Journal Pre-proof

Credit Derivatives and Corporate Default Prediction

Xiaoxia Ye, Fan Yu, Ran Zhao

PII: S0378-4266(22)00018-8
DOI: https://doi.org/10.1016/j.jbankfin.2022.106418
Reference: JBF 106418

To appear in: *Journal of Banking and Finance*

Received date: 10 July 2020
Accepted date: 15 January 2022

Please cite this article as: Xiaoxia Ye, Fan Yu, Ran Zhao, Credit Derivatives and Corporate Default Prediction, *Journal of Banking and Finance* (2022), doi: https://doi.org/10.1016/j.jbankfin.2022.106418

Credit Derivatives and Corporate Default Prediction

Xiaoxia Ye, Fan Yu, and Ran Zhao[1]

This Version: January 6, 2022

[1]Ye is from the University of Liverpool (xiaoxia.ye@liverpool.ac.uk), Yu is from Claremont McKenna College (fyu@cmc.edu), and Zhao is from Claremont Graduate University (ran.zhao@cgu.edu). We would like to thank Lydia Fu for able research assistance, and the associate editor and two referees for constructive comments that helped to improve this paper.

Credit Derivatives and Corporate Default Prediction

Abstract

There have been 128 defaults among U.S. CDS reference entities between 2001 and 2020. Within this sample, the five-year CDS spread is a significant predictor of corporate default in models with equity market covariates and firm attributes. This finding holds for forecast horizons up to 12 months, among financial and non-financial firms, within and without the great financial crisis, and is robust to the inclusion of corporate bond and equity options market information. A decomposition of the CDS spread into liquidity, physical default, and risk premium components shows that most of its predictive power for corporate default comes from the physical default component, both in- and out-of-sample. These results confirm the relevance of information contained in single-name CDS pricing to corporate default prediction.

JEL Classification: G12, G13, G17, G23, G33

Keywords: Credit default swap spread, corporate default prediction, physical default, default risk premium, CDS liquidity.

1 Introduction

The prediction of corporate default is an important topic of interest to both academics and practitioners. The extensive literature on this topic has incorporated a variety of statistical techniques and a wide range of explanatory variables.[1] However, an obvious source of default-relevant information has not been fully exploited by researchers—the credit default swap (CDS) spread. In a CDS contract, the buyer is offered default protection on the underlying corporate bond, while the seller collects a series of payments determined by the CDS spread. Therefore, the CDS spread should reflect market expectation of the likelihood of corporate default.[2]

The main reason why such a simple idea has not been entertained in the default prediction literature is likely pragmatic—firms with CDS trading tend to be large and investment-grade (Saretto and Tookes, 2013; Subrahmanyam et al., 2014). These constitute only a subset of publicly traded firms, and defaults among such firms could be less frequent than in the general population. However, we identify 128 defaults between January 2001 and December 2020 among 760 unique U.S. CDS reference entities with nearly 100,000 firm-month observations. In comparison, Duffie et al. (2007)'s sample consists of 497 defaults among 2,770 industrial firms in CRSP/Compustat, covering 390,000 firm-month observations over 1980-2004. Although our CDS sample is approximately a quarter of the size of the unrestricted sample, the default rates appear to be similar across the two.[3] In any case, the cumulative coverage of CDS firms and their defaults over the last two decades seem substantial enough to warrant a serious look at the issue.

In this paper, we provide comprehensive empirical evidence on the predictive power of

[1]Representatives include Altman (1968)'s discriminant analysis using accounting variables, Shumway (2001)'s hazard model and Campbell et al. (2008)'s dynamic logit model with accounting and market-based variables, and Duffie et al. (2007)'s maximum likelihood estimation of the term structure of conditional default probabilities, incorporating the time-series dynamics of macro and firm-specific covariates.

[2]Longstaff et al. (2005), for example, use default intensities estimated from CDS spreads to help isolate the non-default component of corporate bond yield spreads.

[3]One could argue that even if the default rates are similar, the larger defaults in our CDS sample make a more significant impact on the average bond investor's portfolio, in the same way that larger stocks are more representative of the CAPM market portfolio.

single-name CDS spreads for corporate default and bankruptcy events. We estimate logistic regressions by adding the five-year CDS spread to the set of covariates used in benchmark default prediction models of Duffie et al. (2007) and Duan et al. (2012). We find the CDS spread to be statistically and economically significant in the presence of distance-to-default and other covariates, and it retains its significance during the financial crisis and non-crisis sub-periods as well as within sub-samples of financial and non-financial firms. Compared to the prominent distance-to-default, the CDS spread has similar statistical significance when forecasting defaults one month to 12 months ahead, but its economic significance is smaller. Specifically, when setting all covariates to their sample mean values, the conditional default rate according to our logit model is 0.27 percent per year. Reducing the distance-to-default by one standard deviation raises the conditional default rate to 0.82 percent, by far the largest effect. In comparison, increasing the CDS spread by one standard deviation raises the conditional default rate to 0.32 percent, similar to the effect of another significant covariate, net income over total assets.

We conduct a battery of robustness checks of our main result. First, besides the CDS market, default-relevant information can also be found in the corporate bond and equity options markets, which have the advantage of covering more firms. In the logit analysis, we find no significance for the at-the-money implied volatility or the implied volatility skew in the presence of distance-to-default and the CDS spread. The bond spread is marginally significant, consistent with Han and Zhou (2014).[4] However, the size and significance of the CDS spread coefficient are qualitatively unchanged. Second, when adding other covariates such as the investment-grade CDS index spread, the slope of the CDS curve, or firm leverage and asset volatility in lieu of distance-to-default, the significance of the CDS spread remains robust. Lastly, our main result also survives research designs intended to mitigate the rare event bias found in the estimation of logit models.

[4]Using a larger sample of firms with corporate bonds, Han and Zhou (2014) show that information asymmetry measures extracted from corporate bond transaction data have significant predictive power for corporate default in the presence of distance-to-default.

To further shed light on the source of the CDS spread's predictability for corporate default, we follow the regression-based approach of Dick-Nielsen et al. (2012) and Schwert (2017) to decompose the CDS spreads at different maturities into default and liquidity components, recognizing that CDS liquidity can directly account for a portion of the CDS spread (Bongaerts et al., 2011). Then, we fit a one-factor reduced-form credit risk model to the term structure of the estimated default components, to further break them down into a credit risk premium plus a "pseudo-spread" following Pan and Singleton (2008) and Friewald et al. (2014), where the pseudo-spread is related to the dynamics of the risk-neutral default intensity under the physical measure. Substituting the five-year CDS spread with its three components in the logit regression shows that only the pseudo-spread (or what we call the "physical default component") contributes to default prediction, while the liquidity and risk premium components do not.[5] Out-of-sample, the four covariates in Duffie et al. (2007), including distance-to-default, yield an accuracy ratio of 79 percent for six-month-ahead forecasts. This improves to 83 percent when the physical default component of the CDS spread is included.

The remainder of our paper is organized as follows. Section 2 reviews the relevant literature and outlines our incremental contribution. Section 3 introduces the data sources and summarizes the explanatory variables used in our empirical analysis. Section 4 presents the in-sample estimation and out-of-sample predictive performance of our model including the single-name CDS spread or its components as explanatory variables. Section 5 concludes.

2 Related Literature

The literature on corporate default prediction is too large to be thoroughly reviewed here. Instead, we refer readers to the brief but excellent summary in Duffie et al. (2007). The

[5]Unlike the case of corporate bonds, where bond illiquidity can amplify rollover loss and accelerate endogenous default by equityholders (He and Milbradt, 2014), there is no obvious link between the liquidity of CDSs, a derivative instrument, and the default risk of the reference entity. To the extent that the liquidity and risk premium components make the CDS spread a noisy proxy for default probabilities, the coefficient of the CDS spread in our logit regressions could be downward biased.

two major focuses of the literature have been methodological improvements and the choice of explanatory variables. For instance, Duffie et al. (2007) model corporate defaults as doubly stochastic Poisson processes, and provide maximum likelihood estimates of the default intensity, the intensities of other forms of exits, and the time-series dynamics of the explanatory variables. This information is then used to generate a term structure of conditional default probabilities. Duan et al. (2012) model the "forward default intensity" instead of the instantaneous default intensity in Duffie et al. (2007), and argue that their approach has certain advantages such as computational efficiency and more accurate long-horizon default prediction.

While the first-generation models, such as Altman (1968), offer single-period classifications using mostly accounting information, the more recent default prediction models increasingly rely on the dynamic behavior of market-based variables. For example, Duffie et al. (2007) estimate a parsimonious specification with four covariates: the Merton distance-to-default, the trailing one-year individual stock return and stock index return, and the three-month Treasury bill rate. To this specification, Duan et al. (2012) further add net income over total assets, firm size, cash holdings, market to book, idiosyncratic stock volatility, as well as the trend of some of these variables. The most prominent out of all these explanatory variables is perhaps distance-to-default, due to its theoretical underpinning in the Merton (1974) structural credit risk model. A thorough investigation of the nature of its contribution to default prediction is undertaken by Bharath and Shumway (2008).

Relative to this literature, our paper does not offer any methodological advance. Instead, we note that the market-based variables used in this literature, in particular distance-to-default, contain information exclusively from the equity market. One of our contributions is in the inclusion of credit market information, e.g., CDS spreads and/or corporate bond yield spreads, in routine default prediction models such as conditional logit regressions. Another contribution is the decomposition of the CDS spread into its components and showing that only the physical default component matters to default prediction.

When markets are fully integrated and models are correctly specified, either the equity-based distance-to-default from the Merton model or the CDS-based physical default component from a reduced-form model would suffice for default prediction, with no need for any other covariates. In the real world, however, there is plenty of evidence of model misspecification and segmentation between the equity and CDS markets.[6] As a result, combining equity- and CDS-based default risk measures, as well as other firm characteristics, is likely to be a fruitful approach to corporate default prediction.

