4.679
ESG Social score	4.098	1.842	3.886	4.146	4.080	4.290	4.090
ESG Governance score	5.304	2.016	5.370	5.395	5.294	5.314	5.145

Panel B: Glassdoor Information							
Variable	Mean	Std. dev.	Means at CDS Quintiles				
			Q1	Q2	Q3	Q4	Q5
CDS	137.250	247.438	29.143	50.355	74.874	120.197	411.681
Overall rating	3.521	0.494	3.607	3.581	3.574	3.492	3.351
Business outlook	0.311	0.329	0.394	0.365	0.342	0.297	0.155
CEO rating	0.360	0.315	0.393	0.405	0.401	0.356	0.244
Recommend rating	0.318	0.358	0.388	0.365	0.348	0.302	0.187
Career opportunity rating	2.769	0.485	2.831	2.824	2.788	2.756	2.645
Compensation & Benefits rating	3.000	0.535	3.056	3.034	3.021	3.041	2.849
Culture & Values rating	2.870	0.563	2.899	2.928	2.922	2.889	2.711
Senior management rating	2.546	0.509	2.563	2.592	2.591	2.573	2.412
Work & Balance rating	2.827	0.545	2.832	2.851	2.874	2.869	2.708
Sentiment	0.119	0.134	0.135	0.132	0.133	0.114	0.083
Risk-Related Buzz	0.055	0.008	0.054	0.054	0.054	0.055	0.055

Notes: This table reports summary statistics of firm-week-level variables partitioned by quintiles of CDS spreads. Panel A includes standard financial indicators and ESG scores, while Panel B presents employee-generated information from Glassdoor, including numeric ratings, Sentiment, and Risk-Related Buzz. Quintiles are formed each week based on CDS spread levels, ranging from Q1 (lowest risk) to Q5 (highest risk). All variables are pooled over time and across firms.

its effect is already similar to adding the full ESG composite score. Further adding the written-review-based measures (Sentiment and Risk-Related Buzz) can even improve the model further.

Table 3 reports the distribution of weekly R^2 for nested specifications. Focusing on the Social pillar, the fundamental model augmented by the S score (SCDS) attains a mean cross-sectional R^2 of 75.51%. Using the structured Glassdoor ratings instead of the S score generates a mean cross-sectional R^2 of 76.02%, which already exceeds the contribution of the S score by 0.51 percentage points. Adding the ratings on top of the S score lifts the mean to 76.95% (an increase of 1.44 percentage points relative to SCDS). Including the two orthogonalized written-review-based measures further raises it to 77.95% (an increase of 2.44 percentage points relative to SCDS and 1.00 percentage point beyond SCDS + Glassdoor ratings). For comparison, augmenting FCDS with the composite ESG score produces a mean of 76.06%, which is very close to the FCDS + Glassdoor ratings model (76.02%). Detailed results with the ESG composite score are reported in Appendix B.

Table 3: **Cross-Sectional R^2 from Fundamental Models Augmented with S score and Glassdoor Information**

Variable		Mean	Std. dev.	Min	Max
FCDS		74.55%	0.048	58.05%	82.45%
FCDS	+ Glassdoor ratings	76.02%	0.044	61.84%	83.38%
FCDS	+ Glassdoor ratings + Written-review-based measures	76.99%	0.042	63.15%	85.64%
SCDS		75.51%	0.045	60.35%	83.57%
SCDS	+ Glassdoor ratings	76.95%	0.041	63.59%	84.60%
SCDS	+ Glassdoor ratings + Written-review-based measures (GCDS)	77.95%	0.039	65.30%	86.55%

Notes: This table summarizes the distribution of weekly cross-sectional R^2 from regressions of ln(CDS) on a Fundamental-adjusted specification (FCDS) augmented with different information sets. We compare adding (i) the Glassdoor numeric ratings, and (ii) the ratings plus two written-review-based measures (Sentiment and Risk-Related Buzz) extracted from employee reviews. Then, we add the Social (S) pillar on top of FCDS, resulting in SCDS, and repeat the incremental additions of Glassdoor information. Each week's R^2 is computed from a cross-sectional regression using all firms available in that week.

To illustrate temporal stability, we plot the weekly R^2 paths for all competing layers. Figure 2 shows that FCDS sits well above the benchmark and the MCDS series throughout the sample period. Adding the S score improves fit further (SCDS). The specification that incorporates Glassdoor information (GCDS) is the best-performing model for the majority of weeks. All specifications experience a visible drop in performance in 2020, yet the performance ranking remains unchanged. The time profile therefore corroborates that Glassdoor augments fundamentals at least as well as the S pillar and that text-based channels provide a further stable improvement.

Finally, we evaluate whether richer information produces tighter valuation anchors around which observed spreads correct. Following Bai and Wu (2016), let V_t denote the model-implied anchor for the conditional expectation of $\ln(\text{CDS}_t)$, and define the

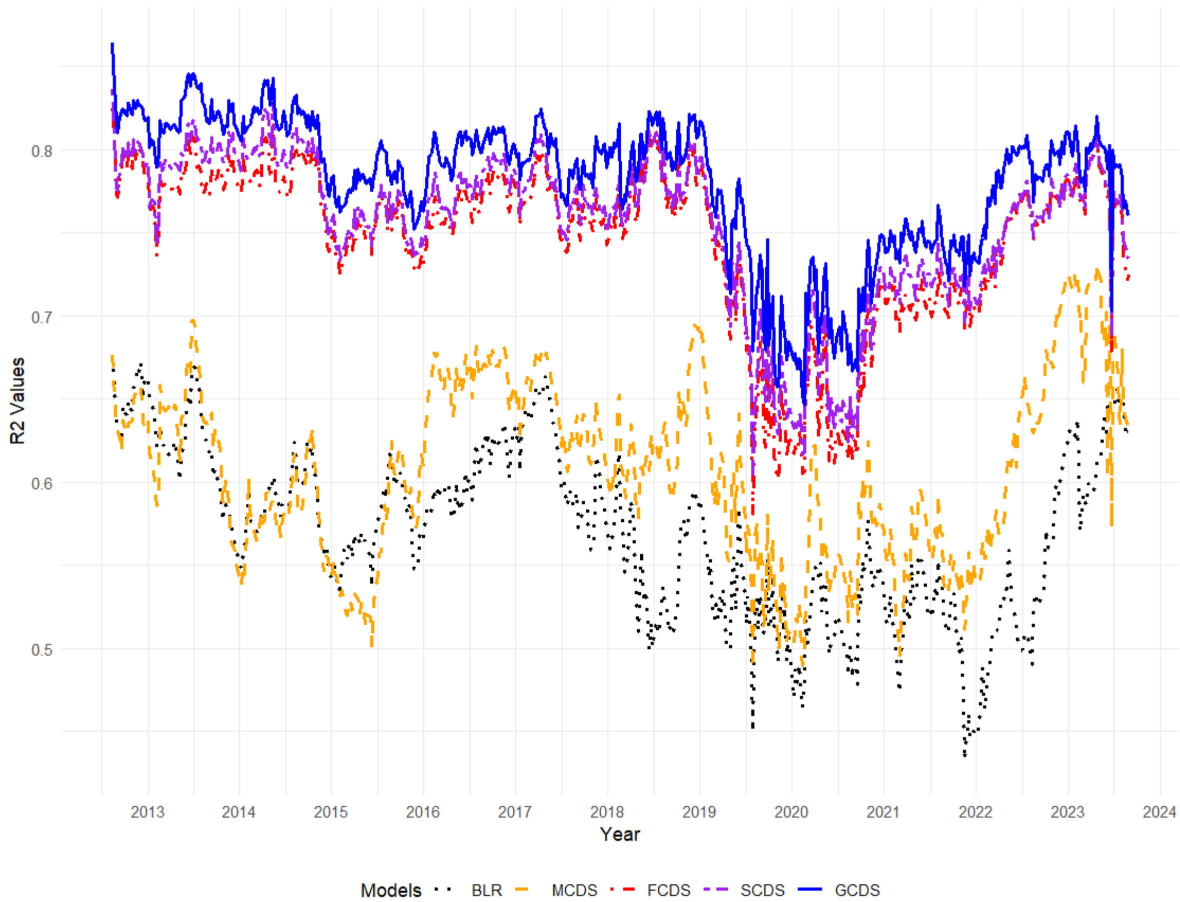


Figure 2: **Time Series of Cross-Sectional R^2 from Competing CDS Valuation Models**

Notes: This figure presents the weekly time series of R^2 values from cross-sectional regressions of market CDS spreads on various model-implied valuations, covering the period from August 2012 to August 2023. The five lines correspond to different levels of informational adjustment: the baseline benchmark model(BLR) using leverage and equity volatility, the Merton-based structural model (MCDS), and versions sequentially augmented with firm fundamentals (FCDS), S score (SCDS), and Glassdoor information (GCDS).

deviation as $e_t = \ln(\text{CDS}_t) - \ln(V_t)$. If V_t is a good proxy for the conditional mean, the transitory component in e_t should dominate and the deviation should follow a partial-adjustment law $e_{t+h} \approx (1 - \kappa h)e_t + \varepsilon_{t+h}$, so that larger κ indicates a stronger pull toward the anchor and, in turn, a higher signal-to-noise ratio of the valuation model. Table 4 reports the cross-sectional mean-reversion estimates. Deviations based on the GCDS model revert the fastest, with an average annualized κ of 1.518 (Newey-West $t = 11.289$) and an implied half-life of 5.48 months. The FCDS and SCDS models deliver similar speeds, with κ being 1.265 and 1.269, and half-life being 6.57 and 6.55 months respectively. The observed true ln(CDS) exhibits only a κ of 0.219 and a half-life of roughly 38.06 months, reflecting the mixture of permanent risk premium and information shocks that Bai and Wu (2016) emphasize will not mean revert at short horizons. Interpreted through their framework, the ordering of κ maps directly to anchor quality: augmenting fundamentals with Glassdoor employee-generated information removes omitted transitory components from ln(CDS), sharpens the anchor $\ln(V_t)$, and yields faster and more precise corrections.

