THE ROLE OF CDS SPREADS IN EXPLAINING BOND RECOVERY RATES

Matteo Barbagli, Pascal François,
Geneviève Gauthier, Frédéric Vrins

REPRINT • 2025 / 01

UCLouvain

LiDAM
LFIN | Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications

Journal of Banking and Finance 174 (2025) 107414

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

The role of CDS spreads in explaining bond recovery rates

Matteo Barbagli [a,*], Pascal François [b], Geneviève Gauthier [c], Frédéric Vrins [a,d]

[a] LFIN/LIDAM, UCLouvain, 1348 Louvain-la-Neuve, Belgium
[b] HEC Montréal, Department of Finance, H3T 2A7 Montréal, Canada
[c] HEC Montréal, Department of Decision Science and GERAD, H3T 2A7 Montréal, Canada
[d] HEC Montréal, Department of Decision Sciences, H3T 2A7 Montréal, Canada

ARTICLE INFO

JEL classification:
G21
G28
G32

Keywords:
Credit risk
Recovery rate
Credit default swap
Corporate bond
Uncertainty

ABSTRACT

We introduce two novel indices built from CDS market data capturing the level and uncertainty information embedded in credit spreads aggregated by industry, and study their role in predicting bonds recovery rates. Analyzing 613 defaulted U.S. corporate bond issues from 2006 to 2019 and using a beta regression model, we find the cross-sectional mean and approximate entropy of CDS spreads aggregated at the sector level to be important predictors of the recovery rates distributions. In the classical beta regression model, both regressors are statistically significant and enhance the pseudo-R^2 by up to 4%. Notably, a forward model selection procedure includes the sector-level regressor before well-known variables such as the bonds' coupon rate or the American default rate. In addition, our sector-uncertainty regressor is the only significant uncertainty variable. These findings offer valuable insights for improving credit risk assessment methodologies and identifying key risk indicators of recovery rates before running prediction models.

1. Introduction

Accurately assessing credit risk remains a key concern for regulators, banks and financial institutions. The creditworthiness of a borrower influences lending decisions and capital requirements, but also impacts the credit spread of debt securities. The Basel Accords quantify this risk through the use of three main parameters: the exposure at default (EAD), the probability of default (PD), and the loss given default (LGD). The latter is defined as 1 minus the recovery rate (RR) and is expressed as a percentage of the nonrecoverable debt related to the EAD (BCBS, 2023a). In this setup, banks are allowed to form their own estimates of the PD under both the foundation and advanced IRB approaches, explaining why PD modeling received particular attention in the literature (Bellalah et al., 2016; Tsai and Chen, 2011). On the other hand, less emphasis has been placed on EAD and LGD modeling. This can be explained by the fact that, under the foundation IRB approach, both act as scaling constants in the computation of regulatory capital (BCBS, 2023b).[1] The EAD is typically known at the time of default, while the LGD – and, consequently, the RR – remains unknown until the end of the liquidation process. This inherent risk has spurred numerous studies into the predictors of recovery rates,[2] with a particular focus on U.S. corporate bonds since the late 1990s.[3] Substantial changes in the dynamics of recovery rates have been observed during the 2008 global financial crisis, affecting both their mean and volatility. Identifying predictors of recovery rates remains therefore an important problem for pricing and risk-management purposes (Jankowitsch et al., 2014). Moreover, it is a key step to build parsimonious and explainable forecasting models avoiding overfitting. In this paper, we contribute to this field by introducing two novel indices capturing the level and uncertainty information embedded in credit default swap (CDS) market data aggregated by industry, and demonstrate that they consistently provide robust predictive power regarding the recovery rates of defaulted corporate bonds.

Initially, linear regression models were used to unveil idiosyncratic factors. Those are static variables characterizing the issue and the

* Correspondence to: UCLouvain, Place des Doyens 1, Boite L2.01.01, Office b 208, 1348 Louvain-la-Neuve, Belgium.
E-mail addresses: matteo.barbagli@uclouvain.be (M. Barbagli), pascal.francois@hec.ca (P. François), genevieve.gauthier@hec.ca (G. Gauthier), frederic.vrins@uclouvain.be (F. Vrins).

[1] The LGD is set equal to a constant under the foundation IRB approach, equal to 40% for senior unsecured corporate bonds and 75% for subordinated corporate bonds (BCBS, 2023b, §32.6–32.7).

[2] Early papers in this stream of literature often use the terms "determinant" or "driver" for variables that exhibit explanatory power regarding recovery rates. In this paper, we use instead "predictors", "regressors" or "factors" to avoid suggesting the existence of a causal relationship.

[3] In contrast, there is only a minority of studies on consumer credits. One can hardly extrapolate the findings related to corporate bonds to consumer loans as the two types of debt are not easily comparable (Distaso et al., 2023).

https://doi.org/10.1016/j.jbankfin.2025.107414
Received 21 February 2024; Accepted 15 February 2025
Available online 3 March 2025

issuer (Shleifer and Vishny, 1992; Altman and Kishore, 1996; Schuermann, 2004). Systematic factors have been identified later on. Those are dynamic indicators capturing the impact of the business and credit cycles on the recovery rates. The most popular indicator is the default rate, first identified in the seminal work of Altman et al. (2005). More recently, financial uncertainty (survey-based, news-based and volatility-based) has been found to be the main systematic predictors of bond recovery rates in a beta regression model (Gambetti et al., 2019). The role of uncertainty on bond recovery rates is intuitive. The market recovery rate (post-default bond price) represents discounted expectations of the ultimate recovery, which depends on the state of the business cycle (Acharya et al., 2007; Altman and Kalotay, 2014). Hence, the higher the uncertainty about the future state of the business cycle, the lower the defaulted bond should be priced. This is consistent with prior studies showing that economy-related uncertainty increases market risk premia (Hansen et al., 1999; Cagetti et al., 2002).

Many recent studies analyze the performance of machine-learning algorithms exploiting hundreds of predictors for the sake of forecasting recovery rates (Yao et al., 2015; Nazemi et al., 2017; Nazemi and Fabozzi, 2018; Nazemi et al., 2018). These models often rely on dimensionality reduction techniques (PCA, kernel PCA, etc.) that jeopardize their interpretability. Consequently, those approaches stand little chances to be approved by regulators on the short term. In this context, Gambetti et al. (2022) show how effective machine and meta-learning algorithms are when trained on a restricted set of well-identified recovery rate regressors. An interesting finding is the recurrence of economic uncertainty as the most important variables. Moreover, low dimensional models based on uncertainty measures outperform higher dimensional ones. This result highlights the importance of conducting inference studies prior to running prediction models.

Our main objective consists in exploiting the CDS market to predict recovery rates. CDS spreads correspond to the fair risk premium the protection buyer should pay to the seller in order to insure the losses associated with the default of a bond up to a given maturity. Therefore, the spread of a given CDS contract is directly related to the expected loss on the underlying bond which, itself, depends both on the PD and the RR. This relationship is evidenced by the celebrated *credit triangle rule* $h \approx \pi/(1 - \mathrm{RR})$ extracted from reduced-form models (Duffie and Singleton, 1999), which gives the fundamental relationship connecting the firm's default intensity (h), the CDS spread (π) and the RR. This suggests that CDS spreads are a natural source of information about recovery rates.

The first stream of literature aiming to capture this information postulates debt structures and default models to disentangle the effect of PD and RR. However, the purpose is not to make inference on the recovery rates of defaulted bonds (no or little variables are used in conjunction with CDS spreads), but is rather to provide market-implied estimations of recovery rates about that specific issue (or issuer) in view of pricing other RR-sensitive contracts such as fixed-recovery rates, recovery swaps (Berd, 2005; Pykhtin, 2003) or credit-valuation adjustment (Gregory, 2012; François and Jiang, 2019). Relying on those approaches in our context is difficult because, as highlighted in our database, only a tiny number of defaulted issues are featured in actual CDS contracts.

Closer to our work is a second type of papers, aiming at studying CDS-based indicators as predictors of the ex-post RR of defaulted bonds. This stream of literature is surprisingly scarce. In fact, the only paper exploiting CDS for the sake of doing inference on RR seems to be Jankowitsch et al. (2014). However, the author does not directly exploit the CDS spreads themselves, but a dummy variable indicating whether a CDS contract is written on that specific issue or not. To the best of our knowledge, there is thus no study investigating the information associated with CDS spreads as a potential predictor of recovery rates of defaulted bonds. The purpose of this paper is to fill this gap.

To circumvent the problem that only a tiny number of defaulted issues are embedded in CDS contracts, we introduce two CDS-based indices defined at the industry level, capturing the sector information embedded in the spread level and variability through time. These CDS-based indices form a new class of predictors defined at the sector level, sitting in between the idiosyncratic (static) variables and the systematic (dynamic) factors identified so far.

To analyze the predictive power of those indices on recovery rates, we collect observations related to defaulted issues of U.S. corporate bonds from the Moody's default and recovery database, while the systematic factors are extracted from specific repositories such as the Federal Reserve Bank of St Louis database. CDS quotes are extracted from Datastream and Thomson Reuters. Combining those databases leads to a dataset of 613 defaulted issues covering the period 2006–2019. In contrast with most studies in the field, we stress the need to reduce the number of categorical variables to avoid data fragmentation and overfitting. We achieve this using a transparent consolidation procedure driven by economic fundamentals. In line with Gambetti et al. (2019), we analyze the predictors of the mean and precision parameters in both classical and variable beta regression models. We define two new variables based on CDS spreads aggregated at the sector level: the cross-sectional mean of the 10-year CDS quotes and the approximate entropy of the 1-year CDS spreads computed on a 12-months rolling window. The former aims at capturing the recovery rate information according to the credit triangle rule, while the latter aims at capturing uncertainty in a more subtle way than volatility.

Our main contributions are as follows. First, we highlight the importance of a consolidation procedure, whose purpose is to decrease the number of categories. While this is often overlooked in similar studies, we find such a pre-processing to be necessary to avoid overfitting and face misleading performance results. Second, we identify for the first time CDS-based metrics as key predictors of recovery rates. Precisely, the cross-sectional mean appears to be informative of the average recovery rates within a sector, while the approximate entropy constitutes a novel proxy of *sector-uncertainty*, complementing the systematic economic uncertainty metrics identified in Gambetti et al. (2019). Third, we find economic uncertainty not to be statistically significant for explaining recovery rates once other systematic controls for the business and credit cycle are involved. Only *sector-uncertainty* appears to be informative to explain both the mean and the dispersion of the recovery rate distributions. An increase of uncertainty within a sector leads to lower recovery rates and greater dispersion around this lower value, on average. This effect is aligned with the existing literature on the impact of uncertainty on the pricing of securities, strengthening the validity of our results. Finally, in contrast with Gambetti et al. (2019) who analyze a wider (1831 observations) and older (1980–2014) sample of defaulted issues, the importance of the American default rate is not supplanted by economic uncertainty proxies, thus regaining its pivotal importance in explaining the recovery rates.

The paper is organized as follows. Section 2 provides a literature review of the key idiosyncratic and systematic factors that have been found to be associated with recovery rates of corporate bonds. The methods connecting CDS spreads to recovery rates are also discussed. The beta regression models, the consolidation procedure and the definition of our new CDS variables are exposed in Section 3. Section 4 introduces the databases, highlights the data fragmentation issue, presents summary statistics of the dependent variable and some information on the regressors. Section 5 presents the empirical results. In Section 6, we investigate the predictive power of our CDS-based regressors. Section 7 concludes.

2. Literature review

Studies on bond recovery rate predictors date from the early 1990's. They have mostly been conducted by academia (Altman and Kishore, 1996; Altman et al., 2005; Caselli et al., 2008; Jankowitsch et al., 2014; Yao et al., 2015) and rating agencies (Van de Castle et al., 2000; Varma and Cantor, 2005; Cantor et al., 2007). Many papers have been published since then, featuring new models and exploiting new data (Mora, 2015; Yao et al., 2015; Nazemi and Fabozzi, 2018; Nazemi et al., 2018; Gambetti et al., 2019). These empirical analyzes mostly deal with the U.S. market and shed the light on both cross-sectional and temporal dispersion. While the former is mainly driven by debt-specific characteristics such as the coupon rate or the industry sector (idiosyncratic factors), the latter results from business cycles such as economic growth and annual default rate (systematic factors). CDS-based indicators defined at the issue level cannot be exploited due to lack of data. By contrast, CDS-based indicators defined at the sector level can be constructed, and can be regarded as variables sitting in between idiosyncratic and systematic factors.

2.1. Idiosyncratic factors

Cross-sectional variations of recovery rates arising from the seniority of the defaulted bond and the industrial sector of the issuer were first highlighted in Altman and Kishore (1996) and later corroborated by Varma and Cantor (2005) and Acharya et al. (2007). These studies show that, on average, bonds with higher seniority or higher degree of collateralisation exhibit higher recovery rates. Likewise, sectors featuring a higher degree of asset redeployability (like the utilities sector) or asset tangibility (mainly property, plants and equipments) present higher recovery rates. Schuermann (2004) corroborates the above evidence identifying higher recoveries in both the utilities and telecommunications sector between 1970 and 2003. Conversely, bonds issued by companies in industries with a high frequency of distress periods display lower recovery rates. During these distressed periods, the disparity between the defaulted assets' supply and demand leads to fire sales, which ultimately results in lower recovery rates for the specific sector (Shleifer and Vishny, 1992). The above reinforces the need to take into account sector specificity, notably through a dedicated categorical variable in regression models.

The nature of the default event also has a significant impact on the recovery rate, as highlighted in Franks and Torous (1994) and later confirmed in Bris et al. (2006). For instance, processes entailing liquidation, such as the Chapter 7 procedure in the U.S., typically yield lower recovery rates. On the other hand, scenarios involving reorganizations (e.g., Chapter 11 procedure in the U.S.) or distressed exchanges, where the financial obligations are renegotiated outside the range of full-scale bankruptcy proceeding generally lead to higher recovery rates (Altman and Karlin, 2009).

2.2. Systematic factors

Systematic factors aim to capture the effect of credit and business cycles, which evolve through time (Acharya et al., 2007; Bruche and González-Aguado, 2010). The most widely known systematic regressor is perhaps the default rate observed in the corresponding market. This was first stressed in the seminal work of Altman et al. (2005), who show a negative relationship between the yearly average recovery rate and yearly average default rate in the U.S. corporate bonds market. This empirical finding is frequently explained by disparities in the demand and supply for tangible assets that defaulted enterprises sell. Other macroeconomic factors have received mixed support. There is no consensus on which one of these variables should be included in both inference or predictive models for recovery rates. Altman et al. (2005) investigate the predictive power of the gross domestic product (GDP) and its returns, and the level of the S&P 500 and its returns. GDP is proven to correlate poorly with average recovery rates; significance improves in the case of GDP returns. However, when included in a multivariate model combined with default rates, these variables prove to be non-significant. Similarly, both the stock market index level and return are reported to be non-significant in multivariate regressions. Contrarily, Mora (2015) report a positive statistically significant relationship between the GDP growth rate and the S&P 500 returns, later confirmed by Nazemi and Fabozzi (2018). Chava et al. (2011) employ five macroeconomic variables to model the impact of economic conditions on bond recovery rates. They find the logarithm of the total defaulted debt amount and the 3-month Treasury bill rate to be both statistically significant in most of their recovery rate models. Jankowitsch et al. (2014) considers four different macroeconomic indicators, and finds a positive relationship between recovery rates and the U.S. Federal Funds rate. This should not come as a surprise because the latter is a critical indicator of monetary policy and influences general credit conditions in the economy. Tobback et al. (2014) investigate the effects of 11 macroeconomic variables on corporate bond recovery rates, and highlight a series of unexpected effects and varied influences depending on the dataset and statistical model used.

