MOODY'S ANALYTICS

RESEARCH FRAMEWORK
MARCH 2018

The Deterioration Probability (DP) Model: Methodology and Validation

Authors

Samuel W. Malone, Ph.D.
+1.212.553.2107
samuel.malone@moodys.com

Irina Baron
+1.212.553.4307
irina.baron@moodys.com

Reginald White
+1.212.553.4184
reginald.white@moodys.com

Contact Us

Americas
+1.212.553.1658
clientservices@moodys.com

Europe
+44.20.7772.5454
clientservices.emea@moodys.com

Asia (Excluding Japan)
+852.3551.3077
clientservices.asia@moodys.com

Japan
+81.3.5408.4100
clientservices.japan@moodys.com

Summary

Ratings downgrades can have material impacts in a variety of buy-side and sell-side client applications. In response to the need for improved early warning of downgrade events, we have developed and validated a transparent, quantitative model for the 1-year probability of downgrade. We call the new measure the Deterioration Probability (DP). This document discusses the model methodology, validation results, and case studies for the new DP metric. The DP provides a model-driven estimate of the probability of downgrade events for rated firms, and of shadow-downgrade events for unrated firms, which is updated daily and has global coverage for the vast majority of firms with publicly traded equity. The principal drivers of the model are drawn from the Moody's Analytics CreditEdge™ database and from historical information on ratings outlooks, ratings histories, and Moody's Analytics' suite of Market Implied Rating (MIR) models.

Table of Contents

1 Introduction

Unanticipated downgrades can have a material impact in a variety of risk management contexts for both the sell-side and the buy-side. Downgrades of borrower and counterparty ratings often induce financial institutions to hold more required reserves and regulatory capital. For fixed income portfolio managers, the impact of unanticipated downgrades on the prices of bonds is also an important source of market risk, which can often be more salient than default risk in portfolios consisting primarily of investment grade credits. In response to the need for improved early warning of downgrade events, therefore, we have developed and validated a transparent, quantitative model for the probability of downgrade, which we call the Deterioration Probability (DP). The DP output ranges from 1 to 70 percentage points.

The model's coverage extends to unrated public firms. For such entities, the DP output should be interpreted as the probability of a "shadow" ratings downgrade, or an estimate of what the probability of a downgrade event might be if the firm were rated by Moody's Investors Service (MIS). We estimate these probabilities exclusively based on the patterns of downgrades observed in data for rated public firms for which a common set of model drivers is available. The DP can therefore be used to rank both rated and unrated firms by their risk of a downgrade-like credit event, based on the statistical patterns observed for rated public firms with similar characteristics.

The new DP model outputs are updated daily and have global coverage for the vast majority of firms with publicly traded equity. The DP is an "unconditional" downgrade event probability metric, analogous to the Expected Default Frequency™ (EDF) measure of default probability in Moody's Analytics CreditEdge database, in that it is not conditioned explicitly on macroeconomic drivers. This stands in contrast to models often used in stress testing applications, which usually must be conditioned on the paths of independent macroeconomic variables.

The drivers used to build the DP are drawn from the CreditEdge Early Warning Toolkit (EWTK, available to clients as an Excel Add-in) as well as from Market Implied Ratings, and data on ratings outlooks and history. The EWTK contains several valuable metrics for assessing credit risk, some of which also prove useful in forecasting downgrades. The EWTK metrics used as drivers of the DP model, which are discussed in detail in Section 3, include: EDF measures, their trigger levels, the slope of the EDF term structure, a firm's industry median EDF growth rate, and the relative EDF level (defined as the firm's EDF divided by the firm's industry median EDF measure). We review the construction of these variables, which are also documented in detail in Ferry and Baron (2016), in Section 3 below.

Among our validation results, a few are worth highlighting here.

» The DP's accuracy ratio[1] in forecasting downgrades is significantly higher than the accuracy ratios obtained for predictive models based on Market-Implied Ratings (MIRs) alone.

» The DP doubles on average during the two years prior to a downgrade event, and tends to peak two months before the downgrade occurs.

» Accuracy ratios for the model typically exceed 50% for both corporate and financial firms, with results that are consistent across most of the spectrum of initial issuer ratings.

» The model exhibits robust out-of-sample performance during the recent bear market in oil prices.

[1] The accuracy ratio (AR), which ranges from 0 to 100%, is a commonly used measure of model performance in binary event prediction. The AR can be calculated as 2*AUC-1, where AUC is the area under the receiver-operating characteristic (ROC) curve that plots the true positive prediction rate against the false positive error rate for a given model. An AR value of 0 corresponds to a model that is no more accurate than random guessing, whereas an AR of 100% corresponds to a model that is perfectly able to differentiate negative (non-default) from positive (default) outcomes of the dependent variable.

» We present three case studies to illustrate how the DP can provide a timely and effective early warning tool for downgrades in the case of a US and two international firms, respectively.

The rest of this paper is organized as follows. Section 2 discusses the analytical framework for forecasting downgrade events and presents a few stylized facts. Section 3 discusses the data and variable definitions, and Section 4 documents the model structure and its validation results. Section 5 concludes.

2 Modeling Downgrade Risk: Analytical Tools and Stylized Facts

To model downgrade risk, we draw upon two sources of information. The first is the CreditEdge™ database, which contains Expected Default Frequency™ (EDF) metrics, their drivers and term structures, and other related variables for over 44,000 publicly traded corporate and financial firms around the world. Moody's Analytics' EDFs are scaled metrics of the probability of default over a given horizon, meaning that their "absolute value has an objective interpretation" (Nazeran and Dwyer, 2015). Based on the premise that default events for rated companies are often preceded by a string of downgrades, an accurate default probability measure such as the EDF is a natural driver to consider when building a downgrade probability model. Holding other factors equal, we would expect higher EDFs to be associated with higher Deterioration Probabilities.

Building upon the CreditEdge data, the Early Warning Toolkit, first documented in Ferry and Baron (2016), recommends tracking five EDF-related metrics, in addition to EDFs themselves, associated with elevated future default risk. We show how these metrics — which include the EDF trigger level, the EDF change, relative EDF level, relative EDF change, and the slope of the term structure — shed additional light on impending credit distress, as measured by a ratings downgrade. We review the definitions of these variables in Section 3.

The second source of information we draw upon to predict downgrades is the history of ratings and rating outlooks. We find evidence of ratings momentum: downgrades become more likely following a downgrade event in the previous twelve months. And we see of course that the current ratings outlook maintained on a company by Moody's Investor Service aids in downgrade prediction. These results are consistent with previous findings by Hamilton and Cantor (2004).