Table 4: **Mean Reversion Estimates for Market CDS and Valuation Deviations**

Statistics	Mean	Std. dev.	Newey–West t-statistics	Half-life (months)
Observed CDS	0.219	0.036	6.088	38.063
BLR	0.537	0.073	7.408	15.477
MCDS	0.621	0.088	7.066	13.397
FCDS	1.265	0.123	10.260	6.573
SCDS	1.269	0.123	10.303	6.554
GCDS	1.518	0.134	11.289	5.481

Notes: This table reports the cross-sectional mean-reversion estimates following Bai and Wu (2016). In each week t, we run a cross-sectional regression of the form $x_{t+h} - x_t = a - \kappa x_t h + \varepsilon_{t+h}$, where x_t denotes the log CDS spread or its value-adjusted deviation at time t, and $h = \frac{1}{52}$ represents a one-week prediction horizon. The parameter κ measures the annualized speed at which the series reverts to its benchmark; larger κ implies faster reversion. We report the time-series mean of κ estimates, their standard deviations, and Newey–West t-statistics across weeks. Half-life (months) converts the annualized speed into the time needed for a deviation to decay by half, computed as $\text{HL} = \ln(2)/\kappa \times 12$. The deviation terms are residuals between market-observed CDS spreads and benchmark CDS valuations from BLR, MCDS, FCDS, SCDS, and GCDS models.

4.2 Additional Validation of Main Results

In this part, we conduct some additional tests to validate the our main results reported above and to demonstrate their economic importance. We first follow Bai and Wu (2016) and assess whether richer information improves the stability of the GCDS model out of sample. For each week t, we randomly split the cross section of firms into a 50% training set and a 50% testing set. On the training half, we estimate an anchor mapping by regressing $\ln(\text{CDS}_i)$ on $\ln(V_i)$ for each specification in BLR, MCDS, FCDS, SCDS, and GCDS. We then form out-of-sample fitted values on the testing half and compute two diagnostics at the weekly cross section: (i) the testing R^2 from regressing ln(CDS) on the out-of-sample fitted values, and (ii) the testing root mean squared error (RMSE), defined as the square root of the cross-sectional mean of squared prediction errors. We repeat this procedure each week and summarize the distribution of weekly out-of-sample R^2 and RMSE.

Table 5 reports the weekly out-of-sample performance. The ranking of models is stable: the baseline BLR model attains a mean cross-sectional R^2 of 57.25%. The FCDS model increases fit markedly to 74.96%. Augmenting fundamentals with the Social pillar (SCDS) yields a modest additional gain to 75.75%. The GCDS model, which adds structured ratings and written-review-based measures from employees, delivers the highest mean out-of-sample R^2 at 78.30%. The dispersion is also lowest for GCDS (weekly standard deviation of 0.054), indicating more stable anchors across weeks. Consistent with these patterns, the out-of-sample RMSE declines monotonically from 0.587 (BLR) to 0.413 (GCDS).

Table 5: **Out-of-sample Weekly Cross-sectional Performance of Different Models**

	BLR	MCDS	FCDS	SCDS	GCDS
Panel A: Out-of-sample Cross-sectional R^2					
Mean	57.25%	60.89%	74.96%	75.75%	78.30%
Std. dev.	0.074	0.077	0.062	0.060	0.054
Min	35.71%	39.56%	51.31%	52.58%	58.91%
Max	77.62%	82.17%	88.43%	88.97%	90.57%
Panel B: Out-of-sample RMSE					
Mean	0.587	0.554	0.444	0.437	0.413
Std. dev.	0.072	0.068	0.068	0.068	0.064
Min	0.395	0.406	0.271	0.265	0.244
Max	0.828	0.788	0.691	0.691	0.666

Notes: This table reports weekly out-of-sample performance for five valuation models: BLR, MCDS, FCDS, SCDS, and GCDS. For each week, firms are randomly split into equally sized training and testing sets with a fixed seed. On the training half, we regress ln(CDS) on the model-implied anchor $\ln(V)$ to obtain a mapping. We then form fitted values on the testing half and compute (i) out-of-sample cross-sectional R^2 from regressing ln(CDS) on the fitted values, and (ii) out-of-sample RMSE as the square root of the mean squared prediction error on the testing half. Reported statistics summarize the cross-week distribution. R^2 entries are in percentage points; RMSE is in log points of CDS spreads.

Next in order to demonstrate the value of Glassdoor information for trading, we conduct an out-of-sample portfolio test and show the value created by our GCDS model. If our model robustly foreshadows subsequent spread corrections, then systematically buying or selling protection on a firm's CDS whenever the observed spread overshoots or undershoots the GCDS level should yield abnormal returns.

For each horizon $h \in \{1, 2, 3, 4\}$ weeks, we form a zero-cost portfolio that goes long or short each CDS contract in proportion to the sign and magnitude of the mismatch between market CDS and the relevant fundamental-based measure. Our approach invests out-of-sample: each day, we calibrate the model on the prior data and apply it to the current cross-section. The profit gain or loss realized over the next h weeks is then evaluated.

Specifically, for each firm at time t, the deviation is computed as the difference between the observed CDS spread and our value-adjusted valuation models. A notional position is then assigned proportionally to this deviation, such that

$$n_{i,t} = c_t \Big(\widehat{\mathrm{CDS}}_{i,t} - \mathrm{CDS}_{i,t} \Big), \tag{7}$$

where c_t is a proportionality coefficient normalized each day so that the total long positions sum to one dollar and the total short positions to minus one dollar, thereby ensuring a market-neutral portfolio. The strategy involves taking these positions at time t and holding them for a fixed horizon h. The profit and loss (PL) for a one-dollar long position is calculated, in the absence of default, using the expression:

$$PL_{i,t,h} = LGD \cdot \Big(\lambda_{i,t+h} - \lambda_{i,t} \Big) \frac{1 - \exp\Big\{ -\big(r_{t+h} + \lambda_{i,t+h} \big)(\tau - h) \Big\}}{r_{t+h} + \lambda_{i,t+h}}, \tag{8}$$

where LGD denotes the loss given default, which we assume fixed at 60% for all contracts (Bai & Wu, 2016), r denotes the continuously compounded benchmark interest rate, which we use the five-year-USD-Treasury interest rate, and $\lambda_{i,t}$ denotes the default arrival rate for the i th-company, which we infer from the corresponding CDS rate by assuming a flat term structure, $\lambda_{i,t} = \frac{\mathrm{CDS}_{i,t}/10000}{LGD}$. In case the company defaults during our investment horizon, the payout for a one-dollar notional long position is given by the loss given default $PL = LGD$. In aggregate, we can regard the dollar profit and loss from the total investment at each date as excess returns on a one-dollar notional long and one-dollar short investment. The investment exercise is purely out-of-sample as the valuations at time t only use information up to time t.

Table 6 summarizes the annualized mean returns, volatility, and Sharpe ratios from such a strategy, comparing results across bivariate linear regressions, MCDS, FCDS, SCDS, and our GCDS. We test horizons from one through four weeks, effectively rebalancing positions at those intervals.

Table 6: **Summary statistics of excess returns from an out-of-sample investment exercise**

Horizon, h weeks	Annual Mean (%)	Std (%)	Annual Min (%)	Annual Max (%)	Sharp Ratio
I. Investments based on the bivariate linear regression					
1	1.525	20.655	-107.391	172.113	0.074
2	2.728	29.211	-151.705	224.259	0.093
3	3.945	36.340	-177.261	261.510	0.109
4	5.049	41.274	-184.899	273.863	0.122
II. Investments based on the MCDS model					
1	1.564	17.715	-73.599	191.974	0.088
2	2.883	27.829	-119.778	275.039	0.104
3	4.205	35.467	-126.423	365.208	0.119
4	5.519	40.831	-149.908	436.710	0.135
III. Investments based on the FCDS model					
1	1.585	17.680	-73.049	195.852	0.090
2	3.044	28.024	-111.991	280.821	0.109
3	4.237	35.717	-133.663	376.915	0.119
4	5.546	40.588	-147.039	452.679	0.137
IV. Investments based on the SCDS model					
1	1.668	18.991	-72.218	207.475	0.088
2	3.253	30.436	-109.060	301.079	0.107
3	4.463	38.925	-130.200	401.368	0.115
4	5.718	44.120	-163.750	487.757	0.130
V. Investments based on the GCDS model					
1	1.706	19.012	-70.433	211.380	0.090
2	3.241	30.336	-112.382	307.876	0.107
3	4.490	38.700	-130.957	413.836	0.116
4	5.830	43.951	-154.450	504.526	0.133

Notes: This table presents summary statistics from an out-of-sample investment experiment that uses observed discrepancies between market CDS quotes and five different CDS valuation approaches (BLR, Merton, FCDS, SCDS, GCDS) to construct a trading strategy. Each panel corresponds to a particular valuation method, and each row reflects a different holding horizon (from one to four weeks). The table reports annualized mean excess returns, return volatility, extreme outcomes (annualized min and max), and the Sharpe ratio. We derive these performance metrics by applying the same notional-allocation framework at each rebalancing date, with the sign and magnitude of the position determined by whether the market CDS is above or below the model-implied level.

The results confirm that the GCDS model not only enhances forecasting metrics but also yields superior economic outcomes. For instance, under the one-week rebalancing horizon, the trading strategy based on the GCDS model generates an annualized mean return of 1.706%, which is higher than those produced by the BLR (1.525%), MCDS (1.564%), and FCDS (1.585%) benchmarks. Although the SCDS model yields a comparable return of 1.668%, the improvement of the GCDS over the structural models is statistically significant at the 5% level based on a t-test. This pattern of economically and statistically significant outperformance persists across longer investment horizons, culminating in the highest four-week mean return of 5.830% among all competing models.

4.3 Decomposition into Sentiment and Risk-Related Buzz

To maintain continuity with the earlier analysis, we begin by quantifying how the two written-review-based measures contribute to the cross-sectional fit using the same weekly R^2 metric. As shown in Table 3, adding the two written-review-based measures on top of the 'SCDC + numeric ratings' specification raises the mean weekly R^2 by about one percentage point (from 76.95% to 77.95%). These patterns indicate that using the two measures jointly delivers incremental information beyond what the ratings summarize, capturing variation in spreads not accounted for by the ratings.

With the overall contribution to R^2 established, we next dissect how the model allocates that gain in R^2 between Sentiment and Risk-Related Buzz after fundamentals, the S score, and numeric ratings are absorbed. Table 7 reports the weekly posterior weights from the Bayesian stacking update described in our methodology: in each week t, we take the residuals from the 'SCDS + numeric ratings' specification, project them onto the two orthogonalized written-review-based measures within a shrinkage framework, and recover posterior combination weights that allocate the marginal explanatory power between Sentiment and Risk-Related Buzz. The weights are estimated week by week using all firms available in that week, so they can be interpreted as time-varying within-layer importance measures rather than fixed coefficients.

According to Table 7, both written-review-based indicators receive positive and stable weights on average. Sentiment has a mean of 0.987 with a standard deviation of

0.196, while Risk-Related Buzz has a mean of 1.023 with a standard deviation of 0.096, indicating that the incremental fit is not driven by a single dimension. The levels are close in magnitude, yet the greater dispersion of Sentiment suggests stronger state dependence relative to the more persistent contribution of Risk-related Buzz.