More recently, Gambetti et al. (2019) emphasize the explanatory power of economic uncertainty (ECU) as a systematic predictor of recovery rates. The authors consider three main categories of time-varying ECU: volatility-based, news-based and survey-based. Volatility-based ECU include the VIX index aiming to gauge stock market volatility, and a one-month financial uncertainty indicator that measures forecast error volatility, as introduced by Jurado et al. (2015) and further developed by Ludvigson et al. (2021). The original economic policy uncertainty index of Baker et al. (2016) is used as a news-based measures of ECU. Finally, the authors employ an inflation uncertainty measure and a proxy of uncertainty relative to both federal and state/local purchases as survey-based proxies. These indicators build on the dispersion of forecasts computed from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters.[4] The authors rely on a beta-regression model and make two contributions. First, all their retained ECU proxies are statistically significant to explain the mean of the recovery rate distributions when no other systematic factors are considered. Second, financial uncertainty, the VIX and their news-based proxy still convey additional information to explain the recovery rates, even when other systematic controls for the business and credit cycle are included in the regression model. Interestingly, the inclusion of multiple ECU proxies supplants the celebrated American default rate from being a key predictor of recovery rates. This shift highlights the pivotal role of economic uncertainty in this context.

2.3. Factors extracted from CDS contracts

The existing literature on the relationship between CDS spreads and RRs predominantly focuses on the extraction of market-implied estimation of recovery rates. Unal et al. (2003) propose a simple approach to infer the risk-neutral density of recovery rates implied by the prices of both junior and senior debt securities of a firm based on an adjusted relative spread statistic. Das and Hanouna (2009) develop a parsimonious model to infer the implied forward curves of default probabilities and recovery rates for a given firm using the term structure of CDS spreads. Jaskowski and McAleer (2013) propose a reduced-form model estimating expected recovery rates and default probabilities from the CDS spread curves. The authors introduce a Bayesian algorithm to estimate the parameters of the model and the latent factors driving recovery risk and default risk. Boudreault et al. (2013) introduce a hybrid credit risk framework combining features of structural and reduced-form models, in which the default probability and the recovery rate are functions of the firm's leverage ratio. Firm-specific leverage ratios are retrieved from CDS spread time series for various maturities. Boudreault et al. (2014) enhance the previous single-firm framework to capture contagion present in a multi-firm framework, illustrating how the contagion effect amongst assets of a

[4] The forecast data is accessible at: https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters.

given bond portfolio changes its risk perception and ultimately impacts the recovery rates. Elkamhi et al. (2014) infers risk-neutral recovery rates from CDS spreads using a quadratic pricing model. Doshi et al. (2018) use information from both senior and subordinate CDSs to isolate the dynamics of recovery rates and study the term structure of expected recovery using a 5-factor reduced-form no-arbitrage model. Recently, François and Jiang (2019) propose a model to extract risk-neutral recovery rates from the term structure of CDS spreads and use them in a model for credit-valuation adjustment computation. François (2019) performs an analysis of the factors driving these market-implied RR and a comparison with the explanatory variables of historical recovery rates.

The purpose of the above studies is to extract market views of recovery rates from the term structure of CDS spreads. In particular, they do not aim to analyze the role played by CDS-based factors on actual recovery rates in an inference problem. To the best of our knowledge, Jankowitsch et al. (2014) is the only study aiming to achieve this goal. The authors show that bonds featured in a CDS contract have significantly higher recovery rates, which can be intuitively explained by the increased demand from protection buyers required to physically deliver the underlying bond. Hence, the impact of CDS on recovery rates is captured by a dummy variable indicating whether there is a CDS market for the corresponding bond, but the effect of the CDS spreads themselves is not investigated.

3. Methodology

As highlighted in the introduction, much of the existing research on the predictors of bond recovery rates relies on standard linear regression techniques (Varma and Cantor, 2005; Altman et al., 2005; Acharya et al., 2007). Yet, this popular benchmark relies on a couple of assumptions that are hardly compatible with the salient features of recovery rates data. For instance, the normal distribution hypothesis is clearly inappropriate to model data bounded between 0 and 1. This sheds some doubts on the validity of the results obtained and explains why alternative approaches, better suited to those characteristics, have been considered recently for inference purposes. A first possibility is to keep the linear regression model but featuring a non-linear "link" transform of the dependent variable. This is for example the approach discussed in Qi and Zhao (2011). However, this methodology has several drawbacks, one of them being that model parameters can hardly be interpreted in terms of the original response. A second route is to rely on more specific regression models whose underlying assumptions better fit with the nature of the dependent variable at hand. Beta regression is a natural and common choice when confronted to variables taking values in the unit interval.

Although the beta distribution is not appropriate to model any type of recovery rates,[5] it is a reasonable assumption when it comes to analyzing recovery rates of corporate bonds (see, e.g., Gambetti et al., 2019 and references therein), a claim that we confirm in the dataset at hand. This motivates our choice to rely on the beta regression model in the context of our application. In this section, we first recall the various specifications of the model, clarify the definition of the dependent variable and introduce the independent variables that will be considered in our empirical study.

3.1. The model

In this work, we rely on the beta regression first introduced by Ferrari and Cribari-Neto (2004) (hereafter named *classical* beta regression) and later extended to the *variable dispersion* beta regression model by Smithson and Verkuilen (2006) and Simas et al. (2010) (referred to as *variable* beta regression in the sequel). This model is specifically designed to deal with rates and proportions, hence, when the dependent variable is measured continuously on $(0,1)$.[6] It postulates that the dependent variable Y_i (which is, in our context, the recovery rate associated with a defaulted corporate bond $i = 1,\dots,N$) follows a beta distribution with specific parameters. Adopting the specification of Simas et al. (2010), we write $Y_i \sim \mathcal{B}(\mu_i,\phi_i)$ where $\mu_i \in (0,1)$ and $\phi_i > 0$ are known as the *mean* and *precision* parameters. The latter can be easily related to the *shape parameters* (α,β) associated with the standard parametrization of the beta law.[7] In this setting, the density of Y_i is

$$f(y|\mu_i,\phi_i) := \frac{\Gamma(\phi_i)}{\Gamma(\mu_i\phi_i)+\Gamma((1-\mu_i)\phi_i)} y^{\mu_i\phi_i-1}(1-y)^{(1-\mu_i)\phi_i-1}, \quad y \in (0,1)\,, \tag{1}$$

whose mean and variance are given by

$$\mathbb{E}[Y_i] = \mu_i \quad \text{and} \quad \mathbb{V}[Y_i] = \mu_i \frac{1-\mu_i}{1+\phi_i}\,. \tag{2}$$

Next, functional relationships are assumed for the parameters. Although other choices are possible, we follow Gambetti et al. (2019) and adopt *logit* and *log* link functions for the mean and precision parameters, respectively :

$$\mu_i = \frac{e^{x_i^\top \theta^\mu}}{1+e^{x_i^\top \theta^\mu}} \qquad \text{and} \qquad \phi_i = e^{z_i^\top \theta^\phi}\,, \tag{3}$$

where $x_i = (1, x_{i,1},\dots,x_{i,J})$ and $z_i = (1, z_{i,1},\dots,z_{i,K})$ are the vectors of observations of J and K covariates explaining the mean and precision parameters ($J+K+2 < N$), and $\theta^\mu = (\theta_0^\mu,\theta_1^\mu,\dots,\theta_J^\mu)$ and $\theta^\phi = (\theta_0^\phi,\theta_1^\phi,\dots,\theta_K^\phi)$.

The estimation of the coefficients θ^μ and θ^ϕ is performed by maximum likelihood, as suggested in Simas et al. (2010). The *classical* beta regression model is retrieved by setting the precision parameter constant: $\phi_i = \phi = e^{\theta_0^\phi}$.

3.2. Dependent variable

In line with the literature, we take as dependent variable the traded price of the bond 30 days after the default event. Specifically, we set $Y_i = P_i(t_i+30)$ where $P_i(t)$ stands for the price of the bond at time t and t_i is the default date of the security i. This definition is a standard convention of the recovery rate of a defaulted bond, which offers a timely and direct reflection of the market's perception of the ultimate recovery rate (Mora, 2015). Indeed, it is acknowledged that the price evolution in this time window is mostly driven by the default event itself (Schuermann, 2004; Mora, 2012; Jankowitsch et al., 2014). For instance, Metz et al. (2012) demonstrate that prices recorded at this interval are better predictors of ultimate recoveries compared to those measured closer to the default date.

3.3. Independent variables

The same set of predictors is used for the mean and precision in the *variable* setup ($x_i = z_i$). As explained in Section 2, we consider three types of explanatory variables: idiosyncratic, systematic and CDS-based predictors.

[5] For instance, Distaso et al. (2023) and Li et al. (2023) highlight that recovery rates associated with defaulted consumer credit exhibit a strong bimodal shape, a feature that cannot be handled by the beta distribution.

[6] We exclude any recovery rate equal to 0 or greater or equal to 1. Inflated beta distributions (Ospina and Ferrari, 2012) can be implemented in order to consider extreme observations.

[7] The density associated with the distribution $\mathcal{B}(\mu_i,\phi_i)$ is simply a re-parametrization of the standard expression of the $Beta(\alpha_i,\beta_i)$ law, where $\mu_i = \alpha_i/(\alpha_i+\beta_i)$ and $\phi_i = \alpha_i+\beta_i$.

Idiosyncratic bond characteristics are associated with the defaulted issue at hand. The key controls identified in the literature are the maturity and coupon rate (both continuous variables), and the debt seniority, the type of the default event and the industry sector (which are all categorical) (Jankowitsch et al., 2014; Nazemi and Fabozzi, 2018; Nazemi et al., 2018; Gambetti et al., 2019). Beyond accounting data pertaining to the issuer of the defaulted bond, it is important to note that most of the additional variables featured in the existing literature are categorical and display a high granularity. For instance Jankowitsch et al. (2014) considers bond covenants (4 categories), and a dummy (2 categories) for the presence of CDS availability. Nazemi and Fabozzi (2018) and Nazemi et al. (2018) consider two additional industry distress dummies based on performance and sales growth in the year preceding the default event, while Gambetti et al. (2019) considers a dummy for the presence of backing guarantees. These categorical variables (coded using several dummies) lead to a substantial fragmentation of the data into many different groups (defined by the combinations of the dummies), each of them counting very few observations (often, none or just one). We could not find any specific discussion about this phenomenon in the literature, but we believe that this is a major concern that calls for specific counter-measures. Indeed, such sparsity is problematic in regression analysis as it triggers identification problems and significantly increases the risk of overfitting, undermining the model's predictive power and generalization's ability. For example, in the limit where there is a single observation in each group, recovery rates can be explained locally using a single constant (defined by the intercept and the sum of the relevant dummy coefficients).

To address this problem, the number of categories needs to be reduced. Two approaches can be considered for this purpose. The first route consists in adopting a statistical technique that would perform consolidation for the sake of maximizing the model performance. An example of such techniques is fused LASSO (Groll et al., 2019) which penalizes for the number of categories within a same group of variables. Although the latter has the advantage of avoiding to make arbitrary choices, it can easily group categories having little in common such as "Arts, Entertainment, and Recreation" and "Mining" industries, which is questionable. In addition, the outcome of the merging procedure is dependent upon the other controls involved, which makes the model difficult to interpret. Alternatively, one can perform a consolidation based on economic and financial considerations, which has the advantage of being open and explainable. This is the approach followed in this paper. For instance, we restrict ourselves to work with 3 consolidate levels of debt seniorities and default types. The sectors are grouped according to the first digit of the North American Industry Classification System (NAICS) code (hereafter *consolidated sectors*). More details about the consolidation procedure are given in Section 4.1.

The second type of variables relies on systematic factors that capture business cycles and the state of the economy. As standard in the field, we consider the American default rate, the industrial production return, the S&P 500 returns, the Federal Funds Rate, as well as the VIX, and the financial and policy uncertainty measures featured in Gambetti et al. (2019). Those indicators vary with time; most of them are disclosed on a monthly basis. For each factor, the corresponding variable used to explain Y_i is chosen to be the last value of the series available at the default date t_i.

Finally, the third kind of factors deals with credit default swaps, aiming to capture the information embedded in the level and uncertainty embedded in the CDS quotes. Ideally, we would like to match each issue with a CDS contract written on the corresponding underlying bond (or, at least, associated with the same issuer). Unfortunately, this is not possible because only a tiny proportion of defaulted bonds' issuers have CDS quotes.[8] We circumvent this challenge by adopting a sector-based approach and exploiting the CDS term structure.

[8] Out of the 348 firms in the CDS dataset detailed in Section 4.3, only 8 are present in the recovery rate dataset in Section 4.1.

Table 1
List of CDS quote variables, the standard deviations of which are used as volatility measures.

Index	Description	Formula
(a)	Average CDS quotes	$\overline{\mathrm{CDS}}_{s,m}(t)$
(b)	Logarithm of the average CDS quotes	$\ln\left(\overline{\mathrm{CDS}}_{s,m}(t)\right)$
(c)	Logarithm of the ratio of current to previous average CDS quotes	$\ln\left(\frac{\overline{\mathrm{CDS}}_{s,m}(t)}{\overline{\mathrm{CDS}}_{s,m}(t-1)}\right)$
(d)	Returns of average CDS quotes	$\frac{\overline{\mathrm{CDS}}_{s,m}(t)-\overline{\mathrm{CDS}}_{s,m}(t-1)}{\overline{\mathrm{CDS}}_{s,m}(t-1)}$

Our first CDS regressor aims to capture the sectors' CDS levels. At each time point, for each available maturity (tenor) and for each sector, we compute the average of the relevant CDS spreads. Mathematically,

$$\overline{CDS}_{s,m}(t) = \frac{1}{N_{s,m}(t)} \sum_{i \in S_{s,m}(t)} CDS_i(t)\,, \tag{4}$$

where $S_{s,m}(t)$ represents the subset of CDS quotes available at time t whose tenor is m and for which the obligor of the underlying bond belongs to sector s, and $N_{s,m}(t)$ is the number of observations in this set. This provides a time series of cross-sectional means for every pair (s, m). The latter index can then be used to define a novel CDS-based sector-level regressor (hereafter called *sector-level regressor*)[9] of the recovery rate. As for the macro factors, we use a monthly frequency, such that all dynamic variables are available at the beginning of each month. Consequently, we take

$$S_{i,m} := \overline{CDS}_{s(i),m}(t_i^-) \tag{5}$$

as regressor of the recovery rate Y_i, where $s(i)$ is the sector associated with issue i and t_i^- stands for the beginning of the month associated with the default date t_i.