Finally we draw upon lessons learned from the structure of Moody's Analytics Credit Transition Model (CTM), which is documented in Metz and Cantor (2007) and Wang, Ding, Pan and Malone (2017). In particular, we see benefits from allowing the coefficients of model drivers to vary by rating.

In terms of functional form, we opt for the logit model specification. Thus, for a given vector of covariates X_{it} observed for firm *i* at time *t* and coefficient vector $\vec{\beta}$, the logistic function *F* defined as follows:

$$F(\vec{X}_{it}) = (1 + \exp(-\vec{\beta}'\vec{X}_{it}))^{-1}$$

The value of $F(\vec{X}_{it})$ lies on the interval (0, 1) by construction and is interpreted as the probability of a "positive" outcome, i.e., of a downgrade in the next 12 months. In particular, we define the outcome variable Y_{it} to be equal to 1 if a downgrade occurs for rated firm *i* during the interval $[t+1, \ldots, t+12]$, and equal to 0 otherwise. For unrated firms, we code the value of Y_{it} as missing.

As we demonstrate in Section 4, model accuracy is enhanced by estimating separate logit model coefficients on different, mutually exclusive subsets of the data. Thus, the predictions of our model use the appropriate sub-model predictions for each subset of the data. These subsets, which we denote with the subscript g, correspond to firms of different sectors (financial vs. corporate) and ratings. Formally, we can express the DP model output using the equation

$$DP(\vec{X}_{it}) = \sum_{g=1}^{G} I_g(i,t) F_g(\vec{X}_{it}),$$

where $F_g(\vec{X}_{it}) = (1 + \exp(-\vec{\beta}_g' \vec{X}_{it}))^{-1}$ is the output of the logit model estimated on subset g of the data, and $I_g(i,t)$ is an indicator function that is equal to one if observation *i* lies in subset g at time *t*, and zero otherwise. For some subsets *g*, we constrain specific model coefficients in the vector $\vec{\beta}$ to be equal to zero in the estimation.

For any time *t* in the period $[1, \ldots, T]$ and observation *i* in the set $[1, \ldots, N(t)]$, the indicator variables sum to one over the set of groups $g \in [1, \ldots, G]$:

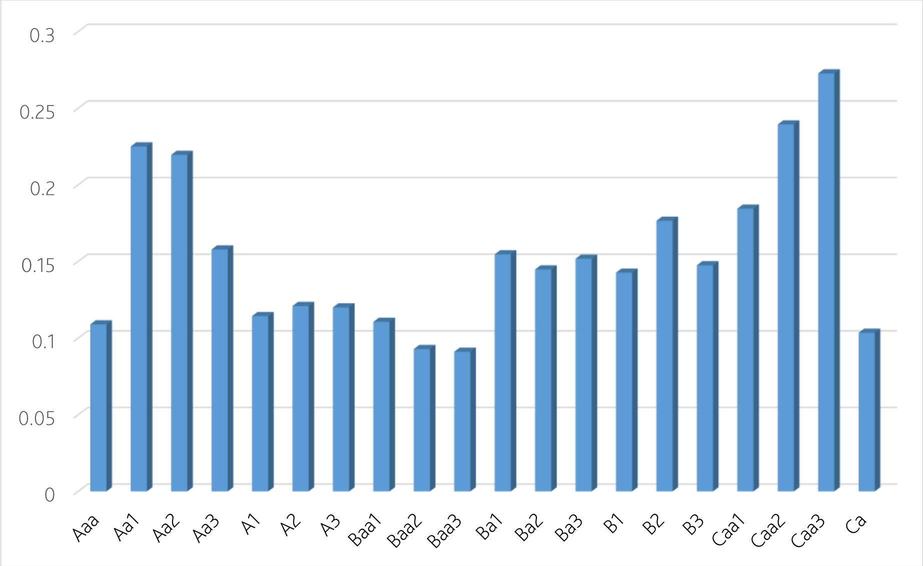
$$\sum_{g=1}^{G} I_g(i,t) = 1, \forall (i,t) \in [1, \ldots, N(t)] \times [1, \ldots, T]$$

Here $N(t)$ denotes the total number of observations at time *t*. The above equation is simply a formal way of stating that we always have a unique set of coefficients $\vec{\beta}_g$ that allow us to compute a value for the logit index $\vec{\beta}_g' \vec{X}_{it}$, which is then mapped to the (0,1) interval using the logit function *F* to compute the DP.

Within this general setup, the next step is specifying the groups. We describe our choices for groups in detail in section 4. The core idea, however, is that downgrade frequencies and their drivers differ meaningfully by the current rating and sector of the issuer.

As initial evidence to support the approach of estimating some model coefficients separately by rating, Figure 1 displays the empirical frequencies of downgrade events by rating for the set of Top 90 firms during the 2004-2016 period. (Top 90 firms, as described in the next section, refers to the set of firms we obtain after sorting firms by size and then selecting those whose liabilities comprise 90% of total liabilities in different sectors and geographies.)

Figure 1. Empirical frequencies of downgrade events by initial rating category



Empirical frequencies of downgrade, as Figure 1 makes clear, can vary substantially by initial rating. Downgrade frequencies are higher on average for companies with sub-investment grade ratings than for those with IG ratings, with a noticeable increase in the frequency of downgrades as we move from the Baa3 rating to the Ba1 rating in particular. This observation is consistent with the notion that membership in the HY universe is more of a "slippery slope" in terms of overall downgrade risk than inclusion in the IG universe. Overall, the downgrade frequency distribution is best described as "bimodal with a jump upward in the average frequency level as we move from IG to HY". In particular, the IG universe exhibits a peak in downgrade frequencies at the Aa1 rating category, whereas the HY universe exhibits a peak at the Caa3 rating category. The existence of different baseline downgrade frequencies by rating is one of the many empirical regularities we exploit in our model design. We now describe the dataset, including the drivers that conform the vector $\vec{X}_{it}$, used to estimate and validate the model.

3 The Dataset and Model Drivers

The primary data for this study are sourced from Moody's Analytics' CreditEdge dataset. For the following analyses, we constructed a sample of global firms with EDF measures from 2004 through 2016 for which we have the most reliable default data. It is reliable in the sense that there should be few "hidden" defaults — defaults that occurred, but that were neither reported nor observed — which could bias our findings. The sample is therefore limited to what we call "Top 90" firms.

The "Top 90" subset corresponds to relatively large firms by region. This subset was formed by sorting firms by size within geographical region each month. Proceeding down this list, we include firms until the point at which the sum total liabilities for included firms account for 90% of the total outstanding liabilities for firms in the region. Further, following the original description of the Top 90 selection procedure in Nazeran and Dwyer (2015), we measure size as the mean of book assets and sales. As of September 2016, the Top 90 set consisted of 4,756 firms, with a total of 10,120 firms appearing in our estimation dataset in at least one month.

