Credit Information in Earnings Calls*

Harry Mamaysky† Yiwen Shen‡ Hongyu Wu§

August 29, 2023

Abstract

We develop a novel technique to extract credit-relevant information from the text of quarterly earnings calls. This information is not spanned by fundamental or market variables and forecasts future credit spread changes. One reason for such forecastability is that our text-based measure predicts future credit spread risk and firm fundamentals. More firm- and call-level complexity increase the forecasting power of our measure for spread changes. Out-of-sample portfolio tests show the information in our measure is valuable for investors. Our results suggest that investors do not fully internalize the credit-relevant information contained in earnings calls.

Keywords: Corporate credit, credit default swaps, return forecasting, NLP

JEL Codes: G11, G12, G14

Online Appendix: https://sites.google.com/view/hmamaysky/research

*We thank Amit Goyal, Xu Guo, Hai Lin, and Avanidhar Subrahmanyam, as well as seminar participants at Balyasny Asset Management, Bank of America, Columbia Business School, CUNY, Vanguard, and the Wolfe NLP/ML Conference for helpful suggestions.

†Columbia Business School, hm2646@columbia.edu.

‡HKUST Business School, yiwenshen@ust.hk.

§Yale School of Management, hw499@yale.edu.

1 Introduction

The U.S. corporate bond market is large and growing, both in absolute terms and relative to GDP, and represents one of the key sources of capital for U.S. corporations.

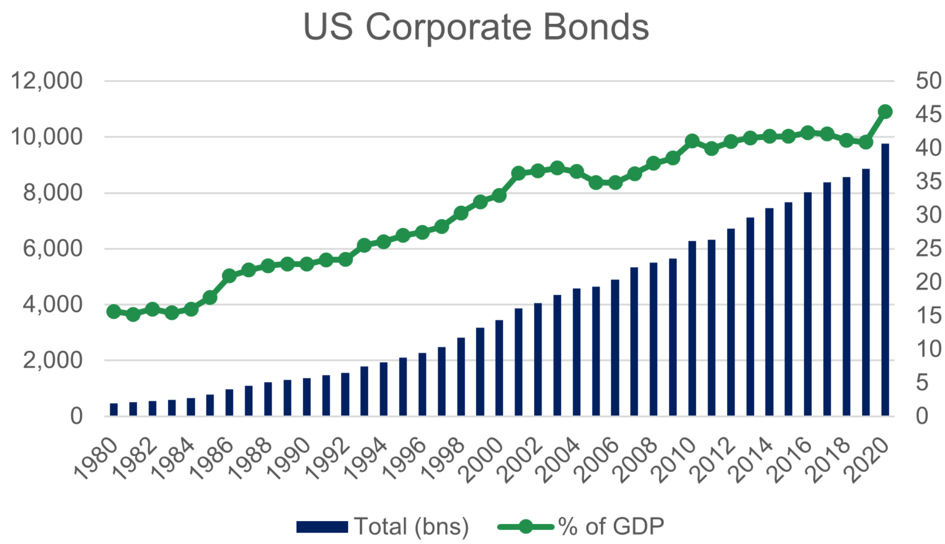

Fig. 1. U.S. corporate bonds amount outstanding in billions (from SIFMA) and as a percent of GDP (using nominal GDP data from FRED).

Corporate bond prices reflect investor assessments of a firm's current and *future* default risk. However, many traditional credit measures, such as debt ratios or firm profitability, are more informative about current rather than future credit risk. Market-based measures of credit risk, like credit spreads or implied volatilities, are, of course, forward looking, but contain other confounding influences, like risk or liquidity premia.

An important source of forward-looking credit information for market participants is communication with management teams. In addition to the publicly disclosed financial metrics that are available from 10-Ks and 10-Qs, management teams convey to investors their thinking about how a company's leverage and balance sheet will appear in the future. Earnings calls represent one – and perhaps the most – important channel for regular communication between investors and management teams. Earnings calls generally occur on a quarterly basis and last between one and one and a half hours. Given that calls are both infrequent and relatively short, if management teams or investors choose to discuss information relevant to a company's credit risk, it is likely because such information is important. In this paper, we show that quarterly earnings calls contain valuable information about the future pricing of credit risk, information that is not already reflected in either credit spreads or in other firm characteristics which have been shown in prior work to forecast corporate bond returns and risk.

One of the challenges with identifying credit-relevant portions of earnings calls is that call transcripts often run into the dozens of pages, many of which contain little credit-relevant information. To address this, we identify a set of credit-related words and phrases, and focus our analysis on the portions of the earnings call transcripts that are in the vicinity of mentions of these credit-related terms. Our sample contains all U.S. earnings calls from 2009–2020. Most of these, and virtually all recent ones, have at least one credit-relevant portion (see Panel A of Figure 2). Using natural language processing (NLP) techniques, we calculate an implied credit spread for each earnings call. The implied credit spread is obtained by regressing (a measure related to) firms' credit default swap (CDS) spreads on the count of words, bigrams and trigrams (collectively, *tokens*) that occur in the vicinity of credit mentions in earnings calls,[1] and then applying the estimated model to the token count of a particular earnings call.

Because of the high dimensionality of this problem – there are tens of thousands of tokens – we use a novel, computationally efficient approach to select the subset of tokens with the highest explanatory power for CDS spreads. For each token, we regress the CDS level prevailing immediately after each call on the number of times that token appears in the credit-relevant portion of the call. We first select the token whose count leads to the highest regression R^2. We then recursively regress the residual from the prior step on each of the still-unselected token counts to find the one with the highest R^2. We implement the above procedure for the panels of investment grade (IG) or high-yield (HY) CDS spreads separately because the credit-relevant language used in IG and HY calls is distinct. Our NLP methodology uses either the full-sample of data or rolling subsamples. In the latter case, we use no forward-looking information in the construction of the implied credit spread. The computational efficiency of our methodology – which results from not having to run multivariate regressions – allows us to conduct fully out-of-sample rolling analysis where new tokens are re-selected in each successive subsample.

We then run a lasso regression of post-call CDS spreads on the selected, high explanatory power tokens. The lasso, or least absolute shrinkage and selection operator, is a modified regression which minimizes the mean squared model error while penalizing the sum of the absolute values of the model coefficients. It is an efficient dimensionality reduction technique; see Hastie, Tibshirani, and Friedman (2009) for details. We include sector dummies in the lasso, and exclude tokens that are concentrated in a specific sector (e.g., "oil wells"). The coefficient estimates from this regression allow us to associate

[1]We discuss our reasons for using CDS rather than corporate bond data below.

words and phrases with either better (lower spreads) or worse (higher spreads) credit news. We do not have to assign an a priori tone to any token as the algorithm learns this tone endogenously. Furthermore, the rolling (out-of-sample) version of our analysis allows the credit tone of tokens to vary over time.

Our measure of the information content of earnings calls is called the *credit score*, defined as the difference between a firm's actual credit spread immediately following the earnings call and the credit spread implied by the lasso model. In the full-sample model, this is just the residual of the lasso regression; in the rolling model, this is the out-of-sample forecast error. A higher credit score indicates the market trades at wider spreads than suggested by the language of earnings calls, and a lower credit score indicates the opposite.[2] The credit score reflects the degree of disagreement between the market's and management team's assessment of corporate creditworthiness.

We show that the lagged credit score negatively forecasts 12-month ahead changes in CDS spreads, even after contemporaneous changes in interest rates, implied volatilities, and firm leverage, as well as a large set of other variables, which we discuss below, are included as controls. Following the logic of Collin-Dufresne, Goldstein, and Martin (2001), the contemporaneous regressors in this specification capture the influences of a Merton (1974)-type model, and thus provide a stringent test of the forecasting power of lagged credit scores for future spread changes. In a pure forecasting regression, after the contemporaneous regressors have been dropped from the right-hand side, the credit score remains a significant and negative forecaster of future CDS changes. Not surprisingly, without the contemporaneous controls, the credit score effect becomes larger. These results hold whether credit score is calculated using the full text sample or in rolling windows. Since the rolling credit score would have been known to investors in real time, this suggests that credit scores contain valuable out-of-sample credit information, something we investigate further below. The forecasting results continue to hold for six-month ahead CDS changes, using both the full-sample and rolling text models.

The evidence thus strongly suggests that implied CDS spreads contain valuable information for forecasting future credit spreads that is not already impounded into post-call CDS levels. To understand this result, we show that credit scores also contain information about future CDS market risk and future corporate fundamentals. Higher credit scores forecast lower CDS risk – across a variety of measures – over the subsequent year. Furthermore, higher credit scores forecast positive changes in future corporate profitability and declines in firm leverage. Importantly, these findings are consistent with the neg-

[2]It is common to refer to decreasing (increasing) credit spreads as *tightening* (*widening*) spreads.

ative forecasting coefficient of credit score for future CDS changes: lower risk, higher profitability, and lower leverage are associated with lower future CDS spreads.

There are two potential channels that can explain the ability of credit scores to forecast CDS spread changes. First, it is possible that market participants are fully aware of the information content of credit scores for future risk and profitability, but that CDS still rationally responds with a lag. Unlike stock prices, which immediately respond to all future anticipated changes in cash flows and discount rates, CDS contracts have a fixed maturity, and so the five-year fixed maturity CDS spread may evolve predictably as a firm's fundamentals slowly change. For example, a firm that plans to delever over the next several years will see a lower CDS spread immediately, but may be expected to see an even lower five-year CDS in one year, and a lower one yet in two years, as credit risk for the company continues to fall and each successive five-year CDS contract thus reflects a lower average credit risk over its life. That credit scores forecast future credit risk and firm fundamentals is consistent with this explanation. The alternative explanation is that market participants do not fully respond to the credit-relevant information content of earnings calls because they are capacity constrained (as in Sims 2011) and do not fully internalize all relevant information. We refer to the two channels as the delayed rational response and the capacity constrained investors hypotheses, respectively.

Under the rational delayed response hypothesis, characteristics that may proxy for call or firm complexity should not impact the degree of predictability from credit scores to future CDS changes. Under the capacity constrained investor hypothesis, more complex calls and calls about more complex firms should result in greater predictability from credit scores to future CDS spread changes. Furthermore, under the rational delayed response hypothesis, knowing the credit score of a firm should not lead to profitable trading strategies because the forecasted CDS spread change is a publicly known, rational reflection of anticipated future changes in firms' risks or fundamentals.

To identify whether complexity amplifies the forecasting power of credit scores, we follow the accounting literature and modify our empirical specification to interact credit score with measures that proxy for the informational environment: the dispersion of analyst earnings forecasts, the number of analysts covering a given firm, a measure of the language complexity of the earnings call itself, and the length of the earnings call transcript.[3] We find that more analyst coverage, which proxies for firm complexity, and

[3]Bhushan (1989) finds that more analysts are associated with larger firms, that have higher stock return volatility and more institutional ownership. Lehavy, Li, and Merkley (2011) show that firms with less readable communications are associated with more analyst coverage. You and Zhang (2009) and Loughran and McDonald (2020) interpret the length of 10-K's as a proxy for either complexity or lack of

greater call transcript length, which proxies for call complexity, increase the predictive power of credit score, by making the credit score coefficient for future spread changes even more negative. The other two interactions also lead to an increase in the magnitude of the credit score coefficient, but the results are not significant. We interpret these results as evidence supporting the capacity constrained investor interpretation.

To assess the practical economic impact of credit scores, we turn to an out-of-sample analysis using the rolling text model, with both the token selection and lasso stages done without using future information. The output of the rolling text model would have been available to market participants in real time. We form long-short credit portfolios that contain the maximally mispriced firms in every month, while maintaining a zero credit exposure (in a sense to be explained in Section 6). Maximally mispriced means that the long side contains firms with the highest credit scores (whose credit spreads are forecasted to tighten) and the short side contains the credits with the lowest credit scores (that are forecasted to widen). Our construction generates a long-short portfolio that isolates differences in credit scores while maintaining minimal overall credit exposure. This portfolio construction, as well as a modified version of our forecasting regression which includes firm fixed effects, point to the importance of cross-sectional variation in the relationship between credit scores and future CDS market outcomes. More general trading strategies may take into account other characteristics shown in the literature to forecast returns, but our approach zeros in specifically on the information content of credit scores.

We evaluate the performance of different parameterizations of this strategy against portfolio simulations conducted under the null of no predictability, which provides a natural benchmark against which to compare our backtested returns. We find that different parameterizations of our trading strategies generate returns that systematically outperform the simulated distribution of these statistics under the no-predictability null for both IG and HY firms. The outperformance is statistically significant and economically large, as the strategies can add 3 to 4% of annualized return to long-only corporate bond portfolios without increasing their credit betas, which in the context of the credit asset class is very large economic effect. This provides further evidence in support of the capacity constrained investor hypothesis, and suggests that the information content of earnings calls would have been a valuable, real-time tool for credit investors.

---

readability.

1.1 Relationship to the Literature

Collin-Dufresne, Goldstein, and Martin (CGM, 2001) focus on explaining changes in credit spreads. They show that several factors suggested by the Merton (1974) model – changing interest rates, stock returns, and changes in implied volatility – can only explain 25% of monthly spread changes of corporate bonds. They show that an aggregate credit factor accounts for much of the variation in the model residuals, and conjecture these movements are caused by common supply-demand conditions in corporate bond markets. Ericsson, Jacobs, and Oviedo (2009) perform an analysis similar to Collin-Dufresne et al. (2001) but use CDS data. They argue that CDS data are a cleaner measure of credit risk than corporate bond spreads, and find that the CGM factors explain a similar mid-20% of the variation of CDS spread changes. While the residuals from the CDS version of the analysis did not have the pronounced common factor found by CGM in corporate bond data, much of the variation in CDS spread changes still went unexplained.[4] Bao, Pan, and Wang (2011) show that illiquidity explains a good deal of time series and cross sectional variation in corporate bond spreads from 2003 to 2009, and its explanatory power rises during the Global Financial Crisis, lending support to the conjectured CGM mechanism. Our results suggest that a portion of credit spread change residuals can be explained by our forward-looking credit score measure, and other control variables introduced since CGM. In our most complete specification in Table 7, the R^2 rises to 38.7%.

There is a large, and growing, literature on the pricing of corporate credit risk. Guo, Lin, Wu, and Zhou (2021, GLWZ) show that corporate bond sentiment is an important forecaster of future returns on corporate bonds. They measure bond sentiment as minus the difference between a bond's current credit spread and the credit spread implied by a fair-value model estimated in rolling windows. High sentiment negatively predicts future returns, and the opposite holds for low sentiment. Bond portfolios that short high- and long low-sentiment bonds earn high risk-adjusted returns. Our credit score-based trading strategy has a similar flavor but uses earnings call implied spreads as the fair-value benchmark. While we intentionally focus only on variation in credit score to isolate the value of this information, our trading strategy can be modified to take into account information from other forecasting variables by using a rolling fair-value approach as in GLWZ.

Bali et al. (2022) apply the dimensionality reduction methods of Gu, Kelly, and Xiu (2020) to a large number of bond characteristics, and show that the most important predictors of month-ahead corporate bond returns are liquidity and downside risk, while bond

[4]Ericsson et al. (2009) and Campbell and Taksler (2003) show there is considerably more explanatory power for the *levels* of CDS spreads or corporate bond yields, as opposed to spread changes.

duration and past returns also matter.[5] They find that imposing dependence between bond and stock characteristics, as suggested by the Merton (1974) model, improves the forecasting performance of pure machine learning approaches. Bali, Subrahmanyam, and Wen (2021) show that corporate bond losers over the past 36 months tend to outperform past corporate bond winners. Bartram, Grinblatt, and Nozawa (2020) document another mean-reversal pattern in corporate bonds, proxied for by the bond book-to-market ratio, defined as the bond price over the bond's face value (and closely related to PVLGD). Chung, Wang, and Wu (2019) study the impact of volatility on the cross-section of corporate bond returns, and show bonds that hedge volatility increases have low expected returns. Cao et al. (2022) find that corporate bonds with large increases in implied volatility over past month have lower future returns relative to bonds with decreases in implied volatility. Bai, Bali, and Wen (BBW, 2019) show that downside risk, measured as the 5th percentile monthly return on a corporate bond over the prior 36 months, is a priced risk factor, with higher historical downside risk forecasting higher future returns. They also find a liquidity premium in corporate bonds, and a short-term reversal effect. Bai, Bali, and Wen (2021) show that the systematic risk implied by the BBW (2019) factor model is priced in the cross section of bond returns, whereas idiosyncratic risk relative to their factor model is not. In addition, corporate bond market returns are predicted by lagged corporate bond return variance. Kelly, Palhares, and Pruitt (2021) use instrumented principal component analysis (Kelly et al. 2020) to jointly estimate factors capable of explaining the cross-section of corporate bond returns and the time-varying loadings of bonds on these factors. They find that the IPCA model is closely approximated by a static five-factor model consisting of the bond spread, bond volatility, duration, and value long-short factors, as well as an equal-weighted corporate bond portfolio.