Table 7: **Descriptive Statistics of Bayesian Weights for Sentiment and Risk-Related Buzz**

Variable	Sentiment	Risk-Related Buzz
Mean	0.986	1.208
t-statistics	151.543***	133.234***
Std. dev.	0.159	0.222
Min	0.500	0.500
p25	0.905	1.125
p50	1.017	1.169
p75	1.110	1.272
Max	1.315	1.697

Notes: This table reports the estimated Bayesian shrinkage weights for two textual information channels derived from Glassdoor reviews: Sentiment and Risk-Related Buzz. These variables are constructed using natural language processing techniques and reflect orthogonal components not captured by numeric ratings. The weights represent the average contribution of each indicator to the model's explanatory power for CDS spreads, with higher values indicating stronger predictive relevance. The t-statistics test the null hypothesis that the mean weight equals 0 (two-sided). All statistics are calculated over the full panel of firm-week observations.

Time variation provides an additional check on whether the model relies on these channels episodically or persistently. Figure 3 shows that both weights remain positive throughout the sample, indicating persistent incremental explanatory power of written-review-based measures. The series comove, rising together around 2020 and subsequently reverting toward their earlier levels. A moderate decline is visible in the mid-sample period prior to 2019. The Sentiment weight is elevated in 2020–2021 relative to earlier periods, and the Risk-Related Buzz weight exhibits a similar temporary increase. The joint increase around 2020 suggests that, when conventional measures are less informative or harder to interpret, both written-review-based measures contribute more to explaining cross-sectional variation in credit spreads. The subsequent reversion indicates that the enhanced role of these two channels is not just a one-time anomaly.

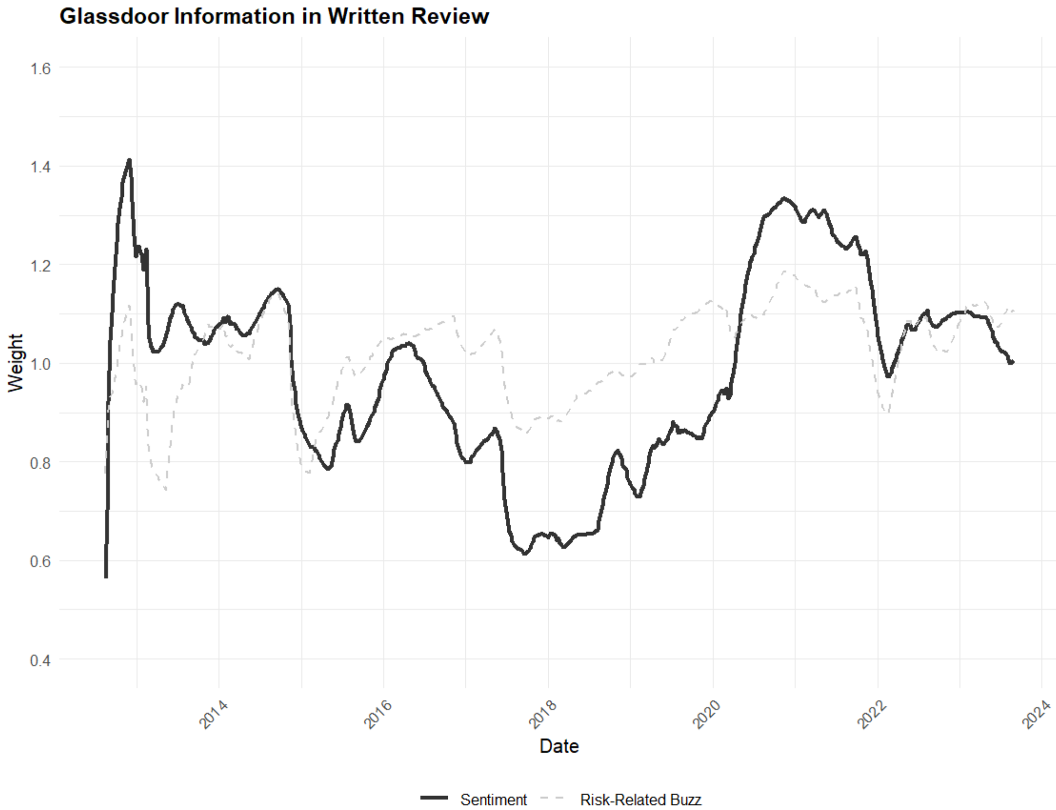


Figure 3: **Bayesian Weights of Sentiment and Risk-Related Buzz**

Notes: This figure displays the weekly Bayesian shrinkage weights of two orthogonal textual components extracted from Glassdoor reviews: Sentiment and Risk-Related Buzz. Both measures are derived from unstructured employee narratives using natural language processing techniques and are orthogonal to numeric ratings.

4.4 Mechanism Tests

Addressing endogeneity of the sentiment channel

In this section we provide further tests for the sentiment and risk-related buzz channels respectively. We begin with the sentiment channel and address endogeneity, because a positive association between sentiment and credit spreads could reflect contemporaneous firm conditions, aggregate mood, or even reverse causality. We address this by adopting an instrumental variable (IV) approach. We instrument quarterly Glassdoor written-review-based Sentiment with two externally timed mood shocks that are orthogonal to firm cash flows: the number of blockbuster movie releases (Hong & Wei, 2025) and global aviation fatalities (Kaplanski & Levy, 2010). Both series are measured at quarter t and used to predict Sentiment at quarter $t+1$ to avoid contemporaneous contamination.[10]

Table 8 reports our IV results. The first stage shows a strong effect (Kleibergen–Paap rk LM = 17.870, p=0.000; Cragg–Donald F=71.853), and the overidentifying restrictions are not rejected (Hansen J p=0.7382). Among the instrumental variables, blockbuster releases are associated with higher Sentiment, whereas global aviation fatalities are associated with lower Sentiment. In the second stage, the instrumented Sentiment loads positively on CDS spreads, implying that a higher Sentiment leads to higher CDS spreads, consistent with standard intuition. These estimates support a causal component in Sentiment beyond the previously absorbed information.

Table 8: **First and Second Stage Regression Results**

	Regression Results	
Variable	First Stage	Second Stage
SCDS	-0.028***	1.028***
	(-61.38)	(445.96)
Sentiment		0.335***
		(4.61)
Risk-Related Buzz	-0.795***	2.389***
	(-17.51)	(7.01)
Release	0.001***	
	(2.62)	
Fatalities	-0.000***	
	(-4.26)	
Constant	0.457***	-0.298***
	(85.54)	(-8.35)
Underidentification test (Kleibergen-Paap rk LM Statistic)	17.870 (p=0.000)	
Weak identification test (Cragg-Donald Wald F statistic)	71.853 25% maximal IV size = 7.25	
Overidentification test (Hansen J statistic)	0.112 (p=0.7382)	

Notes: This table presents results from the instrumental variable (IV) regression analysis, designed to establish a causal relationship between Sentiment and CDS spreads. The first-stage regression results show how two instrumental variables, the number of blockbuster movie releases (*Release*) and global aviation fatalities (*Fatalities*), predict employee sentiment (Sentiment).

[10]Construction details, data sources, and coding rules are provided in Appendix B.

Additional validation of channels

In this section, we conduct further split-sample analyses, which provide additional validation of our two written-review-based channels. We first investigate the effect of labor intensity. A firm's labor intensity may potentially influence the effect of employee sentiment on CDS. This is because our Glassdoor sentiment channel reflects workforce frictions and morale, which influence labor productivity and thus should have a larger effect in labor-intensive firms, because their production relies more on labor. At the same time, the opposite pattern can also be plausible, because such an effect of sentiment on production in labor-intensive firms is potentially already reflected in firm fundamentals. Our Glassdoor sentiment measure is orthogonalized with respect to fundamentals and to the Social pillar in the model, so its influence on labor-intensive firms is perhaps already absorbed by prior layers.

By contrast, the risk-related buzz channel captures forward-looking downside risk information that extends beyond standard ratios and is not inherently tied to the labor–capital mix. Workers in high- or low-labor-intensity firms can all talk about firm risks. Therefore, its contribution need not depend on labor intensity. We next compare the incremental contribution of each channel among high- versus low-labor-intensity firms, and evaluate their interactions with fundamentals and the Social pillar. To measure labor intensity, we follow Donangelo et al. (2019) and compute the labor share,

$$LS_{it} = \frac{\text{Labor costs}_{it}}{\text{Value added}_{it}} = \frac{\text{XLR}_{it}}{\text{OIBDP}_{it} + \Delta\text{INVFG}_{it} + \text{XLR}_{it}},$$

where XLR is Staff expense, OIBDP is Operating income before depreciation, and $\Delta\text{INVFG}_{it} \equiv \text{INVFG}_{it} - \text{INVFG}_{i,t-1}$ is the change in Inventories—finished goods. As in Donangelo et al. (2019), we set $\Delta\text{INVFG}_{it} = 0$ when either INVFG_{it} or $\text{INVFG}_{i,t-1}$ is missing. We compute LS_{it} annually, and carry it forward to the weekly panel, and in each week split the cross-section at that week's median LS_{it} into low and high labor intensity. Table 9 reports the distribution of weekly posterior weights for this split. Level differences are small. Risk-Related Buzz averages 1.0878 for low labor intensity versus 0.9586 for high, and the Sentiment averages 0.9790 versus 0.9096, with comparable dispersion across groups.

To gauge the economic importance, we translate shocks to basis points using the pooled mean spread $\overline{\text{CDS}} = 137$ bps from Table 2. Evaluated at this anchor, a 1-SD increase in the orthogonalized Sentiment correlates with a contemporaneous CDS narrowing of roughly 11 bps in high–labor-intensity firms and about 15 bps in low–labor-intensity firms, a difference of about 31%; whereas Risk-Related Buzz delivers nearly identical widening of about 15 bps in both groups with a difference below 1%.

Table 9: **Bayesian Updating Weights for Sentiment and Risk-Related Buzz by Labor Intensity**

	Low Labor Intensity		High Labor Intensity	
Statistic	Sentiment	Risk-Related Buzz	Sentiment	Risk-Related Buzz
Mean	0.979	1.088	0.910	0.959
Std. dev.	0.112	0.101	0.092	0.113
Min	0.582	0.759	0.515	0.544
Max	1.226	1.372	1.143	1.183

Notes: This table summarizes the distribution of weekly posterior weights from the Bayesian stacking regression that maps the SCDS residuals onto two orthogonalized written-review-based measures from Glassdoor reviews: Sentiment and Risk-Related Buzz. We report the time-varying weights separately for firms with low- vs. high- labor intensity, where labor intensity is classified by a median split of firm-level labor share (following Donangelo et al., 2019). Weights reflect the marginal contribution of each written-review-based measure to explain cross-sectional CDS variation after controlling for fundamentals, S score, and numeric ratings; higher weights indicate greater incremental explanatory power within written reviews. Statistics are computed across all weeks in the sample; each week's weights are obtained from cross-sectional regressions using the firms available in that week.