Global sector-level peer groups fall into two types: the group of non-financial corporates and the group of financial firms. We will refer to these as the two "sector-level" groupings. The explanatory variables in our model include: the EDF, $TrEx$, $Slope$, $RelEDF$, $IndMedEDFGrowth$, a recent downgrade dummy variable labeled $RecentDowngrade$, and positive (POS), stable (STA), and negative (NEG) ratings outlook variables. We now describe each of these variables in detail.

In the DP model, sector-level triggers are used to formulate the dummy variable $TrEx$, which stands for trigger exceedance. The $TrEx$ variable is equal to 1 for a given firm at a given point in time if that firm's EDF exceeds its sector-level trigger at that time, and equal to 0 otherwise. As described in Ferry and Baron (2016), EDF trigger levels are optimized to minimize a forecasting loss function equal to the sum of type 1 (false positive) and type 2 (false negative) error rates in 12-month-ahead default event prediction. We expect EDF trigger exceedance to be associated with higher DPs.

Besides the EDF level, the EWTK variables we use to formulate downgrade prediction models include the $TrEx$ variable, the slope of the EDF term structure, the relative EDF level, and the industry median EDF growth rate. In the next section, we refer to these four latter variables collectively as the "EWTK measures" when discussing the results of alternative models.

As discussed in Ferry and Baron (2016), the EDF term structure, or $Slope$ variable is defined as the value of the 5-year EDF minus the 1-year EDF. Negative values of this variable indicate inverted EDF term structures, and we expect such situations to be associated with a higher risk of downgrade, holding other factors equal.

The relative EDF level variable, $RelEDF$, is equal to a firm's EDF divided by the median EDF of its industry at a given point in time. Country and industry peer groups, which are used to compute the Relative EDF level variable, are defined by country of incorporation and narrower industry category within that country[2]. For example, the industry-level peer group of Ford Motor Co. would be the US Automotive Group. Since we utilize peer group medians, it is necessary for a peer group to have a minimum number of constituents in order for its median to be meaningful. If a company's peer

[2] There are 61 industry classifications.

group has fewer than 10 constituents, we use instead its industry-level peer group defined by industry alone.[3] The country-industry peer group for British American Tobacco, P.L.C., for example, has only four constituents, so its industry peer group is the global tobacco group. We expect that firms with higher relative EDFs will be at higher risk for downgrade, other things equal.

Our fourth EWTK measure is the variable $IndMedEDFGrowth$, which is equal to the growth rate of the industry median EDF for each firm, to measure the impact on downgrade on credit deterioration in a firm's industry. We hypothesize that the coefficient on this variable will be positive in all models.

For additional model drivers, we utilize Moody's Market Implied Ratings platform (MIR), which provides credit risk and relative value signals from three sources: corporate bond markets, credit default swap (CDS) markets, and public firm EDFs. In all cases the input information (bond and CDS spreads and EDFs) is mapped to the Moody's rating scale. For the spread-driven MIRs, the mapping is based on windows around medians of entity-level credit spreads per rating category after adjusting for duration. For EDFs, it is based on median 1-year EDFs per rating category. In the MIR platform market-implied ratings are compared to the entities' senior unsecured or equivalent Moody's ratings.

This last step gives rise to the concept of positive and negative ratings gaps[4]. Take as an example an issuer with a Moody's rating of Baa2. Let's assume further that its 5-year CDS spread is in line with the median CDS spread for all A2 rated issuers, giving it a CDS-implied rating of A2. The issuer's A2 CDS-implied rating is three rating notches above its Baa2 Moody's rating. Thus, in the nomenclature of MIR, the issuer's CDS-implied ratings gap is +3. Similarly, if the issuer's CDS traded in line with the median credit spread for contracts of Ba2 rated issuers, its gap would be -3. The direction of the sign comes from our convention of calculating gaps in terms of "Moody's minus model", and the conversion of Moody's alphanumeric rating scale to a numerical ranking. Finally, if the company's CDS trades in line with the level suggested by its Moody's rating, then the ratings gap is zero. (All references to ratings and rating changes/events in this paper refer to ratings from Moody's Investors Service.)

To build a robust downgrade forecasting model, it is important to account for the fact that company coverage is very different for each of the three market-implied ratings variables. We do this by creating a single variable called the implied rating gap (IRG). The IRG is set equal to the ratings gap from the "best available" market-implied rating (MIR) model. The three MIR datasets have different levels of coverage. The determination of which available model is "best available" for the purposes of calculating the IRG is based on the ordering in-sample, by accuracy ratio, for the logit models that include each individual MIR gap as a driver. When producing the DP metric for unrated firms, we employ a separate model that excludes the IRG and other drivers described below that require the firm to have a rating. The IRG can take integer values in the range from -20 to 20, with negative ratings gaps associated with cases where the model rating is worse than the agency rating for that firm. We expect the coefficient of this variable to be negative and statistically significant in our model.

As final determinants of downgrade risk, we use information from the rating outlook and rating history. Specifically, we include three outlook dummy variables, corresponding to positive (POS), stable (STA), and negative (NEG) outlooks, respectively. We also construct a recent downgrade dummy variable, which is equal to 1 if the firm was downgraded at least once during the past 12 months, and 0 otherwise.

We expect more negative outlooks to be associated with higher downgrade risk, and for downgrade momentum to be present, in the sense that the coefficient of the recent downgrade dummy (labeled $RecentDowngrade$) should be positive and statistically significant. Putting the above variables together, we obtain the following vector of model drivers:

---

[3] This has little impact on companies in North America, Western Europe, and Southeast Asia.

[4] For details, please see the Viewpoints publication Kim and Munves, "Moody's Market Implied Ratings: Description, methodology, and Analytical Applications", July 2016.

$$\vec{X}'_{it} = (1, EDF_{it}, TrEx_{it}, Slope_{it}, RelEDF_{it}, IndMedEDFGrowth_{it}, TrEx_{it} * EDF_{it}, IRG_{it},$$
$$RecentDowngrade_{it}, STA_{it}, POS_{it}, NEG_{it})$$

Consistent with the length of the input vector, each coefficient vector $\vec{\beta}'_g$ has twelve parameters, although some of these parameters, as we mentioned before, are constrained to equal zero for certain groups g. Table 1 below shows which variables from the above subset are included in the various sub-models for corporates by rating category, and Table 2 displays the same information for financial firms.