In a recent study, Avino et al. (2019) construct a sample of annual data on 60 U.S. and European banks with CDS trading over 2004-12, out of which 20 failed in 2008 alone and a total of 31 failed during the sample period. They show that the yearly change of the CDS spread has predictive power for bank failure in the following year. However, the out-of-sample accuracy ratio of a specification that uses distance-to-default alone dominates that of any specification that employs the CDS spread change (even when it is combined with distance-to-default). In comparison, our sample size is much larger and includes both financial and non-financial firms. Our explanatory variables are chosen to be compatible with benchmark studies of U.S. industrial and financial firm defaults, as opposed to those unique to banks.[7] More importantly, the use of data at the monthly frequency allows us to identify the robust predictive power of CDS spreads for corporate default over forecast horizons between one month and a year.

Interestingly, Bharath and Shumway (2008) extract risk-neutral default probabilities from CDS spreads and show that they are correlated with default probabilities implied from various implementations of the Merton model (both the "correct" version and a "naive"

[6]For example, Bharath and Shumway (2008) present evidence showing that the Merton distance-to-default is not a sufficient statistic for the probability of default. Elkhami et al. (2014) demonstrates that assuming a constant 40 percent recovery rate in reduced-form models results in underestimated default probabilities. Hilscher et al. (2015) argue that equity returns generally lead CDS returns due to lower trading costs in the equity market, but Acharya and Johnson (2007) find the opposite ahead of major credit events, which they attribute to insider trading in the CDS market.

[7]For example, Avino et al. (2019) include the tier-one regulatory capital ratio, the ratio between loan loss provisions and the book value of total assets, and the ratio between liquid assets and the sum of total deposits and short-term borrowing in their analysis.

one). This is similar in spirit to but not exactly the same as what we are doing in this paper, which is predicting actual defaults using the CDS spread or the physical default component of the CDS spread. Bharath and Shumway (2008) are further limited by the lack of data in the early phase of the CDS market—they obtain only a small sample of 3,833 firm-months of CDS spreads from December 1998 to July 2003. This is less than four percent of our CDS sample size and insufficient for the accurate estimation of a default prediction model.

3 Data

3.1 CDS Firms and Their Defaults

The major source of default events in our empirical study is Moody's Default and Recovery Database (DRD). The DRD tracks corporate default events at the financial instrument level and issuer level in the U.S. market since 1926. These corporate default events include missed interest or principal payments, distressed exchanges, and bankruptcy filings under Chapter 7 or Chapter 11. We link the DRD records with firm characteristics in other academic databases using exchange ticker symbols and six-digit CUSIP numbers, and confirm that the issuer name in the DRD is consistent with the firm's legal name provided in Compustat. The DRD records are then combined with default events in two other databases, Thomson Reuters' Eikon and Audit Analytics. For the defaults of CDS firms, we cross-validate the DRD records with the Creditfixings website, which provides credit event auctions data, as well as checking that the firms have active CDS trading around the time of default.

[Insert Table 1 here]

The pricing information on single-name CDSs is retrieved from Markit, a leading credit market data service provider. We use the five-year CDS spread as the key independent covariate in the empirical study, as the five-year contract is usually the most liquid maturity on the single-name CDS spread curve. A firm is counted as an actively traded CDS firm in a given month if there are at least 15 daily five-year CDS spreads observed during the month.

From Table 1, the number of actively traded CDS firms reached a peak in 2007 and gradually fell thereafter.[8] We identify and match 128 default/bankruptcy events among these actively traded CDS firms. Table 1 also reports defaulted CDS firms in number and as a percentage of the number of actively traded CDS firms in each calendar year.[9] Evidently, the number of defaulted CDS firms spiked during the financial crisis years of 2007-09, which accounted for 35 percent of all CDS defaults in our sample. It also rose in 2020, perhaps due to the Covid-19 pandemic, comprising 11 percent of all CDS defaults.

3.2 Covariates

In the empirical analysis, we perform logistic regressions with the binary indicator of whether a firm has defaulted in a given month as the dependent variable, and the five-year CDS spread is used as the main source of credit market information for default prediction. For comparison purposes, we borrow the set of independent variables from Duffie et al. (2007) and Duan et al. (2012). These variables are documented below, with the first four from Duffie et al. (2007) and the rest from Duan et al. (2012), who also include distance-to-default.[10]

1. Distance-to-default (DtD): The explanatory and predictive power of this popular covariate for default events has both theoretical justification from the Merton model and empirical support from a large literature including Duffie et al. (2007), Bharath and Shumway (2008), and Duan et al. (2012). We expect it to have a strong negative relation with default events, i.e., a smaller DtD corresponding to a higher default probability. We adopt the standard KMV approach for calculating DtD, following Crosbie and Bohn (2002), Vassalou and Xing (2004), and Hillegeist et al. (2004).

[8] We average the monthly number of active CDS firms to obtain a yearly figure.

[9] Not all of the actively traded CDS firms reported in Table 1 are included in our later empirical analysis because of the requirement of additional variables. Our baseline logit regressions, for example, include around 100,000 firm-month observations with 760 unique CDS firms and all 128 CDS defaults.

[10] In univariate logit regressions with a six-month forecast horizon, all of the listed covariates are significant in predicting default except for the individual stock return, risk-free rate, and firm size. In bivariate regressions with the presence of the five-year CDS spread, all covariates are significant except for the individual stock return, risk-free rate, and firm size. In bivariate regressions with the presence of distance-to-default, all covariates are significant except for the individual stock return and risk-free rate.

2. Individual stock return: This is the monthly return on a firm's stock, which correlates negatively with the market leverage of the firm and its default probability.[11] The firm-level stock return also reflects the idiosyncratic risk of the firm.

3. Stock index return: The monthly S&P 500 index return serves as a systematic risk measure, and is expected to have a negative relation with default events.

4. Risk-free rate: The three-month Treasury bill rate is used to incorporate business cycle information into the empirical framework. It is easier for firms to refinance their existing debt at a lower cost when interest rates are low. On the other hand, lower interest rates may signal recessions, which are coincidental with more frequent default events.

5. Net Income/Asset ratio (ROA): This is the ratio between a firm's net income and total assets, which is negatively related to its default probability.

6. Firm size: This is the logarithm of a firm's total assets, where larger firms tend to be less likely to default.

7. Cash/Asset ratio: The ratio between the sum of cash, cash equivalents, and short-term investments and total assets. We expect it to be negatively related to the default probability.

8. Market-to-book ratio: The ratio of the market value of assets to the book value of assets. Campbell et al. (2008) find that the market-to-book ratios of bankrupt firms are somewhat higher than those of healthy firms, which they attribute to overvaluation.

9. Idiosyncratic stock volatility (IVOL). Following Bali and Cakici (2008), IVOL is estimated by regressing daily individual stock returns on the Fama-French three factors using current month data, and then computing the annualized standard deviation of

[11] Duffie et al. (2007) use the trailing one-year stock return, which is more stable but contains high serial correlation at the monthly frequency due to its overlapping structure. We find that the monthly stock return is more dynamic and better at capturing default-related information from the stock market.

the regression residuals. IVOL is expected to correlate positively with the likelihood of default and bankruptcy.

Besides these main covariates, we use other variables to examine the robustness of the predictive relation between the CDS spread and corporate default. These variables include:

- Corporate bond yield spread: The corporate bond yield spread and the CDS spread both contain information about a firm's credit risk. While corporate bonds have lower liquidity and less favorable tax treatment relative to Treasury bonds (Elton et al., 2001; Longstaff et al., 2005; Bao et al., 2011), causing these non-default factors to enter the corporate bond yield spread, the CDS spread is a direct measure of default risk and thus should contain better information for default prediction. There is also evidence of the CDS spread consistently leading the bond spread in price discovery (Blanco et al., 2005). To construct the corporate bond yield spread, we take corporate bond transaction data from the enhanced TRACE database for the period of July 2002 to December 2020. We follow Bai et al. (2019) to remove bonds with embedded options (e.g., callable, putable, and convertible features) and floating coupon rates, keeping only straight bonds. We estimate monthly corporate spot curves for each issuer, requiring price observations on at least three different bonds each month.[12] The corporate bond yield spread is the difference between the corporate spot rate and the Treasury spot rate at the five-year maturity.

- Option-implied volatility and implied volatility skewness: Information from individual stock options has been shown to explain bond and CDS spreads (Cremers et al., 2008; Cao et al., 2010). Using options data from Bloomberg and Thomson Reuters' Eikon, we

[12]In the early years of TRACE coverage, the number of corporate bonds is small. Therefore, to preserve as much sample period as possible, we estimate the Nelson-Siegel (Nelson and Siegel, 1987) yield curve with its α parameter set to zero when there are price observations on only three different bonds during a month. Note that the Nelson-Siegel spot rate function is $r(m) = \alpha + \beta_1 \left[\frac{1-\exp(-m/\lambda)}{m/\lambda}\right] + \beta_2 \left[\frac{1-\exp(-m/\lambda)}{m/\lambda} - \exp(-m/\lambda)\right]$, where m denotes maturity. The full Nelson-Siegel yield curve is estimated when the number of bonds is greater than or equal to four, and the Nelson-Siegel-Svensson (Svensson, 1994) yield curve is estimated when the number of bonds is greater than or equal to six.

construct at-the-money implied volatility (ATM IV) as the volume-weighted average implied volatility from a firm's call options with moneyness between 0.95 and 1.05. The IV sknewness is the difference between the volume-weighted average implied volatility from a firm's put options with moneyness between 0.8 and 0.95 and the previously defined ATM IV. Both should be positively related to the default probability.

- CDS market index spreads: While single-name CDS spreads contain information about individual firms' credit risk, CDS market indices provide information about systematic credit risk, which can also prove useful to default prediction. Therefore, we calculate equally-weighted averages of five-year CDS spreads of the top 125 firms with S&P's long-term issuer credit rating higher than or equal to BBB- for the investment-grade (IG) CDS index, and the top 100 firms below BBB- for the high-yield (HY) CDS index, where we define a top firm in terms of CDS market depth.[13]

- CDS market depth: This is the number of dealers supplying five-year CDS quotes on a reference entity according to Markit, and is used as a CDS market liquidity measure (Qiu and Yu, 2012). In Section 4.4.1, we will show how to use this variable to estimate the liquidity component of the CDS spread. To the extent that the CDS spread contains a sizable liquidity component, purging it from the CDS spread may enhance the predictive power of the CDS spread for corporate default.