Finally, we provide additional validation of our risk-related buzz channel. We test how the weight on Risk-Related Buzz varies across firms with different levels of credit risk, because a forward-looking risk buzz should receive greater weight before the risk materializes in credit markets. Table 10 classifies firm-weeks by whether week t is labeled a credit-risk event under our baseline 50 bps jump rule as shown in Section 3. It shows that Risk-Related Buzz receives higher average weight and greater variability in the high-risk group (mean 1.0478 versus 1.0192), and the mean difference is statistically significant; whereas Sentiment is only slightly higher in the high-risk group than in the low-risk group (0.9351 versus 0.9262).

Taken together, the IV and split-sample analyses largely support our channel inter-

Table 10: **Bayesian Updating Weights for Sentiment and Risk-Related Buzz by Credit Risk Group**

	Low Credit Risk		High Credit Risk	
Statistic	Sentiment	Risk-Related Buzz	Sentiment	Risk-Related Buzz
Mean	0.926	1.019	0.935	1.048
Std. dev.	0.127	0.111	0.142	0.143
Min	0.564	0.579	0.537	0.542
Max	1.218	1.274	1.304	1.305

Notes: This table reports the distribution of weekly posterior weights from the Bayesian stacking regression that maps MCDS residuals onto two orthogonalized written-review-based measures from Glassdoor reviews: Sentiment and Risk-Related Buzz. In each week t, firms are split ex post by whether their future five-year CDS spread has increased by at least 50 basis points by the end of the sample period: the High Credit Risk group corresponds to 'Jump' observations with $\Delta\text{CDS}_{t\to t+n} \geq 50$ bps, and the Low Credit Risk group corresponds to 'No Jump' observations with $\Delta\text{CDS}_{t\to t+n} < 50$ bps. Weights reflect the marginal contribution of each written-review-based measure to explaining cross-sectional CDS variation after controlling for fundamentals, S score, and numeric ratings. Higher weights indicate greater incremental explanatory power within written reviews. Summary statistics are computed across all weeks.

pretations. The IV design indicates a positive and statistically significant coefficient of instrumented Sentiment on CDS, consistent with a causal component beyond information already absorbed by fundamentals, the Social pillar, and numeric ratings. The cross-sectional splits clarify when these channels are more informative: along the labor-intensity dimension, the incremental contribution of the sentiment channel is larger in low–labor-intensity firms, whereas the risk-related buzz channel is broadly invariant across labor intensities. Along the credit-risk dimension, Risk-Related Buzz receives higher average weight and greater variability in high-risk firm-weeks, in line with the view that forward-looking risk-related content in employee reviews is truly informative about imminent risks. Overall, these tests indicate that Sentiment and Risk-Related Buzz provide incremental information under different conditions, reinforcing the role of Glassdoor employee-generated information as a useful addition to conventional inputs.

4.5 Robustness Checks

We assess robustness along two dimensions that directly affect the written-review-based layer, specifically using alternative measures of Sentiment and of Risk-Related Buzz. Our main results should not be sensitive to measurement differences.

We begin with Sentiment, because the incremental fit attributed to the Glassdoor written-review layer could in principle reflect model-specific tone extraction rather than information in employee narratives. To assess robustness, we construct a dictionary-based benchmark using the Harvard lexicon, which is widely employed in financial text analysis (Gentzkow et al., 2019; X. Huang et al., 2014; Loughran & McDonald, 2011; Tetlock, 2007). Dictionary methods rely on predefined lexicons that map words to affect, offering high interpretability at the cost of weaker handling of context. We apply the Harvard dictionary to the same set of reviews, under the same pre-processing and rolling standardization as the BERT method, and form a review-level score as the difference between the shares of positive and negative tokens:

$$\text{Harvard\_sentiment} = \%\text{Harvard\_positive} - \%\text{Harvard\_negative},$$

and this dictionary-based indicator serves as a direct robustness check for the BERT-based sentiment measure.

We report Figures B.2 and B.3 in Appendix B to benchmark the two sentiment constructions. The figures show that the dictionary-based and BERT-based series comove over time, while the dictionary approach exhibits visibly higher volatility and deeper downside swings. The two measures are largely consistent. In contrast, the BERT-based measure is smoother and more stable.

Building on the evidence that the dictionary- and BERT-based Sentiment measures move in the same direction, we treat the Harvard series as an admissible substitute for the BERT Sentiment and use it for a robustness check. Table 11 shows that the main conclusions are unchanged. In Panel A, mean weekly R^2 rises from 74.55% with fundamentals to 75.51% with S, to 76.95% with ratings, and to 77.97% once text constructed as Harvard Sentiment plus Risk-Related Buzz is added, an increment of about one percentage point over the ratings layer, mirroring the BERT-based result; dispersion also declines to 0.041 in the richest specification. Panel B reports the within-text allocations, with Sentiment weights near one and Risk-Related Buzz weights above one on average, indicating that both channels retain incremental explanatory power once ratings are

Table 11: **Robustness of Cross-Sectional R^2: Replacing BERT Sentiment with Harvard Dictionary**

Panel A: Cross-sectional R^2		
Model	Mean R^2	Std. dev.
FCDS	74.55	0.048
SCDS	75.51	0.045
SCDS + Ratings	76.95	0.043
SCDS + Ratings + Sentiment (Harvard) + Risk-Related Buzz	77.97	0.041
Panel B: Bayesian Weights within the text layer		
Model	Mean R^2	Std. dev.
Sentiment (Harvard)	0.998	0.185
Risk-Related Buzz	1.115	0.157

Notes: This table evaluates the robustness of our valuation fit when the employee sentiment input is constructed with a Harvard Dictionary measure instead of the baseline BERT model; the Risk-Related Buzz construction is unchanged. Panel A reports the time-series mean and standard deviation of weekly cross-sectional R^2 from regressions of ln(CDS) on alternative information sets. Specifications build sequentially from the FCDS model to the SCDS model, then add Glassdoor numeric ratings, and finally add two written-review-based measures, Sentiment (Harvard Dictionary) and Risk-Related Buzz. Panel B summarizes the weekly Bayesian weights within the text layer. Higher R^2 indicates better cross-sectional fit; larger Bayesian weights indicate greater incremental explanatory power of the corresponding written-review-based measure within the text layer.

absorbed.

We then turn to the anchor used to construct Risk-Related Buzz, because the MNIR event label may bias the extracted content toward transitory spikes or toward slow-moving deterioration. Following the forward-labeling logic in Acharya and Johnson (2007), in our main analyses we adopt a jump anchor that labels week t as a risk event if, looking forward from t, there exists at least one trading week s ($s \geq t$) on which the firm's five-year CDS increases by more than 50 bps relative to the week t. As a robustness check, here we use a persistence anchor that flags week t if the firm's CDS level remains at or above 100 bps on every trading week from t until the end of the sample, following the same threshold in Acharya and Johnson (2007). Estimating MNIR under each anchor and comparing results allow us to assess sensitivity to the event definition while holding the modeling pipeline fixed.

Figures B.4 and B.5 in Appendix B compare our two credit-risk event anchors. The 50 bps jump label generates higher and more volatile weekly counts early in the sample, consistent with episodic stress, whereas the 100 bps persistence label yields fewer, smoother, and gradually rising counts indicative of sustained weakness.

Table 12: **Robustness of Cross-Sectional R^2: Risk-Related Buzz Anchored on a 100-bps Persistent CDS Threshold**

Variables	Mean R^2	Std. dev.
Panel A: Cross-sectional R^2		
FCDS	74.55	0.048
SCDS	75.51	0.045
SCDS + Ratings	76.95	0.043
SCDS + Ratings + Sentiment + Risk-Related Buzz (Persistence Label)	78.11	0.040
Panel B: Bayesian Weights within the text layer		
Sentiment	1.016	0.210
Risk-Related Buzz (Persistence Label)	1.130	0.124

Notes: This table examines robustness when the Risk-Related Buzz is re-estimated via MNIR using a forward-looking persistence anchor: at date t, an event is labeled if the firm's five-year CDS remains at or above 100 bps on all trading days from t through the end of the sample period (following the forward-labeling logic of Acharya and Johnson (2007)). Sentiment continues to be extracted with a BERT model. Panel A reports the summary statistics of weekly cross-sectional R^2 from regressions of ln(CDS) on a specification that augments fundamentals (with the Social pillar) and the Glassdoor ratings with the two written-review-based measures. Panel B summarizes the weekly Bayesian stacking weights within the text layer.

To complete the robustness check for different anchoring, we replace the jump-anchor label with the persistence-anchor label. Table 12 reports the results in the same layering used elsewhere. Panel A shows that the fit is at least as strong under the persistence anchor. Relative to the ratings-only layer, mean weekly R^2 rises from 76.95% to 78.11% when we add BERT Sentiment and Risk-Related Buzz, an increment of about 1.16 percentage points. Panel B then opens the written-review layer. Posterior allocations remain positive for both channels, with a mean weight of 1.0159 for Sentiment and 1.1301 for Risk-Related Buzz, and variability that is modest and comparable to the main analysis. The slight tilt toward the risk factor aligns with the persistence anchor's focus on slow-moving deterioration, where risk language carries more marginal information.

Overall, the two robustness checks convey a consistent story. Replacement of the transformer with a dictionary-based sentiment measure preserves the contribution of the sentiment channel, and re-anchoring the risk-related buzz channel in a persistent high-spread state preserves the incremental fit of the text layer while shifting weight toward the risk-related buzz channel.

5 Conclusion

Our results show that employee-generated information significantly contributes to credit risk assessment. Using a layered valuation framework that anchors CDS spreads to fundamentals and then sequentially adds the ESG Social pillar, structured Glassdoor numeric ratings, and two novel written-review-based measures, we find that employee voice, properly decomposed into tone (Sentiment) and content (Risk-Related Buzz), improves cross-sectional fit and sharpens valuation anchors. After orthogonalization with respect to fundamentals and the S score, Glassdoor numeric ratings raise mean weekly cross-sectional R^2 by 1.44 percentage points. Additionally, introducing the two written-review-based measures raises this R^2 by a further 1 percentage point.