Table 1. Downgrade Forecasting Models for Corporates*

Rating	EDF	TrEX	Slope	Relative EDF	Ind Med EDF Growth	TrEX x EDF	IR Gap	Recent Downgrade	Outlook
Aaa	✓	-	✓	✓	✓	-	✓	-	✓
Aa	✓	✓	✓	✓	✓	✓	✓	✓	✓
A	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baa	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ba	✓	✓	✓	✓	✓	✓	✓	✓	✓
B	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caa	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ca-C	✓	-	✓	✓	✓	-	✓	✓	✓
Unrated	✓	✓	✓	✓	✓	✓	-	-	-

* Note that the omission of TrEx and TrEx x EDF variables from the Aaa and Ca-C corporate models simply reflects the fact that these variables were dropped due to lack of in-sample variation. The EDFs of all Aaa rated firms are always below their trigger levels, and the EDFs of all Ca-C rated firms are always above their trigger levels, during the sample period.

Table 2. Downgrade Forecasting Models for Financials

<table>
<tr><th>Rating</th><th>EDF</th><th>TrEX</th><th>Slope</th><th>Relative EDF</th><th>Ind Med EDF Growth</th><th>TrEX x EDF</th><th>IR Gap</th><th>Recent Downgrade</th><th>Outlook</th></tr>
<tr><td>Aaa</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td><td>✓</td><td>-</td><td>-</td></tr>
<tr><td>Aa</td><td rowspan="5">✓</td><td rowspan="5">✓</td><td rowspan="5">✓</td><td>-</td><td rowspan="5">✓</td><td rowspan="5">✓</td><td rowspan="5">✓</td><td rowspan="5">✓</td><td rowspan="5">✓</td></tr>
<tr><td>A</td><td>-</td></tr>
<tr><td>Baa</td><td>-</td></tr>
<tr><td>Ba</td><td>-</td></tr>
<tr><td>B</td><td>-</td></tr>
<tr><td>Caa
Ca-C</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td><td>✓</td><td>-</td><td>-</td></tr>
<tr><td>Unrated</td><td>✓</td><td>✓</td><td>✓</td><td>-</td><td>✓</td><td>✓</td><td>-</td><td>-</td><td>-</td></tr>
</table>

A count of the total number of groups in Tables 1 and 2 for the set of rated firms yields $G = 11$. Note that unrated corporate and financial firms constitute two separate groups that do not form part of this total. Rather, a single model is estimated using data from all corporates of any rating and used to generate the DP for unrated corporates, and similarly for financials. These two model specifications are indicated by the last rows of Tables 1 and 2, respectively.

The availability of significantly more data for corporates gives us additional degrees of freedom with which to estimate model coefficients by rating category, whereas for financials, it is wise to be more parsimonious and estimate fewer models involving less granular rating-based groups. Since the $RelEDF$ variable was not statistically significant for financial firms, we opted to drop that variable from the models used to calculate the DP for financials as well.

Test for multicollinearity among explanatory variables

To test for potential collinearity among our independent variables, we calculated their correlation matrix, as well as the variance inflation factors (VIFs) for each variable. The results of these exercises are contained in Tables A1 and A2 of the Appendix, respectively. All VIFs are less than ten, which provides evidence that multicollinearity is unlikely to be a problem for the set of independent variables used in our logit models.

4 Empirical Results and Validation

In this section, we present the results of model estimation and validation. We follow this with short case studies of the Deterioration Probability for two defaulters, and close by presenting statistics on the distribution of downgrade magnitudes, with a view to helping clients gauge the expected size of downgrade risk for their portfolios.

4.1.1 MODEL COEFFICENTS AND FORECASTING PERFORMANCE

As discussed in Section 2, the functional form of the DP model is logit, and outputs are estimated probabilities of a downgrade in the next 12 months.

Table 3 presents model coefficients and standard errors for the DP model. The in-sample estimation period for the models in this paper, unless otherwise stated, is from January 2004 up to September 2016, and the data frequency is monthly.

From left to right in the Table, the seven model specifications in the table's columns are:

(1) Aaa rated corporates — including ratings-based independent variables

(2) non-Aaa corporates — including ratings-based independent variables

(3) all corporates — excluding ratings-based independent variables

(4) Aaa rated financial firms —including ratings-based independent variables

(5) non-Aaa and non-Caa-Ca-C financial firms — including ratings-based independent variables

(6) Caa-Ca-C rated financial firms — including ratings-based independent variables

(7) all financial firms — excluding ratings-based independent variables

A few comments are in order regarding the models presented in Table 3. First, as noted in the previous section, we found that Aaa-rated firms call for more parsimonious model specifications, both because certain drivers exhibit no variation for Aaa's (e.g., $TrEx$ and the recent downgrade dummy in the case of corporates are always zero), and because the number of observations for such firms is relatively low. These issues guide the model specifications in columns (1) and (5) for Aaa corporates and Aaa financials, respectively.

Second, in the case of financials, we found it necessary to remove Fannie Mae and Freddie Mac from our sample, due to the atypical behavior of the EDFs for these government-sponsored enterprises (GSEs). Thus, the results in column (4) for Aaa financials are based on the sample of financials ex-GSEs.

Third, we found that a univariate model of downgrade risk was best for Aaa and Caa-Ca-C-rated financial firms, and this reflects our choice in columns (4) and (6).

Finally, columns (3) and (7) document the coefficients of the models used to produce "shadow downgrade probabilities" for firms lacking a rating. In order to be fit for this purpose, the models in columns (3) and (7) exclude ratings-based independent variables[5]. The latter set of variables, which are not available for unrated firms, includes outlook dummies, the recent downgrade dummy, and the implied ratings gap. Only the EDF and EWTK-related drivers are used to estimate downgrade risk in these models, and coefficients in columns (3) and (7) reflect the results of estimating the models on the set of rated firms.

To anticipate a potential confusion surrounding the estimation and validation of DP estimates for unrated public firms, a comment is in order. Note that we must estimate DP models for unrated

[5] Another approach that suggests itself for producing DP estimates for unrated firms would be to use the full models estimated on rated firms, which include coefficients for the subset of drivers absent for unrated firms, and simply substitute in the latest cross-section sample average values of each of the missing drivers to produce forecasts each period. We considered this approach, but opted for the somewhat simpler approach described above.

firms on the same population of *rated* public firms used to estimate all of our models. There is no shortcut here. The DP estimate for an unrated firm simply reflects a reasonable, model-driven estimate of what that firm's probability of downgrade would be, if it were currently rated, based on the information contained in its current available EWTK drivers. Should the firm become rated, hence obtaining values for the IRG, outlook, and recent downgrade variables, then our model forecast for the DP will immediately begin using the appropriate, more accurate model that does include those drivers.