Relative to the rest of the corporate credit forecasting literature, our key contribution is to systematically capture credit-relevant information from corporate earnings calls and show that this information is a useful forecaster of future credit spread changes, CDS market risk, and firm fundamentals, even when controlling for an extensive set of other corporate bond forecasting variables, both markets- and fundamentals-based, suggested in the prior work cited above. Our results show that earnings calls are an important source of forward-looking credit information, which is not spanned by other predictors of corporate bond returns and risk, and which is not immediately reflected in market prices.

Our NLP methodology is similar to a recent literature that, rather than specifying word tone a priori, seeks to extract the tonality of words by using market data. A related

[5] van Binsbergen and Schwert emphasize the importance of duration-matching for return measurement.

approach to ours is Manela and Moreira (2017) who use a support vector regression (a close cousin of the lasso) to estimate a mapping from the text of Wall Street Journal articles to the level of the VIX index. Other related papers that use market responses to extract word tone are Jegadeesh and Wu (2013), Ke, Kelly, and Xiu (2021), Garcia, Hu, and Rohrer (2022), and Calomiris, Harris, Mamaysky, and Tessari (2022). Our application of these techniques to the analysis of credit markets is novel in the literature.

We contrast our study with the recent, independent work of Donovan et al. (2021), who construct a measure of credit risk by mapping information from firms' conference calls and the Management Discussion & Analysis sections of 10-Ks to CDS spreads. They show this information predicts future credit events including downgrades, bankruptcy, and the level of credit spreads on private deals. Our work differs from theirs in several important ways. First, we focus on the forecasting power of our credit score measure for *changes* in firms' CDS spreads, which are harder to predict than future spread levels, as we have already pointed out. Second, in assessing the forecasting power of credit score for future CDS spread changes, we control for the information content of current CDS spreads, credit ratings, and many other firm characteristics, whereas Donovan et al. (2021) explicitly focus only on firms where CDS spreads and credit ratings are not available. Third, we investigate potential channels for the forecasting power of our credit measure, and find evidence for both the delayed rational response and constrained investor hypotheses. Such an investigation is not possible in the Donovan et al. (2021) setting because they analyze firms with no traded CDS contracts, and thus have no source of market information to serve as a benchmark forecast. We also develop a portfolio strategy to test the out-of-sample economic significance of our measure. Methodologically, we propose a forward selection algorithm to identify tokens with high explanatory power for firms' credit conditions. This, and our focus on only the parts of earnings calls in the vicinity of credit terms, reduces the dimensionality of our model and facilitates its implementation in real time. Due to the differing focus and set of analyzed companies, our paper and the work of Donovan et al. (2021) are complementary.

The mechanism we document – that credit-relevant information from earnings calls is not fully absorbed by market participants – is similar to Cohen, Malloy, and Nguyen (2020), who show that quarter-over-quarter changes in the text of 10-Qs and 10-Ks predict negative future corporate fundamentals (lower earnings and profitability, higher credit risk) and negative stock returns, but with no announcement day effect. They conclude that investors are inattentive to changes in the language that firms release in their 10-Ks and -Qs. A similar finding was reported by You and Zhang (2009), who found a 12-month stock price drift following 10-K filings, suggesting that investor reaction to the information content of 10-Ks "seems sluggish." Furthermore, You and Zhang (2009) document that more complex 10-K reports are associated with greater underreaction.[6] We provide evidence that investors are similarly inattentive to the credit-relevant portions of earnings calls, and that this information forecasts future risk and fundamentals, and provides information that leads to economically meaningful out-of-sample trading performance.

The rest of the paper proceeds as follows. Section 2 describes our data set and control variable construction. Section 3 describes our text data and our credit score methodology. Section 4 presents the contemporaneous and pure forecasting regression results. Section 5 discusses mechanisms which can explain our findings. Section 6 evaluates our signal's out-of-sample performance. Section 7 concludes.

2 Data

We follow Ericsson, Jacobs, and Oviedo (2009) in using five-year CDS, rather than corporate bond, data in our analysis. Corporate bonds issued by the same company will often have idiosyncratic features – like change of control puts, different call schedules, or different levels of seniority – making comparisons across bonds difficult. Many corporate bonds are very illiquid, and the implicit assumption made in the literature that trading is possible at or near reported trade prices is frequently unwarranted.[7] CDS contracts are less sensitive to the idiosyncrasies of particular bonds because of their cheapest-to-deliver feature: the CDS (or protection) buyer has the right to deliver any of a class of bonds – presumably the cheapest – to the CDS seller in the case of a default event and be paid the bond's full face value. CDS prices, which we obtain from IHS Markit, are composites of end-of-day bid-offer quotes submitted by dealers, and are more reflective of market condi-

[6] Jiang et al. (2019) show that an index of aggregate manager sentiment based on the text of firms' 10-Ks, 10-Qs, and conference calls negatively forecasts stock returns, suggesting stock investors overreact to the information content of management communication. Reconciling the findings of underreaction to specific parts of earnings calls and 10-Ks with this overreaction result to the overall tone of corporate disclosures is an interesting area for future work.

[7] The Trade Reporting and Compliance Engine (TRACE), run by FINRA, reflects all transactions in U.S. corporate bonds. However, many of these transactions, especially those involving retail investors, often happen far away from the prevailing institutional prices. For example, a retail investor buying from a dealer will pay the dealer's offer price, which tends to embed a large transaction cost especially for small trades (Edwards, Harris, and Piwowar 2007). The implicit assumption made in much of the literature that a trading strategy can sell bonds at or near this transacted price overstates the actual profitability from trading corporate bonds. First, non-dealers cannot sell at dealer's offer prices – they can only buy there. Second, the available liquidity at many observed prices, even for a trading strategy willing to buy at this price, is likely minimal as a large fraction of all corporate bonds experience little secondary market trading. Finally, short selling corporate bonds is often not possible because of lack of borrow.

tions than bond trade prices, especially when the latter come from small trades. Oehmke and Zawadowski (2017) argue that speculative credit trading concentrates in CDS markets exactly because their standardization makes CDS contracts more liquid than the underlying corporate bonds.[8] Panel B of Figure 2 shows the number of CDS contracts in our sample over time.[9] We use five-year CDS levels because, as Ericsson, Jacobs, and Oviedo (2009) show, these represent the bulk of outstanding CDS contracts.

We transform CDS spreads to a measure we call *PVLGD*, which calculates the risk-neutral expected present value of the future losses ensured by the CDS contract. The quoted CDS spread S is relates to the PVLGD of a CDS contract via

$$S \times \mathrm{PV01} = \mathrm{PVLGD}. \tag{1}$$

The derivation of (1) is shown in Section A.1 of the Online Appendix. PV01 determines the risk-neutral expected present value of receiving a single basis point annuity over the life of the CDS contract, where the annuity stops paying in the case of default. With complete markets, (1) shows there are two equivalent ways of buying insurance against default via CDS. One is to pay the spread of S basis points over the life of the contract, or until default occurs. The other is to pay PVLGD upfront to the seller of protection. From the seller's point of view, these two income streams are equally valuable.

The PVLGD can most intuitively be interpreted as follows. Consider a five-year Treasury bond with a 3% coupon that trades at par. Now consider a risky corporate bond B, with the same five year maturity, the same 3% coupon, and a price of P_B. Since B is risky, P_B is less than \$100. But how much less? Consider a five-year CDS contract which references bond B and which trades a PVLGD of PV_B. Abstracting away from some modeling and institutional details, buying B and paying PV_B upfront to buy CDS protection provides the equivalent payout to that of the Treasury bond, and therefore

$$P_B + PV_B = \$100. \tag{2}$$

The PVLGD of a CDS can thus be interpreted as the discount from par of a corporate bond with the same maturity and coupon as a par Treasury, and with the same default risk as the CDS contract. In this sense, the PVLGD is very similar to the bond book-to-

[8] While CDS contracts have differing liquidity levels, the variation is less pronounced than for bonds, and dealers are willing to trade at least several million dollars notional at the stated bid-offer prices.

[9] Rarely CDS contracts leave our sample due to either default or M&A activity. Section A.2.4 of the Online Appendix argues that our CDS sample is nevertheless free of survivorship bias as CDS contracts anticipate future credit-relevant events *prior* to the underlying firms exiting the sample.

S	PV01	PVLGD
100	0.045	4.517
200	0.043	8.665
500	0.038	19.192
600	0.037	22.150
2500	0.020	49.619
2600	0.019	50.208

Table 1
This shows the mapping from CDS spread S (in basis points) to PV01 and PVLGD from (1) using assumptions detailed in Section A.1 of the Online Appendix.

market measure of Bartram, Grinblatt, and Nozawa (2020).[10]

Using PVLGD and not S in empirical analysis is important because of the large convexity in CDS spreads. The PV01 of a CDS contact falls quickly with S. Table 1 shows some representative CDS spreads (in basis points), and the associated PV01s and PVLGDs. From (1), to a first-order approximation, the change in the PVLGD of a CDS contract is given by $\Delta\text{PVLGD} \approx \text{PV01} \times \Delta S$. Since a spread increase of 100 basis points has a much larger PVLGD impact starting at a low spread than starting at a high spread (e.g., consider the PVLGD impact of a $100 \rightarrow 200$ spread move versus $2500 \rightarrow 2600$), changes in CDS spreads are a poor measure of the underlying default risk of a corporate bond in (2). Changes in PVLGD provide a much better measure. As we show in Table A.1 and Figure A.1 of the Online Appendix, our PVLGD methodology exactly matches the industry-standard dollar settlement calculation for CDS trades.

We use several sources of company-level data: balance sheet, income, and cash flow data from Compustat; equity price data from CRSP; implied volatility data from Option-Metrics; analyst data from I/B/E/S; and earnings call data from SP Global (the earnings call data are described in Section 3). For Compustat, the data *report dates* are unavailable for roughly half of the dataset. To avoid losing these observations, we timestamp Compustat data using the *data date* plus three months: we assume data date t observations are available only as of ($t + 3$ months) or after. This ensures that we do not use forward-looking information while allowing us to retain the majority of our data. Creating a mapping between all these datasets is an involved process. The mapping methodology and other data issues are discussed in detail in Section A.2 of the Online Appendix.

[10]The PVLGD of a CDS contract is analogous to the implied volatility of an option. It is a model-based transformation of a market price which renders the latter more interpretable. As we show in Section 4.3, replacing PVLGD with log CDS spreads – a model-free measure – leaves our results largely unchanged.

Markit CDS data are classified into ten sectors: basic materials, utility, financials, consumer services, technology, energy, consumer goods, industrial, telecommunications, and others. We drop all financials because many of the controls (discussed below) do not apply to them. Each firm-quarter observation is also classified into IG or HY using the firm's most recent *average rating* field from Markit.[11] IG includes AAA, AA, A, and BBB ratings, and HY includes the others (BB, B, CCC, D, unclassified). In total, there are 9830 monthly observations in the IG group and 4069 in the HY group.

To control for known determinants of corporate bond returns and credit spread changes, we construct an extensive set of predictor variables that have been proposed in the literature. These are summarized in Table 3, and more detailed information about their construction is available in Section A.3 of the Online Appendix. Table 4 contains summary statistics for these control variables. Figure 3 shows the cross-sectional average of firm-level correlation matrices of our control variables, as well as the PVLGD and credit score (defined in Section 3.2). With several exceptions, most correlations between explanatory variables are quite low, suggesting these capture distinct aspects of a firm's credit environment. While PVLGD has several moderate correlations with controls (e.g., it is lower for larger firms and higher for value firms and lower-rated firms), credit score is positively correlated with PVLGD but largely uncorrelated with all other controls.

3 Earnings Calls and Text Model Estimation

Our sample consists of 202,788 earnings call transcripts obtained from *S&P Global* that took place from January 2009 and December 2020. Each transcript undergoes several rounds of revisions, and we use the most recently available version of each transcript, which typically includes corrections to transcription errors that may have occurred in earlier versions.[12] We date an earnings call as having occurred on day t if its announcement date-time took place between 4 PM on day $t-1$ and 4 PM on day t.[13] Earnings calls that take place after 4 PM on day t are therefore dated as of day $t+1$. For our analysis, we need to match a firm's quarterly earnings call with its CDS in that quarter (see Section A.2 of the Online Appendix). Panel C of Figure 2 shows the number of earnings calls that

[11] Markit credit ratings and S&P ratings available from Compustat match very closely.

[12] SP Global delivers earnings calls transcript in four versions. Ranking by how soon a version is available after the call, there are: Spell Checked (minutes after the call), Edited (3 hours after the call), Proofed (2 hours after the Edited copy), Audited (audited after the Proofed copy, no real timing). In our analysis, for each earnings call, we use the version that is the latest available.

[13] Throughout, we refer to business days, not calendar days. For example, if day t is a Friday, day $t+1$ is the subsequent Monday.

can be matched to CDS data in each quarter of our sample. There are a total of 13,899 firm-quarter observations with matched earnings call, CDS, and control variable data.

To extract credit-relevant information from earnings calls we first combine their presentation and Q&A sections into one.[14] Because earnings calls are very long (transcripts are often between 20-30 pages), we need to identify the portions of earnings calls that contain credit relevant information. To do this, we manually collected a list of credit words and phrases that are indicative of discussions about a firm's creditworthiness. We started with a short list of seed words, like "credit," "credit line," "debt," and so on, and then identified frequently co-occurring words and phrases by reading hundreds of call segments containing the initial list of seed words. Sometimes a credit word is used in a non-credit context. For example, "rate" may refer to a company's financing rate, but when used in the phrase "exchange rate", it no longer conveys the correct meaning. To address this, we identified a list of excluded phrases so that any credit word that appears in an excluded phrase does not indicate a credit-relevant part of the earnings call. The credit words and excluded phrases lists are shown in Table A.3 of the Online Appendix.

For a given earnings call, we then identify all credit sentences, which are those containing one or more of the credit words from our list. However, if all the credit words in a sentence come from excluded phrases, that sentence is not identified as a credit sentence. We then take the union of all sentences that occur five sentences before or after credit sentences. All other parts of the earnings call are dropped. We then clean the text in the retained sentences by stemming all words using the Snowball stemmer from Python's `NLTK` package and replacing numbers with tokens that indicate magnitude: `_bln_` for numbers in the billions, `_mln_` for numbers in the millions, and `_num_` for numbers smaller than a million.[15] We generate a document term matrix (DTM) for each earnings call using the cleaned, retained sentences. Each row of the document term matrix corresponds to the earnings calls of firm i on day t, and includes the counts of the unigrams (words), bigrams (two-word phrases), and trigrams (three-word phrases) that appeared in that call (we refer to each as token j). In our analysis, we retain the top $N \in \{2000, 5000\}$ highest-frequency terms once the DTM has been constructed (without this pruning, the DTM would have 947,243 terms). The resultant DTM has 13,899 rows and either 2000 or 5000 columns, is relatively sparse,[16] and reflects *credit-related* language across our earnings call corpus.

[14] We tried a version of our analysis using only the Q&A portion of earnings calls, but found this approach worked less well than using both the presentation and Q&A sections.

[15] Section A.4.1 in the Online Appendix gives details.

[16] For $N = 5000\,(2000)$, 17% (31%) of the DTM entries are non-zero.

3.1 Ranking Tokens

The core of our text methodology is to estimate a mapping from the credit-related language of a day t earnings call to the closest subsequent closing CDS level – measured via the PVLGD transformation from (1). If a call occurs prior to 4 PM on day t, the associated CDS will be the day t close; if the call takes place after 4 PM on day t, the associated CDS will be the closing level on day $t+1$. Estimating a mapping with either 2000 or 5000 tokens is challenging, especially in the out-of-sample version of our analysis where we perform our text analysis in rolling windows (discussed below). One way to reduce the dimensionality of the problem is to focus on a smaller subset of the most frequent tokens in the DTM. The disadvantage of this approach is that some tokens occur frequently, for example the word "said," and yet have very little explanatory power for PVLGDs. Rather than choosing the most frequent terms in the DTM, we focus on selecting the most informative ones using a *forward selection* procedure as follows.

Let $f_{i,t,j}$ denote the number of times token j appears in the earnings call of firm i at time t. The associated PVLGD is $PV_{i,t}$, derived using firm i's closing CDS on day t. We first find the token $j^{(1)}$ with the highest absolute correlation, $|\mathsf{corr}\left(PV_{i,t}, f_{i,t,j^{(1)}}\right)|$, with the set of pooled PVLGDs. Once the first token is selected, we regress its frequency out of the PVLGDs using the pooled regression

$$PV_{i,t} = \alpha^{(1)} + \beta^{(1)} f_{i,t,j^{(1)}} + \xi_{i,t}^{(1)}.$$

Here $\xi_{i,t}^{(1)}$ is the residual once the count of token $j^{(1)}$ has been removed from all PVLGDs. We select the next token $j^{(2)}$ as the one which has the highest correlation with $\xi_{i,t}^{(1)}$, and then regress the count of that token out of $\xi_{i,t}^{(2)}$. The nth token in this process is selected as the one with the highest absolute correlation between its count and the residual from the $n-1$st regression, i.e., the largest $|\mathsf{corr}(\xi_{i,t}^{(n-1)}, f_{i,t,j^{(n)}})|$. We then calculate the residuals $\xi_{i,t}^{(n)}$ from the regression

$$\xi_{i,t}^{(n-1)} = \alpha^{(n)} + \beta^{(n)} f_{i,t,j^{(n)}} + \xi_{i,t}^{(n)}. \tag{3}$$

We repeat the above step until the desired number of tokens is selected.