- CDS slope: This is defined as the ten-year CDS spread minus the one-year CDS spread. We expect the CDS curve to be more positively sloped when the CDS spread level is lower. In other words, there is a negative correlation between the CDS slope and the five-year CDS spread. More importantly, the CDS slope may contain information not found in the spread level that is useful for predicting future default.

- Firm leverage and asset volatility: The procedure used to calculate DtD, as in Crosbie

[13]In the summary statistics and later empirical analysis, we leave out the CDS HY index spread because it has a correlation of 0.93 with the IG index spread. Including both is likely to cause multicollinearity issues.

and Bohn (2002), Vassalou and Xing (2004), and Hillegeist et al. (2004), also produces estimates for firm leverage and asset volatility. Specifically, firm leverage is defined as the sum of short-term debt and 50 percent of long-term debt over the estimated firm value, and asset volatility is simply the annualized standard deviation of the firm value. Instead of combining these two estimates into DtD, one might consider including firm leverage and asset volatility as separate variables for default prediction.

[Insert Table 2 here]

3.3 Summary Statistics

The main covariates are available over approximately 102,000 firm-month observations in our sample and are summarized in Table 2. The five-year CDS spread averages 190 basis points and has a median value of 87 basis points, indicating that the spread is heavily skewed to the right. This positive skewness also characterizes other variables, such as the IG CDS index spread and the corporate bond yield spread, which averages around 200 basis points higher than the CDS spread. The average CDS firm in our sample has total assets of $14.3 billion (the logarithm of total assets, in units of $1 million, is equal to 9.57), a market-to-book ratio of 1.65, net income equal to 4.31 percent of total assets, cash holdings equal to 9.48 percent of total assets, an annualized idiosyncratic stock volatility of 22.2 percent, a monthly stock return of 0.97 percent, and a distance-to-default equal to 4.01. It also has 5.8 dealers providing quotes on its five-year CDS contracts.

[Insert Table 3 here]

Turning now to the correlations among the covariates in Table 3, we first notice that the the CDS spread and the corporate bond yield spread, with a correlation of only 0.35, are separated by a significant "CDS-bond basis" (Bai and Collin-Dufresne, 2019). The CDS slope has a strong negative correlation of -0.32 with the CDS spread, suggesting that the CDS curve tends to be more flat when the credit risk level is higher. The CDS spread is negatively correlated with the key distance-to-default measure, but its largest correlations are found to be with the two equity volatilities, ATM IV and IVOL, at 0.67 and 0.52, respectively. The distance-to-default, on the other hand, has its largest correlation with the market-to-book, at 0.49, while its correlation with ATM IV and IVOL are only −0.40 and −0.30, respectively. Therefore, even among the same set of inputs, different nonlinear functional forms between distance-to-default and the CDS spread can emphasize different inputs, resulting in differential predictive power for default events. Finally, we note that the CDS spread has a correlation with the CDS market depth equal to −0.12, consistent with the early evidence from Qiu and Yu (2012).

4 Empirical Results

4.1 Baseline Regressions

We first conduct logistic regressions to assess the in-sample explanatory power of the main covariates listed in Section 3.2 for corporate default and bankruptcy (1 for a credit event within the month and 0 otherwise). We are particularly interested in whether the addition of five-year CDS spreads increases model performance. Furthermore, we vary the forecast horizon by lagging all independent variables by one, three, six, nine, and 12 months. This is similar to an exercise performed in Campbell et al. (2008).

[Insert Table 4 here]

Table 4 summarizes the coefficient estimates and their statistical significance.[14] The independent variables include the five-year CDS spread combined with the four covariates from Duffie et al. (2007) and the five additional covariates from Duan et al. (2012). Focusing first on Column (3) with a medium forecast horizon of six months, we find that only four covariates—the CDS spread, DtD, NI/TA, and Cash/TA—have statistically significant coefficients. Importantly, the CDS spread is significant in the presence of DtD and other

[14]We also perform Poisson regressions mainly as a robustness check with an alternative functional form for the conditional default probability. Results are qualitatively similar and are available upon request.

covariates. In results not presented here, we compare the default prediction model with the full set of covariates, either with or without the five-year CDS spread. Adding the CDS spread increases the Pseudo R^2 from 0.1296 to 0.1457 at the six-month forecast horizon, and slightly reduces the magnitude and statistical significance of the other covariates.

Turning to the issue of economic significance, we evaluate the marginal effects when all covariates are set to their sample mean values—under this scenario, the annualized probability of default is 0.27 percent, much lower than the unconditional probability of default of 1.57 percent in our sample.[15] Out of the four significant covariates, the DtD has the most economic significance by far—a one standard deviation decrease in DtD more than triples the annualized default probability to 0.82 percent. In comparison, a one standard deviation decrease in NI/TA raises the default probability to 0.34 percent. For the CDS spread, a one standard deviation increase moves the default probability from 0.27 percent to 0.32 percent, a similar effect to the one associated with NI/TA, but much smaller than the effect of a DtD change. In relative terms, this is roughly a 20 percent increase in the conditional default probability.

Looking across the forecast horizon in different columns of Table 4, there are only two covariates with consistently significant coefficients in default prediction—DtD and the CDS spread. Other covariates may have coefficients with the expected sign, but they are significant in at most two or three out of five cases. As the forecast horizon becomes longer, we notice that the pseudo R^2 declines, suggesting that it is more difficult to forecast default events further ahead. At forecast horizons of nine and 12 months, the statistical significance of the CDS spread coefficient is notably less than at shorter horizons, although its magnitude remains fairly constant with respect to the forecast horizon. For the DtD coefficient, both the magnitude and statistical significance are insensitive to the forecast horizon. In the

[15] We divide the number of defaults in our sample, 128, by the total number of firm-months in Column (3) of Table 4, 96,956, resulting in a default probability of $p = 0.00132$ in any given month. The annualized default probability is then $1-(1-p)^{12} = 0.0157$. This is higher than the predicted default probability when covariates are set to their sample means, likely because the default probability is a convex function of the covariates.

remainder of our analysis, we seek to better understand the preceding baseline results. We will focus on a forecast horizon of six months, as it is short enough for the CDS spread to make a significant contribution to default prediction, but not so short as to be compared with "predict(ing) a heart attack by observing a person dropping to the floor clutching his chest" (Campbell et al., 2008).

4.2 Firm Types and Market Environment

A quick browse of the default prediction literature suggests that model specification could be sensitive to the types of firms and the sample period included in the analysis. For example, Milne (2014) points out that the distance-to-default measure failed to predict defaults and stock price declines of financial firms (mainly banks) during the great financial crisis. Duffie et al. (2007) exclude financial firms from their study and estimate their default prediction model using a sample of industrial firms, whereas Duan et al. (2012) find accurate default prediction performance for both industrial and financial firms in the U.S. using their proposed model. Avino et al. (2019) predict U.S. and European bank failures, which occurred mostly during the great financial crisis, using CDS spreads and DtD. Therefore, we check the robustness of our default prediction model with respect to firm types and market conditions in this subsection.

[Insert Table 5 here]

First, we partition the full sample into financial firms (with four-digit SIC codes of 6000-6799) and industrial firms. Columns (1) and (2) in Table 5 report 10 financial defaults and 118 non-financial defaults, with about 12,000 firm-month observations in the financial sub-sample and 85,000 in the non-financial sub-sample. Even though the financial sub-sample is around 12 percent of the total sample size, with only eight percent of the default events, the coefficients of the CDS spread and DtD for financial firms are similar in size to those for non-financial firms, with lower but still significant t-statistics. The pseudo R^2 is much higher among financial firms compared to non-financial firms. These results suggest that we have no reason to preclude financial firms from our analysis.

Second, we divide the sample period into the financial crisis (January 2007 to December 2009) and non-crisis periods. The financial crisis period contains around one third of all default events (45 out of 128 defaults), even though the period only covers about 19 percent of the observations.[16] Columns (3) and (4) present the empirical results for the crisis and non-crisis sub-samples. As it turns out, DtD and the CDS spread are the only covariates to maintain statistical significance across the two sub-samples. A formal test is not able to reject the null hypothesis that these two covariates have equal coefficients between the crisis and non-crisis sub-samples.

Third, we separate the firms according to their S&P's long-term issuer credit ratings into investment-grade and high-yield sub-samples. To the extent that investment-grade firms have low credit risk, their default can be unexpected and difficult to predict. In contrast, high-yield firms may have ongoing concerns about their financial health, implying that there are useful signals from the market about impending failure. In Column (5), we find that for IG firms (around 55 percent of the total sample), there are only nine defaults, reflecting their low ex ante credit risk. Moreover, DtD is the only significant covariate, and the pseudo R^2 is in the single digit, confirming the difficulty in predicting IG defaults. Column (6) paints the opposite picture for HY firms. With only 45 percent of the total sample but virtually all of the defaults, the coefficients are quite similar to the baseline results for the whole sample, presented as Column (0). Notably, both DtD and the CDS spread are highly significant.

Collectively, these findings show that the CDS spread's predictive power for corporate default is quite robust to firm types (financial vs. non-financial) and market conditions (crisis vs. non-crisis). Moreover, it appears to be more effective at predicting HY than IG defaults,

[16]Das et al. (2007) and Duffie et al. (2009) offer evidence, based on a sample period of 1980-2004, that U.S. corporate defaults are more correlated than what can be explained with observable covariates, and proceed to extract a common latent factor driving corporate defaults. The extreme level of default clustering during the great financial crisis offers an opportunity to further validate their approach. However, our focus in this paper is on default prediction, not explaining default correlation. We leave the issues of default clustering and default correlation to future research.

with the caveat that IG defaults are extremely rare (hence difficult to predict).

[Insert Table 6 here]

4.3 Robustness Checks

In this subsection, we conduct a few robustness checks of our baseline regression results, which are presented as Column (0) in Table 6 for reference. These robustness checks utilize additional covariates whose definitions can be found in Section 3.2. The ultimate question is whether the addition of these covariates changes the CDS spread's significant incremental contribution to default prediction.