Table 3. Selected DP Logit Model Coefficients: Downgrade Models with and without Ratings-based Variables*

	Coefficients/*Standard Errors*						
Variable	Aaa Corp	Non-Aaa Corp	All Corp, EWTK	Aaa Fin	Non-Aaa Fin	Caa-Ca-C Fin	All Fin, EWTK
EDF	-2.447**	0.099***	0.327***	--	-0.056*	--	0.447***
	(1.059)	*(0.11)*	*(0.010)*	--	*(0.029)*	--	*(0.025)*
TrEx	--	0.69***	1.33***	--	0.588***	--	1.614***
	--	*(0.032)*	*(0.03)*	--	*(0.081)*	--	*(0.069)*
Slope	-3.76***	-0.063***	-0.015**	--	-0.463***	--	-0.509***
	(1.46)	-0.007	*(0.006)*	--	*(0.026)*	--	*(0.025)*
Relative EDF	6.45***	-0.000	-0.001**	--	--	--	--
	(1.22)	*(0.000)*	*(0.000)*	--	--	--	--
IndMedEDFGrowth	0.001	0.0008***	0.0008***	--	0.002***	--	0.002***
	(0.000)	*(0.000)*	*(0.000)*	--	*(0.000)*	--	*(0.000)*
TrEx x EDF	--	-0.25***	-0.316***	--	-0.239***	--	-0.731***
	--	*(0.011)*	*(0.01)*	--	*(0.032)*	--	*(0.029)*
IR Gap	-0.14***	-0.160***	--	-0.506***	-0.141***	-0.359***	--
	(0.05)	*(0.003)*	--	*(0.11)*	*(0.005)*	*(0.06)*	--
Recent Downgrade	--	0.17***	--	--	0.157***	--	--
	--	*(.02)*	--	--	*(0.043)*	--	--
STA	-1.41***	-1.23***	--	--	-1.01***	--	--
	(0.35)	*(0.02)*	--	--	*(0.034)*	--	--
POS	--	-1.82***	--	--	-1.645***	--	--
	--	*(0.05)*	--	--	*(0.144)*	--	--
NEG	0.41	-0.11***	--	--	0.299***	--	--
	(0.39)	*(0.02)*	--	--	*(0.041)*	--	--
Cons	-1.47***	-1.31***	-2.24***	-2.95***	-1.596***	-1.188***	-1.922***
	(0.33)	*(0.02)*	*(0.01)*	*(0.57)*	*(0.037)*	*(0.135)*	*(0.029)*
N	955	180,008	180,963	148	46,532	430	47,110
LR X^2	109.322***	17,597.61***	6,474.89***	30.73***	4,396.35***	50.38***	1,890.39***
Pseudo R^2	0.1741	0.1282	0.0470	0.1574	0.1234	0.1256	0.0521
AR	65.38%	51.48%	30.64%	51.68%	49.66%	49.04%	34.56%

*The values of coefficients in Table 3 are shown directly above their standard errors, which are included in parentheses. Significance levels for the coefficients are denoted by asterisks, with *** indicating significance at the 1% level, ** indicating significance at the 5% level, and * indicating significance at the 10% level. Most of the coefficients shown are statistically significant at the 1% level.

In the production version of the DP model, we estimate separate model coefficients by rating category for corporate firms. The relatively large size of the rated public corporate universe allows this, and it results in a modest improvement in forecasting accuracy both in- and out-of-sample. The coefficients shown in column (2) of Table 3 are from the pooled non-Aaa corporate data, but the general magnitude and sign patterns observed in that model specification hold across ratings when the coefficient vector is estimated by rating category. These model results for non-Aaa corporates by rating category, which are used in the production version of the model, can be found in Table A3 of the Appendix.

In the tests of model performance that we report in subsection 4.1.2, all performance results are based on the DP model output that uses corporate firm model coefficients estimated by rating category. Outside of column (2), the coefficients in Table 3 reflect those used in the DP model as implemented in production.

The variables used as predictors in Table 3 include the EDF, three additional core EWTK variables (the Slope, the Relative EDF, and the Industry Median EDF Growth), a $TrEx$ dummy indicating whether the EDF is above its sector-level trigger, an interaction term between $TrEx$ and the EDF level, the IRG, three ratings outlook dummy variables, and the recent downgrade dummy. Summary statistics for each model are shown in the last four rows of the table.

The most influential drivers of downgrade risk, in roughly decreasing order of importance, are the ratings outlook, the $TrEx$ variable for EDF trigger exceedance, the implied rating gap (IRG) variable and the EDF, the slope of the EDF term structure, the industry median EDF growth, and the recent downgrade dummy.

The ratings outlook variables and the recent downgrade dummy have statistically significant coefficients, which should come as little surprise given previously documented evidence that these drivers matter for ratings transitions [see, e.g., Metz and Cantor (2007) and Wang, Ding, Pan, and Malone (2017)]. Further, we see that the coefficients of the outlook variables sort as expected, with the coefficient of the NEG variable being greater than the coefficient of the STA variable, which in turn is greater than the coefficient of the POS variable in columns (2) and (5). The coefficient on the recent downgrade dummy is positive and significant, as expected, in columns (2) and (5).

The coefficient on the IRG is negative and statistically significant at conventional confidence levels across models, as expected. This indicates that more negative values for the implied ratings gap, which occur when the agency rating is favorable relative to the Market Implied Rating (MIR), are associated with a higher probability of a downgrade by the rating agency in the next 12 months. The robustness of this indicator, coupled with its availability across firms of all ratings, led us to use it as the sole driver of downgrade risk in models (4) and (6) for Aaa and Caa financial firms.

A negatively sloped term structure is associated with a higher downgrade probability. This makes sense, as negatively sloped EDF term structures are also associated with higher default risk before controlling for the EDF level, as shown in Ferry and Baron (2016), and it is plausible that they should forecast downgrade risk as well.

The only driver that does not seem to be robustly associated with higher downgrade risk is the relative EDF metric from the EWTK, whose coefficient is positive in model (1), negative in model (3), and not statistically significant at conventional levels in the remaining models.