This recursive procedure is similar to a traditional forward selection regression model, where a new token is selected in the nth step as the one which maximizes the incremental R^2 in a specification that includes the prior $n-1$ selected terms. However, this method is orders of magnitude slower than ours because of the need to run a high-dimensional regression for each candidate token in the nth step of the process. Our procedure essentially only involves calculating a correlation between each candidate variable and the residual from the $n-1$st step. The downside of our method is that the residual $\xi^{(n)}$ from the nth step is not guaranteed to be orthogonal to the first $n-1$ selected tokens (though it is orthogonal to token $j^{(n)}$ by construction), whereas in the traditional forward selection approach such orthogonality is guaranteed. However, in unreported results, we verified that correlations between $\xi^{(n)}$ and the already chosen tokens are close to zero. The sizable computational gain of our method far outweighs this slight suboptimality. This computational speedup is less important for the full-sample analysis, but is crucial for out out-of-sample tests in which the forward selection procedure is run in rolling windows.

We implement the above procedure for investment grade and high yield names separately because these groups of companies typically use different language in the credit-related portion of their earnings calls. Table 5 shows the first 20 positive and negative correlation tokens – in terms of the signs of their $\beta^{(n)}$ coefficient from (3) – identified in the forward selection process for the IG and HY samples. For IG some of the top-selected positive correlation tokens include "invest_grade," "coven" (the stemmed form of covenant), and "loan." The presence of these tokens is generally associated with higher PVLGDs (weaker credit). Some of the top-selected negatively correlated tokens, i.e., those associated with lower PVLGDs (stronger credit), are "growth," "sale._num_," and "quarter.increas._num_." For HY, the top-selected positively correlated tokens are "matur," "amend," and "net.loss," all of which are associated with wider credit spreads (worse credit). The top-selected HY tokens associated with lower credit spreads (better credit) are "growth," "share_repurchase," and "earn._num_."[17] We further discuss these selected words and phrases, and explain the Coefs column in Table 5, in Section 3.3.

Some tokens identified in the forward selection process are typically used by firms from a single sector (e.g., "bakken.shale"). Such tokens are functionally equivalent to a sector fixed effect, which is included explicitly in our analysis, as discussed below. We therefore drop from the DTM those tokens that are very prevalent in a single sector. For each token j, we calculate its maximum sector concentration via

$$c_j = \max_k \frac{\sum_{i,t} f_{i,t,j} \mathbf{1}[i \in k]}{\sum_{i,t} f_{i,t,j}}, \tag{4}$$

where $f_{i,t,j}$ is the number of time token j appears in firm i's time t earnings call and $\mathbf{1}[i \in k]$ is an indicator for whether firm i belongs to (the set of firms in) sector k. The c_j measure shows the fraction of all occurrences of token j that took place in the most

[17]Bigrams with a (_) separator are selected credit phrases from Table A.3 of the Online Appendix, while those with a (.) separator were identified in the set of frequently-occurring tokens.

prevalent sector for that token.[18] The subsequent analysis drops tokens with $c_j > t_c$ for some fixed threshold t_c. Since $c_j \leq 1$, when $t_c = 1$ the filter does not drop any tokens.

3.2 Credit Score

The prior section showed how we can rank tokens by their informativeness for explaining the panel of PVLGDs and how we can measure sector concentrations of particular tokens. We now turn to our core task of estimating a mapping from the credit-related language of earnings calls to PVLGDs.

We refer to the DTM whose columns correspond to the retained tokens as L. We then estimate a mapping between PVLGDs and this DTM as follows

$$PV_{i,t} = S_{k(i)} + L_{i,t}\beta + \varepsilon_{i,t}, \tag{5}$$

where $PV_{i,t}$ is the PVLGD of the ith firm on day t (following the pre-/post-4 PM timing convention described at the start of Section 3.1), $S_{k(i)}$ represents the sector fixed effect for sector $k(i)$ of firm i, $L_{i,t}$ is the row of the DTM corresponding to this earnings call, and β is an N_{FS}-dimensional column vector.[19] Even with the dimensionality reduction entailed in the choice of N_{FS}, the regression in (5) is still too high-dimensional to yield stable estimates using traditional methods. Because of this, we estimate (5) using a lasso regression with a penalty parameter selected using 10-fold cross-validation.[20]

The model-implied PVLGD is defined as the fitted value from the regression in (5), i.e., $\hat{PV}_{i,t} = \hat{S}_{k(i)} + L_{i,t}\hat{\beta}$. In the full-sample analysis, we refer to the residual from (5) as the *credit score* associated with the earnings call:

$$CS_{i,t} = PV_{i,t} - \hat{PV}_{i,t} = \hat{\varepsilon}_{i,t}. \tag{6}$$

The credit score is thus the difference between the actual and model-implied PVLGD. Note that the actual PVLGD, $PV_{i,t}$, is observed *after* the earnings call takes place, and

[18] Some examples of full-sample c_j values: bakken 0.958, oil.equival 0.975, medicar.advantag 0.973, eagl.ford 0.972, kroger (a supermarket chain) 0.996, nucor (a steel manufacturer) 0.978, airplan 0.952, nike 0.914, therapi 0.908, restaur 0.900, capital.alloc 0.155, investor 0.163, cash_flow 0.195. Since we have nine sectors, c_j ranges from 1/9 to 1 by construction.

[19] We tried a version of (5) with PVLGD changes, i.e, $PV_{i,t} - PV_{i,t-\delta}$ where $\delta \in \{1, 5, 21, 92, 183, 365\}$ is the number of days from the day t earnings call. We also ran the model with quarter-over-quarter changes in $L_{i,t}$ as the explanatory variable. Neither variant worked as well as the one in (5).

[20] This procedure is implemented using Python's `scikit-learn` toolkit. All regressors in (5) can potentially be excluded by the lasso. A future direction is to explore a neural network approach for estimating (5). One advantage of our lasso approach is greater interpretability of the text model.

should thus reflect the information content of the call. In the rolling analysis (described below), the credit score is still defined as in (6), but the implied PVLGD $\hat{PV}_{i,t}$ is calculated using a model estimated with only historical data applied to the frequency count, $L_{i,t}$, of the present earnings call.

A positive credit score means the actual PVLGD is higher than the one implied by the earnings call, and thus the earnings call contained language that is typically associated with tighter credit spreads than those prevailing in the market immediately after the call. A negative credit score means that the earnings call language called for higher credit spreads than what prevailed in the market immediately after the call. The credit score thus captures information that is present in the firm's earnings call, but not in market CDS prices. The credit score is our main variable of interest in subsequent analysis.

3.3 Text Model Implementation

We perform the forward selection and model estimation from Sections 3.1 and 3.2 both for the full data sample, as well as in rolling windows for out-of-sample analysis. The rolling analysis uses no forward looking information, and produces a text model that would have been available to investors in real time. We perform our main analysis based on full-sample credit scores because these best capture the information in earnings calls that is relevant for future credit outcomes. For robustness, we also run versions of our regressions using the rolling text model. Section 6 develops a trading strategy based on rolling credit scores to show the out-of-sample validity of our approach.

Our estimation has several parameters. One of these is $N \in \{2000, 5000\}$ which determines whether our starting DTM contains the most frequent 2000 or 5000 terms. The second parameter is $N_{FS} \in \{250, 500, 1000\}$ which refers to the number of the most informative words from the recursive forward selection procedure of the prior section that are retained for the mapping. Finally, we allow the concentration threshold to vary with $t_c \in \{1/3, 1/2, 1\}$: any token whose c_j (from eq. 4 estimated using the entire corpus) is greater than t_c is dropped from the analysis and is replaced by the next highest-ranked explanatory token from the recursive selection method with $c_j < t_c$ until N_{FS} tokens are reached. These parameter choices generate 18 possible versions of the analysis, each of which contains its own set of tokens, though there is substantial overlap among the variants; Panel A in Table 2 summarizes the possible parameter combinations.

The 18 model variants described above are used for our full-sample analysis. For the rolling analysis, we add two additional parameters. The first determines the lookback period over which we perform forward selection and over which we estimate the text-

Panel A: Full-sample and rolling parameters	
DTM most frequent words (N)	2000, 5000
Sector concentration threshold (t_c)	1/3, 1/2, 1
Selected words (N_{FS})	250, 500, 1000
Panel B: Rolling parameters	
Annual updating	2-, 3-, 5-year, and expanding windows
Monthly updating	3-year window

Table 2

Specification parameters for the full-sample and rolling text models.

PVLGD mapping from (5). The second parameter, the updating frequency for the model, determines whether our rolling window estimation moves forward by one month or one year at every step. For annual updating, we use four different training horizons in the rolling analysis: 2-, 3- and 5-year rolling windows, and an expanding window from the start of the sample to the present. For monthly updating, we only use 3-year rolling windows because of the computational costs involved. This means that our of out-of-sample analysis has $18 \times 5 = 90$ text model variants, which are summarized in Panel B of Table 2.

For the full-sample and out-of-sample analyses, we estimate the text model separately for the IG and HY samples. We do so because the credit-related language used by investment grade and high yield firms tends to be quite different, as Table 5 demonstrates. For each ratings class (IG or HY), we select the full-sample model (out of the 18 variants) with the highest R^2 from the lasso regression in (5).[21] We then use the implied PVLGDs and credit scores from the selected models in subsequent analysis.

For the rolling analysis, we select in each period the model (out of 90 candidates) with the most predictive credit score in the training window. We run a forecasting regression for 12-month ahead PVLGD changes using data in the traininig window, and ensuring that the dependent variables do not extend outside the training window. The forecasting regression, described in Section 4.2, includes as regressors the credit score from a given text model estimated in the training window, the current PVLGD, and an extensive set of other controls. We select the text model whose credit score coefficient has the largest magnitude t-statistic. We conjecture that text models whose credit scores are good predictors of PVLGD changes in training windows are also the ones whose credit

[21] The lasso R^2 is calculated as $1 - \mathsf{var}(\hat{\varepsilon}_{i,t})/\mathsf{var}(PV_{i,t})$. The highest R^2 full-sample text models for the IG and HY subsamples coincide at $N = 5000$, $N_{FS} = 1000$, and $t_c = 1$.

scores are good out-of-sample predictors.

When the training window moves forward and the forecasting model is re-estimated, we choose a new text model following the same procedure. Each period in our out-of-sample analysis may thus use $\hat{S}_{k(i)}$ dummies and a $\hat{\beta}$ vector obtained from a different text model from the set of 90 variants. We provide further details about the rolling analysis in Section A.4.2 of the Online Appendix, but emphasize that both the model estimation and selection are done without using any future information. The signals resulting from this procedure would be available to investors in real time. Section A.5 of the Online Appendix shows that PVLGDs and implied PVLGDs (both full-sample and rolling) from (5) are related, but imperfectly so, which suggests that deviations between these two measure might be informative, which we begin to investigate in Section 4.

3.4 Word Tone

Table 5 shows the full-sample lasso coefficients associated with the first 20 positive and negative tokens chosen via the recursive forward selection method of Section 3.1 applied to the IG and HY subsamples. In all cases, the full-sample lasso coefficient from (5) has the same signs as the $\beta^{(n)}$ coefficient in (3), which argues that the non-orthogonality of residuals issue in our recursive forward selection method is not problematic in practice. For example, when an IG firm mentions "invest_grade" in its earnings calls, this tends to increase its credit spread since emphasizing its credit rating may indicate a negative signal about its credit condition. On the other hand, mentioning "growth" in the earnings call decreases the credit spread for both IG and HY firms, as it is likely suggesting good prospect for the firm. For IG firms, discussions of sales ("sale._num_") and investor relations ("investor.relat") are interpreted positively and lower credit spreads, while the word "vendor" raises credit spreads, perhaps because it is associated with firms' working capital issues. But the word "amend" (stemmed form of "amendment" or "amending") is very negative for HY credit spreads (i.e., pushes them higher), as the market does not like to hear management teams discuss amending things like bond covenants on conference calls. The same is true of "matur" which may indicate investor nervousness when HY management teams discuss maturity extensions or are overly focused on the maturity structure of their debt. On the other hand, and perhaps surprisingly, "share_repurchas" is associated with lower HY spreads, as the market believes HY companies would only buy back their shares if their management teams were confident in the company's creditworthiness. Our method's ability to endogenously determine the meaning of credit words, and to allow this meaning to change over time in the rolling version of our model, is a key feature of our approach.[22]

3.5 Event Studies

We perform an event study to demonstrate the effectiveness of our credit score measure. We bucket firm-quarter PVLGD and credit score observations into deciles using independent sorts. We analyze four types of non-overlapping *extreme quarters*: where the PVLGD is in the bottom decile (i.e., low) but the credit score is not in the bottom decile; where the credit score is in the bottom decile (i.e., very negative) but the PVLGD is not in the bottom decile; where the PVLGD, but not the credit score, is in the top decile; and where the credit score, but not the PVLGD, is in the top decile. We look at the behavior of firm-level PVLGD in the eight quarters before and after the occurrence of *extreme* quarters. We use the full-sample text model for the analysis. All PVLGDs are normalized by dividing by the extreme quarter PVLGD, so the time zero level is one by construction.

Figure 4 shows the event studies for the four types of extreme events. Each event study is an average across all extreme firm-quarter observations for which all 17 quarters of data exist. The top left chart in the figure shows that extreme low PVLGD quarters are preceded by dropping PVLGDs and then followed by increasing PVLGDs that largely undo the prior eight quarters of tightening. The double sorts ensure that these events are not contaminated by extremely low credit scores, and so emphasize the effect of only having a low PVLGD. The top right chart shows the same analysis for the lowest decile of credit score firm-quarter observations which are not also in the bottom PVLGD decile. Low credit score quarters are also preceded by tightening PVLGDs and then followed by widening PVLGDs. The size of the effect is comparable to the lowest PVLGD decile events (the y-axis range in the two charts is not the same, but the post-event reaction is very similar).

The bottom left chart of Figure 4 shows that the top decile of PVLGD firm-quarter observations, which are not in the top decile of credit score quarters, are not preceded by credit spread widening. The state of having high PVLGDs that are consistent with management language reflected in implied PVLGDs – thus the credit score levels are not extreme – is persistent. But conditional on having extreme high PVLGDs, there is evidence of tightening in the subsequent eight quarters.[23] The bottom right chart of

[22]The coefficients in Table 5 show the marginal impact of each token, but the global impact also depends on the correlation structure of token incidence. For example, if the tokens "faster" and "growth" frequently co-occur, the coefficient of either one in isolation in (5) may be misleading.

[23]In bottom left chart of Figure 4, many events have PVLGDs in the preceding eight quarters that

the figure shows that extreme high credit score firm-quarter observations, which are not extreme high PVLGD quarters, are preceded by credit spread widening, and then followed by strong tightening with a drop in PVLGD of roughly 25% of the event quarter PVLGD in the subsequent two years. This is a much larger effect than the roughly 15% tightening that follows quarters with extreme high PVLGDs but not extreme credit scores.

The evidence from the event studies in Figure 4 suggests that both extreme PVLGD and credit score observations are followed by PVLGD reversals. There is a distinct effect for both extreme credit scores and extreme PVLGDs, and the evidence of mean reversion appears strongest following extreme firm-quarter credit score observations. That future PVLGD changes go in the direction of implied PVLGDs suggests that our text-based implied PVLGD measure contains valuable information that is not already reflected in post-call CDS levels. Section 4 establishes the forecasting ability of credit scores rigorously by controlling for many other predictors of credit returns.

In Table 6, we show one example each of an outlier event in deciles 1 (low credit scores) and 10 (high credit scores) of the extreme credit score, but not-extreme PVLGD, quarters. An outlier credit score often results from a large quarter-over-quarter divergence between implied and actual PVLGDs. The panels in the table correspond to firm-quarter observations, and show the top ten words or phrases sorted by their contribution to the change in that firm's implied PVLGD from the prior to the current (outlier) quarter. For example, if a firm uses the word "liquid" four more times in the outlier quarter relative to the prior one, and if the language model coefficient associated with "liquid" is 0.32, this would contribute to a quarter-over-quarter implied PVLGD increase of 1.28.