First, we consider adding the CDS slope, defined as the 10-year spread minus the one-year spread, to supplement the information provided by the five-year spread. Just as the slope of the Treasury yield curve may help forecast future risk-free interest rates, the slope of the CDS spread curve may contain information about corporate default risk in the future. On the other hand, there is strong empirical evidence that the movement of CDS spread curves is dominated by a single latent factor (Pan and Singleton, 2008). As Column (1) of Table 6 shows, the CDS slope is not significant when added to the baseline specification.

Second, we consider the effect of including systematic credit risk measures in the baseline specification. If the single-name CDS spread explains corporate default mostly because of the systematic credit risk embedded in the CDS spread, then its coefficient would diminish in size and significance in the presence of systematic credit risk variables. Specifically, we incorporate the CDS IG index spread, but Column (2) shows that it is not significant.

Third, we note that CDS spreads are not the only market-based source of information on firm-level credit risk besides DtD—long before the CDS market developed, corporate bonds were traded and they are actually available for a much broader set of firms. Therefore, we examine corporate bond yield spreads as a competing source of default relevant information in the presence of CDS spreads. Column (3) shows that the five-year bond spread is marginally significant, with a coefficient less than one third the size of the five-year CDS spread coefficient, even though the sample standard deviation of the bond spread is comparable to that of the CDS spread. Therefore, the economic significance of the bond spread is weaker, though it does contributes incrementally to default prediction beyond DtD and the CDS spread.

Fourth, by the same logic, we introduce information from the equity options market into the baseline regression. Specifically, we add the ATM IV and IV skewness, since both variables can be informative about the left tail of a firm's stock return distribution. The crucial question here is whether they offer incremental predictive power for corporate default in the presence of the two main covariates, DtD and the CDS spread. Column (4) shows that they do not.

Fifth, we include firm leverage and asset volatility as separate variables, rather than combining them into DtD. Since these two inputs are needed to compute DtD, there is no sample attrition compared to the baseline regression. In Column (5), we find that firm leverage is significant and asset volatility is not, but more importantly, replacing DtD with these two variables in a linear specification results in a smaller pseudo R^2. Apparently, this does not enhance the predictive power of DtD.

In all of these robustness exercises, the CDS spread and DtD coefficients remain qualitatively the same as in the baseline case. The same conclusion holds when we combine all of the additional covariates into one regression in Column (6)—while their t-statistics decline, the CDS spread and DtD coefficients remain significant and their magnitude does not decrease.

[Insert Table 7 here]

Lastly, we address the issue of rare event bias with maximum likelihood estimation of logit models. Three different approaches are used: 1) the bias correction proposed by King and Zeng (2001); 2) the penalized maximum likelihood estimation of Firth (1993); and 3) randomly sampling 10 percent of the non-default observations to artificially boost the frequency of defaults. As Table 7 shows, these approaches do not alter the inference of CDS spread and DtD effects in any significant way compared to the baseline case.

4.4 CDS Spread Decomposition

In this subsection, we dig deeper into the source of the CDS spread's contribution to default prediction. Our specific approach is to decompose the CDS spread into liquidity, physical default, and default risk premium components and examine the separate contribution of each component to default prediction. We describe our two-step decomposition procedure below and then present the empirical results.

4.4.1 Estimating the Liquidity and Default Spreads

In Step 1, we decompose the CDS spread into default and liquidity components follows Dick-Nielsen et al. (2012) and Schwert (2017). Specifically, we construct a standardized illiquidity variable for each CDS firm-month as:

$$\lambda_{i,t} = \frac{L_{i,t} - \mu_L}{\sigma_L},$$

where $L_{i,t}$ is the illiquidity measure for firm i in month t, and μ_L and σ_L are its average and standard deviation over the full sample, respectively. The illiquidity measure is inversely related to the CDS market depth, namely, $L_{i,t} = \bar{N} - N_{i,t}$, where $N_{i,t}$ is the number of dealers providing quotes on five-year CDS contracts for firm i in month t, and $\bar{N}$ denotes the maximum number of quote providers across all firm-months.

In March 2009, a collection of changes in the single-name CDS market were announced together as the so-called "CDS Big Bang Protocol" (Markit Group Limited, 2009), which were intended to standardize CDS contracts in preparation for central clearing.[17] Since the CDS Big Bang significantly changed the liquidity of the CDS market, we divide our sample into two halves. The first half is before March 2009, where the maximum number of quote providers is $\bar{N} = 31$. The second half is after March 2009, where $\bar{N} = 14$.[18] We compute L, μ_L, σ_L, and λ for each sub-sample separately.

[17] Central clearing in the single-name CDS market began in stages, starting from December 2009. See Loon and Zhong (2014) and Marra et al. (2019) for details.

[18] It is likely that the CDS Big Bang and the move toward central clearing raised the cost of market-making,

We then estimate the following CDS spread regression specification:

$$\text{CDS spread}_{i,t} = \alpha + \beta \cdot \lambda_{i,t} + \gamma' \cdot \text{Credit risk controls}_{i,t} + \epsilon_{i,t}. \quad (1)$$

We follow the literature to include the stock return, stock return volatility, firm size, leverage, Altman's Z-score, and ROA as credit risk controls. The above regression is estimated for investment-grade and high-yield CDS reference entities separately, and both before and after the CDS Big Bang in March 2009. We define the CDS liquidity spread as $\beta \cdot (\lambda_{i,t} - \lambda_{5\%,t})$, where $\lambda_{5\%,t}$ is the fifth percentile of the standardized illiquidity variable in month t, representing the time-varying baseline for a highly liquid CDS reference entity. The default spread is calculated by subtracting the liquidity spread from the total CDS spread.

[Insert Table 8 here]

The regression Eq. (1) is estimated for each of the 11 CDS maturities separately (6 months, 1, 2, 3, 4, 5, 7, 10, 15, 20, and 30 years), thus providing the liquidity and default spreads for each maturity. Panel A of Table 8 summarizes the market depth and the results of this estimation for the five-year CDS spread. There is evidently a shift in the distribution of the number of quote providers to the left after the CDS Big Bang. However, the CDS spread is much less sensitive to changes in the standardized illiquidity variable after the CDS Big Bang (the beta coefficient is much smaller), resulting in liquidity spreads that are around three times smaller—the average liquidity spread for IG (HY) firms is 22 (75) bps before the Big Bang and 6 (23) bps after. The percentage of the CDS spread attributed to liquidity also decreases from around 30 percent before the Big Bang to around 15 percent after. These results are consistent with the CDS market becoming more liquid after contract standardization (Loon and Zhong, 2014).

thus favoring large dealers and driving out the small. A similar drop in the number of quote providers was observed in 2006, when the Fed sent a letter to major dealers in the CDS market requiring them to reduce the backlog in trade processing (Qiu and Yu, 2012). Although the number of CDS quote providers decreased after the Big Bang, our analysis below shows that the liquidity component of the CDS spread is much lower during the second half of the sample period, consistent with an improvement in liquidity.

4.4.2 Estimating the Physical Default and Risk Premium Components

In Step 2, we further decompose the estimated default spread into physical default and risk premium components. Specifically, we follow Longstaff et al. (2005), Chun et al. (2019), and Jarrow et al. (2019) to assume that the risk-neutral default intensity h_t of a firm is described by a CIR process:

$$dh_t = (\alpha - \beta h_t)\, dt + \sigma \sqrt{h_t} dw_t^{\mathbb{Q}}, \tag{2}$$

where $w_t^{\mathbb{Q}}$ is a Wiener process under the $\mathbb{Q}$ measure. Adopting the extended affine specification for the market price of risk (Cheridito et al., 2007), we model h_t's $\mathbb{P}$ measure dynamics as

$$dh_t = \left(\alpha^{\mathbb{P}} - \beta^{\mathbb{P}} h_t\right) dt + \sigma \sqrt{h_t} dw_t^{\mathbb{P}}. \tag{3}$$

The relation between the Wiener processes under the two measures is given by

$$w_t^{\mathbb{P}} = w_t^{\mathbb{Q}}(t) + \frac{\alpha - \alpha^{\mathbb{P}}}{\sigma} \int_0^t \frac{1}{\sqrt{h_s}} ds - \frac{\beta - \beta^{\mathbb{P}}}{\sigma} \int_0^t \sqrt{h_s} ds. \tag{4}$$

The CDS spread at time t for protection between t and $t + \tau$ satisfies

$$S_t^{\tau} = \frac{\int_t^{t+\tau} P(t,u)\, \mathbb{E}_2(t,u) \left\{ y - S_t^{\tau} \left(u - \frac{\llcorner 4u \lrcorner}{4}\right) \right\} du}{\frac{1}{4} \sum_{i=1}^{4\tau} P\left(t, t+\frac{i}{4}\right) \mathbb{E}_1\left(t, t+\frac{i}{4}\right)}, \tag{5}$$

where y is the fractional loss given default $(0 < y < 1)$,[19] $P(t,T)$ is the time t price of a risk-free zero-coupon bond that matures at time T, $S_t^{\tau}\left(u - \frac{\llcorner 4u \lrcorner}{4}\right)$ reflects the accrued CDS premium from the previous payment date to the time of default, with $\llcorner 4u \lrcorner$ denoting the largest integer smaller than $4u$, and

$$\mathbb{E}_1(t,u) = \mathbb{E}^{Q}\left[\exp\left(- \int_t^u (c_0 + h_s)\, ds \right) \middle| \mathcal{F}_t \right] = A(u-t) e^{B(u-t)(c_0 + h_t)}, \tag{6}$$

$$\begin{aligned} \mathbb{E}_2(t,u) &= \mathbb{E}^{Q}\left[\exp\left(- \int_t^u (c_0 + h_s)\, ds \right) (c_0 + h_s) \middle| \mathcal{F}_t \right] \\ &= \left[G(u-t) + H(u-t)(c_0 + h_t) \right] e^{B(u-t)(c_0 + h_t)}. \end{aligned} \tag{7}$$

[19] In the estimation, we use one minus the recovery rate provided in the Markit data for y. If the recovery rate is missing, y is set to 60 percent.