Our second class of CDS-based regressors aims to capture the randomness embedded in CDS quotes by computing *sector-uncertainty* measures. In this study, we first consider four volatility indices based on the standard deviation of transforms of $\overline{\mathrm{CDS}}_{s,m}(t)$, defined as in Table 1, over a 12-month rolling window up to time t. Next, similar to the sector-level regressor, we build four *sector-uncertainty regressors* $\mathrm{Vol}^a_{i,m}$, $\mathrm{Vol}^b_{i,m}$, $\mathrm{Vol}^c_{i,m}$ and $\mathrm{Vol}^d_{i,m}$, computed up to time t_i^-. The above regressors correspond to natural dispersion measures. Unfortunately, they will be proven to be either not significant or highly correlated with the sector-level regressor.[10] This calls for an alternative metric to quantify the randomness of CDS spreads' time series. Building on Gambetti et al. (2019)'s findings, we define a novel sector-level uncertainty proxy based on entropy. Entropy is a common measure of uncertainty (Cover and Thomas, 2006) which has been widely used in various financial applications. For instance, maximum entropy is a standard principle for pricing derivative securities in incomplete markets (Stutzer, 1996; Frittelli, 2000) or to analyze contagion in the interbank market (Mistrulli, 2011). It is also at the heart of the concepts of mutual information and independent component analysis (Hyvarinen et al., 2001), which have been used in recent portfolio optimization schemes (Lassance and Vrins, 2021a; Lassance et al., 2021; Lassance and Vrins, 2021b).

In this work, we choose the approximate entropy as uncertainty metric (Pincus, 1991). In a nutshell, the purpose of the latter is to quantify the complexity and unpredictability of a time series by comparing

[9] Even though we define them as sector-level, they are not the same as a static sector-level covariate, say a dummy variable for identifying the sector.

[10] Table A.12 in the Appendix shows that the correlation between the regressors obtained from the CDS averages and the classical volatility indices ranges from 42.87% to 87.38%. The correlation between the regressors derived from the CDS averages and those derived from the approximate entropy varies between 11.40% and 31.34%. The regression tables featuring $\mathrm{Vol}^a_{i,m}$, $\mathrm{Vol}^b_{i,m}$, $\mathrm{Vol}^c_{i,m}$, and $\mathrm{Vol}^d_{i,m}$ are found in the Online Appendix.

Table 2
Mapping of NAICS *sectors* (NAICS 2-digit) into consolidated categories (NAICS 1-digit) for the sectors present in our dataset. The information about the NAICS *sectors* is retrieved from the U.S. Census Bureau website (https://www.census.gov/naics/).

NAICS 1-digit	NAICS 2-digit	Sector title
2	21	Mining
	22	Utilities
	23	Construction
3	31	Manufacturing
	32	Manufacturing
	33	Manufacturing
4	42	Wholesale Trade
	44	Retail Trade
	45	Retail Trade
	48	Transportation and Warehousing
	49	Transportation and Warehousing
5	51	Information
	52	Finance and Insurance
	53	Real Estate Rental and Leasing
	54	Professional, Scientific, and Technical Services
	56	Administrative and Support and Waste Management and Remediation Services
6	62	Health Care and Social Assistance
7	71	Arts, Entertainment, and Recreation
	72	Accommodation and Food Services

the frequency of similar patterns occurring in consecutive periods. A high level of approximate entropy indicates that the time series exhibits a rich diversity of patterns, pointing to increased complexity. Approximate entropy is considered in Gençay and Gradojevic (2017) as a measure of market agents' uncertainty, able to signal extreme market movements, ahead of the 2008 financial crisis and the 1987 market crash.

Transferring the above methodology to our setting, we consider the approximate entropy of the sector-level index of a sector s and maturity m as a potential indicator capturing the uncertainty of that sector by relying on that tenor. This index is noted as $\overline{AECDS}_{s,m}(t)$, and is technically defined as the approximate entropy of the sector-level index $\overline{CDS}_{s,m}(t)$, calculated over a 12-month rolling window up to time t. The index $\overline{AECDS}_{s,m}(t)$ can then be used to derive a novel CDS-based sector-uncertainty regressor (hereafter referred to as the *sector-uncertainty regressor*). Unlike the volatility metrics presented in Table 1, this approach yields a set of regressors exhibiting low correlation to the CDS averages and whose significance does not vanish when both are included in the regression model, thereby complementing the CDS-level information with a CDS-uncertainty metric. A full description of the model is provided in Appendix B. The above procedure yields, for every issue i, $2m$ regressors based on CDS quotes: the last available observations of the corresponding sector-level CDS index (noted $S_{i,m}$) and of sector-uncertainty CDS index noted

$$AES_{i,m} := \overline{AECDS}_{s(i),m}(t_i^-) \tag{6}$$

for each of the available tenors m.

4. Data

Our study features data collected from different sources. The dependent and idiosyncratic variables associated with the defaulted issues are collected from Moody's. Systematic variables and financial uncertainty measures are obtained from specific repositories, the Federal Reserve Bank of St. Louis database and exchanges. Finally, CDS quotes are retrieved from Thomson Reuters and Datastream.

4.1. Moody's default & recovery database

Data related to the U.S. bond recovery rates is sourced from the Moody's Analytics Default and Recovery Database (Moody's DRD). Our dependent variable corresponds to the *default price* entry, defined as the bond price – recorded 30 days following the default event date declared by the rating agency – expressed as a percentage of the bonds' face value (Moody's, 2023b, p. 7). The dataset handled in this study is obtained by applying several filters to the Moody's DRD, which have been used in influential studies in the field such as Nazemi et al. (2017, 2018), Nazemi and Fabozzi (2018), Gambetti et al. (2019, 2022). This leads us to a dataset composed of 613 defaulted issues from 283 issuers, spanning over a 14-year period going from January 2006 through December 2019.

We apply the following filters. First, we exclude *distressed exchange* entries because the definition of the default price is different from all other default types.[11] Second, we focus on dollar-denominated defaulted corporate bonds issued by U.S. companies that have a par value of at least USD 5 million and have no embedded options. Third, we retain only defaulted issues for which extensive data is available. This includes the bonds' *debt seniority*, the bonds' *default event* type, the issuing company's industrial *sector* as reported either by Moody's internal classification system[12] or by the NAICS code, the amount defaulted, the coupon rate (Coupon Rate), the bonds' maturity (Maturity), and the default date itself. Finally, given that our goal is to study the information conveyed by CDS spreads at the sector level, we discard older data or issues related to sectors for which no CDS quote is available; see Section 4.3 for details.

As discussed in Section 3.3, the granular nature of Moody's original classification system results in a problematic fragmentation of data points across the combined categories of *debt seniority* (8 categories), *default event* (10 categories) and the 3-digit NAICS code (39 categories). To overcome this issue, we define 6 *consolidated sectors* by grouping the NAICS *sectors* by their first digit, as shown in Table 2.[13] This consolidation strategy leads to the sector index s in eq. (4) referring to these broader categories. However, this does not solve the sparsity problem

[11] *"The Default Price column contains the trading price of defaulted issue expressed as a percentage of par. For distressed exchanges, this is the price on the day of the exchange. For all others, this is the price 30 days post-default..."*(Moody's, 2023a, p. 20). This explains why those issues are typically removed from the analysis. A noteworthy exception is Jankowitsch et al. (2014), where the prices 30 days post-default of the distressed exchanges are recovered from an additional source.

[12] Both the Moody's 11-code and Moody's 35-code classification systems.

[13] The NAICS *sectors* are defined by the first 2-digit of the NAICS code, which can count up to 6 digits allowing a finer classification of economic activity. The information is retrieved from the U.S. Census Bureau website (https://www.census.gov/naics/).

Table 3
Mapping of *debt seniority* types from Moody's classification to consolidated categories.

Consolidated category	Original category
Senior Unsecured	Senior Unsecured
Subordinated	Senior Subordinated Subordinated Preferred Stock
Senior Secured	Senior Secured Senior Secured: First Lien Senior Secured: Second Lien Senior Secured: Third Lien

Table 4
Mapping of *default type* categories from Moody's classification to consolidated categories.

Consolidated category	Original category
Bankruptcy	Chapter 11 Bankruptcy Receivership Chapter 7 Liquidated
Prepackaged Chapter 11	Prepackaged Chapter 11
Technical Default	Missed interest payment Missed principal and interest payments Missed principal payment Suspension of payments

because, as illustrated in Fig. 1(a), many category combinations still lack observations. For example, there are only 5 observations associated with the consolidated sector "NAICS: 6". Among those, only two are associated with the "Bankruptcy" default type, whose debt seniority level is "Senior Secured: Second Lien".

To address this issue, we aggregate original Moody's debt and default type categories into broader ones, ensuring a faithful representation of the underlying financial phenomenon. With regards to debt seniority, we consider 3 consolidated categories: "Senior Unsecured", "Subordinated", and "Senior Secured". They are constructed according to the standard mapping given in Table 3. Similarly, we reduce the number of default types to 3 consolidated categories, defined according to Table 4. "Bankruptcy" encapsulates various forms of legal insolvency to provide a unified view of default events that ultimately result in a reorganization or liquidation of the issuer. "Prepackaged Chapter 11" remains a distinct category due to its unique pre-negotiated characteristic between debtors and creditors, which often results in different market reactions and recovery rates. Lastly, "Technical Default" groups together events that signify a breach of contractual terms without immediate insolvency implications. Fig. 1(b) shows how the consolidated categories help alleviate the granularity issue.[14] Hereafter, the terms *debt seniority*, *default event* and *sector* are referring to the consolidated categories.

Fig. 2 provides the histogram of recovery rates within our dataset. The summary statistics shown in Table 5 provide a detailed view of the recovery rates sample (the corresponding kernel densities are provided in the Online Appendix).

As detailed in Table 5(a), the average recovery rate stands at 27.05%, while the standard deviation is equal to 25.31%, which is a large value compared to the mean. The distribution appears to be highly right-skewed, with the majority of observations contained within the [0.01,0.1] interval. Summary statistics of recovery rates conditional on either debt seniority, default type, or sector are available in Table 5(b), 5(c), and 5(d). They are consistent with previous findings on recovery rate predictors, as higher degrees of collateralisation and higher levels of seniority generally lead to higher recovery rates. For instance, controlled reorganizations ("Prepackaged Chapter 11") display higher recoveries than uncontrolled reorganizations and asset liquidations ("Bankruptcy"), reflective of the structured and potentially more creditor-friendly nature of these settlements. "Senior Unsecured" debts register the second lowest average recoveries, highlighting their vulnerability in default scenarios. On the opposite end, "Senior Secured" debts display the highest average recovery rate, underscoring the protective effect of collateral and senior claim status, while at the same time being the most dispersed (the same feature is found in Altman and Kalotay, 2014). As expected, "Subordinated" debts exhibit the lowest mean recovery rates, and display the lowest standard deviation. Sector-specific analysis unveils that recovery rates are, on average, higher in those sectors featuring higher asset tangibility. For instance, consolidated sector identified as "NAICS 2" (encompassing "Mining", "Utilities" and "Construction") displays a relatively high median recovery rate compared to sector "NAICS 5" associated mainly with the "Finance and Insurance" and "Information" industries which displays the lowest median recovery rate.

4.2. Uncertainty measures and other systematic variables

As proxies for the business and credit cycle, we include the monthly American default rate (Am. Def. Rate) computed from the Moody's database by dividing the number of monthly defaults by the total number of companies monitored by Moody's during that month. This calculation aims to minimize the impact of any distortions caused by ratings withdrawal. As a proxy of the performance of the economy, we could include a growth rate either computed from the GDP or the U.S. Industrial Production Index (INDPRO *Ret.*), both conveying very similar information. We favor INDPRO over GDP because the former is available monthly, making the economy of needing to specify an interpolation model from the GDP quarterly data. From the Federal Reserve Economic Data (FRED) repository, we extract the monthly Federal Funds rate (Fed. Funds Rate) as a proxy of the credit cycle. Daily data on the Standard & Poor's 500 index is retrieved from the Chicago Board Options Exchange database, and the monthly return is computed (SP500 *Ret.*).

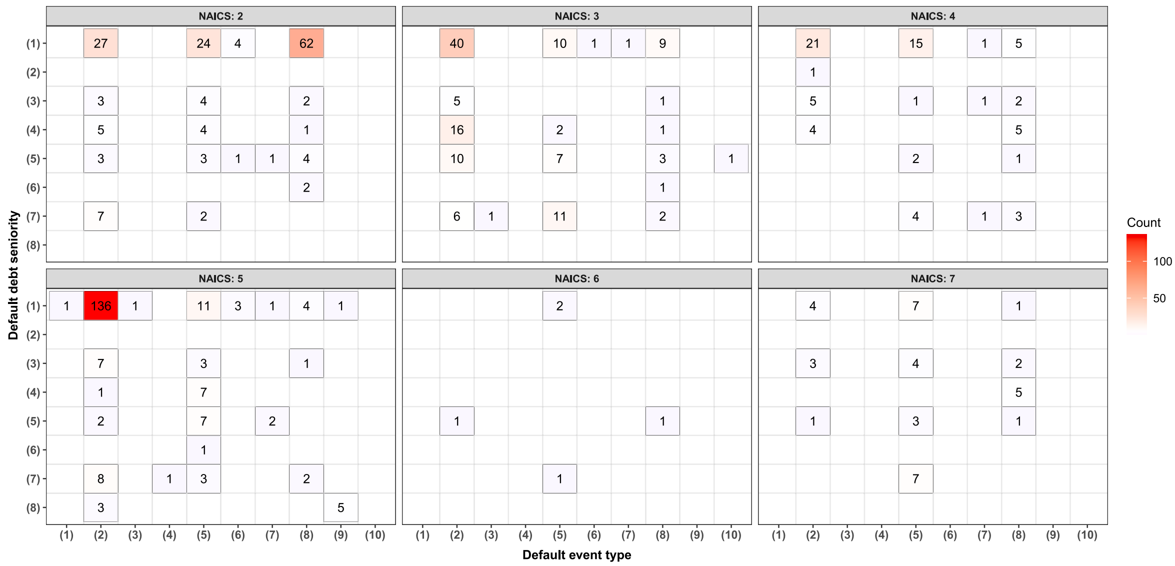
Systematic measures of uncertainty are selected in line with those used by Gambetti et al. (2019), as detailed in Section 2.2. However, we exclude the survey-based metrics drawn from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters, primarily because these are based on the dispersion (interquartile range) of one-year ahead forecasts.[15] The latter is true for both the CPI-based index and the one based on purchases made by U.S. federal and state/local entities. Such a timeline does not align with the temporal framework of the other retained metrics of ECU, which either reflect the current state of uncertainty or offer short-term forecasts. Hence, alongside the VIX sourced from the Chicago Board Options Exchange database, we incorporate the one-month ahead financial uncertainty indicator (Fin. Unc.) by Jurado et al. (2015), retrieved from the authors' website[16] and the U.S. news-based uncertainty proxy (News-based EPU) from the database managed by Baker et al. (2016).[17]

[14] As a robustness check, we have also performed a fused LASSO on the NAICS *sector* categories as detailed in Groll et al. (2019). We found very similar results as when grouping the defaulted issues according to the first digit of the NAICS sector.