The layered design ensures a clean identification of employee voice effects. Numeric ratings convey information close in spirit to what has already been encoded in the S score, but the written reviews provide non-overlapping information that markets appear to incorporate on an ongoing basis. Sentiment captures a behavioral channel, workplace morale that maps into productivity, coordination, retention, and ultimately risk. Risk-Related Buzz reflects an informational channel that flags concrete downside risks before they are fully visible. Consistent with these interpretations, posterior within-layer weights on both written-review-based measures are positive on average and vary over time, with Sentiment being more cyclical and Risk-Related Buzz more persistent. Our additional heterogeneity analyses and robustness tests reinforce the validity of these channels.

These findings speak to research, practice, and policy. For researchers, they demonstrate that high-frequency, bottom-up employee reviews contain important information that can be used to model credit risks, contributing to the understanding firm valuation. The effects of such reviews can be separated into two distinct channels, namely the behavioral channel (sentiment) and the true information channel (risk-related buzz). This goes beyond the single aggregated employee satisfaction measure commonly used in the literature. For practitioners, the results suggest that combining structured ratings with written-review-based measures yields more responsive credit assessments than relying solely on financial fundamentals or ESG metrics. For policymakers and standard setters, the evidence indicates that current S-metrics underweight, or even fail to account for, information that can be extracted from employees about their sentiment or concrete operational and compliance risk, which points to scope for monitoring bottom-up views.

Our research highlights several possible directions for future research. While the IV design and orthogonalization in our study mitigate endogeneity, additional identification strategies, such as regulatory shocks to workplace policies or quasi-experiments around discrete risk events, could further sharpen causal inference. Our measures rely on an anonymous review platform. Extending coverage to non-U.S. settings would test external validity. Finally, mapping these credit-market effects into portfolio and real outcomes will offer a path to quantify welfare and policy implications. Overall, bringing employee voice into credit valuation improves both statistical and economic performance, and it opens a scalable route for further accounting for the role of employees in firm valuation.

References

Abraham, K. G., & Mallatt, J. (2022). Measuring human capital. *Journal of Economic Perspectives*, *36*(3), 103–130.

Acharya, V. V., & Johnson, T. C. (2007). Insider trading in credit derivatives. *Journal of Financial Economics*, *84*(1), 110–141.

Aldatmaz, S., Ouimet, P., & Van Wesep, E. D. (2018). The option to quit: The effect of employee stock options on turnover. *Journal of Financial Economics*, *127*(1), 136–151.

Arora, N., Gandhi, P., & Longstaff, F. A. (2012). Counterparty credit risk and the credit default swap market. *Journal of Financial Economics*, *103*(2), 280–293.

Ashcraft, A. B., & Santos, J. A. (2009). Has the CDS market lowered the cost of corporate debt? *Journal of Monetary Economics*, *56*(4), 514–523.

Augustin, P., Jiao, F., Sarkissian, S., & Schill, M. J. (2020). Cross-listings and the dynamics between credit and equity returns. *The Review of Financial Studies*, *33*(1), 112–154.

Bai, J., & Wu, L. (2016). Anchoring credit default swap spreads to firm fundamentals. *Journal of Financial and Quantitative Analysis*, *51*(5), 1521–1543.

Baker, M., & Wurgler, J. (2002). Market timing and capital structure. *The Journal of Finance*, *57*(1), 1–32.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Fournal of Finance*, *61*(4), 1645–1680.

Bao, J., Pan, J., & Wang, J. (2011). The illiquidity of corporate bonds. *The Journal of Finance*, *66*(3), 911–946.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343.

Barillas, F., & Shanken, J. (2017). Which alpha? *The Review of Financial Studies*, *30*(4), 1316–1338.

Barillas, F., & Shanken, J. (2018). Comparing asset pricing models. *The Journal of Finance*, *73*(2), 715–754.

Barsade, S. G., & Gibson, D. E. (2007). Why does affect matter in organizations? *Academy of Management Perspectives*, *21*(1), 36–59.

Beer, M., & Eisenstat, R. A. (2004). How to have an honest conversation about your business strategy. *Harvard Business Review*, *82*(2), 82–89.

Bernanke, B. S., & Kuttner, K. N. (2005). What explains the stock market's reaction to Federal Reserve policy? *The Journal of Finance*, *60*(3), 1221–1257.

Blanco, R., Brennan, S., & Marsh, I. W. (2005). An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. *The Journal of Finance*, *60*(5), 2255–2281.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, *122*(4), 1351–1408.

Board of Governors of the Federal Reserve System. (2024). *Financial Accounts of the United States (Z.1)* [Accessed: 2025-10-17]. https://www.federalreserve.gov/releases/z1/

Boone, A., Starkweather, A., & White, J. T. (2024). The saliency of the CEO pay ratio. *Review of Finance*, *28*(3), 1059–1104.

Brunetti, C., Carl, M., Gerszten, J., Scotti, C., & Shin, C. (2024). Interconnectedness in the Corporate Bond Market. *Available at SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4821510

Campbell, D. W., & Shang, R. (2022). Tone at the bottom: Measuring corporate misconduct risk from the text of employee reviews. *Management Science*, *68*(9), 7034–7053.

Chava, S., & Roberts, M. R. (2008). How does financing impact investment? The role of debt covenants. *The Journal of Finance*, *63*(5), 2085–2121.

Chemmanur, T. J., Hu, G., & Wei, K. J. (2021). The role of institutional investors in corporate and entrepreneurial finance. *Journal of Corporate Finance*, *66*, 101833.

Coff, R. W. (1997). Human assets and management dilemmas: Coping with hazards on the road to resource-based theory. *Academy of Management Review*, *22*(2), 374–402.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, *1*, 4171–4186.

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, *57*(5), 2113–2141.

Donangelo, A., Gourio, F., Kehrig, M., & Palacios, M. (2019). The cross-section of labor leverage and equity returns. *Journal of Financial Economics*, *132*(2), 497–518.

Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance*, *65*(6), 2213–2253.

Ed, O., & Andrew, R. (2022). *Beware of Junk Employee Engagement Metrics for ESG Reporting* [Accessed: 2025-10-17]. https://www.gallup.com/workplace/392219/beware-junk-employee-engagement-metrics-esg-reporting.aspx

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, *101*(3), 621–640.

Edmans, A., García, D., & Norli, Ø. (2007). Sports sentiment and stock returns. *The Journal of Finance*, *62*(4), 1967–1998.

Eom, Y. H., Helwege, J., & Huang, J.-z. (2004). Structural models of corporate bond pricing: An empirical analysis. *The Review of Financial Studies*, *17*(2), 499–544.

Federal Reserve Bank of St. Louis. (2024). *L.223 Corporate Equities* [Accessed: 2025-10-17]. https://fred.stlouisfed.org/release/tables?eid=808373&rid=52

Fisher, A., Martineau, C., & Sheng, J. (2022). Macroeconomic attention and announcement risk premia. *The Review of Financial Studies*, *35*(11), 5057–5093.

Folloni, G., & Vittadini, G. (2010). Human capital measurement: A survey. *Journal of Economic Surveys*, *24*(2), 248–279.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377.

Friedman, A., & Lev, B. (1974). A surrogate measure for the firm's investment in human resources. *Journal of Accounting Research*, *12*(2), 235–250.

FTSE Russell. (2023). *SOLUTION OVERVIEW: ESG Scores and data model* [Accessed: 2025-10-17]. https://content.ftserussell.com/sites/default/files/ftse-esg-scores-overview-2023.pdf

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574.

Graham, J. R., Kim, H., Li, S., & Qiu, J. (2023). Employee costs of corporate bankruptcy. *The Journal of Finance*, *78*(4), 2087–2137.

Gredil, O. R., Kapadia, N., & Lee, J. H. (2022). On the information content of credit ratings and market-based measures of default risk. *Journal of Financial Economics*, *146*(1), 172–204.

Green, T. C., Huang, R., Wen, Q., & Zhou, D. (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics*, *134*(1), 236–251.

Guiso, L., Sapienza, P., & Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, *117*(1), 60–76.

Hand, J. R., Holthausen, R. W., & LEFTWICH*, R. W. (1992). The effect of bond rating agency announcements on bond and stock prices. *The Journal of Finance*, *47*(2), 733–752.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hirshleifer, D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, *58*(3), 1009–1032.

Hong, S., & Wei, X. (2025). Blockbuster or Bust? Silver Screen Effect and Stock Returns. *Review of Finance*, *29*(2), 603–632.

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806–841.

Huang, K., Li, M., & Markov, S. (2020). What do employees know? Evidence from a social media platform. *The Accounting Review*, *95*(2), 199–226.

Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *The Accounting Review*, *89*(3), 1083–1113.

Ichinowski, C., Shaw, K., & Prennushi, G. (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *American Economic Review*, *87*(3), 291–313.

Investment Company Institute. (2024a). *2024 Investment Company Fact Book: Quick Facts Guide* [Accessed: 2025-10-17]. https://www.icifactbook.org/pdf/2024-factbook-quick-facts-guide.pdf

Investment Company Institute. (2024b). *Ownership of Mutual Funds and Shareholder Sentiment* [Accessed: 2025-10-17]. https://www.ici.org/system/files/2024-10/per30-08.pdf

Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior & Organization*, *176*, 63–90.

Jorion, P., & Zhang, G. (2007). Good and bad credit contagion: Evidence from credit default swaps. *Journal of Financial Economics*, *84*(3), 860–883.

Kaplanski, G., & Levy, H. (2010). Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics*, *95*(2), 174–201. https://doi.org/10.1016/j.jfineco.2009.09.006

Koijen, R. S., & Yogo, M. (2023). Understanding the ownership structure of corporate bonds. *American Economic Review: Insights*, *5*(1), 73–91.

Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, *135*(2), 271–292.

Lajili, K., & Zeghal, D. (2006). Market performance impacts of human capital disclosures. *Journal of Accounting and Public Policy*, *25*(2), 171–194.

Lee, J., Naranjo, A., & Velioglu, G. (2018). When do CDS spreads lead? Rating events, private entities, and firm-specific information flows. *Journal of Financial Economics*, *130*(3), 556–578.

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265–3315.

Longstaff, F. A., Mithal, S., & Neis, E. (2005). Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market. *The Journal of Finance*, *60*(5), 2213–2253.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

LSEG. (2024). *Environmental, Social and Governance scores from LSEG* [Accessed: 2025-10-17]. https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf

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

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance*, *32*(4), 1151–1168.

MSCI Inc. (2024). *MSCI ESG Ratings Methodology* [Accessed: 2025-10-17]. https://www.msci.com/documents/1296102/34424357/MSCI%2BESG%2BRatings%2BMethodology.pdf

Saretto, A., & Tookes, H. E. (2013). Corporate leverage, debt maturity, and credit supply: The role of credit default swaps. *The Review of Financial Studies*, *26*(5), 1190–1247.

Schestag, R., Schuster, P., & Uhrig-Homburg, M. (2016). Measuring liquidity in bond markets. *The Review of Financial Studies*, *29*(5), 1170–1219.