Here, following Duffee (1999), the constant c_0 is added to improve the fit to the data. A, B, G, and H are functions whose formulae can be found on page 2,221 in Longstaff et al. (2005).

This one-factor model is then calibrated to the default spreads at the 11 different maturities. The model parameters (under both $\mathbb{P}$ and $\mathbb{Q}$) along with the default intensity are obtained using the unscented Kalman filter (Christoffersen et al., 2014) in conjunction with maximal likelihood estimation. Then, following Pan and Singleton (2008) and Friewald et al. (2014), taking the filtered default intensity combined with the $\mathbb{P}$ parameters to the CDS pricing function produces the physical default component, while the risk premium component is the difference between the default spread and the physical default component.

4.4.3 Which Component Predicts Default?

Panel B of Table 8 provides the percentage breakdown of the CDS spread, by maturity, into the liquidity, physical default, and risk premium components. Several interesting patterns are worth noting. First, the liquidity component has an inverse-U pattern across maturity, reaching a minimum at the five-year maturity, where the liquidity spread is less than 20 percent of the total CDS spread. In contrast, close to 70 (40) percent of the one-year (30-year) CDS spread is attributed to liquidity. Another interesting pattern is that the percentage of the CDS spread attributed to the default risk premium is rising with maturity. This is because with a one-factor diffusion-based default intensity model, there is greater deviation between physical and risk-neutral default probabilities when the horizon is longer. Lastly, the percentage of the CDS spread due to physical default reaches a maximum at the five-year maturity, lending support to our choice of using the five-year spread in default prediction (assuming that the physical default component is what provides the predictive power).

[Insert Table 9 here]

To see each component's contribution to default prediction, we refer to Table 9. In

Column (1), we include the default and liquidity spreads as separate regressors. The results show that only the default spread is significant, with coefficient similar to the CDS spread coefficient in the baseline case in Column (0). In Column (2), we include all three components as separate regressors. Here, only the physical default component is significant, while the liquidity and risk premium components are not. Compared to the baseline case, allowing the three components to be separate contributors to default prediction increases the pseudo R^2 from 0.1457 to 0.1675. These results underscore the importance of extracting the "right" information from CDS spreads for predicting corporate default.

4.5 Out-of-Sample Performance

In this last subsection, we perform out-of-sample forecasting of default and bankruptcy events using the CDS spread and other covariates. The training dataset is constructed on a 60-month rolling basis, starting from January 2001. The out-of-sample forecasting begins in January 2006 and ends in December 2020. There are 108 default events during this 180-month sample period. We rank the firms by their predicted one-, three-, six-, nine-, or 12-month forward default probabilities from the logit model. For a given percentile of ranked firms with the worst predicted n-month forward default probabilities, we ask what fraction of default events that actually occurred n-month ahead are included within this subset. The plot of this fraction of defaults as a function of the percentile of predicted default probability is called the power curve. The area under this curve and above the 45 degree line multiplied by two is defined as the accuracy ratio for default prediction. What we summarize below either graphically or in table format are the power curves and accuracy ratios averaged over all rolling five-year training periods.

[Insert Figure 1 here]

Figure 1 shows the power curve across different models and forecast horizons. Four logit regression models are presented here—the first using the four conventional covariates of DtD, firm stock return, S&P 500 index return, and three-month Treasury bill rate from Duffie et al.

(2007) (hereafter DSW07), the second adding the five-year CDS spread, the third replacing the CDS spread with the default spread, and the last replacing the default spread with the physical default component. Graphically, we can see that the out-of-sample performance of the model improves the most with the addition of the physical default component to the DSW07 four-covariate specification.

[Insert Table 10 here]

Table 10 provides more details about the out-of-sample performance. It shows the accuracy ratio as well as several points on the power curve at the 10th, 20th, 30th, and 40th percentiles. At a medium six-month forecast horizon, the accuracy ratio of the DSW07 specification is 79 percent. This contrasts with a higher accuracy ratio of 88 percent for one-year default prediction reported by DSW07. The large difference could be attributed to a number of factors. First, the sample periods are different—we examine 2006-20 while DSW07 1993-2004. Second, we use a five-year rolling window to estimate the model, while DSW07 take all of the data from 1980 up to the point of a forecast. Perhaps most importantly, our sample consists of CDS reference entities, which tend to be larger firms whose default is more infrequent and more difficult to predict, while DSW07's sample contains all publicly traded firms.

Overall, the numbers presented in Table 10 confirm that a larger fraction of default events are correctly classified by our predicted default probabilities when the model specification includes CDS-related information, in particular the physical default component of the CDS spread. Most of the improvement occurs at a lower percentile of the power curve, representing firms with higher predicted default probabilities, the cohort we actually care about. At the six-month forecast horizon, the accuracy ratio increases from 79 percent for the DSW07 specification to 83 percent when the model includes the physical default component of the CDS spread, a four percentage point increase. This level of improvement in the accuracy ratio remains robust across the range of forecast horizons we have considered.

5 Conclusion

We provide the first comprehensive empirical study of the relevance of CDS market information to corporate default prediction using data from 2001-20. The baseline logit regressions show that, across forecast horizons from one month to 12 months, the CDS spread and distance-to-default (DtD) are the only two covariates that consistently maintain their statistical significance. While the economic significance of DtD is the largest by far among all covariates, a one-standard-deviation increase in the CDS spread raises the conditional default probability by about 20 percent. The significance of the CDS spread in default prediction is robust to different firm types and market conditions. It also survives the inclusion of systematic credit risk factors, corporate bond yield spreads, and equity options market information. A decomposition of the CDS spread into liquidity, physical default, and risk premium components confirms that, as expected, only the physical default component contributes significantly to default prediction both in- and out-of-sample. In summary, CDS market information ought to be incorporated into default prediction models whenever such information is available.

References

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

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

Avino, D. E., Conlon, T., and Cotter, J. (2019). Credit default swaps as indicators of bank financial distress. *Journal of International Money and Finance*, 94:132–139.

Bai, J., Bali, T. G., and Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3):619–642.

Bai, J. and Collin-Dufresne, P. (2019). The CDS-bond basis. *Financial Management*, 48(2):417–439.

Bali, T. G. and Cakici, N. (2008). Idiosyncratic volatility and the cross section of expected returns. *Journal of Financial and Quantitative Analysis*, 43(1):29–58.

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

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

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

Bongaerts, D., De Jong, F., and Driessen, J. (2011). Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market. *Journal of Finance*, 66(1):203–240.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *Journal of Finance*, 63(6):2899–2939.

Cao, C., Yu, F., and Zhong, Z. (2010). The information content of option-implied volatility for credit default swap valuation. *Journal of Financial Markets*, 13:321–343.

Cheridito, P., Filipović, D., and Kimmel, R. L. (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, 83(1):123–170.

Christoffersen, P., Dorion, C., Jacobs, K., and Karoui, L. (2014). Nonlinear kalman filtering in affine term structure models. *Management Science*, 60(9):2248–2268.

Chun, A. L., Namvar, E., Ye, X., and Yu, F. (2019). Modeling municipal yields with (and without) bond insurance. *Management Science*, 65(8):3694–3713.

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

Crosbie, P. and Bohn, J. (2002). Modeling default risk. *Technical Report, KMV, LLC.*

Das, S. R., Duffie, D., Kapadia, N., and Saita, L. (2007). Common failings: How corporate defaults are correlated. *Journal of Finance*, 62(1):93–117.

Dick-Nielsen, J., Feldhütter, P., and Lando, D. (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics*, 103(3):471–492.

Duan, J.-C., Sun, J., and Wang, T. (2012). Multiperiod corporate default predictionA forward intensity approach. *Journal of Econometrics*, 170(1):191–209.

Duffee, G. R. (1999). Estimating the price of default risk. *Review of Financial Studies*, 12(1):197–226.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009). Frailty correlated default. *Journal of Finance*, 64(5):2089–2123.

Duffie, D., Saita, L., and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635–665.

Elkhami, R., Jacobs, K., and Pan, X. (2014). The cross section of recovery rates and default probabilities implied by credit default swap spreads. *Journal of Financial and Quantitative Analysis*, 49(1):193–220.

Elton, E. J., Gruber, M. J., Agrawal, D., and Mann, C. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance*, 56(1):247–277.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

Friewald, N., Wagner, C., and Zechner, J. (2014). The cross-section of credit risk premia and equity returns. *Journal of Finance*, 69(6):2419–2469.

Han, S. and Zhou, X. (2014). Informed bond trading, corporate yield spreads, and corporate default prediction. *Management Science*, 60(3):675–694.

He, Z. and Milbradt, K. (2014). Endogenous liquidity and defaultable bonds. *Econometrica*, 82(4):1443–1508.

Hillegeist, S. A., Keating, E. K., Cram, D. P., and Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1):5–34.

Hilscher, J., Pollet, J. M., and Wilson, M. (2015). Are credit default swaps a sideshow? evidence that information flows from equity to cds markets. *Journal of Financial and Quantitative Analysis*, 50(3):543–567.

Jarrow, R., Li, H., Ye, X., and Hu, M. (2019). Exploring mispricing in the term structure of CDS spreads. *Review of Finance*, 23(1):161–198.

King, G. and Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2):137–163.

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

Loon, Y. C. and Zhong, Z. K. (2014). The impact of central clearing on counterparty risk, liquidity, and trading: Evidence from the credit default swap market. *Journal of Financial Economics*, 112(1):91–115.

Markit Group Limited (2009). The CDS Big Bang: Understanding the changes to the global CDS contract and North American conventions.

Marra, M., Yu, F., and Zhu, L. (2019). The impact of trade reporting and central clearing on CDS price informativeness. *Journal of Financial Stability*, 43:130–145.

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

Milne, A. (2014). Distance to default and the financial crisis. *Journal of Financial Stability*, 12:26–36.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60(4):473–489.

Pan, J. and Singleton, K. J. (2008). Default and recovery implicit in the term structure of sovereign CDS spreads. *Journal of Finance*, 63(5):2345–2384.

Qiu, J. and Yu, F. (2012). Endogenous liquidity in credit derivatives. *Journal of Financial Economics*, 103(3):611–631.