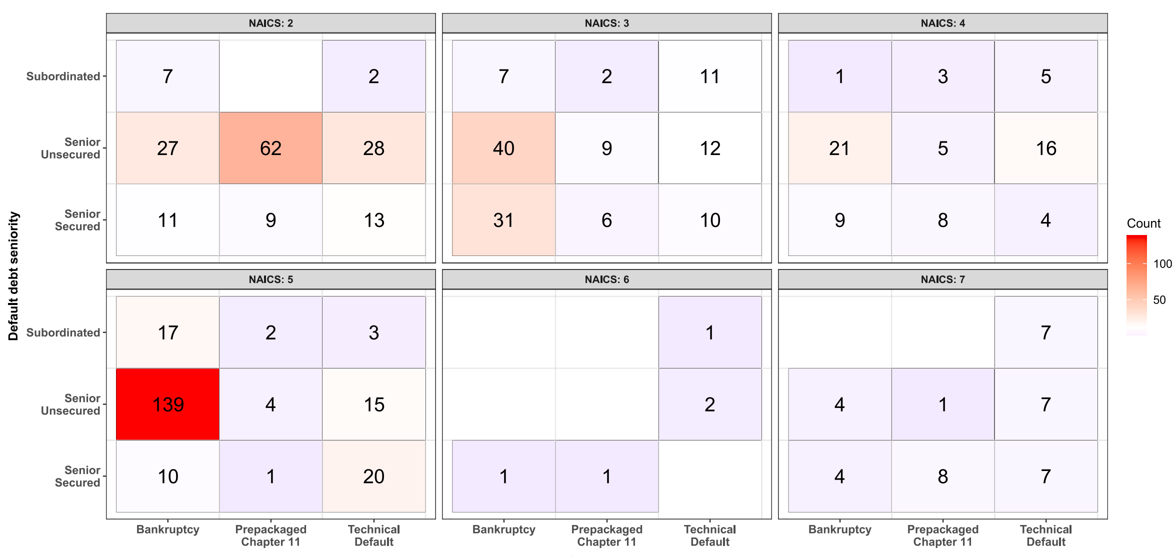
[15] On the section dedicated to detailing their data we find: "...*For each series, we look at the quarterly forecasts for one year in the future*" (see: https://www.policyuncertainty.com/us_monthly.html).

[16] See: https://sydney-ludvigson.squarespace.com/macro-and-financial-uncertainty-indices.

[17] Data and methodology are available at: https://www.policyuncertainty.com/us_monthly.html.



(a) Original classification system as in Moody's DRD. The *default event* types are depicted on the horizontal axis : Bankruptcy (1), Chapter 11 (2), Chapter 7 (3), Liquidated (4), Missed interest payment (5), Missed principal and interest payments (6), Missed principal payment (7), Prepackaged Chapter 11 (8), Receivership (9), Suspension of payments (10). The *debt seniority* types are depicted on the vertical axis: Senior Unsecured (1), Preferred Stock (2), Senior Secured (3), Senior Secured: First Lien (4), Senior Secured: Second Lien (5), Senior Secured: Third Lien (6), Senior Subordinated (7), Subordinated (8).



(b) Consolidated classification system. The vertical and horizontal axes show the consolidated categories for the *debt seniority* and the *default event* type, respectively.

Fig. 1. Numbers of observations (defaulted issues) by default type and debt seniority using (a) the original Moody's DRD classification system and (b) the consolidated debt seniorities and default types as defined in Tables 3 and 4, respectively. The sectors are defined based on the consolidated (first digit) NAICS codes.

4.3. CDS spreads dataset

We collect from Datastream and Thomson Reuters a sample of seven-maturities ($m = 1,2,3,4,5,7,10$ years) CDS quotes associated with 348 firms[18] for which the NAICS code can be retrieved. All CDS contracts are U.S. dollar denominated and the seniority of the underlying bond is senior unsecured. The sample period begins in January 2005 and ends in December 2019 (180 months), which results in 2,436 time series, each of which representing the CDS spread over time for a specific maturity and a given firm. Additionally, we compiled 62,640 term structures, where each firm provides one term structure for each month in the 180-month period. Our sector-specific CDS indicators, introduced in Section 3.3, are computed using the consolidated sectors s defined in Section 4.1. We then connect these CDS indices to the RRs by aligning the first digit of the respective NAICS codes.

Figs. 3 and 4 show the evolution of the sector-level and sector-uncertainty indices $\overline{CDS}_{s,m}(t)$ and $\overline{AECDS}_{s,m}(t)$ for $m = 1,5,10$. For the sector-level indices, we observe spikes of various intensities across sectors during the 2008 financial crisis. The most pronounced spikes are observed for sectors "NAICS: 7", "NAICS: 4" and "NAICS: 5". This underlines the strong impact of the crisis on mainly the trade, financial and entertainment industries. Additionally, we observe a generally upward sloping term structure, except during spike periods. For the

[18] Before September 2010, the data was provided by CMA via Datastream. After that date, the data is provided by Thomson Reuters. The data is combined from the two providers by using the function "SPLC" of Datastream.

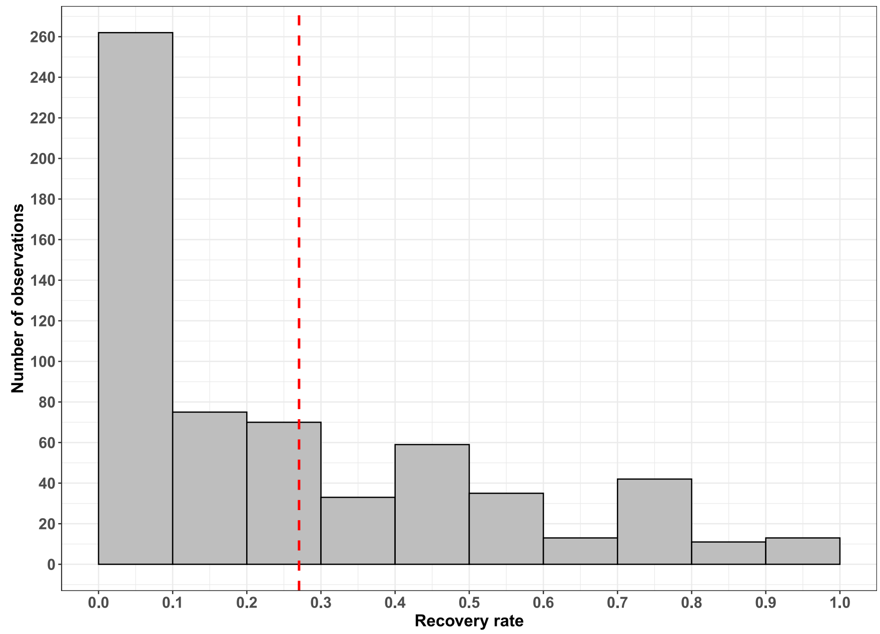


Fig. 2. Histogram of recovery rates in our sample. The vertical dashed line shows the average RR.

Table 5
Summary statistics of the recovery rate sample. All displayed values are expressed as percentages except the sample size (N) and skewness. Pctl(0.25) and Pctl(0.75) denote the 25th and 75th percentiles, respectively.

(a) Whole sample set

	N	Min	Mean	Median	Max	St.dev.	Pctl(0.25)	Pctl(0.75)	Skewness
Recovery rate	613	0.01	27.05	17.50	99.50	25.31	8.50	42.75	1.06

(b) Per consolidated default type

Default event type	N	Min	Mean	Median	Max	St.dev.	Pctl(0.25)	Pctl(0.75)	Skewness
Bankruptcy	329	0.01	20.84	10.00	97.00	22.52	8.50	25.00	1.77
Technical Default	163	0.01	32.56	26.00	96.00	26.85	9.12	50.75	0.61
Prepackaged Chapter 11	121	0.36	36.51	41.75	99.50	25.88	14.00	54.50	0.40

(c) Per consolidated debt seniority

Default debt seniority	N	Min	Mean	Median	Max	St.dev.	Pctl(0.25)	Pctl(0.75)	Skewness
Senior Unsecured	392	0.03	21.62	10.00	95.38	19.75	8.50	30.94	1.41
Senior Secured	153	0.01	46.84	47.50	99.50	29.40	19.25	74.00	0.03
Subordinated	68	0.01	13.81	5.00	80.50	19.19	0.50	18.50	1.87

(d) Per consolidated sector (NAICS 1-digit code)

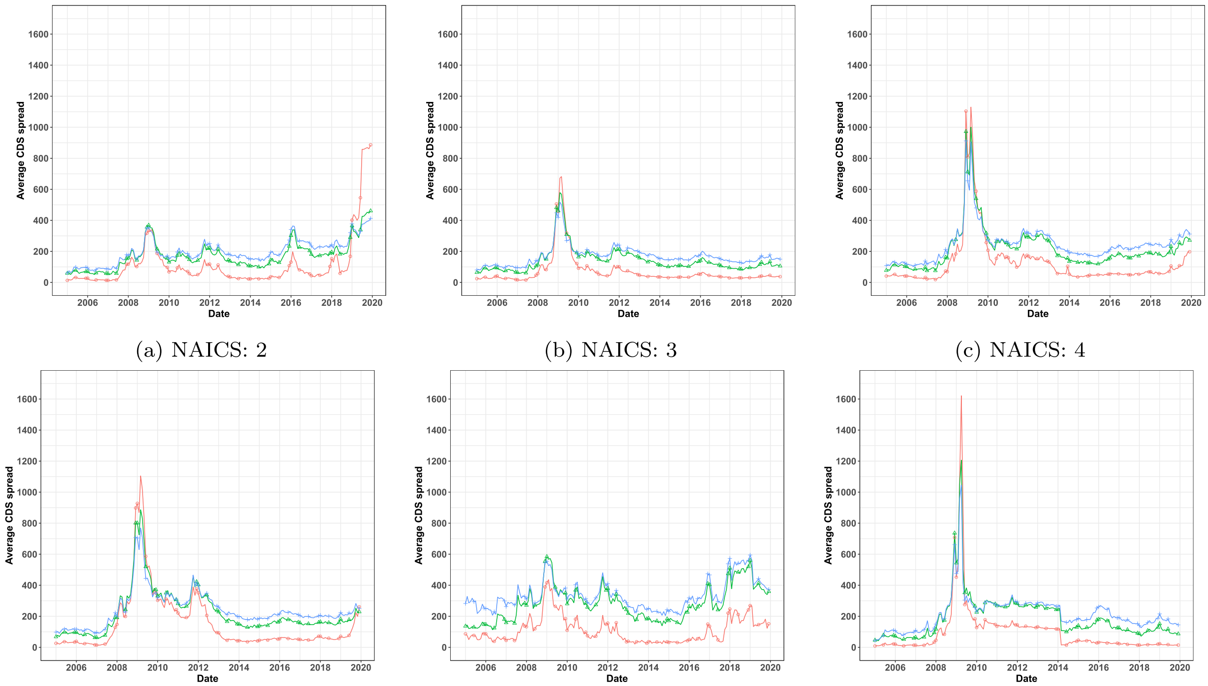
Sector identifier	N	Min	Mean	Median	Max	St.dev.	Pctl(0.25)	Pctl(0.75)	Skewness
NAICS: 5	211	0.01	18.45	10.00	95.38	22.04	8.50	12.50	2.00
NAICS: 2	159	0.01	29.22	24.50	95.00	22.10	8.00	43.38	0.57
NAICS: 3	128	0.01	29.12	17.50	96.00	27.59	9.00	47.81	0.96
NAICS: 4	72	0.88	38.92	29.00	99.50	28.62	19.37	59.13	0.61
NAICS: 7	38	2.00	33.06	23.72	83.00	26.56	11.00	55.38	0.55
NAICS: 6	5	37.00	50.80	49.00	68.00	11.63	45.00	55.00	0.41

sector-uncertainty indices, we observe a volatile approximate entropy, which spikes at different times depending on the underlying CDS maturity. This highlights the fact that uncertainty may behave differently on short-term and long-term tenors.

Note that merging the Moody's and the CDS data leads to discard a few observations: we loose 7 RR observations due to missing NAICS sectors in the CDS spread dataset,[19] and 22 RR observations as we have to wait the first 12 months to compute the first approximate entropy metric, leading to the 613-points recovery rate dataset discussed in Section 4.1.

We conclude this section by providing the full correlation matrix of our numerical independent variables in Table 6. Lines (3) to (9) provide the complete correlation matrix between the systematic regressors. Interestingly, as in Gambetti et al. (2019), the maximum correlation is found between the VIX index and the financial uncertainty indicator, with a value of 90.10%. This high value suggests considering these two variables separately in our regression exercise. Additionally, we find high values (all above 70%) between: (i) the systematic proxies of ECU and the industrial production return and (ii) the financial uncertainty indicator and the American default rate. We do not identify other colinearity issues through pairwise correlations. A finer analysis

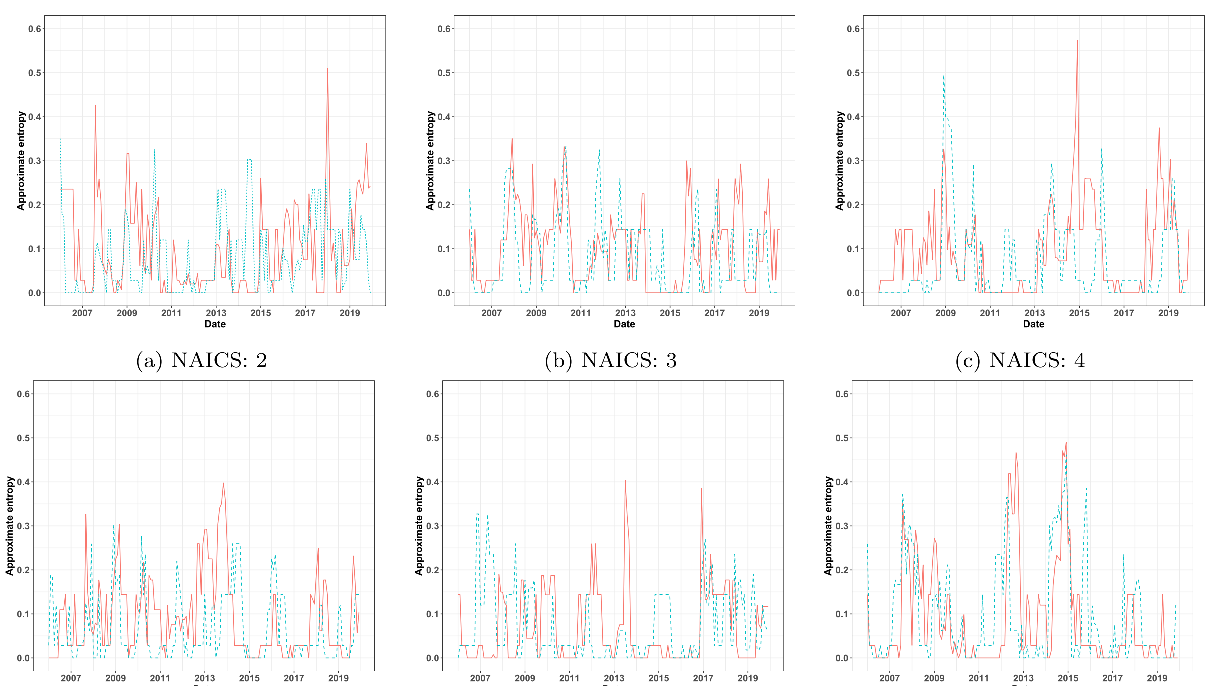
[19] In the Moody's dataset we have two additional sectors: *Agriculture, Forestry, Fishing and Hunting* (NAICS: 11), and *Public Administration* (NAICS: 92).