Shan, C., & Tang, D. Y. (2023). The value of employee satisfaction in disastrous times: Evidence from COVID-19. *Review of Finance*, *27*(3), 1027–1076.

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35–55.

S&P Global. (2025). *S&P Global ESG Scores: Methodology* [Accessed: 2025-10-17]. https://portal.s1.spglobal.com/survey/documents/spglobal_esg_scores_methodology.pdf

Taddy, M. (2013). Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, *108*(503), 755–770.

Taddy, M. (2015). Distributed multinomial regression. *The Annals of Applied Statistics*, *9*(3), 1394–1414.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

U.S. Securities and Exchange Commission. (2024). *Securities and Exchange Commission Office of the Whistleblower Annual Report to Congress for Fiscal Year 2024* [Accessed: 2025-10-17]. https://www.sec.gov/files/fy24-annual-whistleblower-report.pdf

Widener, S. K. (2006). Human capital, pay structure, and the use of performance measures in bonus compensation. *Management Accounting Research*, *17*(2), 198–221.

A Bayesian Shrinkage Estimation

In our research, we use a robust methodology proposed by Bai and Wu (2016) . By leveraging the Merton (1974) structural model, we incorporate a comprehensive set of firm-specific fundamental characteristics, as well as non-financial indicators to investigate the additional explanatory power of Glassdoor information. This methodology involves a series of steps, including the conversion of distance-to-default measures into raw CDS valuations, the correction of valuation biases via local quadratic regression, and the integration of additional firm fundamentals, S score and Glassdoor information, using a Bayesian shrinkage method. The results are weighted average CDS valuation, which demonstrates superior cross-sectional explanatory power and stability over time.

A.1 Valuing CDS spreads based on firm fundamentals

To generate valuations on the five-year CDS spread, we start with the classic structural model of Merton (1974). Merton (1974) assumes that the total asset value (A) of a company follows a geometric Brownian motion with instantaneous return volatility σ_A, the company has a zero-coupon debt with a principal value D and time-to-maturity T, and the firm's equity (E) is a call option on the firm's asset value with maturity equal to the debt maturity and strike equal to the principal of the debt. We compute the distance-to-default measure from the Merton model using the total debt to market capitalization ratio and the stock return realized volatility as inputs:

$$\text{Distance to default} = \frac{\ln(\frac{A}{D}) + (r - \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}}. \tag{A.1}$$

To compute a firm's distance to default, we take the company's market capitalization as its equity value E, the company's total debt for the zero-coupon bond D, and the one-year realized stock return volatility as an estimator for stock return volatility σ_A. We further assume zero interest rates ($r = 0$) and fix the debt maturity at $T = 10$ years for all firms:

$$E = A \cdot N(d + \sigma_A \sqrt{T}) - D \cdot N(d), \tag{A.2}$$

$$\sigma_E = N(d + \sigma_A \sqrt{T}) \sigma_A A / E. \tag{A.3}$$

We solve for the firm's asset value A and asset return volatility σ_A from the two equations in (A.2) and (A.3) via an iterative procedure, starting at $A = E + D$. With the solved asset value and asset return volatility, we compute the standardized distance to default according to equation (A.1). Then we convert the distance-to-default into a raw CDS valuation based on a constant hazard rate assumption and a 40% recovery rate:

$$RCDS = -6000 \cdot \ln(N(d))/T, \tag{A.4}$$

where d is distance-to-default we calculated from equation (A.1), and $RCDS$ is a raw credit default spread measure based on Bai and Wu (2016), to retain the key contributions of the Merton model while avoiding its limitations in predicting actual defaults. The fixed 40% recovery rate is a standard simplifying assumption in the CDS literature. To the extent that the recovery rate can also vary across firms, this simple transformation does not capture such variation.

To explain the cross-sectional variation of market CDS observations, at each date we estimate the RCDS on the whole universe of chosen companies, and map the RCDS to the corresponding oobserved CDS via a cross-sectional local quadratic regression:

$$\ln(CDS) = f(\ln(RCDS)) + R, \tag{A.5}$$

where CDS denotes observed CDS from S& P Global, $f(\cdot)$ denotes the local quadratic transformation of the RCDS value, and R denotes the regression residual from this mapping. We use $RCDS$ rather than distance-to-defualt directly is because the transformation in (A.4) moves the distance-to-default measure closer to the observed CDS so that the local quadratic regression in (A.5) becomes more stable numerically. Meanwhile, the local quadratic regression has significant advantages in handling nonlinear relationships and data with complex structures over the ordinary regression. By fitting a quadratic polynomial near each data point, it can better capture local features of the data, providing more accurate and flexible fitting. And we use a Gaussian kernel for the local quadratic regression and set the bandwidth to twice as long as the default choice to reduce potential overfitting. Finally, we label the local-quadratic transformed Merton model-based CDS valuation as MCDS, $\ln(\text{MCDS}) = \hat{f}(\ln(\text{RCDS}))$.

Next, we use a long list of firm fundamental characteristics mentioned in section 2 that are not included in the Merton-based valuation but have been shown to be informative about a firm's credit spread. We use a Bayesian shrinkage method to combine the Merton-based valuation with the information from this long list of additional fundamental characteristics to generate a weighted average CDS valuation.

Formally, let F_t denote an $(N \times K)$ matrix for N companies and K additional credit-risk informative firm fundamental characteristics at date t. At each date, we first regress each characteristic cross-sectionally against MCDS to orthogonalize its contribution from the Merton prediction:

$$F_t^k = f_k(\ln(MCDS_t)) + x_t^k, \qquad k = 1, 2, \ldots\ldots, K, \tag{A.6}$$

where $f_k(\cdot)$ denotes a local linear regression mapping and x_t^k denotes the orthogonalized component of F_t^k, which means that after removing the part related to $MCDS$, the remaining part of each feature in the current layer is orthogonal to other features in the previous layer. This orthogonalization process can effectively reduce the multi-collinearity problem between different layers. And we use the local linear regression to accommodate potential nonlinearities in the relation further.

Second, we regress the Merton prediction residual, $R\_merton_t = ln(CDS_t/MCDS_t)$, cross-sectionally against each of the K orthogonalized characteristic x_t^k via another local linear regression:

$$R\_merton_t^k = f_k(x_t^k) + e_t^k, \qquad k = 1, 2, \ldots\ldots, 9. \tag{A.7}$$

Through this local linear regression, we generate a set of K residual predictions, $R\_m\hat{erton}_t^k, k = 1, 2, \ldots\ldots, K$, from the K fundamental characteristics. The two local linear regressions in (A.6) and (A.7) remove the potential nonlinearity in the relations and orthogonalize each characteristic's contribution to the original Merton valuation.

Third, we stack the K predictions to an N × K matrix, $X_t = [R\_m\hat{erton}_t^1, R\_m\hat{erton}_t^2, \ldots, R\_m\hat{erton}_t^K]$, and estimate the weight among them via the following linear cross-sectional relation:

$$R\_merton_t = X_t\, W_t + e_t, \tag{A.8}$$

where W_t denoting the weights on the K firm fundamental characteristics.

To perform the stacking regression in (A.8), we need all K predictions to be available. However, for a given company, it is possible that only a subset of the K characteristics, and hence only a subset of the K predictions, are available. We fill the missing predictions with a weighted average of the other predictions on the firm, where the relative weights are determined by the R-squared of the regressions in (A.7) for each available variable following the method by Bai and Wu (2016):

$$R\_merton_t^{i,j} = \sum_{k=1}^{\tilde{K}} w_t^k R\_merton_t^{i,k}, \tag{A.9}$$

$$w_t^k = e^\top (ee' + \text{diag}\langle 1 - R^2 \rangle)^{-1}, \tag{A.10}$$

where $R\_merton_t^{i,j}$ represents the missing residual prediction for firm i from the j-th variable at time t; $\tilde{K}$ denotes the subset of available residual predictions on the firm i; w_t^k represents the weight for the k-th variable at time t; e is a vector of error term from equation (A.7); and R^2 values are the values from regressions of each characteristic in equation (A.7). This weighting scheme is motivated by the Bayesian principle, where the prior prediction is set to zero, and the relative magnitude of the measurement error variance for each available residual prediction is proportional to $1 - R^2$. This method helps in managing missing data by effectively borrowing strength from available data while accounting for the reliability of the predictions based on their R^2 values.

Once the missing values are replaced by a weighted average, the time-t weightings (W_t) among the K predictions in equation (A.8) can be estimated in principle via a simple least square regression; however, to reduce the potential impact of multi-collinearity and to increase intertemporal stability to the weight estimates, we perform a Bayesian regression update by taking the previous day's estimate as the prior:

$$\hat{W}_t = (X_t^T X_t + P_{t-1})^{-1}(X_t^T R\_merton_t + P_{t-1}\hat{W}_{t-1}), \tag{A.11}$$

$$P_t = \mathrm{diag}\langle (X_t^T X_t + P_{t-1})\phi \rangle, \tag{A.12}$$

where ϕ controls the degree of intertemporal smoothness that we impose on the weights. We start with a prior of equal weighting and choose $\phi = 0.98$ for intertemporal smoothing.

In the final step to combine firm fundamental characteristics with Merton CDS, we add the weighted average prediction of the residual back to the MCDS valuation to generate a new CDS valuation, so that we isolate its effect from the factors that are already considered in the MCDS model. And we label the new value-adjusted CDS as FCDS:

$$ln(FCDS)_t = ln(MCDS)_t + X_t\hat{W}_t. \tag{A.13}$$

A.2 Valuing CDS Spreads based on ESG and Glassdoor ratings

To assess the additional explanatory power of S score and Glassdoor information, we extend our framework by sequentially incorporating S score and Glassdoor information. Each layer is orthogonalized to the prior information layers:

$$S_t = f(\ln(FCDS_t)) + s_t, \tag{A.14}$$

$$R\_fcds_t = f(s_t) + e_t, \tag{A.15}$$

$$\ln(SCDS)_t = \ln(FCDS)_t + R\_\hat{f}cds_t\hat{W}_t, \tag{A.16}$$

where $f(\cdot)$ denotes a local linear regression mapping and s_t denotes the orthogonalized component of S_t from MSCI ESG S pillar, with W_t denoting the weight on S score. Because our sample excludes firms without ESG scores, there is no need to adjust for missing values in S score, and we use the similar Bayesian regression update in equation (A.11) and (A.12).