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

Schwert, M. (2017). Municipal bond liquidity and default risk. *Journal of Finance*, 72(4):1683–1722.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74(1):101–124.

Subrahmanyam, M. G., Tang, D. Y., and Wang, S. Q. (2014). Does the tail wag the dog? The effect of credit default swaps on credit risk. *Review of Financial Studies*, 27(10):2927–2960.

Svensson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical report, IMF Working Paper 94/114.

Vassalou, M. and Xing, Y. (2004). Default risk in equity returns. *Journal of Finance*, 59(2):831–868.

Table 1: Corporate default/bankruptcy statistics

The total number of default and bankruptcy events for actively traded CDS firms each year over the sample period of January 2001 to December 2020. A firm is counted as an actively traded CDS firm in a given month if there are at least 15 daily five-year CDS spreads observed during the month. The number of actively traded CDS firms each year is obtained by averaging over all months of the year.

Year	Actively traded CDS firms	Defaults/bankruptcies	(%)
2001	259	1	0.39%
2002	443	3	0.68%
2003	605	4	0.66%
2004	743	1	0.13%
2005	803	11	1.37%
2006	842	3	0.36%
2007	889	9	1.01%
2008	880	16	1.82%
2009	843	20	2.37%
2010	798	7	0.88%
2011	796	7	0.88%
2012	776	4	0.52%
2013	743	3	0.40%
2014	741	3	0.40%
2015	725	6	0.83%
2016	717	4	0.56%
2017	702	2	0.28%
2018	692	4	0.58%
2019	676	6	0.89%
2020	646	14	2.17%

Table 2: Summary statistics for covariates

The summary statistics for the five-year CDS spread of the month's last observation (CDS 5Y), the five-year corporate bond yield spread fitted through the Nelson-Siegel-Svensson approach (Corp Bond 5Y), the 10-year minus one-year CDS spread (CDS Slope), the Merton distance-to-default (DtD), the monthly stock return, (Stock Return), the monthly S&P 500 index return (Index Return), the three-month Treasury bill rate (Treasury Rate 3M), the ratio of net income to total assets (NI/TA), the logarithm of the firm's total assets in $ million (Size), the ratio of cash and short-term investments to total assets (Cash/TA), the market-to-book ratio (MtB), the idiosyncratic stock volatility estimated from the Fama-French three-factor model (IVOL), the CDS market depth measured by the number of dealers providing five-year CDS quotes (Market Depth), the investment-grade CDS index spread constructed as the equally-weighted average of five-year CDS spreads of 125 top investment-grade firms (CDS-IG), firm leverage (Firm Leverage) and asset volatility (Asset Volatility) used in the DtD calculation, the at-the-money option-implied volatility (Option ATM IV) and implied volatility skewness (Option IV Skewness) from individual stock options. For more details of these covariates, refer to Section 3.2.

	N	Mean	Std. Dev.	5%	25%	50%	75%	95%
CDS 5Y	102,325	0.0190	0.0426	0.0020	0.0046	0.0087	0.0195	0.0594
Corp Bond 5Y	90,535	0.0385	0.0575	0.0044	0.0143	0.0255	0.0420	0.0977
CDS Slope	88,751	0.0084	0.0927	0.0006	0.0039	0.0072	0.0135	0.0342
DtD	102,325	4.0126	3.8658	0.3414	1.4901	3.4603	6.0733	11.0991
Stock Return	102,325	0.0097	0.0994	-0.1475	-0.0388	0.0107	0.0584	0.1609
Index Return	102,325	0.0057	0.0423	-0.0748	-0.0170	0.0111	0.0313	0.0701
Treasury Rate 3M	102,325	0.0139	0.0159	0.0002	0.0010	0.0092	0.0217	0.0496
NI/TA	102,325	0.0431	0.0882	-0.0757	0.0104	0.0592	0.0829	0.1472
Size	102,325	9.5667	1.4622	7.5108	8.5411	9.4279	10.4263	12.2016
Cash/TA	102,325	0.0948	0.1016	0.0050	0.0258	0.0604	0.1309	0.2929
MtB	102,325	1.6536	0.9018	0.9281	1.0901	1.3703	1.8868	3.3098
IVOL	102,325	0.2222	0.1726	0.0778	0.1219	0.1746	0.2612	0.5169
Market Depth	102,324	5.7697	3.9584	2.0000	3.0000	5.0000	7.0000	13.0000
CDS-IG	102,325	0.0084	0.0040	0.0047	0.0058	0.0069	0.0101	0.0169
Firm Leverage	102,325	0.1905	0.1798	0.0291	0.0641	0.1672	0.2336	0.3441
Asset Volatility	102,325	0.4599	0.2212	0.1934	0.2959	0.4125	0.5775	0.9018
Option ATM IV	91,155	0.3339	0.1844	0.1580	0.2182	0.2860	0.3897	0.6689
Option IV Skewness	91,155	0.0092	0.0093	0.0042	0.0065	0.0082	0.0105	0.0192

Table 3: Correlations for covariates

The pairwise Pearson correlations for the five-year CDS spread of the month's last observation (CDS 5Y), the five-year corporate bond yield spread fitted through the Nelson-Siegel-Svensson approach (Corp Bond 5Y), the 10-year minus one-year CDS spread (CDS Slope), the Merton distance-to-default (DtD), the monthly stock return, (Stock Return), the monthly S&P 500 index return (Index Return), the three-month Treasury bill rate (Treasury Rate 3M), the ratio of net income to total assets (NI/TA), the logarithm of the firm's total assets in $ million (Size), the ratio of cash and short-term investments to total assets (Cash/TA), the market-to-book ratio (MtB), the idiosyncratic stock volatility estimated from the Fama-French three-factor model (IVOL), the CDS market depth measured by the number of dealers providing five-year CDS quotes (Market Depth), the investment-grade CDS index spread constructed as the equally-weighted average of five-year CDS spreads of 125 top investment-grade firms (CDS - IG), firm leverage (Firm Leverage) and asset volatility (Asset Volatility) used in the DtD calculation, the at-the-money option-implied volatility (Option ATM IV) and implied volatility skewness (Option IV Skewness) from individual stock options. For more details of these covariates, refer to Section 3.2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1) CDS 5Y	1.00																	
(2) Corp Bond 5Y	0.35	1.00																
(3) CDS Slope	-0.32	-0.09	1.00															
(4) DtD	-0.38	-0.18	-0.08	1.00														
(5) Stock Return	-0.04	-0.03	0.02	-0.01	1.00													
(6) Index Return	-0.04	-0.04	0.01	-0.01	0.50	1.00												
(7) Treasury Rate 3M	-0.12	-0.08	-0.04	0.10	-0.02	-0.03	1.00											
(8) NI/TA	-0.33	-0.14	0.02	0.37	0.03	0.01	0.06	1.00										
(9) Size	-0.15	-0.05	-0.10	-0.07	-0.01	0.01	-0.08	-0.02	1.00									
(10) Cash/TA	0.01	-0.01	-0.02	0.13	0.01	0.01	-0.06	0.12	0.02	1.00								
(11) MtB	-0.18	-0.11	-0.05	0.49	-0.02	-0.01	0.08	0.44	-0.15	0.26	1.00							
(12) IVOL	0.52	0.24	-0.13	-0.30	-0.05	-0.11	-0.04	-0.25	-0.17	0.03	-0.14	1.00						
(13) CDS Market Depth	-0.12	-0.05	-0.05	0.08	-0.02	-0.04	0.39	0.03	0.16	-0.06	0.00	-0.03	1.00					
(14) CDS-IG	0.24	0.22	-0.15	-0.12	-0.05	-0.16	-0.41	-0.11	-0.03	-0.02	-0.13	0.35	-0.14	1.00				
(15) Firm Leverage	0.28	0.11	0.12	-0.45	0.00	0.01	-0.06	-0.15	-0.15	-0.16	0.03	0.13	-0.05	-0.01	1.00			
(16) Asset Volatility	0.23	0.11	0.11	-0.51	-0.01	-0.01	-0.02	-0.09	-0.12	0.09	-0.05	0.26	-0.14	0.03	0.10	1.00		
(17) Option ATM IV	0.67	0.33	-0.15	-0.40	-0.10	-0.17	-0.11	-0.32	-0.21	0.05	-0.22	0.73	-0.08	0.53	0.16	0.32	1.00	
(18) Option IV Skewness	0.06	0.06	-0.05	-0.05	-0.05	-0.07	-0.02	-0.04	-0.01	-0.02	-0.05	0.13	-0.01	0.23	-0.01	0.01	0.17	1.00

Table 4: Effect of CDS spread on corporate default prediction

The baseline regressions examine the conditional effect of the CDS spread on corporate default prediction for different forecast horizons. Model specifications are logistic regressions with standard errors clustered at the firm level. The forecast horizons are one month, three months, six months, nine months, and 12 months. All independent variables are lagged by the corresponding forecast horizon in the empirical analysis. ***, **, and * represent statistical significance at the one, five, and 10 percent level, respectively. t-statistics are in parentheses.