(a) NAICS: 2   (b) NAICS: 3   (c) NAICS: 4

(d) NAICS: 5   (e) NAICS: 6   (f) NAICS: 7

Fig. 3. Evolution of the sector-level index $\overline{CDS}_{s,m}(t)$ for every *sector* (panels) and *maturity* (markers). For conciseness, we focus on $m = 1$ (○), 5 (△) and 10 (+).



(a) NAICS: 2   (b) NAICS: 3   (c) NAICS: 4

(d) NAICS: 5   (e) NAICS: 6   (f) NAICS: 7

Fig. 4. Evolution of the sector-uncertainty index $\overline{AECDS}_{s,m}(t)$ for every *sector* (panels) and *maturity* (line type). For conciseness, we focus on $m = 1$ (solid) and 10 (dashed).

Table 6
Lower triangle of the correlation matrix between our numerical independent variables. We display the correlations among: idiosyncratic variables (first two rows), systematic measures (next seven rows), and the CDS-based regressors (last two rows). All values are expressed as percentages.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Coup. Rate	100										
(2)	Maturity	−41.52	100									
(3)	Am. Def. Rate	−40.22	32.54	100								
(4)	INDPRO *Ret.*	56.81	−50.67	−60.58	100							
(5)	SP500 *Ret.*	41.64	−33.83	−15.58	66.56	100						
(6)	Fed Fund Rate	−31.05	24.44	−0.27	−32.45	−29.63	100					
(7)	Fin. Unc.	−46.24	44.74	72.30	−79.99	−53.35	20.50	100				
(8)	VIX	−39.62	38.98	62.36	−72.95	−63.31	4.72	90.10	100			
(9)	News-based EPU	−49.75	40.20	37.30	−71.16	−66.87	28.90	64.09	63.12	100		
(10)	$S_{i,10}$	−21.57	17.96	59.13	−25.17	8.33	−21.02	52.97	50.18	22.70	100	
(11)	$AES_{i,1}$	13.54	−19.38	−9.17	35.48	30.23	−0.17	−23.94	−31.34	−22.84	13.50	100

is conducted on the regression models featured in Section 5, using the variance inflation factors (VIFs).

5. Empirical results

We investigate the effect of the regressors introduced above on the mean and precision parameters of a beta distribution. In Section 5.1, we present the results for the *classical* beta regression model (with fixed dispersion ϕ). In Section 5.2, we perform model selection on the classical beta regression models, and in Section 5.3 we present the outcomes of a series of robustness checks. The results for the *variable* beta regression model are discussed in Section 5.4.

5.1. Predictors of expected recovery rates

We denote the models as $\mathcal{M}_i^\mu$, where $i = \{1, \ldots, 8\}$ stands for the identifier of the model and the superscript μ indicates that only the mean of the distribution varies across issues.

The results are shown in Table 7. They illustrate how the performance of a model varies as a result of the addition of security-specific characteristics and systematic factors ($\mathcal{M}_2^\mu$) and economic uncertainty measures ($\mathcal{M}_{3-5}^\mu$) as well as the dynamic sector-specific factors ($\mathcal{M}_{6-8}^\mu$). As suggested in Ferrari and Cribari-Neto (2004), we compare model performances in terms of pseudo-R^2 (R_p^2), defined as the squared sample correlation coefficient between the linear predictor $x^\top\theta^\mu$ and the logit link-transformed response $\ln(y/(1-y))$. Additionally, we include both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), to measure the trade-off between goodness-of-fit and model complexity. Even though both metrics serve a similar purpose, the penalization scheme for model complexity differs (with BIC generally being more conservative). Reporting both measures thus offers a more comprehensive analysis of the model's performance.

In Model $\mathcal{M}_1^\mu$, we incorporate consolidated categorical variables such as the seniority of defaulted bonds, the industrial sector of issuers, and the default type, as outlined in Table 5. Our benchmark model is $\mathcal{M}_2^\mu$. It integrates all security-specific variables (categorical and numerical) alongside controls for business and credit cycles, detailed in Section 4.2. This model achieves an R_p^2 of approximately 30.9% and supports previous findings. Notably, it confirms a significant inverse relationship between default rate and recovery rates (Altman et al., 2005), a positive relation with stock market returns (Mora, 2015) as well as a positive link with the Federal Funds rate (Jankowitsch et al., 2014). Furthermore, the model reveals that higher coupon rates are associated with lower average recovery rates.

Models $\mathcal{M}_3^\mu$ to $\mathcal{M}_8^\mu$ feature the same controls as in our benchmark model $\mathcal{M}_2^\mu$. In $\mathcal{M}_3^\mu$ to $\mathcal{M}_5^\mu$, we sequentially introduce proxies of systematic uncertainty. These results partially corroborate Gambetti et al. (2019)'s findings; the retained ECU proxies are statistically significant only in models that do not feature additional systematic variables (see Table D.14 in the Appendix for details). Note that these models suffer from potential multicollinearity issues, as highlighted in the fourth column of Table 6. For the given dataset and time horizon, there is a high correlation (above 70%) between these systematic uncertainty metrics and our business and credit cycle controls, as explained at the end of Section 4.2. This is the reason why we do not investigate further the joint effect of these systematic uncertainty proxies.

Models $\mathcal{M}_6^\mu$ to $\mathcal{M}_8^\mu$ incorporate our CDS-based covariates, first separately and then jointly. In $\mathcal{M}_6^\mu$, we include the cross-sectional average of the CDS spreads ($S_{i,m}$). All maturities of $S_{i,m}$ provide a statistically significant[20] and informative regressor as shown in Table C.13 in the Appendix, we find only a marginal difference in terms of coefficients magnitude and R_p^2 contribution going from 32.06% (for the 1-year) to 32.39% (for the 10-year). Overall, higher sector CDS spreads are associated with lower means of recovery rates. We present in Table 7 only the marginally most informative $S_{i,m}$, which is the 10-year tenor.[21] Model $\mathcal{M}_7^\mu$ includes our proxy for sector-level uncertainty. As shown in the last three lines of Table C.13, we find that only the approximate entropy computed from the average CDS spreads with a 1-year underlying maturity ($AES_{i,1}$) is statistically significant in explaining the RR. It appears that it is the approximate entropy of the most sensitive time series (most short-term looking) at our disposal that conveys the highest information regarding the impact of uncertainty on the mean of the recovery rate distributions. In essence, the higher the entropy, the greater the unpredictability embedded in the sector, which in turn impacts the expected recovery rate. As in Gambetti et al. (2019), uncertainty is negatively related to the mean of the recovery rate distribution. It considerably improves the model explanatory power, leading to the highest R_p^2 so far, approximately equal to 34.5%. Nevertheless, in contrast with proxies of systematic uncertainty, only the sector-level uncertainty is statistically significant. The different behavior of systematic and sector-uncertainty variables suggests that they convey different information about the parameters of the recovery rate distributions.

Model $\mathcal{M}_8^\mu$ includes both CDS-based regressors. This model emerges as the most effective in explaining the mean of the distributions of recovery rates, as highlighted by its R_p^2, AIC, and BIC values. The proposed sector-level regressors are statistically significant and exhibit coefficients of comparable or greater magnitude than other covariates.

5.2. Model selection

Beta regression is not a generalized linear model, it belongs to the family of generalized additive models for location, scale and shape

[20] We rely on the standard level of 95% confidence level to identify statistical significance.

[21] Similar results are obtained when considering the $m = 5$-year tenor, usually considered as the most liquid CDS maturity. The R_p^2 of model drops from 32.39% to 32.18% and the magnitude of the estimated coefficient changes from −0.2271 to −0.2114.

Table 7
Classical beta regression models for the mean of the recovery rate distribution. All models include consolidated categorical variables for controlling for bonds' seniority, default type and industrial sector of the bond issuer as depicted in Table 5. The variables $S_{i,m}$ and $AES_{i,m}$ are defined in eq. (5) and eq. (6), respectively.

	Dependent variable:							
	RR distributions - Mean model							
	$\mathcal{M}_1^\mu$	$\mathcal{M}_2^\mu$	$\mathcal{M}_3^\mu$	$\mathcal{M}_4^\mu$	$\mathcal{M}_5^\mu$	$\mathcal{M}_6^\mu$	$\mathcal{M}_7^\mu$	$\mathcal{M}_8^\mu$
Constant	−1.3757***	−1.1217***	−1.1247***	−1.1214***	−1.1244***	−1.1197***	−1.2921***	−1.2708***
	(0.0789)	(0.1105)	(0.1106)	(0.1105)	(0.1106)	(0.1098)	(0.1135)	(0.1132)
Coupon Rate		−0.0954*	−0.0952*	−0.0957*	−0.0960*	−0.1248**	−0.1051**	−0.1252**
		(0.0539)	(0.0540)	(0.0542)	(0.0540)	(0.0541)	(0.0532)	(0.0536)
Maturity		−0.0484	−0.0445	−0.0483	−0.0494	−0.0244	−0.0442	−0.0291
		(0.0481)	(0.0483)	(0.0482)	(0.0482)	(0.0483)	(0.0478)	(0.0481)
Am. Def. Rate		−0.2324***	−0.1992***	−0.2320***	−0.2419***	−0.1121*	−0.1682***	−0.1002
		(0.0582)	(0.0685)	(0.0583)	(0.0661)	(0.0646)	(0.0586)	(0.0642)
INDPRO *Ret.*		0.0334	0.0039	0.0318	0.0376	0.0385	0.1023	0.0977
		(0.0905)	(0.0969)	(0.0928)	(0.0917)	(0.0899)	(0.0904)	(0.0902)
SP500 *Ret.*		0.2212***	0.2094***	0.2197***	0.2313***	0.2785***	0.2253***	0.2624***
		(0.0598)	(0.0618)	(0.0639)	(0.0699)	(0.0606)	(0.0591)	(0.0601)
Fed. Funds Rate		0.1636***	0.1642***	0.1637***	0.1676***	0.1187***	0.2139***	0.1793***
		(0.0438)	(0.0438)	(0.0438)	(0.0455)	(0.0449)	(0.0442)	(0.0463)
Fin. Unc.			−0.0748					
			(0.0842)					
News-based EPU				−0.0044				
				(0.0622)				
VIX					0.0241			
					(0.0801)			
$S_{i,10}$						−0.2271***		−0.1450**
						(0.0566)		(0.0585)
$AES_{i,1}$							−0.2534***	−0.2198***
							(0.0478)	(0.0495)
Observations	613	613	613	613	613	613	613	613
Debt seniority dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry sector dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Default event dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.2578	0.309	0.3095	0.3091	0.3091	0.3239	0.3446	0.3494
AIC	−567.97	−607.06	−605.79	−605.07	−605.15	−619.8	−634.08	−637.65
BIC	−519.37	−531.95	−526.26	−525.54	−525.62	−540.27	−554.55	−553.71
Log Likelihood	294.9853	320.5317	320.8942	320.5342	320.5762	327.8993	335.0417	337.827

Note: *p<0.1; **p<0.05; ***p<0.01.

(GAMLSS). Therefore, we perform variable selection using backward and forward stepwise algorithms based on the generalized Akaike information criterion (step-GAIC), as outlined in Stasinopoulos and Rigby (2008).[22]

The results for the backward and forward procedures are given in Tables 8 and 9, respectively. In the backward step-GAIC, we start from $\mathcal{M}_8^\mu$ and decide, at each iterative step, which variable to remove – if any – in order to minimize the AIC of the resulting model. This allows us to assess the contribution of each variable in the presence of all the other relevant ones, and helps identifying the important variables when accounting for multicollinearity. The first two variables to be removed are Maturity and INDPRO *Ret.*, in this order. The algorithm then stops as it fails to improve the AIC of the resulting model by removing another explanatory variable. In the forward step-GAIC, we set $\mathcal{M}_1^\mu$ as baseline model (featuring just a constant and the dummies) and the target model as $\mathcal{M}_8^\mu$. At each iterative step, the inclusion of an additional covariate – among those in $\mathcal{M}_8^\mu$ – is determined by its contribution in minimizing the AIC of the augmented model. The order of the variables shown in Table 9 coincides with that in which the algorithm adds them sequentially to the baseline model.

Interestingly, both forward and backward approaches converge to the same final model, suggesting that $\mathcal{M}_8^{\mu*}$ is well-specified and its explanatory variables are consistently relevant.[23] Notably, the second place of $S_{i,10}$ in Table 9 shows that our new CDS-based regressor is the most informative. In particular, it is prioritized over other variables previously identifies ad key predictors of RR, being either idiosyncratic (such as coupon rate) or systematic (such as the default rate). This robustness check yields a parsimonious model, $\mathcal{M}_8^{\mu*}$, which effectively balances model complexity with explanatory power, as reflected in its R_p^2, AIC and BIC.

[22] As explained in Mantel (1970), backward stepwise selection is preferred, particularly in contexts with a smaller number of regressors relative to observations, to mitigate multicollinearity effects. This approach, by evaluating variables within the full model context, can better discern the unique contributions of each regressor, enhancing model reliability and interpretability (Heinze et al., 2018).

[23] Notice that in both cases, the deleted variables from $\mathcal{M}_8^\mu$ are amongst the ones that were not statistically significant already in $\mathcal{M}_8^\mu$.

Table 8
Backward step-GAIC on $\mathcal{M}_8^\mu$.

	Dependent variable:
	RR distributions (Mean submodel) $\mathcal{M}_8^{\mu^*}$
Constant	-1.3273^{***} (0.1042)
Coupon Rate	-0.1162^{**} (0.0531)
Am. Def. Rate	-0.1391^{**} (0.0563)
SP500 *Ret.*	0.3003^{***} (0.0531)
Fed. Funds Rate	0.1635^{***} (0.0445)
$S_{i,10}$	-0.1547^{***} (0.0581)
$AES_{i,1}$	-0.2119^{***} (0.0489)
Observations	613
Debt seniority dummies	Yes
Industry sector dummies	Yes
Default event dummies	Yes
Pseudo R-squared	0.347
AIC	−639.96
BIC	−564.85
Log Likelihood	336.98

Note: *p<0.1; **p<0.05; ***p<0.01.