The decile 1 example is the New York Times (NYT) credit score from July to November 2017. The implied PVLGD of the company rose by over eight points, while its PVLGD remained almost unchanged. Contributing to the increase in implied PVLGD was the less frequent use of the phrase "revenue expectations" which is associated with lower PVLGD scores (hence the negative coefficient in the table), more frequent use of the word "obligation" (root "oblig") which is associated with higher PVLGD scores, and less frequent use of words like "user," "growth," "workforce," and "acquire" all of which are associated with lower PVLGDs. Based on our full-sample text model this difference in conference call credit-related language should have been associated with a large rise in NYT's PVLGD. This likely did not happen because, at the time of this conference call, NYT has no bonds that were deliverable into the CDS contract (the latter can still trade

are also in the top PVLGD decile. But these preceding quarters happen in the early part of our sample (which starts during the global financial crisis of 2007-2009) and thus do not show up as events because they lack the required 17 quarters of data to be included.

at positive spreads in anticipation of bond issuance) with the last NYT bond having matured in December 2016. So even though management language was credit-negative, we conjecture there was no CDS reaction because market participants did not anticipate imminent bond issuance from the company.

The decile 10 example is KB Home (KBH), a publicly traded home builder. From January 2016 to March 2016, KBH's PVLGD fell by three points, but its implied PVLGD fell by over nine points. The largest contributors to the fall in implied PVLGD were more frequent use of the phrases "share repurchase" which is associated with lower credit spreads (perhaps because companies repurchase shares when their balance sheets are strong), "share," "revenue expectation," "operating income margin," and less frequent use of the word "liquid" (or "liquidity") which is associated with higher credit spreads. Managerial tone during KBH's conference call suggested that credit spreads should have fallen more quarter-over-quarter than they actually did.

4 Empirical Results

Our empirical analysis consists of two parts. First, we analyze a specification that regresses changes in PVLGD on contemporaneous changes in several key explanatory variables. We show that the presence of lagged credit score improves the explanatory power in this specification, which is surprising given that credit score is a lagged covariate. We then drop the contemporaneous terms and analyze a pure forecasting version of the regression. We again conclude that credit score is an important forecasting variable even after controlling for an extensive collection of other credit forecasting variables. We start off using the full-sample text model, which best reflects the actual information content of earnings calls, and then test the robustness of our results using the rolling model in Section 4.3.

In the subsequent analysis, we drop 267 observations with PVLGD larger than 30, corresponding to a CDS spread higher than 2350 basis points on average. These outliers account for 1.9% of our final sample. They are usually companies in very distressed conditions, for which our credit score may not be an informative measure.

4.1 Contemporaneous Analysis

We first analyze the degree to which changes in PVLGD can be explained by contemporaneous variables that are suggested by Merton (1974); this connects our paper to the work on explaining contemporaneous changes in credit spreads done by Collin-Dufresne, Goldstein and Martin (2001) and Ericsson, Jacobs and Oviedo (2009). Because we use credit score as a regressor, we restrict the sample to firm-months pairs (i, t) where an earnings call exists for firm i in month t. $PV_{i,t}$ denotes firm i's PVLGD level in month t. Here and in subsequent analysis, we use t to refer to months, and not the days of earnings calls as before. We consider the 12-month changes in PVLGD: $\Delta PV_{i,t+\ell} = PV_{i,t+\ell} - PV_{i,t}$ with $\ell = 12$. To explain $\Delta PV_{i,t+\ell}$, we first use the contemporaneous changes in three key variables: firm i's leverage $LEV_{i,t}$, defined as its long-term debt to long-term debt plus market capitalization ratio, the risk-free rate $R_t^{(f)}$, and the implied volatility of firm i's options $IV_{i,t}$ (see Table 3 for variable definitions). Denote the 12-month changes in these variables by $\Delta LEV_{i,t+\ell}$, $\Delta R_{t+\ell}^{(f)}$, and $\Delta IV_{i,t+\ell}$, respectively. These three variables follow from the structural model in Merton (1974), which shows a firm's credit spread can be fully explained by the firm's leverage, asset volatility, and the risk-free rate. Cremers, Driessen, Maenhout, and David (2008) further find the option implied volatility is more effective in explaining the CDS spread than realized stock volatility.

To capture potential momentum or mean-reversion in CDS spreads, we include the current PVLGD level $PV_{i,t}$ in the regression. We use the credit score $CS_{i,t}$ defined in (6) to test if it can explain future changes in CDS spreads. Finally, we control for a set of fundamental, equity, and credit factors ($X_{i,t}$) that have been shown to explain bond yields or credit spreads in different settings; these are listed in Table 3. The full specification of the contemporaneous regression is given by:

$$\begin{aligned}\Delta PV_{i,t+\ell} = \alpha &+ \beta_{lev} \Delta LEV_{i,t+\ell} + \beta_{rf} \Delta R_{t+\ell}^{(f)} + \beta_{vol} \Delta IV_{i,t+\ell} \\ &+ \beta_{pv} PV_{i,t} + \beta_{cs} CS_{i,t} + \beta_{fac}^{\top} X_{i,t} + \varepsilon_{i,t}.\end{aligned} \tag{7}$$

Our interest is in β_{cs}, which reflects the impact of the credit score, after controlling for many other forecasting variables. We cluster standard errors in (7) at the firm-month level. Table 4 shows summary statistics for the variables involved in this analysis.

The regression results for 12-month ahead PVLGD changes are reported in Table 7. The t-statistics of the coefficients are shown in parentheses, and only control variables that are significant in at least one specification are retained. This regression uses the full-sample credit score (results using the rolling text model are discussed in Section 4.3). In the first column, we only include the contemporaneous changes in the three variables: risk-free rate, implied volatility, and leverage. We see all the three variables are statistically significant with the expected signs in the full sample. Specifically, PVLGD decreases in contemporaneous changes in the risk-free rate, but increases with implied volatility and leverage ratio. The R^2 from the regression is 26.1%, suggesting a large proportion of the variation in credit spread remains unexplained by the structural model. This is consistent with the findings in Collin-Dufresne, Goldstein, and Martin (2001) and Ericsson, Jacobs, and Oviedo (2009). The results are similar when we add other known predictors of credit spread changes, as shown in the second column.

In column (3), we add the current PVLGD level to the regression. We find the coefficient of current PVLGD is negative and statistically significant, suggesting PVLGD has a mean-reverting tendency: a one point increase in current PVLGD is associated with a 0.223 point decrease in PVLGDs over a 12-month horizon. Finally, the specification in column (4) further adds the credit score. We find it has a negative coefficient of -0.177 and is statistically significant with a t-statistics of -9.75. Recall the credit score is defined as the actual PVLGD minus the implied PVLGD from the earnings call. Thus, the negative coefficient of credit score suggests that when the implied PVLGD is lower than the actual PVLGD (credit score is positive), the PVLGD tends to decrease in the next 12 months. That is, the PVLGD moves *towards* the level implied by the earnings call over a time-horizon of 12 months. A negative credit score (implied higher than actual) will tend to increase PVLGDs over the next 12 months by 18% of the deviation of implied from actual PVLGD at the time of the call, which is a very large impact. This suggests that the credit score contains information that is not spanned by other market variables – including *future* changes in firm i's leverage, implied volatility, and interest rates – and which is useful for forecasting future changes in PVLGD in the direction of the current earnings-call implied PVLGD.

Our results contribute to the resolution of – though do not fully resolve – a long-standing question in the credit literature: What drives time-series *changes* in credit spreads? Part of what drives changes in credit spreads is a mean-reverting tendency of PVLGDs and another part has to do with communication from management teams about the creditworthiness of their firms. The incremental contribution to R^2 of this information, as well as of the other control variables (going from column 1 to 4 in Table 7) is 12.6%, bringing the full model R^2 to 38.7%.

4.2 Forecasting Analysis

The regression in (7) – which simply mirrors the specification in Collin-Dufresne, Goldstein, and Martin (2001) and Ericsson, Jacobs, and Oviedo (2009) – provides an interesting correlation analysis between changes in credit spreads and contemporaneous changes in other firm-level and market variables. However, it cannot be used to forecast PVLGD as the contemporaneous changes $\Delta LEV_{i,t+\ell}$, $\Delta R^{(f)}_{i,t+\ell}$, and $\Delta IV_{i,t+\ell}$ are obviously not known ex-ante. Furthermore, the regression in (7) suffers from potential endogeneity issues. For example, the increase in a firm's implied volatility may be caused by the deterioration in its credit condition, instead of the other way around.

It is cleaner to drop the contemporaneous terms from (7), which then leads to a pure forecasting specification, with no econometric issues:

$$\Delta PV_{i,t+\ell} = \alpha + \beta_{pv} PV_{i,t} + \beta_{cs} CS_{i,t} + \beta_{fac}^{\top} X_{i,t} + \varepsilon_{i,t}. \tag{8}$$

We retain the control vector $X_{i,t}$ in the forecasting model since these values are known at time t. The coefficients β_{pv} and β_{cs} capture the impact on future PVLGD changes from the current PVLGD and full-sample credit score, respectively.

The estimation result for (8) is reported in column (5) of Table 7.[24] The R^2 of the regression is 18%, which is not surprising given that we dropped three contemporaneous market-based covariates. It shows our model is indeed effective in forecasting changes in firm's credit spread. The coefficients on current PVLGD and credit score are still negative and statistically significant; in fact, the magnitude of the credit score coefficient slightly increases. For PVLGD, $\beta_{pv} = -0.206$, meaning that a one point increase in current PVLGD is associated with 0.206 points decrease in PVLGD over the next 12 months. The coefficient for the credit score is $\beta_{cs} = -0.197$.[25] Thus, the PVLGD is expected to retrace just under one fifth of its deviation away from the earnings call implied PVLGD level over the next 12 months. Both the PVLGD and credit score effects are larger than those in the contemporaneous specification in (7). The information content of earnings calls is thus very useful for forecasting future changes in corporate credit spreads.

4.3 Robustness Checks

We next perform several robustness checks. First, we replace 12-month ahead PVLGD changes with six-month ahead changes as the dependent variable. Table A.4 in the Online Appendix shows that credit score is a robust forecaster of six-month ahead changes in credit spreads as well. The coefficient of credit score is −0.121 in the forecasting regression, statistically significant at the 1% level.

[24]To save space, Table 7, as well as Tables 8 and 9 (to be explained shortly), only report results for control variables in $X_{i,t}$ that are significant in at least one specification, although all controls in $X_{i,t}$ are included in the regressions.

[25]In standard deviation terms, a one standard deviation increase in credit score is associated with a 0.18 ($-0.197\times 2.41/2.60$) standard deviation decrease in future PVLGDs.

Next, we obtain consistent findings when using 12-month ahead changes in log CDS spreads as the dependent variable in equations (7) and (8), instead of changes in PVLGD. Unlike PVLGD, the CDS spreads are directly observed in the data without any model assumptions. The results are reported in Table A.5 of the Online Appendix. In the forecasting regression, the coefficient of credit score is -0.027 and statistically significant at the 1% level, suggesting a positive credit score forecasts a decrease in CDS spreads.[26]

We next add both firm and time (year-month) fixed effects into regressions (7) and (8). The results are reported in Table A.6 of the Online Appendix. We find they are largely similar to our main specification in Table 7. Specifically, in the forecasting regression (8), the credit score coefficient is -0.095 with a t-statistics of -5.77. The coefficient, which reflects only time-series mean-reversion due to the presence of a firm fixed effect, is roughly half the size of the credit score coefficient in our base regressions in Table 7. This suggests that a good portion of the predictability from credit score to PVLGD changes comes from cross-sectional effects. Another difference is in the PVLGD coefficient, which goes from -0.206 to -0.526 (both highly significant) when fixed effects are included. This increase occurs because of pronounced cross-sectional persistence in PVLGDs (i.e., some firms trade at higher credit spreads than others), but after controlling for this with a firm fixed effect, there is considerable mean reversion for within-firm spreads.

We then examine the rolling credit score versions of regressions (7) and (8) for 12-month ahead PVLGD changes. These use a credit score model that would have been known to investors in real time, in the sense that $\hat{S}_{k(i)}$ and $\hat{\beta}$ from (5) are estimated using only information available prior to the current conference call. We run OLS regressions to estimate (7) and (8) using the same set of regressors as in the full sample analysis, except that credit score now comes from our rolling text models. Tables A.7 and A.8 of the Online Appendix report the results of these regressions. The rolling credit score remains a significant forecaster of future PVLGD changes, though the magnitude of the effect is weaker than in Tables 7 and A.4. For example, the coefficient of credit score in the rolling text model regression for 12-month PVLGD change -0.041, which is statistically significant at the 1% level.

Our results, using the full-sample and rolling credit score measures, suggest that both PVLGD and credit score are useful predictors of future changes in PVLGD. This confirms past findings of a price-to-book (proxied for by the PVLGD) value effect in corporate credit (Bartram, Grinblatt, and Nozawa, 2020; Bali, Subrahmanyam, and Wen, 2021).

[26]Note that the units of the PVLGD coefficients in Table 7 (PVLGD changes) and those of the log CDS coefficients in Table A.5 (log CDS changes) are not directly comparable.

Further, we find the surprising result that textual information from earnings calls is not fully incorporated into CDS prices, and is useful for forecasting CDS spread changes over the ensuing 6 to 12 months. The forecasting success of the rolling version of the text model, which is free of forward-looking information, suggests that credit scores may contain economically useful information for investors, a topic we return to in Section 6. But first, we explore potential mechanisms underlying our results.

5 Channels

Why does the post-call PVLGD level not fully reflect the credit-relevant information content of earnings calls? Two channels may be at play. First, it is possible that credit score forecasts of future PVLGD changes are consistent with their forecasts of future credit market risk or fundamentals. Then the ability of credit score to forecast future PVLGD changes may be due to a rational, but slow, market response to new information. We refer to this as the delayed rational response hypothesis and discuss the underlying mechanism below. An additional possibility is that earnings calls contain credit-relevant information, but investors only internalize this information with a lag. We refer to this as the constrained information processing hypothesis. We find evidence supportive of both hypotheses, though, surprisingly, the constrained information hypothesis appears empirically more important.

Why would CDS spreads react to publicly available credit-relevant information with a lag? Since our CDS contracts reflect anticipated default risk and risk premia over the next five years, on-the-run CDS spreads (i.e., those of the most recent five-year contract) may react to public fundamental information slowly, if that information changes the firm's future, but not present, credit risk. For example, consider a company that announces it will reduce its credit risk by issuing equity and buying back debt in year six. The current five-year CDS may not change at all because it still covers five credit-risky years. But the new five-year CDS in one year, which will cover the firm's high credit risk for four years and one year of low credit risk, will predictably trade at a lower spread than the current five-year CDS. In two years, the new five-year CDS contract will predictably trade at a still lower spread as its ratio of high- to low-risk years is even lower, and so on. CDS contracts may thus rationally react more slowly to public information than stock prices, because the latter reflect *all* future cash flows, and not only those over the next five years.

We use the full-sample text model in this section. Investors who fully paid attention to the credit relevant portions of earnings calls would have been able to extract much more information than our full-sample text model can capture; but this model is as close as we can get to the full information set that would have been available to attentive investors.

5.1 Delayed Rational Response Hypothesis

We first check whether the credit score forecasts changes in future credit risk in a way consistent with the delayed rational response hypothesis. We consider three market-based risk measures. The first one is the realized volatility of monthly PVLGD changes in the next 12 months, defined the same as RVCredit (see Table 3), except that it is calculated based on observations 12-month ahead. In addition, we consider the maximum monthly change and maximum cumulative change in a firm's PVLGD over the next 12 months. Recall an increase in the PVLGD means an increase in perceived credit risk or default risk premium, and a decrease in the value of the associated corporate bonds. We forecast credit risk using the following regression:

$$R_{i,t+12} = \alpha + \beta_{pv} PV_{i,t} + \beta_{cs} CS_{i,t} + \beta_{fac}^{\top} X_{i,t} + \varepsilon_{i,t}, \tag{9}$$

where $R_{i,t+12}$ is one of the three risk measures: realized volatility (RiskVol, i.e., RVCredit for the next 12 months), maximum monthly change (RiskMaxIncr), or maximum cumulative change (RiskCumSum). $X_{i,t}$ contains the same control variables as in equation (7). In particular, to adjust for serial correlation in risk measures, lagged realized volatility is included in the controls via the RVCredit variable.[27] If the ability of credit score to negatively forecast PVLGD changes stems from its ability to forecast future risk, we would expect the coefficient β_{cs} in (9) to be negative.

The regression results for (9) are reported in the first three columns of Table 8. The credit score, PVLGD and control variables are standardized so each coefficient estimate shown in the table denotes the change in the dependent variable in units of its interquartile range (IQR) due to a unit IQR change in the independent variable. The IQR, defined as the difference between the 75th and 25th percentile of a variable, is a measure similar to standard deviation (in a normal distribution, the IQR equals 1.35 standard deviations), but the IQR is much less sensitive to outliers. Standard errors are clustered by month and firm. All data have been winsorized at the 1% and 99% levels. Only those control variables that are significant in at least one specification are shown in the table.