4.1.2 THE DP AS A NON-LINEAR FUNCTION OF THE EDF

The net effect of the EDF on the log odds ratio index used to compute the DP is nonlinear, owing to the multiple channels through which it acts on the index. From the logs odds index function,

$$\begin{aligned}\log\left(\frac{\mathrm{DP}_{it}}{1-\mathrm{DP}_{it}}\right) = {} & \beta_0 + \beta_1 EDF_{it} + \beta_2 \mathrm{TrEx}_{it} + \beta_3 \mathrm{Slope}_{it} + \beta_4 RelEDF_{it} \\ & + \beta_5 \mathrm{IndMedEDFGrowth}_{it} + \beta_6 \mathrm{TrEx}_{it}\,\mathrm{EDF}_{it} + \beta_7 IRG_{it} \\ & + \beta_8 RecentDowngrade_{it} + \beta_9 STA_{it} + \beta_{10} POS_{it} + \beta_{11} NEG_{it}\end{aligned}$$

we see that an increase in the EDF will affect the DP estimate through the following five terms, holding the *Slope* variable constant:

$$\beta_1 EDF_{it} + \beta_2 \mathrm{TrEx}_{it} + \beta_4 RelEDF_{it} + \beta_6 \mathrm{TrEx}_{it}\,\mathrm{EDF}_{it} + \beta_7 IRG_{it}.$$

Note that, for simplicity of exposition, we suppress the dependence of the beta coefficient estimates above on the rating of the firm and its status as a financial or corporate entity. Differentiating the above expression to obtain the "net" EDF coefficient, we obtain:

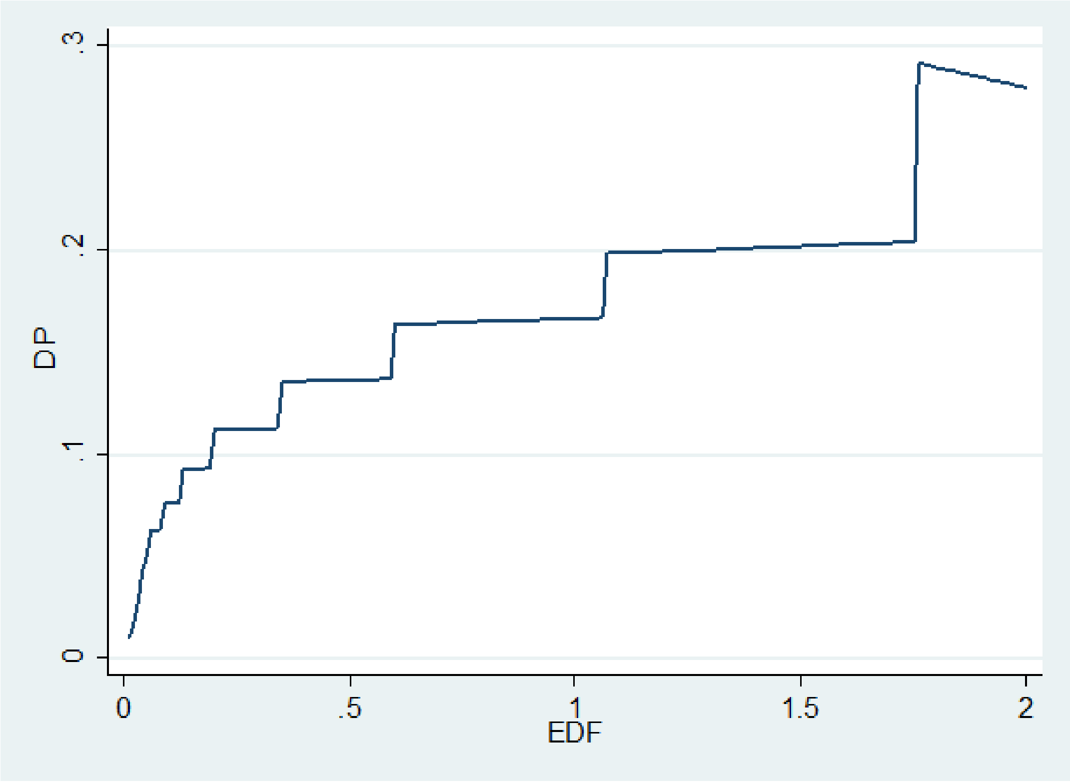
$$\beta_1 + \beta_4/\mathrm{IndMedEDF_{it}} + \beta_6 \mathrm{TrEx}_{it}$$

The DP function also contains discontinuities at points where the IRG increases by one due to the EDF crossing the boundary into the next implied rating category, as well as at the sector-level trigger value at which $TrEx$ increases from zero to one. Between these points of discontinuity, the slope of the DP as a function of the EDF is given by the expression above, which also depends on the value of $TrEx$ itself.

For most sub-models based on rating, we find that the "net EDF coefficient" for TrEx_{it} = 0 is positive for non-Aaa corporates and negative for non-Aaa/non-Caa-Ca-C financials. When TrEx_{it} = 1, however, the net EDF coefficient becomes negative for non-Aaa corporates as well as financials with agency ratings between Baa1 and B3, inclusive.

To illustrate the net effect of the EDF acting through all five relevant channels, we will plot the DP versus the EDF level for one corporate and one financial firm as examples. First, consider the case of the non-financial firm LabCorp in November, 2017. The value of the *Slope* variable for this observation is 0.122, the industry median EDF growth is 19.1 percent, the recent downgrade dummy equals zero, and the firm has a stable ratings outlook. Holding these factors constant, Figure 2 plots the DP as a function of the EDF for EDF in the range from 0.01 percent to 2 percent, using coefficients from the model for a corporate firm with a rating of Baa2, as was LabCorp's at the time. For the purposes of this exercise, we assume that the IRG is equal throughout to the value of the EDF-IR, which itself is a non-decreasing function of the EDF. The five aforementioned effects through which the EDF affects the DP manifest in Figure 2 as follows.

Figure 2. Downgrade Probability vs. EDF function for LabCorp (November 2017)



First, the DP rises in a "stair-step" pattern that reflects the effect of the EDF-IR incrementing discretely as the EDF increases and passes through the thresholds that define different implied rating levels. Between these discrete jumps, the line segments have a small positive slope in the region where $\mathrm{TrEx}_{it} = 0$, which reflects a positive net EDF coefficient for Baa2-rated corporates, given the value for LabCorp's industry median EDF of 0.33. The "net EDF coefficient" for LabCorp in November 2017 evaluates to 0.081 + (-0.01)/0.33 = 0.051 > 0 when $\mathrm{TrEx}_{it} = 0$ and 0.081 + (-0.01)/0.33 -.29 = -0.24 < 0 when $\mathrm{TrEx}_{it} = 1$.

Finally, if the EDF breaches the applicable trigger level, $TrEx$ will change values from zero to one, causing the index function to increase by a value equal to that variable's coefficient. The $TrEx$ coefficient is positive and large in nearly all the models, including the one that applies to this example. Thus, the EDF's crossing of the trigger level would have resulted in a large discrete jump upwards in the DP from around 0.2 to just under 0.3 for LabCorp in November 2017.

In general, we find for corporates that higher EDFs are associated with higher downgrade risk up until the point at which the EDF crosses its trigger level. After that point, downgrade risk jumps significantly higher, but the marginal effect of additional increases in the EDF becomes negative, excepting discontinuous jumps upward due to the IRG variable. The net result of these latter two effects is an upward sloping "sawtooth" pattern in the relationship between the DP and the EDF for values of EDF greater than its trigger level.