	(1) 1M	(2) 3M	(3) 6M	(4) 9M	(5) 12M
CDS 5Y	3.986***	4.104***	4.063***	3.927***	3.841***
	[6.694]	[5.912]	[6.104]	[4.288]	[3.954]
DtD	-0.320***	-0.209***	-0.288***	-0.281***	-0.228***
	[-5.863]	[-4.230]	[-5.523]	[-4.850]	[-4.870]
Stock Return	-3.045***	-0.526	-0.464	-1.785**	-2.227*
	[-2.964]	[-0.669]	[-0.385]	[-2.085]	[-1.932]
Index Return	1.743	-4.296	-2.618	2.070	6.557*
	[0.697]	[-1.643]	[-0.995]	[0.844]	[1.688]
Treasury Rate 3M	7.380	8.001	9.027	14.091**	19.351***
	[0.784]	[0.897]	[1.217]	[2.029]	[2.786]
NI/TA	-1.921**	-1.158	-2.752***	-0.512	-0.704
	[-2.430]	[-1.362]	[-3.112]	[-0.508]	[-0.493]
Size	-0.087	-0.025	-0.063	-0.091	-0.085
	[-1.029]	[-0.301]	[-0.807]	[-1.229]	[-1.154]
Cash/TA	-3.963**	-2.963	-4.977**	-4.712*	-3.330
	[-1.984]	[-1.325]	[-2.048]	[-1.897]	[-1.531]
MtB	-0.343	-0.736*	-0.666	-0.751*	-0.653*
	[-0.900]	[-1.712]	[-1.519]	[-1.653]	[-1.704]
IVOL	0.099	1.446***	0.028	0.286	0.676
	[0.215]	[3.192]	[0.054]	[0.499]	[1.056]
Intercept	-5.887***	-6.511***	-5.411***	-5.139***	-5.669***
	[-6.204]	[-6.324]	[-5.521]	[-5.546]	[-6.027]
N	100,584	98,836	96,956	95,088	93,371
# Defaults	128	128	128	128	127
Pseudo R^2	0.2336	0.2046	0.1457	0.1138	0.1060
Clustered SE	Y	Y	Y	Y	Y

Table 5: Effect of CDS spread on corporate default prediction: Sub-sample analysis
The regressions examine the conditional effect of the CDS spread on corporate default prediction for different sub-samples. The forecast horizon is six months. Column (0) presents the baseline regression in Table 4. The sub-samples are financial firms vs. industrial firms (Columns (1) and (2)), crisis period from 2007-09 vs. the non-crisis period (Columns (3) and (4)), and investment-grade firms vs. high-yield firms (Columns (5) and (6)). ***, **, and * represent statistical significance at the one, five, and 10 percent level, respectively. t-statistics are in parentheses.

	(0) Base	(1) F	(2) NF	(3) C	(4) NC	(5) IG	(6) HY
CDS 5Y	4.063***	4.372**	3.939***	4.996***	3.986***	9.602	3.751***
	[6.104]	[2.471]	[5.334]	[3.680]	[4.752]	[0.785]	[5.419]
DtD	-0.288***	-0.347**	-0.292***	-0.313***	-0.265***	-0.409***	-0.222***
	[-5.523]	[-1.982]	[-4.714]	[-3.251]	[-4.423]	[-2.630]	[-3.679]
Stock Return	-0.464	-0.799	-0.434	-0.981	-0.134	2.415	-0.477
	[-0.385]	[-0.263]	[-0.329]	[-0.487]	[-0.095]	[0.591]	[-0.382]
Index Return	-2.618	5.281	-3.742	0.765	-5.566**	-6.826	-2.643
	[-0.995]	[0.492]	[-1.460]	[0.149]	[-1.963]	[-1.041]	[-0.884]
Treasury Rate 3M	9.027	-15.252	11.984	-4.937	10.61	5.263	11.209
	[1.217]	[-0.637]	[1.561]	[-0.461]	[0.959]	[0.360]	[1.294]
NI/TA	-2.752***	1.054	-3.110***	0.232	-4.076***	1.077	-2.798***
	[-3.112]	[0.495]	[-3.662]	[0.124]	[-4.394]	[0.372]	[-3.070]
Size	-0.063	0.113	0.023	-0.079	-0.061	-0.072	0.038
	[-0.807]	[0.362]	[0.237]	[-0.559]	[-0.658]	[-0.365]	[0.405]
Cash/TA	-4.977**	-6.573	-3.589	-8.593	-3.584	-11.846	-4.212
	[-2.048]	[-1.300]	[-1.289]	[-1.504]	[-1.346]	[-1.427]	[-1.587]
MtB	-0.666	-0.002	-0.844*	-0.439	-0.867	0.205	-1.024**
	[-1.519]	[-0.001]	[-1.744]	[-0.757]	[-1.381]	[0.502]	[-1.998]
IVOL	0.028	1.909***	-0.396	-0.270	0.001	-2.550	0.091
	[0.054]	[2.959]	[-0.617]	[-0.275]	[0.002]	[-0.797]	[0.166]
Intercept	-5.411***	-9.361**	-5.902***	-4.708***	-5.378***	-6.293***	-5.763***
	[-5.521]	[-1.974]	[-5.408]	[-2.858]	[-4.225]	[-2.831]	[-4.841]
N	96,956	11,941	85,373	18,255	79,059	53,856	43,458
# Defaults	128	10	118	45	83	9	119
Pseudo R^2	0.1457	0.2617	0.1457	0.1522	0.1494	0.0856	0.1305
Clustered SE	Y	Y	Y	Y	Y	Y	Y

Table 6: **Effect of CDS spread on corporate default prediction: Robustness checks**
The regressions examine the conditional effect of the CDS spread on corporate default prediction when adding alternative risk measures. The forecast horizon is six months. Column (0) presents the baseline regression in Table 4. Column (1) adds the CDS slope; Column (2) adds the CDS IG index spread; Column (3) adds the five-year corporate bond spread; Column (4) adds the ATM IV and IV skewness from equity options; Column (5) replaces DtD with firm leverage and asset volatility; Column (6) incorporates all variables mentioned above. ***, **, and * represent statistical significance at the one, five, and 10 percent level, respectively. t-statistics are in parentheses.

	(0)	(1)	(2)	(3)	(4)	(5)	(6)
CDS 5Y	4.063***	5.433***	3.931***	5.339***	4.206***	4.584***	5.061**
	[6.104]	[3.655]	[5.912]	[5.138]	[5.565]	[6.699]	[2.513]
CDS Slope		2.968					0.494
		[1.036]					[0.201]
CDS-IG			19.063				65.769
			[1.592]				[0.848]
Corp Bond 5Y				1.779*			2.946*
				[1.890]			[1.945]
Option ATM IV					0.794		0.692
					[1.369]		[0.944]
Option IV Skewness					-4.052		-2.136
					[-0.748]		[-0.239]
Firm Leverage						3.232***	0.728
						[4.925]	[0.547]
Asset Volatility						-0.650	-1.118
						[-1.163]	[-1.488]
DtD	-0.288***	-0.250***	-0.289***	-0.300***	-0.302***		-0.246***
	[-5.523]	[-4.619]	[-5.459]	[-5.374]	[-5.122]		[-3.008]
Stock Return	-0.464	-0.650	-0.410	-0.788	-1.058	-0.743	-2.027
	[-0.385]	[-0.517]	[-0.346]	[-0.574]	[-0.814]	[-0.568]	[-1.133]
Index Return	-2.618	-2.956	-3.659	-2.733	-2.354	-2.100	-3.471
	[-0.995]	[-1.003]	[-1.273]	[-0.947]	[-0.773]	[-0.775]	[-0.709]
Treasury Rate 3M	9.027	10.465	3.831	11.659	10.841	11.062	10.372
	[1.217]	[1.324]	[0.481]	[1.502]	[1.284]	[1.490]	[0.960]
NI/TA	-2.752***	-2.489**	-2.903***	-1.895*	-2.204**	-3.139***	-1.359
	[-3.112]	[-2.540]	[-3.257]	[-1.745]	[-2.224]	[-3.636]	[-0.956]
Size	-0.063	-0.032	-0.062	-0.137	-0.133	0.048	-0.185
	[-0.807]	[-0.383]	[-0.791]	[-1.474]	[-1.532]	[0.611]	[-1.558]
Cash/TA	-4.977**	-5.006**	-4.877**	-4.924*	-3.520	-2.868	-2.852
	[-2.048]	[-2.109]	[-2.044]	[-1.830]	[-1.546]	[-1.216]	[-1.223]
MtB	-0.666	-0.885	-0.686	-0.645	-0.437	-1.888***	-0.698
	[-1.519]	[-1.637]	[-1.565]	[-1.344]	[-1.126]	[-3.610]	[-1.238]
IVOL	0.028	0.176	0.298	-0.584	-0.695	0.401	-0.841
	[0.054]	[0.298]	[0.613]	[-0.886]	[-0.936]	[0.741]	[-1.039]
Intercept	-5.411***	-5.619***	-4.980***	-4.737***	-5.313***	-6.601***	-3.926**
	[-5.521]	[-5.203]	[-4.877]	[-4.078]	[-5.240]	[-6.223]	[-2.131]
N	96,956	85,391	97,314	86,382	88,655	96,956	84,594

# Defaults	128	128	128	98	128	128	128
Pseudo R^2	0.1457	0.1455	0.1481	0.1610	0.1500	0.1375	0.1578
Clustered SE	Y	Y	Y	Y	Y	Y	Y

Table 7: Effect of CDS spread on corporate default prediction: Rare event bias
The regressions estimate the conditional effect of the CDS spread on corporate default prediction with corrections for potential rare event bias. The forecast horizon is six months. Column (0) presents the baseline regression in Table 4. Column (1) uses the bias correction in King and Zeng (2001). Column (2) uses the penalized MLE approach when estimating the logit regression. Column (3) combines all default events with randomly sampling 10 percent of the non-default observations. ***, **, and * represent statistical significance at the one, five, and 10 percent level, respectively. t-statistics are in parentheses.