Our first finding is that the current PVLGD level forecasts higher future risk. This is

[27]Replacing RVCredit with maximum monthly or cumulative PVLGD changes in the prior 12 months as controls leaves our results unchanged.

to be expected since higher CDS spreads reflect higher credit, and market, risk. Consistent with the delayed rational response hypothesis, we find that β_{cs} is significantly negative for all three risk measures, even after controlling for the effects captured by $X_{i,t}$. The effect, in units of IQR, ranges from -0.081 to -0.280, which is a large impact, and in line with the predictive power of credit score for future PVLGD changes (see footnote 25). When the call-implied PVLGD is lower than the market PVLGD, i.e., when credit score is positive, we expect lower risk in the future. A delayed, rational reaction to this decreased risk level – which in unreported results persists longer than one-year – may partially explain why credit score negatively forecasts future changes in PVLGD.

Another possible channel for the forecasting power of credit score is through firm fundamentals. If a higher credit score forecasts improved firm fundamentals, it can lead to predictably lower future credit spreads according to the delayed rational response hypothesis, as long as this change in fundamentals takes place over a long time period. We now rerun the specification in (9) but where the dependent variable is the change in one of the following fundamental measures: profitability as defined in Fama and French (2008), the firm's market leverage (LEV), defined as long-term debt divided by the sum of long-term debt and the market value of equity, and the logarithm of firm's total assets. We calculate the changes in these variables over the next eight quarters to allow for the rational delayed response to play out, starting with the first post-call quarterly observation of each dependent variable and ending with the eighth.[28] In the last three columns of Table 8, we report the standardized coefficients – the change in units of IQR of each dependent variable for a one unit IQR change in forecasting variable – for the impact of PVLGD and credit score on these three fundamental measures. The winsorization and standard errors are calculated in the same way as with the risk measures.

Profitability, defined as the ratio of net income to sales as in Gompers, Ishii, and Metrick (2003), is positively forecasted by credit score (at the 5% level), suggesting that when implied PVLGD is lower than PVLGD (i.e., management conversation is more sanguine than market CDS spreads) this is good news for future profitability. This holds even after controlling for a large set of other forecasting variables, and is consistent with the delayed rational response hypothesis. The economic magnitude of this effect is 0.075 in standardized terms, which is in line with, though smaller than, the impact of credit score on future PVLGD changes. Credit score also negatively forecasts firms' market leverage – a positive for creditworthiness – with a smaller magnitude (-0.033 in unit of IQR) and lower statistical significance (at the 10% level). Profitability and market leverage are

[28]The results are qualitatively similar when using one- to five-year ahead windows.

not forecastable by PVLGD, again pointing to the unique information content of credit scores. Credit score does not forecast log asset growth, while PVLGD forecasts asset growth negatively, though the effect is not significant.

We therefore find evidence that credit scores negatively forecast future market risk and firm leverage but positively forecast future firm profitability. These results are consistent with the delayed rational response hypothesis, with lower risk, higher profitability, and lower leverage being associated with gradual declines in future PVLGD.

5.2 Constrained Information Hypothesis

We now test the hypothesis that credit scores forecast future PVLGD changes because CDS spreads do not fully react to the information content of earnings calls. To do so, we investigate how the forecastability of PVLGD by credit scores depends on a firm's information environment. We consider four information measures, each of which proxies for how difficult it may be for capacity constrained investors, as in Sims (2011), to fully absorb all credit-relevant information from earnings calls.

First, we measure the length of each earnings call (TransLen) using the number of words in the call transcript. The longer the call, the harder it is for capacity-constrained investors to extract the credit-relevant portions of the call. Thus, such credit-relevant information may be incorporated into prices more slowly for longer calls. Next, we calculate the Flesch-Kincaid grade level (FKGrade) for each call. This is a heuristic measure of the reading difficulty of English articles expressed as the number of years of education generally needed to understand the text. The Flesch-Kincaid score is given by

$$\text{FKGrade} = 0.39\left(\frac{\text{total words}}{\text{total sentences}}\right) + 11.8\left(\frac{\text{total syllables}}{\text{total words}}\right) - 15.59.$$

We conjecture that the higher the Flesch-Kincaid score of a call transcript, the harder it will be for capacity constrained investors to fully react to its information content, which will lead to more predictability of PVLGD changes by credit scores. Loughran and McDonald (2020) argue that Flesch-Kincaid-type scores are imperfect measures of readability, and suggest using text length instead. In light of this, we expect call length to be a better proxy for call complexity.

Next, we look at two measures related to analyst coverage of firms. We obtain from the I/B/E/S database the number of analysts (NumAnlst) following each firm's stock. We define $\text{NumAnlst}_{i,t}$ as the total number of individual analysts that have given at least one price forecast for firm i within the 12 months preceding month t. We set $\text{NumAnlst}_{i,t}$ to zero if we cannot find any analyst with a price forecast for firm i in the 12 months preceding t. Finally, we calculate the I/B/E/S analyst dispersion (DispAnlst) in price forecasts for a firm's stock. $\text{DispAnlst}_{i,t}$ is defined as the standard deviation of stock price forecasts for firm i divided by the mean of the forecasts in the year preceding month t.[29] Lehavy, Li, and Merkley (2011) show that firms with more complex investor communications attract more analyst coverage; we therefore interpret the number of analysts covering a firm as a proxy for the firm's business complexity. After all, employing an analyst is costly, and if a firm is easy to understand, it is inefficient for many analysts to cover this firm only to produce non-differentiated research; large analyst coverage is only justified if each analyst produces unique insights, which is possible if the firm being analyzed is complex. Analyst disagreement may also proxy for firm complexity. Under the capacity constrained investor thesis, earnings calls of more complex firms are harder to fully understand, and thus it should take the market longer to respond to the information content of calls of firms with more analysts and with higher analysis disagreement.

Each of the four information measures is classified into a decile bin based on the cross-section of observations over the past 12 months. For example, we compare the number of analysts covering firm i's stock in the year prior to month t to all $\text{NumAnlst}_{j,s}$ observations for $s \in \{t-11, t\}$ and then assign $\text{NumAnlst}_{i,t}$ to a decile bin based on these observations. This binning controls for outliers and potential nonlinearities in the raw measures. Section A.6.2 of the Online Appendix gives more details on the procedure. We then interact the decile form of each variable with the credit score in the forecasting regression (8) for future PVLGD changes. The interaction coefficient indicates whether a given measure of the firm's information environment increases or decreases the forecasting power of credit score for future PVLGD changes. The interacted version of (8) is

$$\Delta PV_{i,t+\ell} = \alpha + \beta_{pv} PV_{i,t} + \beta_{cs} CS_{i,t} + \beta_{md} MD_{i,t} + \beta_{cs\times md} CS_{i,t} \times MD_{i,t} + \beta_{fac}^{\top} X_{i,t} + \varepsilon_{i,t}, \quad (10)$$

where $MD_{i,t}$ is the demeaned decile variable, using the full-sample mean of all decile levels. Our focus is on $\beta_{cs\times md}$ which measures the impact of the information environment on the forecasting power of credit score.[30]

The results for regression (10) are in Table 9. The decile variable in the interaction term in (10) is labeled as TransLen, FKGrade, NumAnlst, and DispAnlst respectively. For comparison, we also report the regression results without the interactions, i.e., dropping the terms $\beta_{md} MD_{i,t}$ and $\beta_{cs\times md} CS_{i,t} \times MD_{i,t}$ in (10). The regression without the inter-

[29] Section A.6.1 of the Online Appendix details the calculation of analyst dispersion.
[30] The results with change in log CDS spread as dependent variable in (10) are qualitatively similar.

action term for each information measure is constrained to have the same observations as the with-interaction regression, which means the sample size changes slightly for the four different specifications. The factors $X_{i,t}$ from (8) are included in all the specifications, but only ones that are significant in at least one specification are shown in the table.

We see that the information interaction term $\beta_{cs \times md}$ is insignificant for the Flesch-Kincaid grade and for analyst dispersion, though both terms are negative suggesting that higher analyst disagreement and higher Flesch-Kincaid scores increase the impact of credit score for future PVLGD changes. The insignificance of the Flesch-Kincaid interaction may be due to the fact that Flesch-Kincaid-type measures are poor representations of readability, as Loughran and McDonald (2020) argue.

However, the credit score-information interaction terms are significantly negative for the number of analysts (-0.026) and for transcript length (-0.033), both are significant at the 1% level. These effects are economically large. For example, if the transcript length of an earnings call is in the top decile – so $MD_{i,t} \approx 4.5$ since it is demeaned – the effect of credit score on future PVLGD changes equals $\beta_{cs} + \beta_{cs \times md} MD_{i,t} = -0.204 - 0.033 \times 4.5 = -0.352$, which is 72% larger than the average effect $\beta_{cs} = -0.204$. The number of analysts has a similarly large impact on the forecasting power of credit score.

These results support the constrained information hypothesis for the forecastability of PVLGD by credit scores. They show that the market may not fully incorporate credit-relevant information from the earnings calls of firms where the call is long (and thus difficult to fully digest) or where the firm's business is very complex (as proxied by analyst coverage). Thus, automated measures of the credit impact of earnings calls help address the capacity constraint faced by investors and may be of great practical value. Our information constraint result is most closely related to the underreaction mechanism in You and Zhang (2009) and Cohen, Malloy, and Nguyen (2020), who provide evidence that stock investors are inattentive to the language in firms' 10-K filings. Other related papers document stock price underreaction to the information content of news (Tetlock, Saar-Tsechansky, and Macskassy, 2008; Heston and Sinha, 2017; Ke, Kelly, and Xiu, 2021; Glasserman, Li, and Mamaysky, 2022).

Overall, our results suggest that the rational delayed response of fixed-maturity CDS spreads to public news plays a role in explaining why credit scores forecast future PLVGD changes. However, the evidence also points to informationally constrained investors who do not respond to the totality of credit-relevant information in earnings calls. Under the rational delayed response hypothesis, the information in credit scores should not lead to profitable trading strategies because, controlling for other firm characteristics, the risk-reward embedded in CDSs with far-from-zero and close-to-zero credit scores should be the same. However, under the informationally constrained investor hypothesis, knowing credit scores should lead to profitable trading strategies because this information has not yet been fully reflected in market prices. To better understand the economic magnitude of – and the underlying mechanism behind – our findings, we next turn our attention to an out-of-sample analysis of PVLGD predictability by credit scores.

6 Out-of-Sample Tests

We use trading simulations to show the out-of-sample economic impact of our credit score measures. Recall the credit score represents the discrepancy between the market PVLGD and the model-implied PVLGD (see equation 6). It therefore is a measure of disagreement between the market and the model. We found in Section 4 that higher credit scores forecast future declines in a firm's PVLGD. Thus, on average, when the model and the market disagree, the market tends to gravitate in the direction of the model over time. We now use the rolling text model (see Section 3.3), which employs a credit score that would have been known to investors in real time because it only uses historical information. The rolling text model allows us to evaluate the out-of-sample benefit of using credit score information. Our trading strategies seek to isolate the credit score dimension of information. Richer trading strategies may use other information shown in the literature to forecast credit returns, but our focus is narrower than this and we analyze strategies keyed only off credit score.

Before detailing our portfolio construction, a brief aside on nomenclature: In our trading strategy, we go long firms whose CDSs have positive credit scores, in anticipation of dropping future PVLGDs, and we go short firms with negative credit scores. Going long credit through CDS means selling protection, and thus betting against default. If protection is sold at a CDS spread corresponding to a 10 PVLGD, and the credit spread of the firm improves (i.e., falls), the PVLGD will drop meaning that the CDS can be bought back at a lower price than 10 and a gain is realized. Going short credit via CDS means buying protection, and betting on increased default risk. If one buys protection via CDS at a PVLGD of 10 and the credit spread of the firm increases, the PVLGD will increase and the CDS contract can be unwound by selling at a price higher than 10, thus realizing a profit. This is the opposite of what happens with corporate bonds, where going long means buying bonds and going short means borrowing and selling bonds.

Given the tendency of future PVLGDs to trend in the direction of the earnings call-implied PVLGDs, we interpret the credit score as a measure of mispricing. We then seek to construct maximally mispriced portfolios – i.e., those with the highest possible weighted credit score – subject to certain constraints. In each month t, our portfolio consists of three equally weighted sub-portfolios constructed at the end of months $t-1$, $t-2$, and $t-3$. We use three sub-portfolios to allow time for markets to react to earnings call news. At the end of month t, the positions in the sub-portfolio from month $t-3$ are liquidated and a new, month t sub-portfolio is added.

At the end of month t, we select companies that had at least one earnings call in the past three months.[31] For purposes of the trading strategy, each company's credit score is calculated by (6) using its PVLGD at the end of month t and the implied PVLGD from its most recent earnings call (which could be up to three months ago). Index the companies by $i = 1, 2, \ldots, n_t$ and denote their weights in the portfolio by $w_{i,t}$. We allow both long and short positions: a positive $w_{i,t}$ means we go long the credit, and thus sell protection via CDS; and a negative $w_{i,t}$ means we go short the credit by buying CDS. Given credit scores for month t – defined as above – we would like to maximize the total mispricing of the portfolio, given by $\sum_{i=1}^{n_t} w_{i,t} CS_{i,t}$, subject to following constraints: First, we require the portfolio to be self-financing by having a zero total PVLGD (more on this below); second we impose concentration limits by setting a lower and upper bound for each individual weight $w_{i,t}$; finally, we add a constraint on the total long and short position of the portfolio, which limits the leverage allowed for the portfolio.

Formally, the portfolio optimization problem is given by

$$\max \quad \sum_{i=1}^{n_t} w_{i,t} CS_{i,t} \qquad \text{(Total mispricing)} \tag{11}$$

$$\text{s.t.} \quad \sum_{i=1}^{n_t} w_{i,t} PV_{i,t} = 0, \tag{12}$$

$$w_{i,t} \leq u, \;\; w_{i,t} \geq l, \;\; \forall i, \tag{13}$$

$$\sum_{i=1}^{n_t} w_{i,t} \mathbf{1}[w_{i,t} > 0] \leq U, \;\; \sum_{i=1}^{n_t} w_{i,t} \mathbf{1}[w_{i,t} < 0] \geq L. \tag{14}$$

The constraint in (12) can be thought of as controlling for other important CDS characteristics, to the extent that these are captured by a firm's PVLGD. The constants l and u in (13) denote the lower and upper bound for the individual positions; U and L in

[31] In order to calculate strategy returns, we only consider firm-month observations for which we have CDS data in months t through $t+3$. We address any potential survivorship bias due to firms leaving our sample in Section A.2.4 of the Online Appendix.

(14) denote the limit for the total long and short position of the portfolio. The constraint (14) is non-linear as the indicator function $\mathbf{1}[\cdot]$ is involved. In Section A.7 of the Online Appendix we show how this constraint can be linearized. This optimization is solved using the `gurobi` package in Python. The resultant portfolio isolates firms' credit score exposure while ensuring minimal exposure to overall credit markets. We also investigated a version of the above portfolio problem which constrains portfolio turnover, and found the results to be qualitatively similar to the unconstrained results that we report below.

We assume that \$100 of capital supports the portfolio resulting from the above optimization. Since PVLGD in (1) is calculated for CDS contracts with \$100 principal, the weights in (11-14) give the fraction of the \$100 that is invested in a single CDS contract; l, u, L, U should be interpreted in this context. It will simplify matters to assume that all CDS contracts trade with a coupon of zero, which means that the buyer of protection needs to pay, and the seller thus receives, an upfront amount equal to the PVLGD from (1). With this convention, the portfolio is self-financing via the constraint in (12) because any short CDS position is purchased via PVLGD received from the sale of protection in a long position. The \$100 capital is assumed to earn the risk-free rate, which can be ignored when calculating excess returns.

For example, consider a sub-portfolio with $\{w_1 = 0.5, w_2 = 0.5, w_3 = -2\}$ where $u = 0.5$, $U = 1$, $l = L = -2$, and the PVLGDs are $\{10, 10, 5\}$ respectively. This sub-portfolio satisfies the constraints of the optimization problem in (11–14). Consider the excess return of the portfolio if the PVLGDs in the next month become $\{8, 8, 4.5\}$ respectively. The two long positions in securities 1 and 2 (long positions are indicated via positive weights) will earn \$1 each because the CDS was sold at a PVLGD of 10 and now trades at a PVLGD of 8, for a total gain from the long side of the portfolio of \$2. The short side of the portfolio will experience a loss of \$0.5 for each of two contracts (CDS was bought for 5 and now trades at 4.5), leading to a total loss from the short side of \$1. Therefore the portfolio will gain \$1 on a capital base of \$100. This portfolio will thus experience a 1% excess return.