Table 9
Forward step-GAIC from $\mathcal{M}_1^\mu$ to $\mathcal{M}_8^\mu$.

	Dependent variable:
	RR distributions (Mean submodel) $\mathcal{M}_8^{\mu^*}$
Constant	-1.3273^{***} (0.1042)
$S_{i,10}$	-0.1547^{***} (0.0581)
SP500 *Ret.*	0.3003^{***} (0.0531)
$AES_{i,1}$	-0.2119^{***} (0.0489)
Fed. Funds Rate	0.1635^{***} (0.0445)
Am. Def. Rate	-0.1391^{**} (0.0563)
Coupon Rate	-0.1162^{**} (0.0531)
Observations	613
Debt seniority dummies	Yes
Industry sector dummies	Yes
Default event dummies	Yes
Pseudo R-squared	0.347
AIC	−639.96
BIC	−564.85
Log Likelihood	336.98

Note: *p<0.1; **p<0.05; ***p<0.01.

A preliminary check based on the correlation matrix in Table 6 leads us to exclude some pairs of variables because of obvious multicollinearity issues. To further capture the interactions between *combinations* of variables, we rely on the generalized variance inflation factors (GVIF) (Fox and Monette, 1992). The GVIF is an extension of the VIF designed to better handle categorical variables (displaying $n > 1$ categories coded using $\mathrm{Df} = n - 1$ dummies, each of which having its own scaling coefficient). Multicollinearity issues can be ruled out if all variables in the model satisfy $\mathrm{GVIF}^{1/\mathrm{Df}} < \epsilon$ where Df is the degrees of freedom and ϵ is some reasonable threshold. This inequality can be interpreted in a similar fashion as $\mathrm{VIF} < \epsilon$ for continuous variables. A typical value for ϵ is 5 (Fisch and Momtaz, 2020; Tajvarpour and Pujari, 2022). In $\mathcal{M}_8^{\mu*}$, all retained covariates meet this inequality for all $\epsilon \geq 2.14$.

5.3. Robustness check

We have conducted a series of additional robustness checks on the models featured in Table 7 to ensure the reliability and validity of our findings.[24] In this section, we present the key findings, while the complete analysis along with the corresponding regression tables are provided in the Online Appendix.

First, we investigate the impact of controlling for sector-specific default rates by replacing our U.S. aggregate default rate (Am. Def. Rate) with Expected Default Frequencies (EDFs) indices obtained from Moody's. Our best model remains $\mathcal{M}_8^\mu$, where $S_{i,10}$ and $AES_{i,1}$ are highly significant, with coefficients of similar magnitudes to those in Table 7. Moreover, the R_p^2 of our best models are about the same, suggesting that the information provided by the EDF indices aligns closely with that of the aggregate American default rate. The EDF indices are thus complementary to, rather than replacements for, our sector-based CDS signals.

[24] We are grateful to an anonymous referee for valuable suggestions in this respect.

Table 10
Variable beta regression models for the mean and precision of the recovery rates. Models feature a logit link for the mean submodel and log-link for the precision submodel. Variables have been selected with backward step-GAIC. The starting model for the variable selection procedure generating $\mathcal{M}_8^{\mu,\phi^*}$ is model $\mathcal{M}_8^{\mu}$ presented in Table 7. All models include consolidated categorical variables for controlling for bonds' seniority, default type and industrial sector of the bond issuer as depicted in Table 5. The variables $S_{i,m}$ and $AES_{i,m}$ are defined in eq. (5) and eq. (6), respectively.

	Dependent variable:	
	RR distributions (Mean model) $\mathcal{M}_8^{\mu,\phi^*}$	RR distributions (Precision submodel) $\mathcal{M}_8^{\mu,\phi^*}$
Constant	−1.5730*** (0.1085)	1.8878*** (0.1337)
Coupon Rate	−0.1096** (0.0512)	
Am. Def. Rate	−0.2409*** (0.0566)	0.3739*** (0.0625)
SP500 *Ret.*	0.3524*** (0.0552)	−0.3645*** (0.0690)
Fed. Funds Rate	0.1375*** (0.0470)	
$S_{i,10}$	−0.1143** (0.0569)	
$AES_{i,1}$	−0.1361** (0.0540)	−0.1691*** (0.0619)
Observations	613	
Debt seniority dummies	Yes	
Industry sector dummies	Yes	
Default event dummies	Yes	
Pseudo R-squared	0.3264	
AIC	−764.8	
BIC	−636.67	
Log Likelihood	411.4018	

Note: *p<0.1; **p<0.05; ***p<0.01.

Second, we perform a horse race to compare the explanatory power of CDS spreads against corporate bond market spreads (ICE indices). To that end, following the same methodology as in Section 3.3, we replace $S_{i,10}$ and $AES_{i,1}$ by comparable signals built from ICE indices. Our results are robust to the credit-risk index being chosen, as both bonds and CDS signals convey relevant and consistent information about recovery rates. Interestingly, CDS spreads exhibit greater explanatory power. In fact, for a fixed specification, models using *sector-level* and *sector-uncertainty* regressors derived from CDS spreads consistently achieve a higher R_p^2 than those based on ICE indices.

Third, we extend the sample of recovery rates (RR) beyond the criteria introduced in Section 4.1 to assess the impact of the USD 5 million cutoff and the exclusion of embedded options. The findings remain consistent, with no significant changes in the estimated coefficients or their statistical significance. In fact, for this extended sample, the R_p^2 further improves in model $\mathcal{M}_8^{\mu}$ to 35.83% *vs.* 34.94% in Table 7.

Fourth, we consider alternative proxies for the credit cycle that capture medium to long-term financing conditions and market risk appetite, rather than relying solely on the Federal Funds rate. While R_p^2 slightly varies with the choice of proxy, both $S_{i,10}$ and $AES_{i,1}$ remain significant, with coefficients close in magnitude to those in Table 7.

These results demonstrate the robustness of the CDS-based measures in predicting recovery rates.

5.4. Predictors of dispersion of the recovery rates

This section evaluates the effectiveness of covariates considered in Section 5.1 for the sake of explaining both the mean and dispersion of the beta distributions associated with recovery rates. These models are denoted as $\mathcal{M}_8^{\mu,\phi*}$.

As discussed in Simas et al. (2010), the *variable* beta regression model requires estimating more parameters ($J + K + 2$ instead of $J+2$). To deal with the trade-off between creating an over-parametrized model and understanding which factors are the most important to explaining the shape of recovery rate distributions, we perform model selection, similarly as in Section 5.2. As usual for this type of models, we first select the relevant variables for the mean submodel, and proceed next with the selection of the variables for the precision submodel. Hence, $\mathcal{M}_8^{\mu}$ is our starting model for the backward step-GAIC selection procedure which now aims at selecting the relevant variables for the precision parameter among those used for the mean. Table 10 confirms that the optimal mean submodel $\mathcal{M}_8^{\mu,\phi*}$ is close to $\mathcal{M}_8^{\mu*}$, albeit with slightly different mean coefficients due to joint parameter estimation with the precision submodel. The results for the forward step-GAIC are presented in Table D.15 of the Appendix and depict similar outcomes.

Our findings reinforce the statistical significance of our novel CDS-based metrics in explaining the mean of recovery rates; both covariates are retained in the joint parameters estimation. In particular, our two CDS-based regressors remain statistically significant to explain the mean of the beta distribution, although at 5% confidence level. In addition, our CDS-based uncertainty regressor is also significant to explain the precision of the distribution, at 1% confidence level. The sign of this coefficient is also intuitive: for a given μ, an increase in $AES_{i,1}$ leads to a lower precision, ϕ. This means that when CDS uncertainty increases within a sector, the recovery rate is expected to diminish on average, but also tends to exhibit more dispersion. The same kind of relationship between uncertainty and precision was found Gambetti et al. (2019) when considering their news-based ECU proxy. The latter, however, only quantifies a systematic phenomenon, not a sectorial effect. In addition, for the given dataset and time horizon, only our sector-level uncertainty appears to be statistically significant to explain recovery rates: when controlling for business and credit cycles, all the other systematic uncertainty proxies become irrelevant. Similar conclusions were reached for the classical beta regression models presented in Section 5.1.

In contrast, we observe a different behavior for our systematic controls. For both the American default rate and the S&P 500, we observe an opposite impact on the mean and the dispersion of the recovery rate distributions. Notably, this implies that in high default periods, recovery rates tend to be lower with a lower variance, indicating a concentration of the density on the left; vice-versa for the stock-market returns. Both effects align with previous empirical evidence (Altman et al., 2005; Mora, 2015).

6. Predictability of mean recovery rates

The goal of this section is to investigate the predictive power of our new CDS-based regressors. To this end, we focus specifically on lagging the CDS-based variables while keeping the other regressors at their values closest to the default event. This approach allows us to isolate and examine the role of CDS spreads and entropy measures in predicting mean recovery rates when we are ahead of the default event.

As explained in Section 3.3, we denote the latest available time before default as t_i^- (i.e., the beginning of the month associated with the default date t_i). For this analysis, we lag our CDS-based regressors by k-months, where $k = 1, 2, 3$, and compare the predictability of mean recovery rates across these lag periods. For consistency with Section 5.1, we only show the results for our most informative model $\mathcal{M}_8^\mu$ and verified that all other tenors combinations lead to a model featuring a lower R_p^2. Table 11 presents the regression results, showing the performance of the sector-level CDS predictor ($S_{i,10}$) and the entropy-based uncertainty measure ($AES_{i,1}$) at various lags.

The results in Table 11 demonstrate that both CDS-based regressors remain statistically significant across all k periods, though their coefficients decrease slightly as the lag increases. For instance, the coefficient for $S_{i,10}$ declines from −0.1450 at (t_i^-) to −0.1341 at $(t_i^- - 3)$, while the coefficient for $AES_{i,1}$ decreases from −0.2198 to −0.0949 over the same period. This suggests that although more recent CDS data is indeed more informative, these measures continue to provide valuable predictive insights even when sampled months before the default event.

Furthermore, the overall model fit, as indicated by R_p^2, decreases slightly as the CDS-based regressors are lagged further from the default date, ranging from 34.94% at (t_i^-) to 31.16% at $(t_i^- -3)$. This is expected, as more recent data typically captures the most relevant information. Despite this, the consistent significance of $S_{i,10}$ and $AES_{i,1}$ across all lags underscores their robustness in predicting the mean of the RR. Additionally, macroeconomic variables such as the SP500 *Ret.* and the Federal Funds Rate consistently exhibit strong significance, reinforcing their role in influencing recovery outcomes.

These findings highlight the predictive power of CDS-based regressors for explaining the mean of the recovery rate distributions, even when sampled months before the default event.

7. Conclusion

Previous studies relying on machine learning techniques or generalized regressions have highlighted the importance of idiosyncratic characteristics such as the debt seniority or the coupon rate, as well as of systematic factors such as the default rate. More recently, systematic indicators of financial and policy uncertainty have also been proven to matter. In this paper, we introduce two novel regressors exploiting the information conveyed by CDS quotes gathered at a consolidated sector level defined by the first digit of NAICS code. These factors capture the level and uncertainty embedded in CDS quotes for each consolidated sector, as measured by the cross-sectional mean of average CDS at the default date and the approximate entropy of the average CDS computed over a 12-month window preceding the default date. As such, they capture the information associated with the sectors dynamics, and sit in between idiosyncratic and systematic variables considered so far. To the best of our knowledge, this is the first attempt to investigate the power of CDS quotes to explain recovery rates of defaulted bonds.

To analyze the usefulness of those indicators, we conduct an extensive empirical study featuring more than 600 defaulted bonds issued by U.S. corporates extracted from the authoritative Moody's default and recovery database. The latter is then crossed with other databases such as FRED, Datastream, Thomson Reuters, from which systematic variables and CDS quotes are recovered. A careful analysis of the dataset at hand reveals that special actions need to be taken to avoid the trap of overfitting. This can be explained by the sparsity resulting from the data fragmentation associated with the numerous groups defined by the combination of categorical variables. Although we could not find a similar discussion in earlier papers, this is a material problem calling for specific measures, as most groups feature none or just one observation. To circumvent this problem, a consolidation procedure that merges dummies and defines consolidated categories is proposed. This procedure is based on financial considerations and has the advantage to be transparent and explainable. Interestingly, robustness checks reveal that alternative data-driven consolidation procedures such as fused LASSO lead to similar conclusions. Our results suggest that our CDS-based metrics are relevant predictors of recovery rates and should be considered as material factors in inference and forecasting exercises.

Beta regression models fitted to our dataset indicate that our CDS-based variables can be considered as significant predictors. For instance, the sector-level CDS regressor associated with the 10-year maturity is the first variable being selected in a backward step-GAIC exercise featuring a mean-model, suggesting a material explanatory power. The sector-uncertainty CDS regressor associated with the 1-year tenor is also material, and is detrimental to the systematic measures of uncertainty. As for the variable beta regression model – which explains both the mean and precision parameters – both the sector-level and sector-uncertainty CDS regressors matter to explain the mean parameter, but only the latter matters in the precision sub-model. Interestingly, all the systematic measures of financial uncertainty become irrelevant when business cycles and the sector-uncertainty variables are included.

CRediT authorship contribution statement

Matteo Barbagli: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pascal François:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Geneviève Gauthier:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Frédéric Vrins:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work of Frédéric Vrins is supported by the Fonds de la Recherche Scientifique - (F.S.R.-FNRS) grant CDR J.0037.18 and by the grant ARC 18-23/089 of the Belgian Federal Science Policy Office, Belgium. Geneviève Gauthier is supported by NSERC, Canada (RGPIN-2024-03791), a professorship funded by HEC Montréal, Canada, and the HEC Montréal Foundation. Pascal François is supported by a fellowship from the Canadian Derivatives Institute, Canada.

Table 11
Classical beta regression models for the mean of the recovery rates, with lagged CDS-based regressors. The columns represent different lag periods, where t_i^- denotes the closest available time before the default event, and $(t_i^- - k)$ corresponds to lagged regressors k-months prior to t_i^-. Hence, in this table, $S_{i,m} := \overline{CDS}_{s(i),m}(t_i^- - k)$ and $AES_{i,m} := \overline{AECDS}_{s(i),m}(t_i^- - k)$, for $k = 0, 1, 2, 3$. All models include consolidated categorical variables for controlling for bonds' seniority, default type and industrial sector of the bond issuer as depicted in Table 5.