Next, we apply the same process to Glassdoor information, which includes numeric ratings, Sentiment, and Risk-Related Buzz:

$$Glassdoor_t = f(\ln(SCDS_t)) + g_t, \tag{A.17}$$

$$R\_scds_t = f(g_t) + e_t, \tag{A.18}$$

$$\ln(GCDS)_t = \ln(SCDS)_t + \hat{R\_scds_t}\hat{W}_t, \tag{A.19}$$

where $Glassdoor_t$ denotes both numeric ratings and two written-review-based measures, and g_t denotes the orthogonalized component of $Glassdoor_t$ from Glassdoor employee reviews, with W_t denoting the weight on each components in the reviews. The final valuation, GCDS, incorporates fundamental, S score and Glassdoor information.

B Additional results

B.1 Validation of Risk-Related Buzz Extraction

To validate that our Risk-Related Buzz measure captures meaningful risk-related content from employee reviews, we examine the vocabulary with the highest MNIR loadings (ϕ coefficients) across the three review sections. Table B.1 presents representative terms strongly associated with subsequent credit risk events, along with illustrative examples from actual Glassdoor reviews.

The extracted vocabulary reveals several coherent themes related to organizational risk factors. In the Pros section (Panel A), terms like *induction*, *accommodation*, and *referral* appear with high positive loadings, suggesting that when employees emphasize formal processes and benefits in positive contexts, it may signal bureaucratic rigidity or compensation structures that correlate with future credit stress. Similarly, terms like *crisis* and *monotonous* in positive contexts may indicate underlying tensions.

The Cons section (Panel B) contains more explicit risk signals. Words like *comprehend* and *automated* appear in contexts describing complex business models and impersonal systems, potentially indicating operational opacity. Terms such as *nerve* and *unwillingness* directly reflect employee anxiety and organizational resistance to change—factors that may precede financial distress.

Most revealing is the Advice to Management section (Panel C), where employees offer constructive criticism. High-loading terms like *purse* (in contexts criticizing resource constraints), *understaffed*, and *suddenly* (describing abrupt organizational changes) provide direct insights into operational weaknesses that could manifest in credit markets. The appearance of *SOP* (Standard Operating Procedures) and *evolution* suggests employees perceive excessive bureaucracy or insufficient innovation.

The examples demonstrate that our MNIR approach successfully identifies linguistically nuanced risk signals without imposing ex ante assumptions about which terms should matter. The coherence of the extracted themes,spanning operational rigidity, resource constraints, management quality, and strategic direction, supports the construct validity of Risk-Related Buzz as capturing meaningful precursors to credit deterioration.

B.2 Descriptive results

Before turning to identification, we compare model–implied valuations with observed CDS to assess alignment in levels and comovement. Table B.2 contrasts the MCDS model with specifications that sequentially add fundamentals, S score, and Glassdoor information. Correlations with ln(CDS) rise monotonically as the information set is enriched. The MCDS correlates at 0.785, adding fundamentals raises the correlation to 0.866, and including S score increases it to 0.874. Incorporating Glassdoor information lifts the correlation further, to 0.881 when only numeric ratings are used and to 0.887 when two written-review-based measures are added.

We then summarize cross-sectional fit for the benchmark model, the MCDS model and the FCDS model. Table B.3 reports weekly R^2 for the three models. The mean R^2 is 56.86% for the benchmark, 60.74% with the Merton input, and 74.55% once fundamentals are included. Panel B reports average improvements: the MCDS model improves on the benchmark by 3.88 percentage points, and the FCDS model improves upon the MCDS model by 13.82 percentage points. Paired t-tests indicate that these gains are systematic. These results benchmark our evidence against Bai and Wu (2016), who study U.S. listed firms over 2003–2009 and document a large, systematic step-up when fundamentals are layered on a structural anchor. Our estimates show the same ordering and a comparable jump from the MCDS to the FCDS model, indicating that the explanatory power of fundamentals in anchoring CDS spreads is stable across samples and periods.

B.3 Overview of the IV design

This section documents how we construct and time our instrumental variables (IV) used to identify the causal component of the link between employee sentiment and CDS spreads. We (i) define two externally timed mood shocks, weekly blockbuster movie releases and global aviation fatalities, (ii) describe their data sources and sample coverage,

Table B.1: **Representative Vocabulary with High Risk-Related Buzz Loadings (ϕ)**

Vocabulary	ϕ (MNIR)	N_{docs}	Brief example
Panel A: Pros			
induction	3.617	16	Very good induction process.
accommodation	2.671	18	Some of the employee accommodations are favorable.
referral	1.596	32	Good discounts, job referral bonuses; Easy to hit referral goals.
bathroom	1.476	33	Clean bathrooms and break room; Wages are low but they have money to refurbish the bathrooms.
crisis	1.250	39	Assist associates and management when having personal crisis.
monotonous	1.124	17	The job is never monotonous or stale.
attain	0.906	29	The company celebrates the success that is attained.
driving	0.822	20	Hard-driving culture; CEO has the charismatic driving force.
basketball	0.776	25	On-site fitness center, gourmet cafeteria, basketball.
Sunday	0.580	48	Extra paid days off, always off on Sundays.
Panel B: Cons			
comprehend	3.856	16	The business model is "so unique" that one cannot possibly comprehend it without starting at the bottom.
compassionate	3.554	21	Upper management not very compassionate about family matters.
automated	3.121	16	The award/recognition system is almost "automated" and can feel impersonal.
abide	1.730	27	Management set poor examples, employees must abide by Core Values at all times.
nerve	1.208	34	RIFs and an up-and-down market make it nerve-wracking to know if you'll be employed.
shelf	1.137	54	Almost every year projects are thrown back on the shelf.
columbus	1.115	36	They want to stay equal to other places in Columbus instead of paying really well.
unwillingness	1.084	28	Slow to change; unwillingness to innovate.
bread	1.078	20	Recognized as our bread and butter; "Bread winner" pay is not enough.
residential	0.950	27	Business customers aren't going to sign up if residential support is bad.
Panel C: Advice to Management			
purse	5.569	16	Keeping a tight purse hurts opportunities for employees to develop.
attend	1.262	65	Make sure all managers attend trainings on best practices.
selfish	1.118	30	Be a team with a common purpose instead of a selfish goal.
laugh	1.066	27	Take leadership more seriously otherwise initiatives become a laughing stock.
SOP	1.056	30	Reduce non-necessary procedures and SOP to increase efficiency.
enthusiastic	1.033	24	Managers should be enthusiastic and informative.
tremendously	0.853	13	It would help tremendously to have more techs working.
evolution	0.853	20	Be bolder about product evolution.
understaffed	0.844	42	Branches are so understaffed that managers must run a teller window and the platform.
suddenly	0.833	13	Don't suddenly fire/retire 75% of senior technical leaders.

Notes: This table lists representative tokens with high MNIR loadings on the Risk-Related Buzz extracted from Glassdoor written reviews. ϕ is the token-level coefficient from the MNIR projection (higher ϕ indicates a stronger association with risk-related content). N_{docs} counts distinct reviews in which the token appears within the firm-quarter level documents. Section indicates the review segment (pros, cons, or advice). Example excerpts are lightly truncated for space.

Table B.2: **Summary Statistics and Correlation with Market CDS for Alternative Valuation Measures**

Variable	Mean	Std. dev.	Min	Max	Correlation
ln(CDS)	4.426	0.883	0.883	9.175	1.000
ln(MCDS)	4.426	0.700	3.131	8.646	0.785***
ln(FCDS)	4.455	0.756	2.713	8.646	0.866***
ln(SCDS)	4.411	0.756	2.732	8.684	0.874***
ln(SCDS + Numeric Ratings)	4.442	0.761	2.661	8.622	0.881***
ln(GCDS)	4.416	0.762	2.683	8.649	0.887***

Notes: This table reports the distributional characteristics and Pearson correlations between market CDS spreads (log-transformed) and a series of model-implied CDS valuations. The alternative measures include MCDS, FCDS, SCDS, and Glassdoor-adjusted CDS using both numeric ratings and written-review-based measures. All statistics are computed on the pooled panel of weekly firm-level observations. Correlation coefficients reflect the alignment between each valuation model and observed market CDS spreads, with significance levels denoted by *** for p-values below 0.01.

Table B.3: **Cross-Sectional Explanatory Power of Alternative CDS Valuation Models**

Panel A: Cross-sectional R^2

Variable	**Benchmark**	**MCDS**	**FCDS**
Mean	56.86%	60.74%	74.55%
Std. dev.	0.050	0.054	0.048
Min	43.26%	48.71%	58.05%
Max	67.58%	73.18%	82.45%

Panel B: R^2 differences

Variable	**MCDS − Benchmark**	**FCDS − MCDS**
Mean	3.88%	13.82%
Std. dev.	0.040	0.042
Min	-7.34%	4.03%
Max	15.50%	24.52%
Paired t-test	22.4581 ($p < 0.0001$)	78.2317 ($p < 0.0001$)

Notes: This table reports the distributional characteristics of the cross-sectional R^2 from weekly regressions of log observed CDS spreads on different valuation models. Panel A presents the summary statistics of R^2 across the benchmark model, the MCDS model, and the FCDS model. Panel B shows the pairwise differences in R^2 between the benchmark and MCDS models, and between the MCDS and FCDS models. Statistics are computed from the pooled panel of firm-week observations.

and (iii) summarize identification logic and diagnostic tests. Diagnostic tests for instrument validity, including under-identification, weak-identification, and over-identification, are all passed. Detailed first- and second-stage estimates and test statistics are reported in the main content in Table 8.

In order to address potential endogeneity in the relationship between employee sentiment and CDS spreads, we use two externally timed mood shocks: the number of blockbuster movie releases (*Release*) and global aviation fatalities (*Fatalities*). Blockbuster releases generate positive sentiment shocks by enhancing general mood and enjoyment, whereas aviation fatalities induce immediate negative emotional reactions through widespread media coverage and heightened anxiety.

We utilize weekly counts of blockbuster movie releases from 2012 to 2023 as the first instrument, sourced from Box Office Mojo. Hong and Wei (2025) find a significant positive correlation between blockbuster movie releases and U.S. stock market returns in the subsequent week, consistent with films improving investors' mood through enjoyment and escapism. In the same spirit, we posit that major releases can lift employees' affect and shift reported workplace sentiment, while not directly affecting an individual firm's valuation outside the entertainment sector. Following Hong and Wei (2025), we define a blockbuster as a film released in at least 4,000 theaters. Under this threshold, the sample contains 169 unique movies, about 14 per year.