	(0) Baseline	(1) Bias Correction	(2) Penalized MLE	(3) Reduced Sample
CDS 5Y	4.063***	3.100***	3.094***	4.531***
	[6.104]	[5.676]	[4.839]	[4.790]
DtD	-0.288***	-0.266***	-0.267***	-0.321***
	[-5.523]	[-5.904]	[-4.452]	[-5.506]
Stock Return	-0.464	-2.181**	-2.181***	-3.152***
	[-0.385]	[-2.566]	[-2.839]	[-3.351]
Index Return	-2.618	2.286	2.280	1.695
	[-0.995]	[0.969]	[0.913]	[0.564]
Treasury Rate 3M	9.027	10.648	10.618	16.022
	[1.217]	[1.062]	[1.074]	[1.537]
NI/TA	-2.752***	-3.197	-3.186	-6.682*
	[-3.112]	[-1.161]	[-0.929]	[-1.715]
Size	-0.063	-0.051	-0.051	-0.056
	[-0.807]	[-0.652]	[-0.526]	[-0.655]
Cash/TA	-4.977**	-2.915	-2.914	-2.378
	[-2.048]	[-1.469]	[-1.431]	[-1.211]
MtB	-0.666	-0.257	-0.248	-0.175
	[-1.519]	[-0.753]	[-0.609]	[-0.481]
IVOL	0.028	1.685***	1.685***	0.827*
	[0.054]	[4.362]	[4.047]	[1.941]
Intercept	-5.411***	-6.965***	-6.978***	-4.577***
	[-5.521]	[-7.664]	[-6.261]	[-4.461]
N	96,956	96,956	96,956	8,287
# Defaults	128	128	128	128
Pseudo R^2	0.1457	0.17750	0.2074	0.2973
Clustered SE	Y	Y	Y	Y

Table 8: CDS spread decomposition

Panel A summarizes the estimation of Eq. (1) across IG and HY firms before and after the CDS Big Bang, using five-year CDS spreads. Min, Max, and Mean Market Depth are the minimum, maximum, and mean value of the number of CDS quote providers, respectively. $\hat{\beta}$ is the estimated coefficient of the standardized CDS illiquidity variable. The mean and standard deviation of the CDS liquidity spread are presented, as is the mean of the CDS liquidity spread as a percentage of the total CDS spread. Panel B presents the mean and standard deviation of the percentage of the total CDS spread attributed to liquidity (% Liquidity), physical default (% P Default), and the default risk premium (% Risk Premium) across different maturities.

Panel A	*Before CDS Big Bang*		*After CDS Big Bang*	
	Investment-grade	High-yield	Investment-grade	High-yield
Min Market Depth	2.000	2.000	2.000	2.000
Max Market Depth	28.000	31.000	14.000	14.000
Mean Market Depth	6.897	6.070	4.411	4.592
$\hat{\beta}$	0.000905	0.002758	0.000518	0.001353
Mean(Liquidity Spread)	0.218%	0.749%	0.063%	0.229%
SD(Liquidity Spread)	0.162%	0.483%	0.060%	0.160%
Mean(% Liquidity)	30.118%	30.632%	16.886%	15.311%

Panel B	CDS 1Y	CDS 3Y	CDS 5Y	CDS 7Y	CDS 10Y	CDS 20Y	CDS 30Y
Mean(% Liquidity)	68.54%	40.14%	18.80%	21.86%	25.77%	32.81%	41.49%
SD(% Liquidity)	48.46%	36.77%	28.95%	25.34%	22.49%	19.56%	18.46%
Mean(% P Default)	11.51%	34.56%	50.27%	42.46%	33.15%	16.34%	6.91%
SD(% P Default)	39.51%	34.71%	31.34%	29.63%	28.60%	27.98%	27.09%
Mean(% Risk Premium)	19.95%	25.31%	30.94%	35.67%	41.08%	50.85%	51.60%
SD(% Risk Premium)	35.48%	33.11%	30.05%	28.77%	27.69%	26.57%	25.49%

Table 9: Effect of CDS spread components on corporate default prediction

The regression examines the conditional effect of the CDS spread components on corporate default prediction. The forecast horizon is six months. Column (0) presents the baseline regression in Table 4. Column (1) replaces the five-year CDS spread with its default and liquidity components. Column (2) further splits the default component into the physical default and risk premium components. ***, **, and * represent statistical significance at the one, five, and 10 percent level, respectively. t-statistics are in parentheses.

	(0)	(1)	(2)
CDS 5Y	4.063***		
	[6.104]		
Default 5Y		3.959***	
		[5.986]	
Physical Default 5Y			5.749***
			[3.479]
Risk Premium 5Y			0.993
			[0.532]
Liquidity 5Y		21.101	43.552
		[1.067]	[0.580]
DtD	-0.288***	-0.282***	-0.258***
	[-5.523]	[-5.174]	[-4.926]
Stock Return	-0.464	-0.449	-0.022
	[-0.385]	[-0.374]	[-0.017]
Index Return	-2.618	-2.953	-3.324
	[-0.995]	[-1.132]	[-1.089]
Treasury Rate 3M	9.027	9.969	8.518
	[1.217]	[1.300]	[1.147]
NI/TA	-2.752***	-2.785***	-2.287**
	[-3.112]	[-3.135]	[-1.982]
Size	-0.063	-0.054	-0.087
	[-0.807]	[-0.653]	[-1.028]
Cash/TA	-4.977**	-4.962**	-4.790*
	[-2.048]	[-2.041]	[-1.893]
MtB	-0.666	-0.674	-0.817
	[-1.519]	[-1.529]	[-1.612]
IVOL	0.028	0.069	-0.399
	[0.054]	[0.134]	[-0.616]
Intercept	-5.411***	-5.647***	-4.944***
	[-5.521]	[-5.420]	[-4.771]
N	96,956	96,956	96,956
# Defaults	128	128	128
Pseudo R^2	0.1457	0.1463	0.1675
Clustered SE	Y	Y	Y

Table 10: Out-of-sample performance of default prediction models

Summary statistics of out-of-sample performance of logistic default prediction models using 1) the four DSW07 covariates (labeled "DtD"); 2) DSW07's covariates plus the CDS spread (labeled "CDS Spread"); 3) DSW07's covariates plus the default spread (labeled "CDS Default"); and 4) DSW07's covariates plus the physical default component of the CDS spread (labeled "CDS P Default"). "Area" is two times the area under the power curve and above the 45 degree line, measuring the accuracy ratio of overall default prediction performance with the chosen forecast horizon ("Lag") and covariate combination. Also presented are points on the power curve at the 10th, 20th, 30th, and 40th percentile, respectively.

Lag	Statistics	CDS P Default	CDS Default	CDS Spread	DtD
1 month	Area	87%	85%	84%	84%
	10th	68%	62%	62%	62%
	20th	78%	72%	72%	70%
	30th	80%	76%	78%	74%
	40th	84%	84%	82%	82%
3 months	Area	84%	81%	80%	79%
	10th	64%	56%	56%	52%
	20th	72%	62%	62%	58%
	30th	76%	72%	70%	68%
	40th	80%	82%	78%	78%
6 months	Area	83%	80%	80%	79%
	10th	64%	56%	54%	60%
	20th	70%	62%	62%	62%
	30th	72%	66%	66%	66%
	40th	78%	78%	76%	76%
9 months	Area	83%	80%	81%	79%
	10th	66%	56%	58%	58%
	20th	70%	64%	64%	62%
	30th	72%	66%	66%	64%
	40th	80%	76%	74%	76%
12 months	Area	83%	80%	80%	80%
	10th	66%	56%	56%	60%
	20th	68%	62%	60%	62%
	30th	74%	66%	64%	64%
	40th	78%	76%	74%	76%

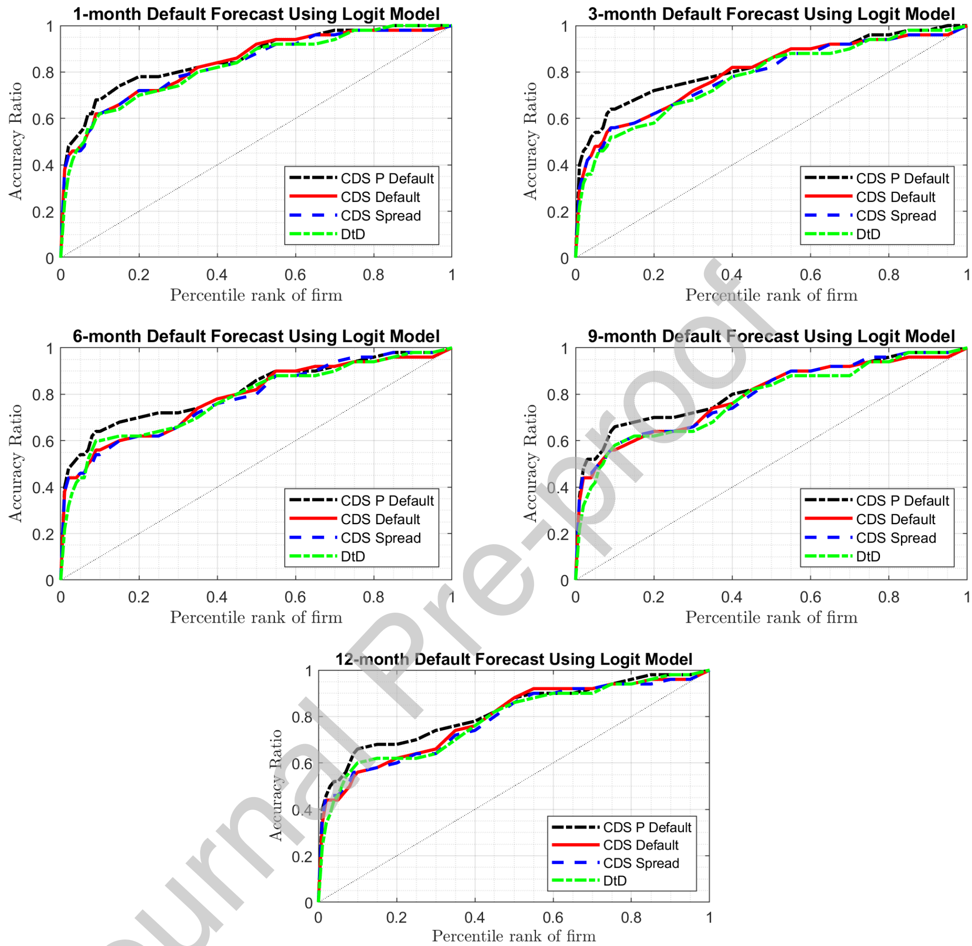

Figure 1: **Out-of-sample power curves of default prediction models.** This figure presents the average out-of-sample power curves starting in January 2006, using a five-year rolling-window training dataset, and ending in December, 2020. Forecast horizons are one, three, six, nine and twelve months. The predictive model specifications are logistic regressions using 1) the four DSW07 covariates (labeled "DtD"); 2) DSW07's covariates plus the CDS spread (labeled "CDS Spread"); 3) DSW07's covariates plus the default spread (labeled "CDS Default"); and 4) DSW07's covariates plus the physical default component of the CDS spread (labeled "CDS P Default").

Credit Author Statement

Xiaoxia Ye, Fan Yu, and Ran Zhao contributed equally as co-authors to this paper.