	Dependent variable:			
	Recovery rate distributions - Mean model			
	(t_i^-)	$(t_i^- - 1)$	$(t_i^- - 2)$	$(t_i^- - 3)$
Constant	−1.2708***	−1.1506***	−1.1411***	−1.1379***
	(0.1132)	(0.1106)	(0.1125)	(0.1122)
Coupon Rate	−0.1252**	−0.1216**	−0.1023*	−0.1000*
	(0.0536)	(0.0542)	(0.0539)	(0.0544)
Maturity	−0.0291	−0.0419	−0.0529	−0.0634
	(0.0481)	(0.0480)	(0.0480)	(0.0485)
Am. Def. Rate	−0.1002	−0.1105	−0.1191*	−0.0961
	(0.0642)	(0.0672)	(0.0720)	(0.0761)
INDPRO *Ret.*	0.0977	0.0833	0.1215	0.1593
	(0.0902)	(0.0905)	(0.0943)	(0.0994)
SP500 *Ret.*	0.2624***	0.2174***	0.2144***	0.2198***
	(0.0601)	(0.0616)	(0.0600)	(0.0608)
Fed. Funds Rate	0.1793***	0.1488***	0.1534***	0.1331***
	(0.0463)	(0.0462)	(0.0458)	(0.0448)
$S_{i,10}$	−0.1450**	−0.1509**	−0.1142*	−0.1341**
	(0.0585)	(0.0588)	(0.0599)	(0.0657)
$AES_{i,1}$	−0.2198***	−0.1276***	−0.0950*	−0.0949**
	(0.0495)	(0.0439)	(0.0485)	(0.0469)
Observations	613	613	613	613
Debt seniority dummies	Yes	Yes	Yes	Yes
Industry sector dummies	Yes	Yes	Yes	Yes
Default event dummies	Yes	Yes	Yes	Yes
Pseudo R-squared	0.3494	0.3317	0.3175	0.3116
AIC	−637.65	−619.54	−610.89	−610.14
BIC	−553.71	−535.59	−526.94	−526.28
Log Likelihood	337.827	328.77	324.4463	324.0691

Note: *p<0.1; **p<0.05; ***p<0.01.

Appendix A. Correlations between sector-level and sector-uncertainty regressors

See Table A.12.

Appendix B. Approximate entropy

Approximate entropy is a statistical tool used to quantify the level of complexity or unpredictability present in a time series (Pincus, 1991). This method hinges on comparing the frequency of similar patterns, referred to as templates, that occur in consecutive segments of a time series. When a time series displays a high level of approximate entropy, it signifies a rich diversity of patterns, pointing to increased complexity or unpredictability in the data. Importantly, approximate entropy is a statistic that is not influenced by the distribution of the data and is resilient to the effects of outliers. It provides a non-negative numerical value to the time series, with higher values indicating a higher degree of randomness or variability.

To calculate approximate entropy, the first step is to select a time sequence $x = \{x_1, \ldots, x_T\}$ of length T, then for any $k < T$ let

$$u_k(i) = \{x_i, x_{i+1}, \ldots, x_{i+(k-1)}\} \tag{B.1}$$

be a k-dimensional vector, starting at the i-the term in the sequence x.

Let $u_k(i)$ and $u_k(j)$ be two vectors. The distance between the two vectors $u_k(i)$ and $u_k(j)$ of length k is defined as

$$d\left(u_k(i), u_k(j)\right) = \max\left\{ |x_{i+h} - x_{j+h}| \,:\, 0 \le h \le k-1 \right\} . \tag{B.2}$$

Two vectors $u_k(i)$ and $u_k(j)$ are similar to each other, if for some $r > 0$,

$$d\left(u_k(i), u_k(j)\right) \le r \,, \tag{B.3}$$

where r denotes the specified tolerance, that clearly determines the maximum distance between the two vectors. In order to determine the number of similar vectors $u_k(j)$ for each of the $T-(k-1)$ vectors $u_k(i)$, the respective distances between the vectors must be measured.

Let $n_k(i)$ denote the number of vectors $u_k(j)$ similar to $u_k(i)$ for a fixed i and k, where

$$1 \le n_k(i) \le T - (k-1) \,.$$

The relative frequency of finding a vector $u_k(j)$ in the sequence x that is similar to $u_k(i)$, with tolerance level r, is defined as

$$C_k\left(u_k(i)|x, r\right) := \frac{n_k(i)}{T-(k-1)} \tag{B.4}$$

and the average frequency of the natural logarithm of eq. (B.4) is defined as

$$\psi_k(r) := \frac{1}{T-(k-1)} \sum_{i=1}^{T-(k-1)} \ln\Big(C_k\left(u_k(i)|x, r\right) \Big) \,, \tag{B.5}$$

For a finite sequence $x = \{x_1, \ldots, x_T\}$ of length T, the approximate entropy is estimated using the following statistic:

$$ApEn_k(T, r) := \psi_k(r) - \psi_{k+1}(r) \,, \tag{B.6}$$

where k denotes the length of the vectors $u_k(i)$ and r is the specified tolerance level. Following the work of Gradojevic and Caric (2017), we

Table A.12
First three columns of the correlation matrix between our *sector-level* regressors and the *sector-uncertainty* regressors defined in Section 3.3. The first three lines feature the sector-level regressors $S_{i,m}$. Lines (4) to (15) feature the sector-uncertainty regressors based on the volatility (12-month rolling standard deviation) of the CDS quotes in Table 1. The last four lines feature the sector-uncertainty regressors $AES_{i,m}$ based on the approximate entropy. All values are expressed as percentages. The variables $S_{i,m}$ and $AES_{i,m}$ are defined in eq. (5) and eq. (6), respectively, and the variables $\mathrm{Vol}^x_{i,m}$ are introduced in Section 3.3. .

	(1)	(2)	(3)
(1) $S_{i,1}$	100		
(2) $S_{i,5}$	91.01	100	
(3) $S_{i,10}$	87.07	98.88	100
(4) $\mathrm{Vol}^a_{i,1}$	87.38	81.61	76.25
(5) $\mathrm{Vol}^a_{i,5}$	72.71	83.18	79.76
(6) $\mathrm{Vol}^a_{i,10}$	67.90	80.41	77.94
(7) $\mathrm{Vol}^b_{i,1}$	68.59	63.36	58.59
(8) $\mathrm{Vol}^b_{i,5}$	61.62	72.25	67.49
(9) $\mathrm{Vol}^b_{i,10}$	62.62	76.40	72.63
(10) $\mathrm{Vol}^c_{i,1}$	42.87	45.75	43.87
(11) $\mathrm{Vol}^c_{i,5}$	48.80	58.68	52.58
(12) $\mathrm{Vol}^c_{i,10}$	51.51	64.06	57.44
(13) $\mathrm{Vol}^d_{i,1}$	63.39	59.94	56.78
(14) $\mathrm{Vol}^d_{i,5}$	54.67	64.94	58.74
(15) $\mathrm{Vol}^d_{i,10}$	54.75	67.82	61.57
(16) $AES_{i,1}$	25.59	11.40	13.50
(17) $AES_{i,5}$	15.46	18.24	19.43
(18) $AES_{i,10}$	31.34	27.93	30.61

set $k = 2$ and $r = 0.2 \times \sigma_x$, where σ_x represents the standard deviation of the time series x. In our analysis, T denotes the length of our rolling window of 12 months.

Appendix C. Regression models with various CDS tenors

See Table C.13.

Appendix D. Additional regressions

D.1. Predictors of expected recovery rates

See Table D.14.

D.2. Predictors of dispersion of the recovery rates

See Table D.15.

Appendix E. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbankfin.2025.107414.

This section contains the Online Appendix.

Data availability

Data will be made available on request.

Table C.13
Classical beta regression models for the mean of the recovery rates. All models include consolidated categorical variables for controlling for bonds' seniority, default type and industrial sector of the bond issuer as depicted in Table 5. The variables $S_{i,m}$ and $AES_{i,m}$ are defined in eq. (5) and eq. (6), respectively.

	Dependent variable:						
	Recovery rate distributions - Mean model						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	−1.1541***	−1.1214***	−1.1197***	−1.2921***	−1.1245***	−1.1422***	−1.2708***
	(0.1106)	(0.1099)	(0.1098)	(0.1135)	(0.1107)	(0.1113)	(0.1132)
Coupon Rate	−0.1213**	−0.1232**	−0.1248**	−0.1051**	−0.0948*	−0.0985*	−0.1252**
	(0.0541)	(0.0541)	(0.0541)	(0.0532)	(0.0539)	(0.0539)	(0.0536)
Maturity	−0.0270	−0.0250	−0.0244	−0.0442	−0.0468	−0.0338	−0.0291
	(0.0486)	(0.0484)	(0.0483)	(0.0478)	(0.0482)	(0.0486)	(0.0481)
Am. Def. Rate	−0.1419**	−0.1141*	−0.1121*	−0.1682***	−0.2279***	−0.2134***	−0.1002
	(0.0633)	(0.0657)	(0.0646)	(0.0586)	(0.0587)	(0.0588)	(0.0642)
INDPRO *Ret.*	0.0439	0.0352	0.0385	0.1023	0.0361	0.0521	0.0977
	(0.0901)	(0.0900)	(0.0899)	(0.0904)	(0.0906)	(0.0907)	(0.0902)
SP500 *Ret.*	0.2626***	0.2717***	0.2785***	0.2253***	0.2215***	0.2224***	0.2624***
	(0.0607)	(0.0606)	(0.0606)	(0.0591)	(0.0598)	(0.0598)	(0.0601)
Fed. Funds Rate	0.1726***	0.1311***	0.1187***	0.2139***	0.1624***	0.1666***	0.1793***
	(0.0436)	(0.0443)	(0.0449)	(0.0442)	(0.0438)	(0.0437)	(0.0463)

(*continued on next page*)

Table C.13 (*continued*).

	Dependent variable:						
	Recovery rate distributions - Mean model						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$S_{i,1}$	−0.1706*** (0.0513)						
$S_{i,5}$		−0.2114*** (0.0579)					
$S_{i,10}$			−0.2271*** (0.0566)				−0.1450** (0.0585)
$AES_{i,1}$				−0.2534*** (0.0478)			−0.2198*** (0.0495)
$AES_{i,5}$					−0.0211 (0.0410)		
$AES_{i,10}$						−0.0817* (0.0421)	
Observations	613	613	613	613	613	613	613
Debt seniority dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry sector dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Default event dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.3206	0.3218	0.3239	0.3446	0.3096	0.3146	0.3494
AIC	−615.87	−617.61	−619.8	−634.08	−605.32	−608.97	−637.65
BIC	−536.34	−538.08	−540.27	−554.55	−525.79	−529.44	−553.71
Log Likelihood	325.9347	326.8053	327.8993	335.0417	320.6612	322.4838	337.827

Note: *p<0.1; **p<0.05; ***p<0.01.

Table D.14
Classical beta regression models for the mean of the recovery rates. All models include categorical variables for controlling for bonds' seniority, default type and industrial sector of the bond issuer as depicted in Table 5.

	Dependent variable:				
	RR distributions - Mean model				
	(1)	(2)	(3)	(4)	(5)
Constant	−1.37574*** (0.07887)	−1.35879*** (0.09329)	−1.20098*** (0.09827)	−1.28176*** (0.09912)	−1.18262*** (0.09938)
Coupon Rate		−0.03972 (0.05139)	−0.09033* (0.05223)	−0.06104 (0.05288)	−0.07255 (0.05157)
Maturity		−0.06500 (0.04823)	−0.02651 (0.04856)	−0.05685 (0.04847)	−0.03908 (0.04803)
Fin. Unc.			−0.26156*** (0.05733)		
News-based EPU				−0.10894** (0.05423)	
VIX					−0.26014*** (0.05631)
Observations	613	613	613	613	613
Debt seniority dummies	Yes	Yes	Yes	Yes	Yes
Industry sector dummies	Yes	Yes	Yes	Yes	Yes
Default event dummies	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.2578	0.2598	0.2809	0.2667	0.2823
AIC	−567.97	−566.24	−583.87	−568.02	−583.9
BIC	−519.37	−508.8	−522.01	−506.16	−522.04
Log Likelihood	294.9853	296.1208	305.9329	298.0105	305.9505

Note: *p<0.1; **p<0.05; ***p<0.01.

Table D.15
Variable beta regression models for the mean and precision of the recovery rates. Models feature a logit link for the mean submodel and log-link for the precision submodel. Variables have been selected with forward step-GAIC. The starting model for the variable selection procedure generating $\mathcal{M}_8^{\mu,\phi^*}$ is model $\mathcal{M}_1^{\mu}$ presented in Table 7. All models include consolidated categorical variables for controlling for bonds' seniority, default type and industrial sector of the bond issuer as depicted in Table 5. The variables $S_{i,m}$ and $AES_{i,m}$ are defined in eq. (5) and eq. (6), respectively.

	Dependent variable:	
	RR distributions (Mean submodel) $\mathcal{M}_8^{\mu,\phi^*}$	RR distributions (Precision submodel) $\mathcal{M}_8^{\mu,\phi^*}$
Constant	−1.5654*** (0.1092)	1.8694*** (0.1372)
Coupon Rate	−0.0984* (0.0541)	−0.0457 (0.0698)
Am. Def. Rate	−0.2378*** (0.0569)	0.3607*** (0.0772)
SP500 *Ret.*	0.3518*** (0.0557)	−0.3522*** (0.0726)
Fed. Funds Rate	0.1387*** (0.0470)	
$S_{i,10}$	−0.1174** (0.0569)	−0.0043 (0.0785)
$AES_{i,1}$	−0.1366** (0.0541)	−0.1651*** (0.0632)
Observations	613	
Debt seniority dummies	Yes	
Industry sector dummies	Yes	
Default event dummies	Yes	
Pseudo R-squared	0.3265	
AIC	−761.23	
BIC	−624.26	
Log Likelihood	411.6161	

Note: *p<0.1; **p<0.05; ***p<0.01.

References

Acharya, V.V., Bharath, S.T., Srinivasan, A., 2007. Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. J. Financ. Econ. 85 (3), 787–821. http://dx.doi.org/10.1016/j.jfineco.2006.05.011.

Altman, E.I., Brady, B., Resti, A., Sironi, A., 2005. The link between default and recovery rates: Theory, empirical evidence, and implications. J. Bus. 78 (6), 2203–2228. http://dx.doi.org/10.1086/497044.

Altman, E.I., Kalotay, E.A., 2014. Ultimate recovery mixtures. J. Bank. Financ. 40, 116–129. http://dx.doi.org/10.1016/j.jbankfin.2013.11.021.

Altman, E.I., Karlin, B., 2009. The re-emergence of distressed exchanges in corporate restructurings. J. Credit. Risk 5 (2), 43–55. http://dx.doi.org/10.21314/JCR.2009.087.

Altman, E.I., Kishore, V.M., 1996. Almost everything you wanted to know about recoveries on defaulted bonds. Financ. Anal. J. 52 (6), 57–64. http://dx.doi.org/10.2469/faj.v52.n6.2040.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. Q. J. Econ. 131 (4), 1593–1636. http://dx.doi.org/10.1093/qje/qjw024.

BCBS, 2023a. Basel Framework - Calculation of RWA for Credit Risk - IRB Approach: Minimum Requirements to Use IRB approach. Technical Report December, Bank for International Settlements, Basel, Switzerland, URL https://www.bis.org/basel_framework/chapter/CRE/36.htm.

BCBS, 2023b. Basel Framework - Calculation of RWA for Credit Risk - IRB Approach: Risk Components. Technical Report December, Bank for International Settlements, Basel, Switzerland, URL https://www.bis.org/basel_framework/chapter/CRE/32.htm?inforce=20230101&published=20200327.