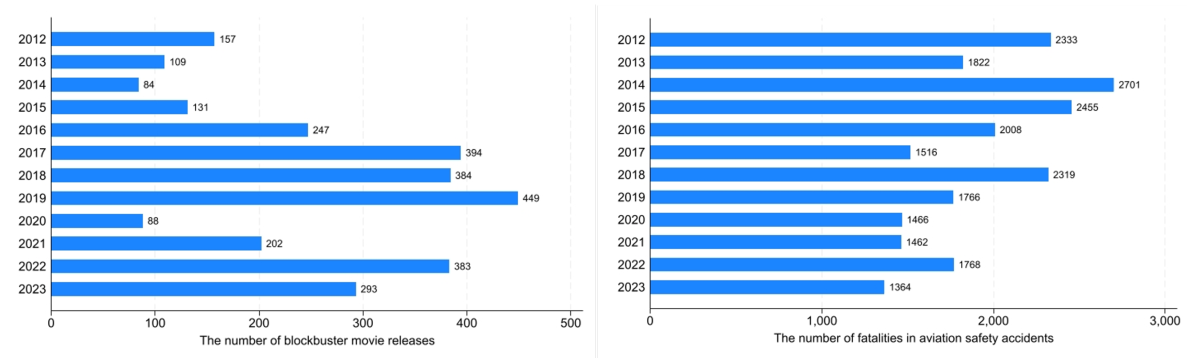


Figure B.1: **Annual Frequency of Exogenous Sentiment Shocks Used as Instruments**

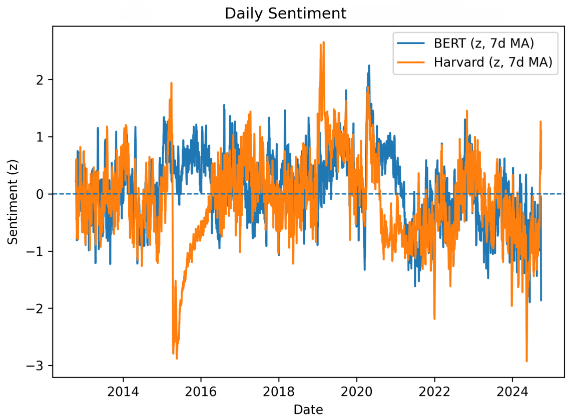
Notes: This figure shows the annual counts of two instrumental variables used to identify exogenous variation in employee sentiment: the number of blockbuster movie releases (left panel) and the number of fatalities in aviation safety accidents (right panel). These instruments proxy for positive and negative shifts in public mood, respectively, and are employed in the instrumental variable strategy to address endogeneity in the relationship between Glassdoor sentiment and CDS spreads. The time variation in these events provides plausibly exogenous shocks that influence employee sentiment without directly affecting firm fundamentals.

Figure B.1 shows the number of blockbuster releases per year, revealing COVID-19's transient suppression (2020–2021) followed by recovery, with a mean of 2 movies per week in 2020 and 5 for the whole sample. We align week-t movie releases with week-t+1 Sentiment to capture the natural delay in mood propagation, which means that films are typically viewed on weekends, and the affect persists into workdays. This one-week lag reduces contemporaneous contamination of firm operations while isolating sentiment transmission. For robustness, we re-run the IV tests on the 2012–2019 subsample; results are not significantly different from the full sample.

The second instrument uses weekly global aviation fatality counts from the Aviation Safety Network. Prior research shows that aviation disasters trigger immediate and widespread public anxiety (Kaplanski & Levy, 2010). These emotional responses extend beyond passengers or the aviation industry to the broader public, including employees across sectors. Such events are unexpected, externally driven, and independent of firm-specific conditions outside aviation and closely linked industries. They therefore provide plausibly exogenous negative mood shocks suitable for identifying variation in employee sentiment unrelated to firm fundamentals. Figure B.1 also plots annual fatality totals during our sample, highlighting both periods of relative stability and spikes associated with major incidents. On average, there are 28 fatalities per week, indicating a recurring source of emotionally salient events that can influence public and workforce sentiment.

B.4 Robustness Check

This section establishes robustness along two dimensions: (i) sentiment measurement and (ii) event labeling. First, we benchmark the transformer-based (BERT Model) sentiment series against a dictionary-based (Harvard Dictionary) alternative to verify directional agreement while documenting systematic differences in volatility and amplitude. Second, we re-define stress episodes using alternative anchors (high-frequency 50 bps jumps vs. sustained $\geq$100 bps persistence) and trace their implications for the MNIR Risk-Related Buzz. Across both robustness checks, the qualitative conclusions of the analysis in our main content are unchanged, which we show in the Table 11 and Table 12 in the main content.



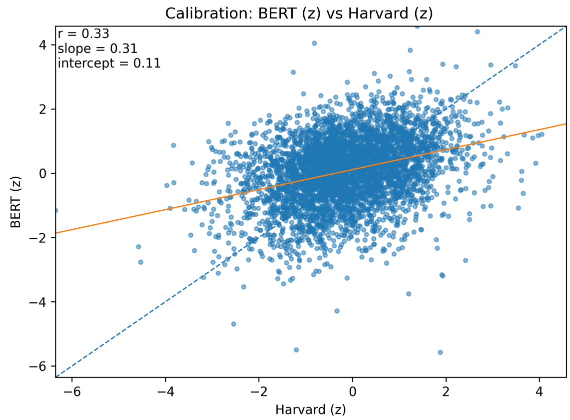


Figure B.2: **Daily Employee Sentiment: BERT vs Harvard (Time-Series Comparison)**

Notes: This figure compares daily employee sentiment extracted from Glassdoor reviews using a transformer-based model (BERT) versus a dictionary-based measure (Harvard). For each calendar day, we aggregate review-level scores and apply a 365-day rolling standardization to obtain z-scores; the plotted series are 7-day moving averages. Review-level outliers are winsorized at the 1st/99th percentiles, and days with fewer than a minimum number of reviews are excluded to mitigate small-sample noise.

Figure B.3: **Calibration of Alternative Sentiment Measures: BERT (z) vs Harvard (z)**

Notes: This figure plots the cross-method calibration between the dictionary-based Harvard sentiment (x-axis) and the transformer-based BERT sentiment (y-axis), using contemporaneous daily z-scores constructed via a 365-day rolling mean–variance standardization. Points represent daily aggregates after review-level winsorization (1%/99%); the 45° line (grey dashed) indicates equality, and the solid line is the fitted OLS relation reported on the plot (intercept and slope).

We show Figures B.2 and B.3 for two reasons: first, to document that the BERT- and Harvard-dictionary-based sentiment measures move in broadly the same direction over time, which justifies using the dictionary series as a robustness check; second, to illustrate that the dictionary approach is more volatile, supporting the choice of the transformer model for the main analysis.

In Figure B.2, both lines are seven-day smoothed daily z-scores, so the vertical axis can be read as a deviation from each firm-day's recent history. The two series comove around zero, rise into 2018–2020, and ease thereafter, indicating agreement on the cycle phase of workplace sentiment. At higher frequency, the Harvard line displays deeper troughs in downswings, consistent with greater tail sensitivity to negative terms, whereas the BERT line compresses extremes and is therefore more stable. Figure B.3 provides the corresponding cross-method calibration by regressing BERT z on Harvard

z using contemporaneous daily aggregates. The fitted relation has a slope near 0.31, a small positive intercept around 0.11, and a correlation of about 0.33. Reading the figure, points lie broadly along an upward-sloping cloud below the 45° line, which shows that the two measures align in direction but differ in amplitude, with BERT attenuating movements that Harvard records more strongly, especially on the left tail.

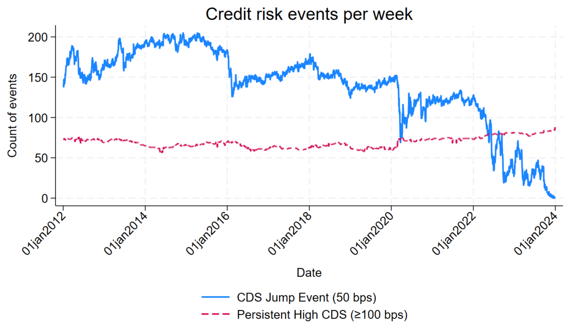


Figure B.4: **Credit-Risk Event Definitions: Time-Series Counts (50 bps Jump vs 100 bps Persistence)**

Notes: This figure compares the frequency of two alternative credit-risk event definitions used to anchor the MNIR-based Risk-related buzz indicator. Following Acharya and Johnson (2007), at each date t we assign an event label by looking forward from t to the sample end. The Jump series (solid line) flags t if there exists a trading day $s \geq t$ on which the firm's five-year CDS increases by more than 50 bps relative to $s-1$. The Persistence series (dashed line) flags t if the CDS level remains at or above 100 bps on all trading days from t through the sample end. For each date, event counts are aggregated across firms.

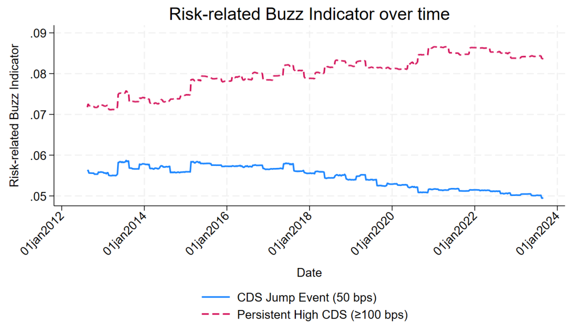


Figure B.5: **Risk-Related Buzz (MNIR) Under Alternative Anchors: Cross-Sectional Means Over Time**

Notes: This figure plots the cross-sectional average of the MNIR-based Risk-related buzz indicator under two anchoring schemes for event labels. The solid line uses the original 50-bps Jump definition (an event at t if $\exists\, s \geq t$ with a next-day CDS increase > 50 bps), while the dashed line uses the Persistence definition (an event at t if the firm's CDS stays ≥ 100 bps on all trading days from t to the sample end). For each date, we compute the firm-level Risk-related Buzz indicator given the specified anchor and then take the cross-sectional mean.

We show Figure B.4 and Figure B.5 in order to document that our alternative event labels select different but economically reasonable states without altering the main conclusions. Following the forward-labeling logic in Acharya and Johnson (2007), in our main analyses, we adopt a jump anchor that labels week t as a risk event if, looking forward from t, there exists at least one trading week s ($s \geq t$) on which the firm's five-year CDS increases by more than 50 bps relative to the prior week $s-1$. As a robustness check, we use a persistence anchor that flags week t if the firm's CDS level remains at or above 100 bps on every trading week from t until the end of the sample. Figure B.4 reports weekly counts under the two anchors: the 50 bps jump series is higher and more volatile early on and declines after 2020, whereas the $\geq$ 100 bps persistence series is lower, smoother, and drifts up, indicating episodic stress versus sustained weakness. Figure B.5 plots the corresponding cross-sectional means of the MNIR Risk-Related Buzz: the persistence-anchored series lies above and steps up gradually, while the jump-anchored series is lower and trends down, reflecting the different event sets rather than a change in methodology.