6.1 Evaluating Strategy Performance

At the end of every month t, we solve the optimization in (11–14) to generate a new sub-portfolio and use it to replace the month $t-3$ sub-portfolio that is being dropped. The change in the PVLGD for the overall portfolio in month t is the sum over the three equal-weighted sub-portfolios:

$$\Delta PV_t^{(tot)} = \frac{1}{3} \times \sum_{j=1}^{3} \Delta PV_{t-j,t}^{(sub)}, \tag{15}$$

where $\Delta PV_{t-j,t}^{(sub)}$ denotes the month–t change in the total PVLGD of the sub-portfolio constructed at the end of month $t-j$, given by

$$\Delta PV_{t-j,t}^{(sub)} = \sum_{i=1}^{n_{t-j}} w_{i,t-j}(PV_{i,t} - PV_{i,t-1}).$$

Here firm i's weight $w_{i,t-j}$ was calculated at the end of month $t-j$. For the first month in the sample, all three portfolios are set to the sub-portfolio from the initial optimization.

The portfolio's excess return in month t is $R_t = -\Delta PV_t^{(tot)}/100$, where the denominator reflects the \$100 capital supporting the CDS positions. The minus sign reflects the convention that a long position (indicated with a positive weight $w_{i,t-j}$) makes money when PVLGD declines, and short positions (indicated with negative weights) make money when the PVLGD increases. The annualized excess return of the strategy is

$$\bar{R} = \Big[\prod_{t=1}^{T}(1 + R_t)\Big]^{12/T} - 1, \tag{16}$$

where T is the total number of months over which we run the strategy.

We implement our strategy separately for investment grade (IG) and high yield (HY) names, based on the firm's ratings immediately preceding the earnings call. We set the limits on total long and short positions as $(L, U) = (-4, 4)$ for both IG and HY. This allows four-fold leverage for the portfolio, which is not extreme since our portfolio has a zero total PVLGD and thus little exposure to the aggregate credit market. We consider 25 representative specifications of individual weight limits (l, u), reflecting different degrees of single-name concentration allowed in the portfolio. Both symmetric and asymmetric weight limits are considered. We show the portfolio test results in Table 10, with Panels A and B for the IG and HY groups, respectively. Each cell reports the annualized return from the portfolio associated with a weight limit (l, u), with the values of l and u given in the table's rows and columns respectively.

We test the statistical significance of the annualized return using 100 simulated portfolios under the null hypothesis of no-predictability, for the IG and HY portfolios for each (l, u) combination listed in the table. Details of the simulation strategy are in Section

A.7.1 of the Online Appendix. The 1%, 5%, and 10% significance levels of the strategy return are represented by ***, **, and *, respectively. For example, the strategy return is significant at 1% level if at most one of the 100 simulated strategies has a higher average annualized return.

Before discussing the results in Table 10, we make several observations. First, our portfolio construction methodology, and in particular the constraint in (12), isolates credit score exposure without having an exposure to aggregate credit markets indexes.[32] Thus our portfolio strategy hinges on the cross-sectional relationship between credit scores and future CDS market outcomes. In practice, combining credit scores with other factors for forecasting credit returns may further improve portfolio performance. Second, compared to much of the corporate bond anomaly literature, our trading strategy is more readily implementable since CDS contracts for a single firm are fungible and can be easily bought (going short) or sold (going long). Footnote 7 discusses some limitations of simulated strategies using corporate bond data. Finally, because the CDS market is generally much more liquid than any one corporate bond, there are fewer off-market CDS prices (Markit reports CDS spreads that are the consensus marks across all dealers trading that contract) than there are untradable bond prices on TRACE. Our portfolio simulations establish a natural benchmark under the null of no predictability, while controlling for the underlying data quality, against which our CDS trading strategies can be compared.[33]

We find our trading strategy delivers a positive return in all cases. For example, with a -0.3 limit on short and a 0.1 limit on long positions, the trading strategy has an annual return of 2.934% for IG and 4.135% for HY. The p-values of both results relative to portfolio returns simulated under the null of no predictability are better than 1%, which means that at most one of the 100 simulations had higher average annualized returns. Looking at the significance levels for the returns of the IG and HY strategy variants (across the cells of Table 10), we see that 24 of the 25 IG strategies are significant at the 5% level or better, while 12 (14) of 25 HY strategies are significant at the 5% (10%) level or better. Specifically, the trading strategy for HY generally delivers large and significant returns when the limit on the short positions is large (e.g., the bottom two rows with $l = -0.3$ or -0.4). This suggests that when trading HY, the investor should hold large short positions on CDS with negative credit scores, essentially betting their credit spreads

[32]For the 30 portfolio outcomes shown in Table 10, the range of monthly portfolio return betas to either the IG or HY indexes is $(0.01, 0.15)$. The IG and HY aggregate credit indexes are obtained from FRED website.

[33]In future research, it would be interesting to apply our simulation approach to TRACE bond data and check whether some of the reported returns from long-short corporate bond portfolios remain significant.

will further widen.

We also test the joint statistical significance of the 25 trading strategies for IG/HY. When evaluating the joint significance, we need to account for the fact that the returns of the 25 trading strategies are not independent. The details of the statistical test are included in Section A.7.2 of the Online Appendix. The collective performance of our trading strategies is statistically highly anomalous: Under the null hypothesis of no-predictability, the probability to achieve the joint performance of our 25 trading strategies is smaller than 0.5% for both IG and HY.

To gauge the economic importance of our trading strategies, consider a long-only IG (HY) credit portfolio. According to data obtained from FRED, since 1972 (1986) the average annualized return of the IG (HY) long-only corporate bond portfolio has been 7.14% (7.55%). Consider supplementing these returns with those of the $(l, u) = (-0.3, 0.1)$ strategies. The IG and HY returns would rise to 10.07% and 11.69% respectively. Such gains would place an IG or HY credit fund into the upper echelon of its competitors. Using the $100 of corporate bonds as collateral, the CDS long-short strategies could be readily implemented by institutional investors. Of course, institutions could simply tilt existing long-only portfolios by overweighting bonds with positive credit scores and underweighting those with negative credit scores, which would produce more modest performance improvements, but without requiring the use of any leverage.

We conclude that credit scores contain useful information for forecasting credit spread changes, which translates to statistically and economically significant gains in an out-of-sample strategy that can be readily implemented by institutional investors. This evidence lends further weight to the capacity constrained investor hypothesis. If the predictability of credit score for future PVLGD changes only reflected a delayed but rational response to the information content of credit scores, such a rational reaction would not lead to profitable trading strategies, as the risk-reward embedded in CDSs with large positive or negative credit scores would be identical to that of CDSs with close-to-zero credit scores, once other firm characteristics are controlled for. However, we find this is not the case.

7 Conclusion

In this paper we introduce credit score, a novel measure of corporate creditworthiness that we extract from the text of quarterly earnings calls. Our measure is straightforward to implement, can be calculated in real time, and endogenously determines the credit impact of words used in earnings calls to discuss corporate credit. Our measure forecasts future changes in corporate credit spreads, even after controlling for an extensive set of predictors for corporate bond and CDS returns identified in the literature. We have presented evidence that credit score forecasts future firm-level outcomes, such as market risk, profitability, and firm leverage. Surprisingly, our complexity-based tests suggest that market participants do not fully incorporate the credit-relevant information in earnings calls into their credit assessments. In out-of-sample trading tests, we show that credit scores lead to economically and statistically significant gains that can be achieved by institutional investors trading in long-short CDS portfolios. This finding supports the informationally constrained investor hypothesis and suggests that algorithmically-generated credit scores should be of interest to both academic researchers and to credit investors.

In order to offer an economic assessment of the usefulness of our earnings call measure, our trading strategies intentionally focus only on cross-sectional information embedded in credit scores. We do not claim that this trading strategy optimally uses all available information that may be useful for forecasting credit returns. For example, future work can consider potential time-series predictability from credit scores, which was not our focus. Another interesting area for future work is to extend our portfolio construction methodology to take into account other firm characteristics that have been shown to forecast credit returns, and to assess the performance of this expanded trading strategy on CDS and corporate bond data.

Our paper contains several methodological innovations. First, we argue PVLGD changes are a cleaner measure of changes in market-perceived creditworthiness than are changes in credit spreads. Second, the implied PVLGD measure that we obtain from the text of earnings calls is novel and provides a template that can be applied in other contexts (for example, to assess the impact of earnings calls on firm-level implied volatility). As part of obtaining an implied PVLGD, we introduce a highly computationally efficient technique to identify important explanatory words for PVLGDs. Third, our methodology for simulating portfolio returns under the null of no predictability is novel, and provides a natural benchmark that controls for underlying data quality. Finally, as in Donovan et al. (2021), our implied PVLGD measure can be used to extract CDS-like information for firms with earnings-calls, but without active CDS markets. We hope that future researchers will find these tools useful.

References

Altman, E., 1968, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *Journal of Finance*, 23 (4), 589–609.

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

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

Bali, T., A. Goyal, D. Huang, F. Jiang, and Q. Wen, 2022, "Predicting corporate bond returns: Merton meets machine learning," working paper.

Bali, T., A. Subrahmanyam, and Q. Wen, 2021, "Long-term reversals in corporate bond markets," *Journal of Financial Ecomomics*, 139, 656–677.

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

Banz, R. W, 1981, "The Relationship between Return and Market Value of Common Stocks," *Journal of Financial Economics*, 9 (1), 3–18.

Bartram, S., M. Grinblatt, and Y. Nozawa, 2020, "Book-to-market, mispricing, and the cross-section of corporate bond returns," working paper.

Bernard, V. and J. Thomas, 1989, "Post-earnings-announcement drift: Delayed price response or risk premium?" *Journal of Accounting Research*, 27, 1–36.

Bharath, ST. and T. Shumway, 2008, "Forecasting default with the Merton distance to default model," *Review of Financial Studies*, 21(3), 1339–69.

Bhushan, R., 1989, "Firm characteristics and analyst following," *Journal of Accounting and Economics*, 11, 255–274.

Calomiris, C., J. Harris, H. Mamaysky, and C. Tessari, 2022, "Fed implied market prices and risk premia," working paper.

Campbell, J. and G. Taksler, 2003, "Equity volatility and corporate bond yields," *Journal of Finance*, 58 (6), 2321–2350.

Cao, J., A. Goyal, X. Xiao, and X. Zhan, 2022 "Implied volatility changes and corporate bond returns," *Management Science*, forthcoming.

Chordia, T., A. Goyal, Y. Nozawa, A. Subrahmanyam, and Q. Tong, 2017, "Are capital market anomalies common to equity and corporate bond markets? An empirical investigation," *Journal of Financial and Quantitative Analysis*, 52 (4), 1301–1342.

Chung, K., J. Wang, and C. Wu, 2019, "Volatility and the cross-section of corporate bond returns," *Journal of Financial Economics*, 133 (2), 397–417.

Cohen, L., C. Malloy, and Q. Nguyen, 2020, "Lazy prices," *Journal of Finance*, 75 (3), 1371–1414.

Collin-Dufresne, P., R. Goldstein, and J.S. Martin, 2001, "The determinants of credit spread changes," *Journal of Finance*, 56 (6), 2177–2207.

Cremers, M., J. Driessen, P. Maenhout, and D. Weinbaum, 2008, "Individual stock-option prices and credit spreads," *Journal of Banking & Finance*, 32(12), 2706–2715.

Donovan, J., J. Jennings, K. Koharki, and J. Lee., 2021, "Measuring credit risk using qualitative disclosure," *Review of Accounting Studies*, 26 (2), 815–863.

Edwards, A., L. Harris, and M. Piwowar, 2007, "Corporate bond market transaction costs and transparency," *Journal of Finance*, 62 (3), 1421–1451.

Ericsson, J., K. Jacobs, and R. Oviedo, 2009, "The determinants of credit default swap premia," *Journal of Financial and Quantitative Analysis*, 44 (1), 109–132.

Fama, E. F., and K. R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47 (2), 427—465.

Fama, E. F., and K. R. French, 2008, "Dissecting Anomalies," *Journal of Finance*, 63 (4), 1653—1678.

Garcia, D., X. Hu, and M. Rohrer, 2022, "The color of finance words," working paper.

Gompers, P., J. Ishii, and A. Metrick, 2003, "Corporate governance and equity prices," *Quarterly Journal of Economics*, 118 (1), 107–155.

Gu, S., B. Kelly, and D. Xiu, 2020, "Empirical asset pricing via machine learning," *Review of Financial Studies*, 33, 2223–2273.

Guo, X., H. Lin, C. Wu, and G. Zhou, 2021, "Investor sentiment and the cross-section of corporate bond returns," working paper.

Hastie, T., R. Tibshirani, and J. Friedman, 2009, *The elements of statistical learning: Data mining, inference and prediction*, Springer.

Heston, S. and N. Sinha, 2017, "News vs. sentiment: Predicting stock returns from news stories," *Financial Analysts Journal*, 73 (3), 67–83.

Jegadeesh, N., 1990, "Evidence of predictable behavior of security returns," *Journal of Finance*, 45 (3), 881–898.

Jegadeesh, N. and D. Wu, 2013, "Word power: A new approach for content analysis," *Journal of Financial Economics*, 110, 712–729.

Jegadeesh, N., and S. Titman, 1993, "Returns to buying winners and selling losers: implications for stock market efficiency," *Journal of Finance*, 48 (1), 65–91.

Jiang, F., J. Lee, X. Martin, and G. Zhou, 2019, "Manager sentiment and stock returns," *Journal of Financial Economics*, 132 (1), 126–149.

Jostova, G., S. Nikolova, A. Philipov, and C. W. Stahel, 2013, "Momentum in corporate bond returns," *Review of Financial Studies*, 26 (7), 1649–1693

Kelly, B. T., Pruitt, S., and Su, Y, 2020, "Instrumented principal component analysis," working paper.

Ke, Z., B. Kelly, and D. Xiu, 2021, "Predicting returns with text data," working paper.

Kelly, B., D. Palhares, and S. Pruitt, 2023, "Modeling corporate bond returns," *Journal of Finance*, 78 (4), 1967–2008.

Lehavy, R., F. Li, and K. Merkley, 2011, "The effect of annual report readability on analyst following and the properties of their earnings forecasts," *Accounting Review*, 86 (3), 1087–1115.

Lehmann, B., 1990, "Fads, martingales, and market efficiency," *Quarterly Journal of Economics*, 105 (1), 1–28.

Lee, M., P. Meyer-Brauns, S. Rizova, and S. Wang, 2020, "The cross-section of corporate bond returns," *Dimensional Fund Advisors working paper.*

Glasserman, P., F. Li, F., and H. Mamaysky, 2022, "Time variation in the news-returns relationship," working paper.

Loughran, T. and B. McDonald, 2020, "Textual analysis in finance," *Annual Review of Financial Economics*, 12, 357–375.

Manela, A. and A. Moreira, 2017, "News implied volatility and disaster concerns," *Journal of Financial Economics*, 123 (1), 137–162.

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

Nozawa, Y., 2017, "What drives the cross-section of credit spreads?: A variance decomposition approach," *Journal of Finance*, 72 (5), 2045–2071.

Oehmke, M. and A. Zawadowski, 2017, "The anatomy of the CDS Market," *Review of Financial Studies*, 30 (1), 80–119.

Penman, S., F. Reggiani, S. A. Richardson, I. Tuna, 2014, "An accounting-based characteristic model for asset pricing", working paper.

Polbennikov, S., M. Dubois, and A. Desclée, 2020, "Have systematic credit factors lived up to expectations? Performance through Covid-19," Barclays Credit Research, April 2020.

Sims, C., 2011, "Rational inattention and monetary economics," Chapter 4 in *Handbook of Monetary Economics*, Vol. 3A, Elsevier.

Tetlock, P., M. Saar-Tsechansky, and S. Macskassy, 2008, "More than words: Quantifying language to measure firms' fundamentals," *Journal of Finance*, 63 (3), 1437–1467.

van Binsbergen, J. and M. Schwert, 2022, "Duration-based valuation of corporate bonds," working paper.

You, H. and Z.-j. Zhang, 2009, "Financial reporting complexity and investor underreaction to 10-K information," *Review of Accounting Studies*, 14, 559–586.