Turning next to the case of a financial firm, consider the insurer Aetna in November 2017. It's agency rating at that time, like that of LabCorp, was Baa2. The values of Aetna's *Slope* and industry median EDF growth variables were 0.542 and -7.41, respectively, its recent downgrade dummy was equal to zero, and its outlook was stable.

Figure 3. Downgrade Probability vs. EDF function for Aetna (November 2017)

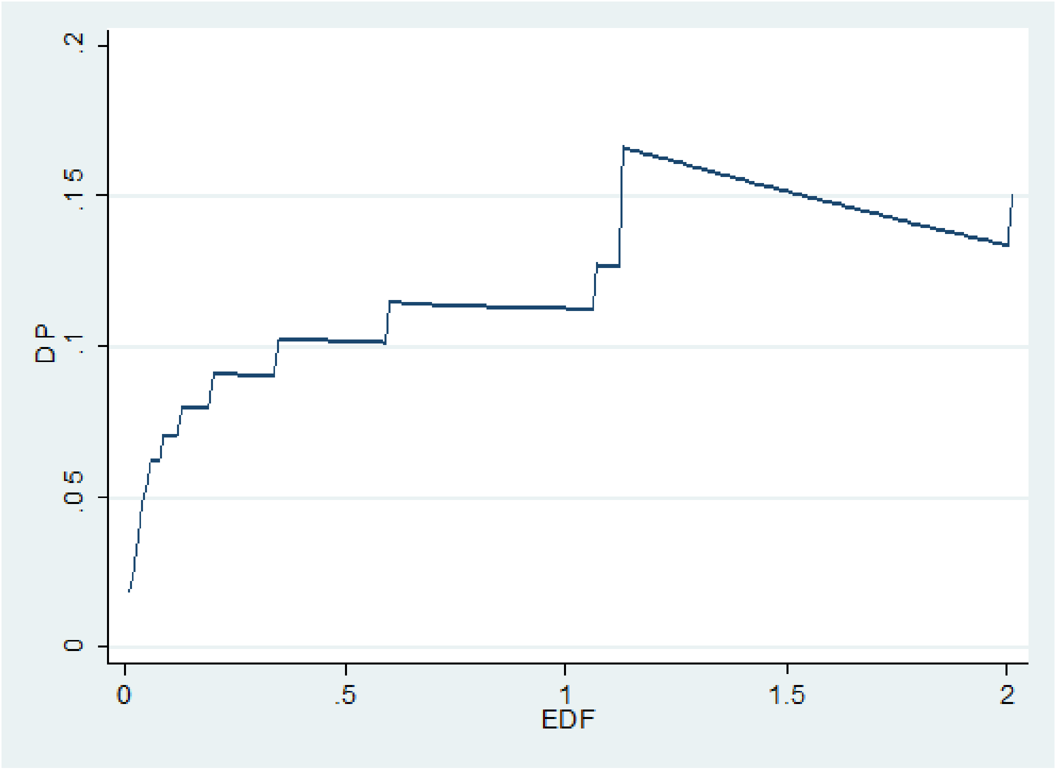


Figure 3 plots the DP as a function of the EDF with these factors held constant, following the same methodology we used to plot the relationship for non-financial firms. Note that, in all financial firm models, we impose the coefficient restriction $\beta_4 = 0$ due to the fact that the $RelEDF$ variable did not prove to be a statistically significant driver for such firms. Thus, the net EDF coefficient for

Aetna is $\beta_1 = -.0562 < 0$ when $\text{TrEx}_{it} = 0$ and $\beta_1 + \beta_6 = -.0562 + (-.2386) = -.2948 < 0$ for $\text{TrEx}_{it} = 1$. These values for the betas are taken directly from column (5) of Table 3.

Similar to the case of LabCorp, we see that the DP rises in a stair-step pattern until the point where the EDF reaches the November 2017 trigger level for global financials of 1.2 percent, at which point the DP function experiences a somewhat large discrete jump upwards before falling into a "sawtooth" pattern for values of the EDF over the trigger level.

Our findings on the nonlinear effect of the EDF on the probability of downgrade for both corporates and financials can be summarized as follows. For low EDFs, increases in the DP are driven primarily by deterioration in the implied ratings gap (IRG) until the point where the EDF crosses the trigger level for its sector. At that point, the DP undergoes a relatively large jump upwards. For values of the EDF greater than the trigger, a "sawtooth" pattern emerges.

Note that we have deliberately kept the value of the *Slope* variable constant in the above analysis. However, the slope of the term structure has a relatively large negative correlation with the EDF level; at -0.88, it is the largest correlation by absolute value between any of our variables. Thus, given the negative and somewhat large coefficient for the *Slope* variable in our models for non-Aaa/non-Caa-Ca-C financial firms in particular, on average we are likely to see increases in the EDF associated with modest but smooth increases in the DP between changes in the implied rating for those firms as the EDF deteriorates. In other words, as the 1-year EDF rises, a simultaneous flattening of the EDF term structure often contributes modestly to the increase in the DP as well. When a recent downgrade has happened or when the outlook worsens, the curves shown above shift up accordingly.

4.1.3 MODEL ACCURACY AND MODEL COMPARISON

To assess the predictive accuracy of the DP, Table 4 reports the Brier scores and accuracy ratios of the full DP model for all firms, the set of non-financial corporates, and the set of financial firms, respectively. We use the version of the model that includes rating-based drivers. All accuracy ratios shown in Table 4 are statistically different from zero at the 1% level.

Table 4. Brier scores and Accuracy Ratios for DP Model

Prob. of Downgrade	DP (all)	DP (Corporates)	DP (Financials)
Brier Score	0.0970	0.0964	0.0995
Accuracy Ratio	0.5346	0.5436	0.4986
p statistic	0.0000	0.0000	0.0000

While the Brier scores'[6] in Table 4 for each group of firms are substantially greater than the equivalent scores obtained by EDFs in default prediction on the same sample, and the accuracy ratios somewhat lower, the fact that the ARs are substantially above zero indicates that the model has meaningful predictive power. Figure 4 displays the ROC curves[7] for each of the four broad DP variations whose coefficients are reported in Table 3.

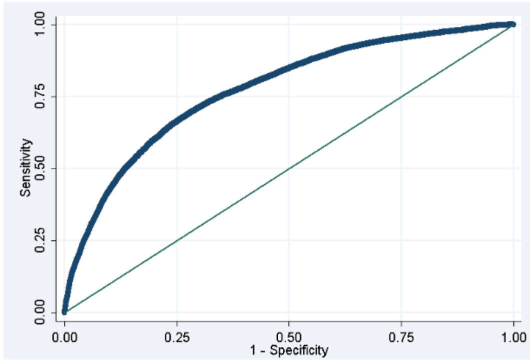
As expected, the version of the DP model with ratings-based drivers (see Figures 4A and 4C) outperforms the model that omits the outlook and recent downgrade information (see Figures 4B and 4D). The downgrade model slightly outperforms for corporates vis-à-vis financials.