Bellalah, M., Zouari, S., Levyne, O., 2016. The performance of hybrid models in the assessment of default risk. Econ. Model. 52, 259–265. http://dx.doi.org/10.1016/j.econmod.2014.10.051.

Berd, A., 2005. Recovery swaps. J. Credit. Risk 1 (3), 61–70. http://dx.doi.org/10.21314/JCR.2005.020.

Boudreault, M., Gauthier, G., Thomassin, T., 2013. Recovery rate risk and credit spreads in a hybrid credit risk model. J. Credit. Risk 9 (3), 3–39. http://dx.doi.org/10.21314/JCR.2013.164.

Boudreault, M., Gauthier, G., Thomassin, T., 2014. Contagion effect on bond portfolio risk measures in a hybrid credit risk model. Financ. Res. Lett. 11 (2), 131–139. http://dx.doi.org/10.1016/j.frl.2013.07.005.

Bris, A., Welch, I., Zhu, N., 2006. The costs of bankruptcy: Chapter 7 liquidation versus Chapter 11 reorganization. J. Financ. 61 (3), 1253–1303. http://dx.doi.org/10.1111/j.1540-6261.2006.00872.x.

Bruche, M., González-Aguado, C., 2010. Recovery rates, default probabilities, and the credit cycle. J. Bank. Financ. 34 (4), 754–764. http://dx.doi.org/10.1016/j.jbankfin.2009.04.009.

Cagetti, M., Hansen, L.P., Sargent, T., Williams, N., 2002. Robustness and pricing with uncertain growth. Rev. Financ. Stud. 15 (2), 363–404, URL https://www.jstor.org/stable/2696781.

Cantor, R., Emery, K., Keisman, D., Ou, S., 2007. Moody's Ultimate Recovery Database. Technical Report April, Moody's Investors Service, Special Comment, New York, United States of America, URL https://www.moodys.com/sites/products/defaultresearch/2006600000428092.pdf.

Caselli, S., Gatti, S., Querci, F., 2008. The sensitivity of the loss given default rate to systematic risk: new empirical evidence on bank loans. J. Financ. Serv. Res. 34 (1), 1–34. http://dx.doi.org/10.1007/s10693-008-0033-8.

Chava, S., Stefanescu, C., Turnbull, S., 2011. Modeling the loss distribution. Manag. Sci. 57 (7), 1267–1287. http://dx.doi.org/10.1287/mnsc.1110.1345.

Cover, T., Thomas, J., 2006. Elements of Information Theory. Wiley, http://dx.doi.org/10.1002/047174882X.

Das, S.R., Hanouna, P., 2009. Implied recovery. J. Econom. Dynam. Control 33 (11), 1837–1857. http://dx.doi.org/10.1016/j.jedc.2009.04.005.

Distaso, W., Roccazzella, F., Vrins, F., 2023. Business Cycle and Realized Losses in the Consumer Credit Industry. Discussion Paper, LIDAM/LFIN 2023/07, UCLouvain, Louvain-la-Neuve, Belgium, URL http://hdl.handle.net/2078.1/281562.

Doshi, H., Elkamhi, R., Ornthanalai, C., 2018. The term structure of expected recovery rates. J. Financ. Quant. Anal. 53 (6), 2619–2661. http://dx.doi.org/10.1017/S0022109018000558.

Duffie, D., Singleton, K.J., 1999. Modeling term structures of defaultable bonds. Rev. Financ. Stud. 12 (4), 687–720, https://www.jstor.org/stable/2645962.

Elkamhi, R., Jacobs, K., Pan, X., 2014. The cross section of recovery rates and default probabilities implied by credit default swap spreads. J. Financ. Quant. Anal. 49 (1), 193–220. http://dx.doi.org/10.1017/S0022109014000088.

Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31 (7), 799–815. http://dx.doi.org/10.1080/0266476042000214501.

Fisch, C., Momtaz, P.P., 2020. Institutional investors and post-ICO performance: an empirical analysis of investor returns in initial coin offerings (ICOs). J. Corp. Financ. 64, 101679. http://dx.doi.org/10.1016/j.jcorpfin.2020.101679.

Fox, J., Monette, G., 1992. Generalized collinearity diagnostics. J. Amer. Statist. Assoc. 87 (417), 178–183. http://dx.doi.org/10.2307/2290467.

François, P., 2019. The determinants of market-implied recovery rates. Risks 7 (2), 57. http://dx.doi.org/10.3390/risks7020057.

François, P., Jiang, W., 2019. Credit value adjustment with market-implied recovery. J. Financ. Serv. Res. 56, 145–166. http://dx.doi.org/10.1007/s10693-018-0298-5.

Franks, J.R., Torous, W.N., 1994. A comparison of financial recontracting in distressed exchanges and Chapter 11 reorganizations. J. Financ. Econ. 35 (3), 349–370. http://dx.doi.org/10.1016/0304-405X(94)90037-X.

Frittelli, M., 2000. The minimal entropy martingale measure and the valuation problem in incomplete markets. Math. Finance 10, 39–52. http://dx.doi.org/10.1007/s10479-019-03364-2.

Gambetti, P., Gauthier, G., Vrins, F., 2019. Recovery rates: Uncertainty certainly matters. J. Bank. Financ. 106, 371–383. http://dx.doi.org/10.1016/j.jbankfin.2019.07.010.

Gambetti, P., Roccazzella, F., Vrins, F., 2022. Meta-learning approaches for recovery rate prediction. Risks 10 (6), 124. http://dx.doi.org/10.3390/risks10060124.

Gençay, R., Gradojevic, N., 2017. The tale of two financial crises: An entropic perspective. Entropy 19 (6), 244. http://dx.doi.org/10.3390/e19060244.

Gradojevic, N., Caric, M., 2017. Predicting systemic risk with entropic indicators. J. Forecast. 36 (1), 16–25. http://dx.doi.org/10.1002/for.2411.

Gregory, J., 2012. Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets. John Wiley & Sons, http://dx.doi.org/10.1002/9781118673638.

Groll, A., Hambuckers, J., Kneib, T., Umlauf, N., 2019. LASSO-type penalization in the framework of generalized additive models for location, scale and shape. Comput. Statist. Data Anal. 140, 59–73. http://dx.doi.org/10.1016/j.csda.2019.06.005.

Hansen, L.P., Sargent, T.J., Tallarini, Jr., T.D., 1999. Robust permanent income and pricing. Rev. Econ. Stud. 66 (4), 873–907, URL https://www.jstor.org/stable/2566924.

Heinze, G., Wallisch, C., Dunkler, D., 2018. Variable selection–a review and recommendations for the practicing statistician. Biom. J. 60 (3), 431–449. http://dx.doi.org/10.1002/bimj.201700067.

Hyvarinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley, http://dx.doi.org/10.1002/0471221317.

Jankowitsch, R., Nagler, F., Subrahmanyam, M.G., 2014. The determinants of recovery rates in the US corporate bond market. J. Financ. Econ. 114 (1), 155–177. http://dx.doi.org/10.1016/j.jfineco.2014.06.001.

Jaskowski, M., McAleer, M., 2013. Estimating Implied Recovery Rates from the Term Structure of CDS Spreads. Discussion Paper 2013-005/III, Tinbergen Institute, pp. 1–33, URL https://tinbergen.nl/discussion-paper/3827/13-005-iii-estimating-implied-recovery-rates-from-the-term-structure-of-cds-spreads.

Jurado, K., Ludvigson, S.C., Ng, S., 2015. Measuring uncertainty. Am. Econ. Rev. 105 (3), 1177–1216. http://dx.doi.org/10.1257/aer.20131193.

Lassance, N., DeMiguel, V., Vrins, F., 2021. Optimal portfolio diversification via independent component analysis. Oper. Res. 70 (1), 55–72. http://dx.doi.org/10.1287/opre.2021.2140.

Lassance, N., Vrins, F., 2021a. Minimum Rényi entropy portfolios. Ann. Oper. Res. 299, 23–49. http://dx.doi.org/10.1007/s10479-019-03364-2.

Lassance, N., Vrins, F., 2021b. Portfolio selection with parsimonious higher comoments estimation. J. Bank. Financ. 126, 106115. http://dx.doi.org/10.1016/j.jbankfin.2021.106115.

Li, A., Li, Z., Bellotti, A., 2023. Predicting loss given default of unsecured consumer loans with time-varying survival scores. Pacific- Basin Financ. J. 78, 101949. http://dx.doi.org/10.1016/j.pacfin.2023.101949.

Ludvigson, S.C., Ma, S., Ng, S., 2021. Uncertainty and business cycles: exogenous impulse or endogenous response? Am. Econ. J.: Macroecon. 13 (4), 369–410. http://dx.doi.org/10.1257/mac.20190171.

Mantel, N., 1970. Why stepdown procedures in variable selection. Technometrics 12 (3), 621–625. http://dx.doi.org/10.1080/00401706.1970.10488701.

Metz, A., Sorensen, S., Keisman, D., Chiu, D., 2012. Trading Prices as Predictors of Ultimate Corporate Recovery Rates. Technical Report March, Moody's Investors Service, New York, United States of America, URL https://www.moodys.com/research/doc–PBC_139896.

Mistrulli, P., 2011. Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns. J. Bank. Financ. 35 (5), 1114–1127. http://dx.doi.org/10.1016/j.jbankfin.2010.09.018.

Moody's, 2023a. Default and Recovery Database (DRD) Technical Specifications. Technical Report January, Moody's Analytics, New York, United States of America, URL https://www.moodys.com/sites/products/ProductAttachments/DRD%20Documentation%20v2/DRDV2_Technical_Specifications.pdf.

Moody's, 2023b. Default and Recovery Database (DRD) User Guide. Technical Report February, Moody's Analytics, New York, United States of America, URL https://www.moodys.com/sites/products/ProductAttachments/DRD%20Documentation%20v2/DRDV2_User_Guide.pdf.

Mora, N., 2012. What determines creditor recovery rates? Fed. Reserv. Bank Kans. City Econ. Rev. 79 (2), 109–123, URL https://www.kansascityfed.org/documents/1381/2012-What%20Determines%20Creditor%20Recovery%20Rates%3F.pdf.

Mora, N., 2015. Creditor recovery: The macroeconomic dependence of industry equilibrium. J. Financ. Stab. 18, 172–186. http://dx.doi.org/10.1016/j.jfs.2015.04.004.

Nazemi, A., Fabozzi, F.J., 2018. Macroeconomic variable selection for creditor recovery rates. J. Bank. Financ. 89, 14–25. http://dx.doi.org/10.1016/j.jbankfin.2018.01.006.

Nazemi, A., Heidenreich, K., Fabozzi, F.J., 2018. Improving corporate bond recovery rate prediction using multi-factor support vector regressions. European J. Oper. Res. 271 (2), 664–675. http://dx.doi.org/10.1016/j.ejor.2018.05.024.

Nazemi, A., Pour, F.F., Heidenreich, K., Fabozzi, F.J., 2017. Fuzzy decision fusion approach for loss-given-default modeling. European J. Oper. Res. 262 (2), 780–791. http://dx.doi.org/10.1016/j.ejor.2017.04.008.

Ospina, R., Ferrari, S.L., 2012. A general class of zero-or-one inflated beta regression models. Comput. Statist. Data Anal. 56 (6), 1609–1623. http://dx.doi.org/10.1016/j.csda.2011.10.005.

Pincus, S.M., 1991. Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. 88 (6), 2297–2301. http://dx.doi.org/10.1073/pnas.88.6.2297.

Pykhtin, M., 2003. Unexpected recovery risk. Risk Mag. 16 (8), 74–78, https://www.risk.net/risk-management/credit-risk/1530263/unexpected-recovery-risk.

Qi, M., Zhao, X., 2011. Comparison of modeling methods for loss given default. J. Bank. Financ. 35 (11), 2842–2855. http://dx.doi.org/10.1016/j.jbankfin.2011.03.011.

Schuermann, T., 2004. What Do We Know About Loss Given Default?. Working Paper, WhartonFinancial Institutions Center No. 04-01, The Wharton School, Philadelphia, United States of America, pp. 1–53. http://dx.doi.org/10.2139/ssrn.525702.

Shleifer, A., Vishny, R.W., 1992. Liquidation values and debt capacity: A market equilibrium approach. J. Financ. 47 (4), 1343–1366. http://dx.doi.org/10.1111/j.1540-6261.1992.tb04661.x.

Simas, A.B., Barreto-Souza, W., Rocha, A.V., 2010. Improved estimators for a general class of beta regression models. Comput. Statist. Data Anal. 54 (2), 348–366. http://dx.doi.org/10.1016/j.csda.2009.08.017.

Smithson, M., Verkuilen, J., 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol. Methods 11 (1), 54. http://dx.doi.org/10.1037/1082-989X.11.1.54.

Stasinopoulos, D.M., Rigby, R.A., 2008. Generalized additive models for location scale and shape (GAMLSS) in R. J. Stat. Softw. 23 (7), 1–46. http://dx.doi.org/10.18637/jss.v023.i07.

Stutzer, M., 1996. A simple nonparametric approach to derivative security valuation. J. Financ. 51 (5), 1633–1652. http://dx.doi.org/10.2307/2329532.

Tajvarpour, M.H., Pujari, D., 2022. The influence of narrative description on the success of crowdfunding campaigns: The moderating role of quality signals. J. Bus. Res. 149, 123–138. http://dx.doi.org/10.1016/j.jbusres.2022.05.012.

Tobback, E., Martens, D., Van Gestel, T., Baesens, B., 2014. Forecasting loss given default models: impact of account characteristics and the macroeconomic state. J. Oper. Res. Soc. 65, 376–392. http://dx.doi.org/10.1057/jors.2013.158.

Tsai, M.-S., Chen, L.-C., 2011. The calculation of capital requirement using Extreme Value Theory. Econ. Model. 28 (1–2), 390–395. http://dx.doi.org/10.1016/j.econmod.2010.08.010.

Unal, H., Madan, D., Güntay, L., 2003. Pricing the risk of recovery in default with absolute priority rule violation. J. Bank. Financ. 27 (6), 1001–1025. http://dx.doi.org/10.1016/S0378-4266(02)00255-8.

Van de Castle, K., Keisman, D., Yang, R., 2000. Suddenly Structure Mattered: Insights into Recoveries of Defaulted Debt. Technical Report May, Standard & Poor's Credit Week, New York, United States of America, pp. 62–68, URL http://www.financerisks.com/filedati/WP/ALGO_PAPER/ch2_suddenly.pdf.

Varma, P., Cantor, R., 2005. Determinants of recovery rates on defaulted bonds and loans for North American corporate issuers: 1983–2003. J. Fixed Income 14 (4), 29–44. http://dx.doi.org/10.3905/jfi.2005.491110.

Yao, X., Crook, J., Andreeva, G., 2015. Support vector regression for loss given default modelling. European J. Oper. Res. 240 (2), 528–538. http://dx.doi.org/10.1016/j.ejor.2014.06.043.