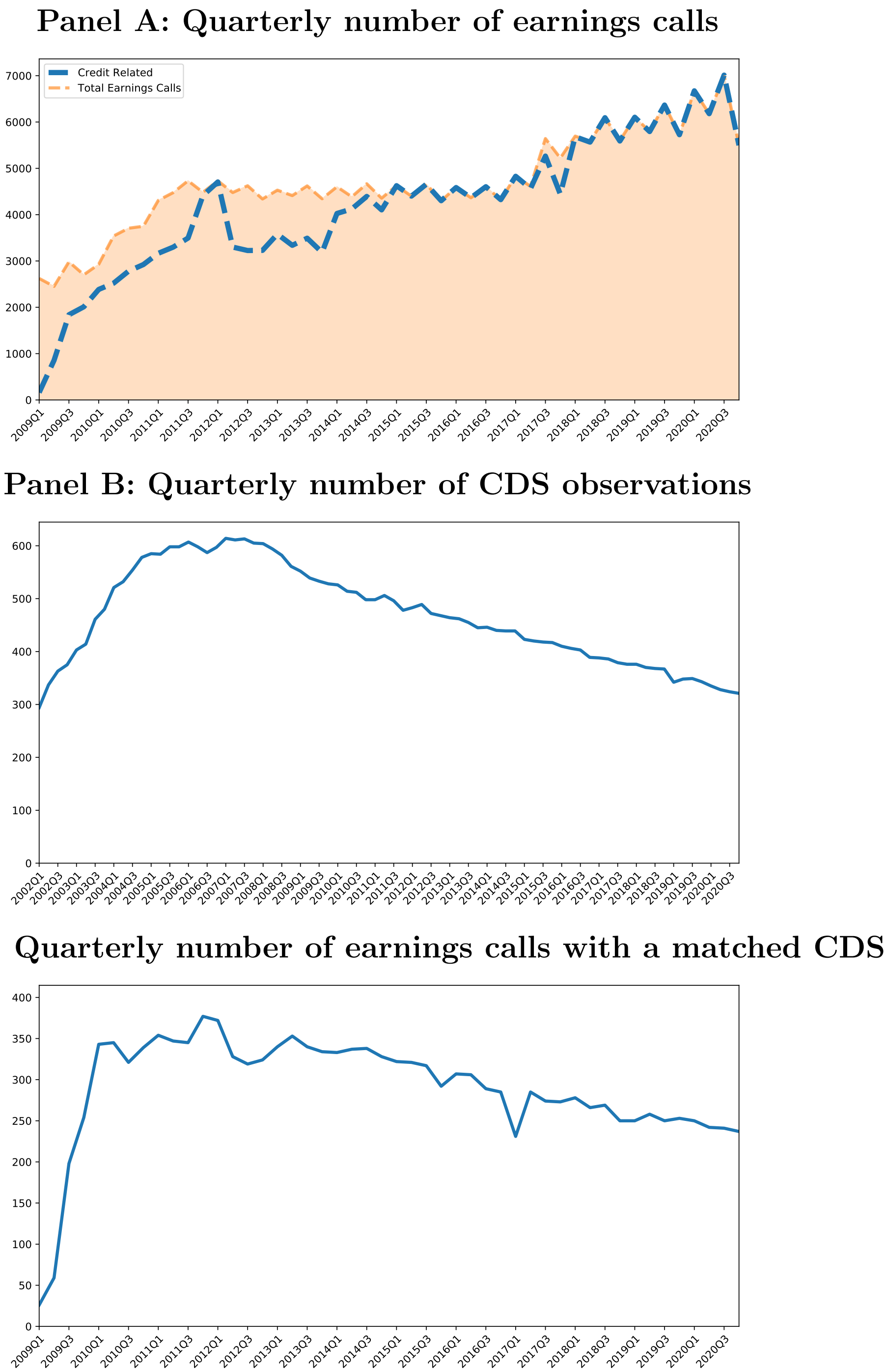

Fig. 2. The top panel shows the number of quarterly earnings calls in the data set, as well as the number of calls that have a credit-related discussion. The middle panel shows the number of quarterly credit default swap observations (we count each firm at most once within each quarter). The bottom panel shows the number of quarterly earnings calls that can be matched to a credit default swap in that quarter.

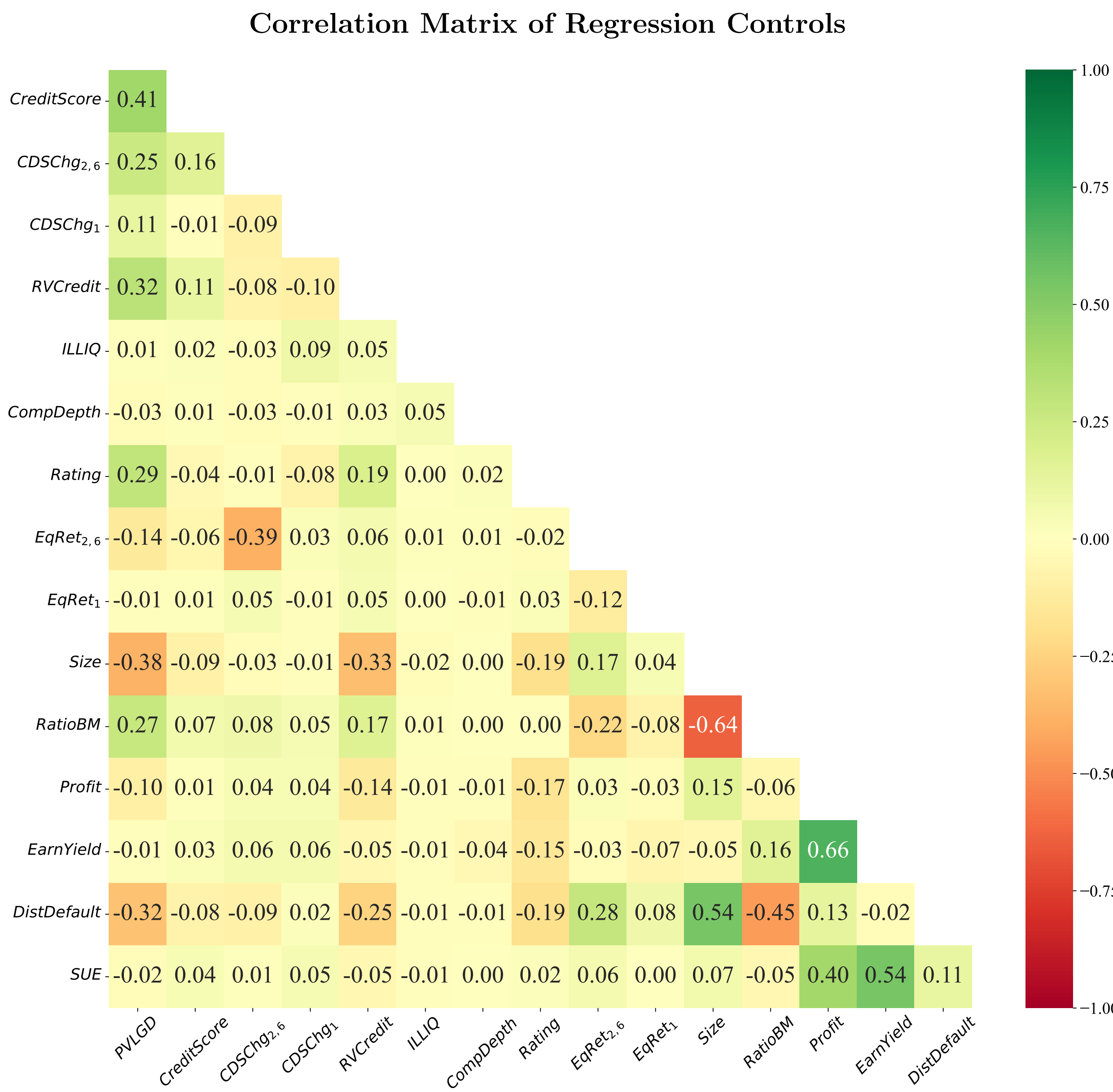

Fig. 3. This figure shows the correlation of the regression control variables calculated as the mean across firms of the time series correlations of variables for each firm using the full sample of data. The variables are shown in Table 3 and described fully in Section A.3 of the Online Appendix. All statistics are based on full sample series winsorized at 1% and 99% percentiles.

Event studies around PVLGD and credit score outliers

Panel A: Low PVLGD and credit score studies

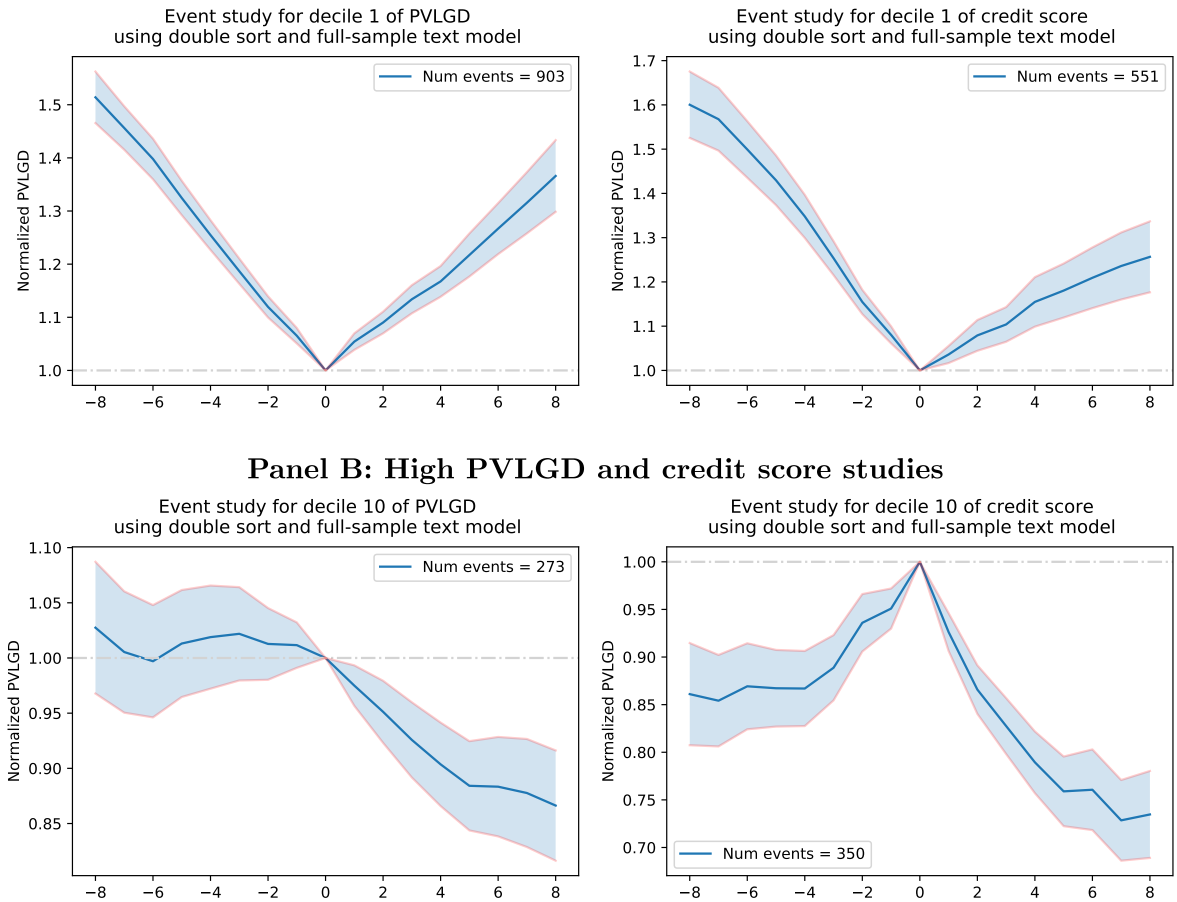

Fig. 4. The left plot of Panel A shows the normalized PVLGD response in eight quarter windows around name-date pairs whose PVLGD falls into the bottom decile of all name-date PVLGD observations, but whose credit score does not fall into the bottom decile of all name-date credit score observations. The right plot of Panel A shows the analogous normalized PVLGD event study for bottom decile name-date credit score pairs whose PVLGD is not in the bottom decile of name-date PVLGD observations. The left plot of Panel B shows the analogous PVLGD response for name-date observations in the top decile of PVLGD but not in the top decile of credit score. The right plot of Panel B shows the analogous PVLGD response for name-date observations in the top decile of credit score that are not also top decile observations of PVLGD. The number of events in each event study is shown in the legend; an event enters the study sample only if the full 17 quarters of data are available. The 95% confidence bands assume independence.

Control Variable Definitions

Variable	Description	Relevant Literature
$\text{CDSChg}_{2,6}$	CDS momentum in month $t-6$ to $t-1$	Jostova et al. (2013), Lee et al. (2021)
CDSChg_{1}	CDS reversal in the most recent one month	Bartram, Grinblatt, and Nozawa (2020)
RVCredit	CDS realized volatility in prior 12 months	Chordia et al. (2017)
ILLIQ	CDS illiquidity, measured by the negative of covariance of daily change in PVLGD and its one-day lag	Bao, Pan, and Wang (2011)
CompDepth	Composite depth as a proxy for CDS market depth, another measure for market liquidity	Bao, Pan, and Wang (2011)
Rating	Average credit rating from Markit	Bali et al. (2022), Guo, Lin, Wu, and Zhou (2021), Bao, Pan, and Wang (2021)
$\text{EqRet}_{2,6}$	Equity momentum in month $t-6$ to $t-1$	Jegadeesh and Titman (1993)
EqRet_{1}	Equity reversal in the most recent one month	Jegadeesh (1990), Lehmann (1990), Chordia et al. (2017)
Size	Firm size, logarithm of market value of firm equity	Fama and French (1992), Chordia et al. (2017)
RatioBM	Firm value, book to market ratio	Fama and French (2008)
Profit	Firm profitability, ratio of equity income to sales (quarterly)	Gompers et al. (2003)
EarnYield	Earnings yield, ratio between diluted earnings per share and stock price	Penman, Richardson, Reggiani, and Tuna (2014), Bartram, Grinblatt, and Nozawa (2020)
DistDefault	Distance-to-default measure in Bharath and Shumway (2008)	Bharath and Shumway (2008)
SUE	Standardized unexpected earnings in Bernard and Thomas (1989)	Bartram, Grinblatt, and Nozawa (2020)
R^f	Risk-free rate as the 10-year US treasury bond interest rate	Collin-Dufresne, Goldstein, and Martin (2001)
LEV	Firm leverage, long-term debt divided by long-term debt plus firm's market value	Ericsson, Jacobs, and Oviedo (2009)
IV	Average implied volatility of firm's 30-day put and call options	Cremers et al. (2008)

Table 3

This table summarizes the regression control variables used in our study. Equity measures refer to the stock of the firm whose bonds are being referenced by the CDS. All measures are calculated as of month-end. R^f, LEV, and IV are only used as covariates in the contemporaneous regressions of Section 4. Section A.3 of the Online Appendix provides detailed explanations of all of these variables.

Summary Statistics

Variable	Count	Mean	Std	0.05	0.25	0.5	0.75	0.95
PVLGD	13517	6.46	5.64	1.31	2.57	4.35	8.17	19.63
CreditScore	13817	−0.01	2.41	−3.67	−1.27	−0.19	1.03	4.40
R^f	13517	2.26	0.70	0.87	1.82	2.26	2.71	3.56
IV	13512	0.29	0.13	0.15	0.20	0.26	0.35	0.55
LEV	13517	0.44	0.18	0.17	0.29	0.41	0.57	0.77
$\text{CDSChg}_{2,6}$	13140	−0.12	1.94	−3.17	−0.60	−0.03	0.34	2.82
CDSChg_1	13478	−0.08	0.88	−1.55	−0.25	−0.02	0.12	1.17
RVCredit	13225	0.72	0.83	0.05	0.18	0.41	0.93	2.56
ILLIQ ($\times 10^{-6}$)	11072	1.51	22.46	−24.01	−1.93	0.03	4	32.74
CompDepth	13517	5.05	2.61	2	3	5	7	10
Rating	13517	4.10	1.01	3	3	4	5	6
$\text{EqRet}_{2,6}$	13375	0.04	0.18	−0.25	−0.06	0.04	0.13	0.33
EqRet_1	13474	0.01	0.08	−0.12	−0.03	0.01	0.06	0.15
Size (log)	13517	23.19	1.42	20.77	22.19	23.22	24.17	25.67
RatioBM	12903	−1	0.83	−2.48	−1.41	−0.91	−0.45	0.17
Profit	13516	0.07	0.14	−0.11	0.02	0.07	0.12	0.27
EarnYield	13515	0.01	0.03	−0.03	0.01	0.01	0.02	0.04
DistDefault	13223	4.28	3.51	−0.27	1.70	3.62	6.20	11.18
SUE ($\times 10^{-2}$)	13219	0	0.04	−0.04	−0	0	0.01	0.04
PVLGDChg_{1y}	11374	−0.15	2.60	−4.43	−0.91	−0.07	0.55	3.84
RiskVol	12002	0.65	0.77	0.05	0.17	0.36	0.83	2.28
RiskMaxIncr	12002	1.31	1.83	0.07	0.27	0.62	1.50	4.95
RiskCumSum	12002	1.60	3.03	−0.42	0.09	0.52	1.72	7.89
LEVChg_{2y}	9557	−0	0.09	−0.15	−0.06	−0.01	0.05	0.16
ProfChg_{2y}	9561	−0	0.17	−0.21	−0.02	0	0.03	0.18
AssetChg_{2y} (log)	9557	0.09	0.21	−0.21	−0.02	0.07	0.18	0.45
NumAnlst	12698	15.89	7.62	5	11	15	20	31
DispAnlst	12693	0.15	0.09	0.06	0.09	0.13	0.18	0.31
FKGrade	13277	10.24	1.85	8.10	9.20	10.20	11.10	12.60
TransLen	13277	8675	2116	5071	7584	8762	9737	12015

Table 4

This table includes all variables in our analysis and their summary statistics. We annotate actual scales of some variables if necessary. PVLGD is the measure of the price of a CDS contract as defined by equation (1), CreditScore is the measure of creditworthiness as in equation (6). See Section A.3 for detailed definitions of interest rate R^f, implied volatility IV, leverage LEV, CDS reversal CDSChg_1, CDS momentum $\text{CDSChg}_{2,6}$, realized volatility RVCredit, illiquidity ILLIQ, market depth CompDepth, average credit rating Rating, equity momentum $\text{EqRet}_{2,6}$, equity reversal EqRet_1, firm size Size, firm value RatioBM, firm profitability Profit, earnings yield EarnYield, distance-to-default DistDefault, and standardized unexpected earnings SUE. PVLGDChg_{1y} is the yearly change of PVLGD, RiskVol, RiskMaxIncr, RiskCumSum is the realized volatility, maximum monthly change, and maximum cumulative changes of PVLGD in next 12 months, respectively; LEVChg_{2y}, ProfChg_{2y}, and AssetChg_{2y} (log) denote the two-year ahead changes in market leverage ratio, profit as in Gompers, Ishii, and Metrick (2003), and the logarithm of total asset, respectively. NumAnlst, DispAnlst, FKGrade, and TransLen are information proxies defined as in Sections 5.1 and 5.2. All statistics are based on full sample series winsorized at 1% and 99% percentiles.

Forward Selected Words and Full-Sample Lasso Regression Coefficients

Panel A: Investment Grade

Words	Coefs	Words	Coefs
invest_grade	0.316	growth	−0.020
num._mln_	0.030	_num_._bln_	−0.051
coven	0.510	train	−0.118
loan	0.103	sale._num_	−0.044
missouri	0.173	quarter.increas._num_	−0.137
suppli	0.092	weather	−0.038
chief.oper	0.361	worldwid	−0.094
price.environ	0.130	impact	−0.022
pound	0.195	investor.relat	−0.228
year.year.basi	0.136	realize	−0.105
report.form	0.429	ratio._num_._num_	−0.151
loss._num_._num_	0.210	gas	−0.029
vendor	0.235	net.interest_expens	−0.168
current.quarter	0.299	suppli.chain	−0.114
adjust.oper.margin	0.168	truck	−0.081
softwar	0.051	file.sec	−0.315
list	0.146	credit_rate	−0.072
ford	0.091	year.quarter	−0.087
mln.cash	0.173	recycl	−0.096
liquid	0.059	jay	−0.108

Panel B: High Yield

Words	Coefs	Words	Coefs
matur	0.722	growth	−0.100
amend	1.044	share_repurchas	−0.384
net.loss	1.413	earn._num_	−0.567
game	0.329	strong.balanc_sheet	−0.960
liquid	0.320	oper.rate	−0.455
transform	0.406	averag	−0.149
reduct	0.222	pandem	−0.583
atlant	0.891	extend	−0.322
airlin	0.500	share	−0.115
factor.caus	0.668	user	−0.564
trend	0.142	residenti	−0.599
cash	0.168	invest_grade	−0.501
freight	0.540	matur._num_	−0.445
fresh	1.065	_num_.present	−0.835
repres	0.276	great	−0.195
mail	0.514	net.debt	−0.508
hill	1.128	incom	−0.116
bundl	0.734	long.term.sharehold	−1.203
network	0.195	revenu.expect	−1.628
trail._num_	0.798	load	−0.249

Table 5
This table shows the top twenty words identified by the correlation-based forward selection method of Section 3.1, sorted by the order of selection. The sign of the variables is determined from their $\beta^{(n)}$ coefficient from (3). Also shows are the corresponding coefficients from the full-sample lasso regression from 3.2. Bi- and trigrams are indicated by a period (.) separating the individual words. Bigrams with an underscore (_) separator indicate that the bigram is from the credit words list in Table A.3 in the Online Appendix.

Extreme Credit Score Examples: Deciles 1 and 10

Decile 1 Example

Company: New York Times Co. (NYT)
Date: 2017.07.27 to 2017.11.01
PVLGD: 2.303 to 1.873
Implied PV: 7.323 to 15.706

Word	Contribution	Coefficient
revenu.expect	4.883	−1.628
oblig	2.483	0.497
user	1.692	−0.564
growth	0.797	−0.100
workforc	0.563	−0.563
quarter.year	0.532	−0.133
foundat	0.471	−0.471
acquir	0.420	−0.210
reduc	0.405	0.135
plan._num_._num_	0.402	0.402

Decile 10 Example

Company: KB Home (KBH)
Date: 2016.01.07 to 2016.03.23
PVLGD: 19.155 to 16.185
Implied PV: 17.529 to 8.304

Word	Contribution	Coefficient
share_repurchas	−3.840	−0.384
share	−1.952	−0.115
revenu.expect	−1.628	−1.628
oper.incom.margin	−1.377	−0.689
statement.mean	−1.303	−1.303
liquid	−1.280	0.320
presid.chief.execut	−0.934	−0.934
today.discuss	−0.780	−0.780
profit	−0.742	0.046
situat	−0.726	0.363

Table 6

This table shows an example of credit scores in deciles 1 and 10 of the extreme credit score but not-extreme PVLGD double sorts described in Section 3.5. The implied PVLGDs come from the full-sample model. Each row corresponds to a token that is a top implied PVLGD change contributors in the quarter. The contribution column shows the magnitude of the contribution (change in token frequency times token coefficient), and the coefficient column shows the full-sample implied PVLGD model coefficient associated with the token. For example, tokens with negative coefficients will contribute positively to credit score if they occur less frequently.

Dependence of 12-month PVLGD Changes on Forecasting Variables: Full Sample Text Model

	(1)	(2)	(3)	(4)	(5)
R^f	−0.423	−0.451	−0.321	−0.307	
	(−4.73)	(−4.59)	(−3.73)	(−3.77)	
IV	4.625	4.179	3.995	3.951	
	(5.98)	(5.53)	(5.36)	(5.49)	
LEV	12.349	11.538	11.671	11.616	
	(13.59)	(12.14)	(12.55)	(12.76)	
PVLGD			−0.223	−0.180	−0.206
			(−10.62)	(−8.89)	(−7.94)
CreditScore				−0.177	−0.197
				(−9.75)	(−9.35)
$\text{CDSChg}_{2,6}$		−0.128	−0.057	−0.035	−0.016
		(−3.23)	(−1.62)	(−1.01)	(−0.35)
CompDepth5y		0.002	−0.065	−0.063	−0.096
		(0.15)	(−4.08)	(−4.08)	(−4.51)
Rating		−0.085	0.257	0.123	0.296
		(−1.56)	(3.93)	(2.13)	(3.16)
EqRet_1		−0.758	−0.942	−0.920	−0.606
		(−1.78)	(−2.29)	(−2.30)	(−1.07)
Size		−0.080	−0.186	−0.187	−0.085
		(−1.80)	(−3.88)	(−4.10)	(−1.62)
Profit		0.310	0.357	0.318	0.479
		(1.46)	(1.67)	(1.57)	(2.20)
RVCredit		−0.618	0.036	0.029	−0.057
		(−5.01)	(0.33)	(0.27)	(−0.46)
DistDefault		−0.022	−0.052	−0.050	−0.049
		(−1.67)	(−4.19)	(−4.14)	(−3.31)
R^2	0.261	0.300	0.368	0.387	0.180
N	11387	9470	9470	9470	9470
CS Series	full	full	full	full	full
With Factors		Yes	Yes	Yes	Yes

Table 7
This table reports regressions results of 12-month PVLGD changes on the contemporaneous changes in company market leverage, risk-free rate, and option implied volatility, adding PVLGD level, implied credit score and factors. The full specification is given in equation (7). The table reports results based on the full sample implied credit score. Column (1) presents results with contemporaneous changes, column (2) adds control factors, column (3) adds PVLGD levels, column (4) adds credit scores, and column (5) drops contemporaneous changes. We cluster standard errors by entity and month, the t-statistics are reported in parenthesis. Only control variables that are statistically significant in at least one specifications are included. All statistics are based on full sample series winsorized at 1% and 99% percentiles.

Forecast Regressions on Risk and Fundamental Measure Changes: Full Sample Text Model

	RiskVol	RiskMaxIncr	RiskCumSum	ProfChg_{2y}	LEVChg_{2y}	AssetChg_{2y} (log)
CreditScore	−0.081	−0.141	−0.280	0.075	−0.033	0.006
	(−4.98)	(−6.13)	(−7.78)	(2.01)	(−1.79)	(0.31)
PVLGD	0.324	0.198	0.010	−0.078	−0.046	−0.073
	(5.91)	(2.95)	(0.13)	(−0.86)	(−0.99)	(−1.40)
CDSChg_1	0.034	0.041	0.051	−0.002	0.016	0.001
	(2.34)	(1.75)	(1.64)	(−0.08)	(1.60)	(0.13)
RVCredit	0.379	0.419	0.365	0.081	−0.035	0.061
	(8.24)	(6.81)	(4.97)	(1.25)	(−1.07)	(1.78)
CompDepth5y	0.177	0.174	0.069	−0.060	−0.039	−0.121
	(4.02)	(2.99)	(0.90)	(−0.70)	(−1.12)	(−2.54)
Rating	0.097	0.167	0.267	−0.104	0.141	0.091
	(3.28)	(3.87)	(4.54)	(−1.04)	(3.36)	(1.81)
$\text{EqRet}_{2,6}$	−0.043	−0.086	−0.077	−0.019	0.034	0.067
	(−1.97)	(−2.69)	(−1.61)	(−0.32)	(1.53)	(3.97)
EqRet_1	−0.038	−0.041	−0.107	0.011	0.014	0.054
	(−1.60)	(−1.12)	(−1.91)	(0.17)	(0.54)	(2.90)
Size	0.085	0.064	0.140	0.131	0.135	−0.037
	(2.16)	(1.12)	(1.77)	(1.07)	(2.85)	(−0.58)
RatioBM	−0.003	−0.003	0.036	−0.203	−0.021	−0.105
	(−0.13)	(−0.09)	(0.75)	(−2.51)	(−0.69)	(−2.37)
Profit	−0.023	−0.021	−0.024	−1.164	0.034	0.104
	(−1.51)	(−1.01)	(−0.86)	(−10.15)	(2.15)	(5.28)
EarnYield	0.018	0.029	0.015	−0.115	0.004	0.013
	(1.35)	(1.39)	(0.53)	(−2.91)	(0.33)	(0.87)
DistDefault	−0.149	−0.178	−0.255	0.240	0.128	0.119
	(−5.32)	(−4.51)	(−4.56)	(2.83)	(4.14)	(2.52)
SUE	−0.003	−0.006	0.002	−0.062	0.011	−0.015
	(−0.57)	(−0.65)	(0.14)	(−2.96)	(1.86)	(−2.72)
R^2	0.457	0.277	0.143	0.292	0.063	0.057
N	9759	9759	9759	8131	8131	8125

Table 8

This table reports regression results of the 12-month risk and 2-year changes in fundamental measures on full sample implied credit score, PVLGD, and control variables. Coefficients reflect the change in the dependent variable in units of its interquartile range (IQR) due to a unit IQR change in the independent variable. T-statistics are reported in parentheses. We cluster standard errors by entity and month. The first three dependent variables reflect 12-month ahead outcomes and the last three reflect 24-month ahead ones. RiskVol is defined as the realized volatility of PVLGD, RiskMaxIncr is the maximum monthly PVLGD change, RiskCumSum is the maximum cumulative PVLGD change over the next year, ProfChg_{2y} is the two-year change in profitability defined as in Gompers, Ishii, and Metrick (2003), LEVChg_{2y} is the two-year change in the market leverage ratio, AssetChg_{2y} (log) is the two-year change of the logarithm of total assets. Only control variables that are statistically significant in at least one specifications are shown. We winsorize both dependent and independent variables at 1% and 99% percentiles.

Forecast Regressions on 12-Month PVLGD Changes with Interactions: Full Sample Text Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CreditScore	−0.193	−0.185	−0.190	−0.189	−0.193	−0.210	−0.190	−0.204
	(−8.93)	(−8.69)	(−8.90)	(−8.88)	(−8.89)	(−10.20)	(−8.92)	(−9.74)
PVLGD	−0.193	−0.204	−0.193	−0.193	−0.193	−0.198	−0.192	−0.197
	(−7.50)	(−7.83)	(−7.44)	(−7.45)	(−7.50)	(−7.90)	(−7.43)	(−7.78)
DispAnlst		0.065						
		(3.71)						
CS_DispAnlst		−0.002						
		(−0.27)						
FKGrade				0.007				
				(0.59)				
CS_FKGrade				−0.007				
				(−1.03)				
NumAnlst						0.045		
						(2.13)		
CS_NumAnlst						−0.026		
						(−3.42)		
TransLen								0.017
								(1.22)
CS_TransLen								−0.033
								(−4.63)
CompDepth	−0.095	−0.098	−0.094	−0.093	−0.096	−0.096	−0.094	−0.093
	(−4.39)	(−4.52)	(−4.33)	(−4.32)	(−4.40)	(−4.48)	(−4.33)	(−4.31)
Rating	0.275	0.263	0.274	0.274	0.274	0.291	0.275	0.293
	(3.02)	(2.83)	(3.04)	(3.01)	(3.03)	(3.46)	(3.04)	(3.41)
Size	−0.074	−0.069	−0.081	−0.080	−0.071	−0.145	−0.081	−0.095
	(−1.43)	(−1.34)	(−1.56)	(−1.55)	(−1.36)	(−2.41)	(−1.56)	(−1.84)
Profit	0.513	0.557	0.546	0.545	0.515	0.512	0.526	0.537
	(2.27)	(2.54)	(2.39)	(2.41)	(2.30)	(2.28)	(2.28)	(2.41)
DistDefault	−0.052	−0.048	−0.050	−0.050	−0.053	−0.052	−0.049	−0.049
	(−3.51)	(−3.27)	(−3.32)	(−3.32)	(−3.56)	(−3.48)	(−3.30)	(−3.32)
R^2	0.165	0.17	0.165	0.165	0.164	0.17	0.164	0.172
N	9006	9006	9089	9089	9002	9002	9096	9096

Table 9

This table reports regression results of the 12-month PVLGD changes on the PVLGD level, full sample implied credit score, control variables and their interactions with credit score. Even columns report results with factors and interections, and odd columns report corresponding results without interactions. (Full specification is: $\Delta PV_{i,t+\ell} = \alpha + \beta_{pv}PV_{i,t} + \beta_{cs}CS_{i,t} + \beta_{cv}CV_{i,t} + \beta_{cs\times cv}CS_{i,t} \times CV_{i,t} + \beta_{fac}^{\top}X_{i,t} + \varepsilon_{i,t}$). (1)-(2) are for results when the control variable is analyst dispersion during the past 12 months; (3)-(4) for Flesch-Kincaid Grade; (5)-(6) for number of analysts that have made at least 1 estimate during the past 12 months; (7)-(8) for transcript length of earnings calls. We cluster standard errors by firm and month, the t-statistics are reported in parenthesis. Only control variables that are statistically significant in at least one specifications are included. All statistics are based on full sample series winsorized at 1% and 99% percentiles.

Trading Strategy Simulations using Rolling Text Model

Panel A: Investment Grade

	0.01	0.02	0.05	0.1	0.2	0.3	0.4
−0.01	–	–	–	0.865***	0.615***	–	–
−0.02	–	–	–	1.669**	1.747***	–	–
−0.05	–	–	1.806***	2.406***	2.141***	1.662***	1.289
−0.1	1.228***	1.626***	2.241***	2.779***	2.716***	2.180***	1.897***
−0.2	1.460***	2.001***	2.447***	3.073***	3.413***	–	–
−0.3	–	–	2.546***	2.934***	–	–	–
−0.4	–	–	2.626***	2.867***	–	–	–

Panel B: High Yield

	0.01	0.02	0.05	0.1	0.2	0.3	0.4
−0.01	–	–	–	0.235**	0.249**	–	–
−0.02	–	–	–	0.369	0.463**	–	–
−0.05	–	–	1.327	1.500	0.944	0.505	0.655
−0.1	1.203***	1.714***	2.608**	2.687	2.630*	2.071	1.727
−0.2	1.105	2.400***	3.437**	3.360	3.014*	–	–
−0.3	–	–	3.691**	4.135***	–	–	–
−0.4	–	–	4.027***	4.366**	–	–	–

Table 10

This table reports the annualized return from each portfolio test specification and its statistical significance under the null hypothesis of no predictability. Panel A and B report results for IG and HY samples respectively. The leverage constraints in (14) for the two samples are set as $(-4, 4)$. Each cell corresponds to a weight limit specification $\{l, u\}$ in (13), with the value of l and u given in the row and column respectively. P-values indicating 1%, 5%, and 10% significance levels are represented by ***, **, and *, respectively. More details of the simulation are in Section A.7.1 of the Online Appendix.