[6] The Brier score for binary event prediction is equal to the mean of the squared differences between event probabilities and outcome variables. Lower Brier scores indicate greater forecasting ability.

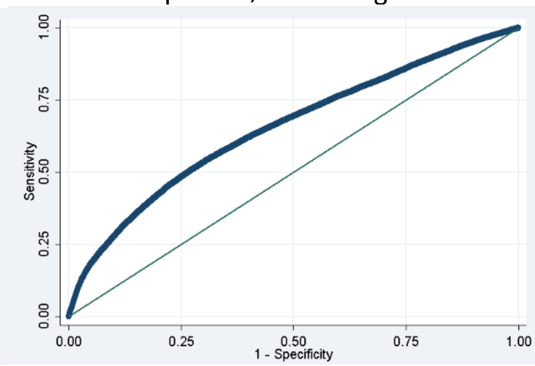
[7] The ROC (receiver operating characteristic) curve for a binary event prediction metric plots the rates of true positives attained for different cutoffs of the metric against the corresponding rates of false positives. The rate of true positives is often referred to as the metric's sensitivity, and the rate of false positives is equal to one minus the metric's specificity. The accuracy ratio for the metric is calculated as twice the area under the ROC, minus one.

Figure 4. ROC Curves

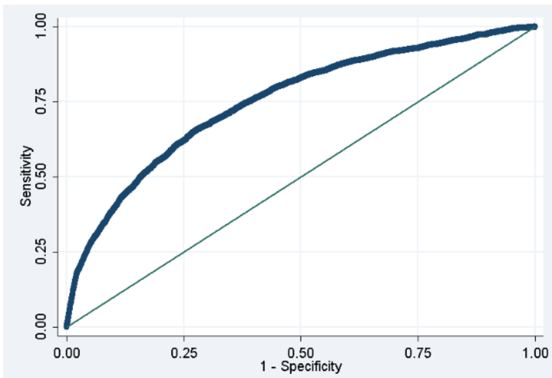
4A. DP Corporates



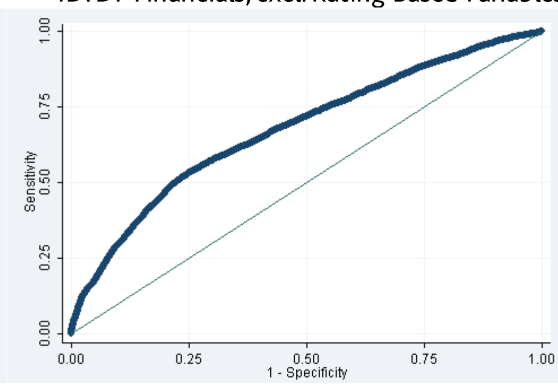
4B. DP Corporates, excl. Rating-Based Variables



4C. DP Financials



4D. DP Financials, excl. Rating-Based Variables



For context, it is useful to compare the DP's performance to that of alternative models built for the same purpose. In particular, we are aware that some clients use the gap between the agency rating and the various Market-Implied Ratings (MIRs) to monitor the risk of downgrade. The utility of this approach is corroborated by the negative and statistically significant coefficient on the IRG variable in the DP model equations shown in Table 3.

We estimate a series of seven models to illustrate more clearly the value-added of the DP for downgrade prediction vis-à-vis alternatives. The models include: logit models with the EDF as the only driver, with the IRG as the only driver, with the EDF plus the IRG, with these drivers plus the EWTK measures, a model that adds the recent downgrade dummy, a model that further adds the outlook dummies, and finally the DP model itself. Accuracy ratios obtained in-sample for each of these models are displayed in Table 5.

Table 5. In-Sample Accuracy Ratios for Downgrade Forecasting Models

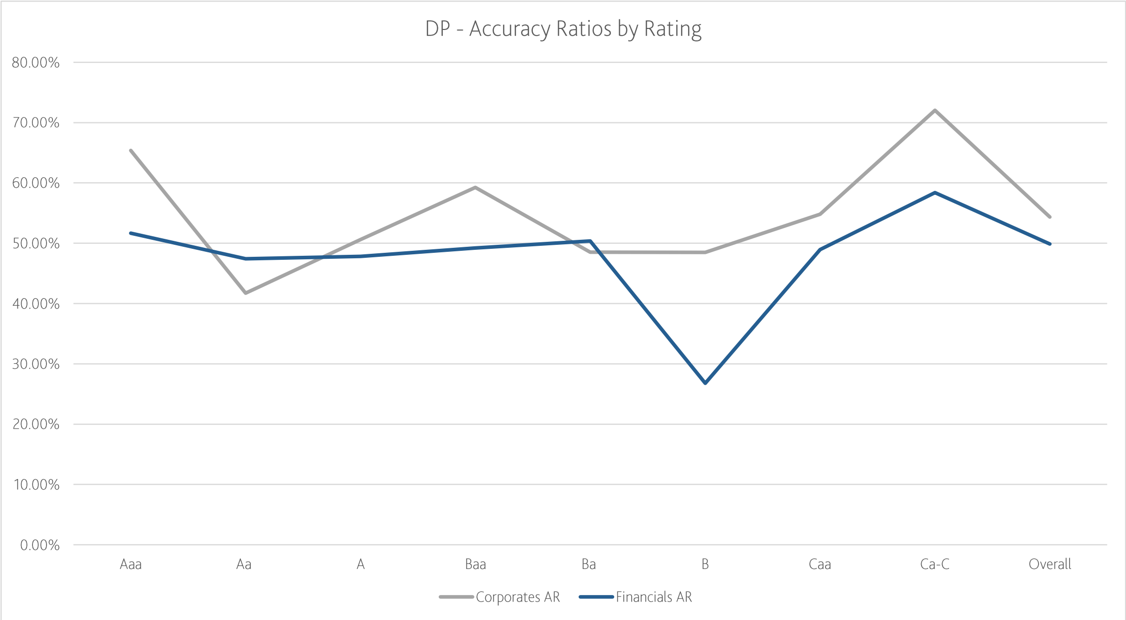
Inputs	AR Corporates	AR Financials
EDF	28.96%	24.96%
IR Gap	37.94%	30.36%
EDF + IR Gap	39.66%	29.82%
EDF + EWTK Measures + IR Gap	41.52%	40.14%
EDF + EWTK Measures + IR Gap + Recent Downgrade	41.92%	41.18%
EDF + EWTK Measures + IR Gap + Recent Downgrade + Outlook	51.78%	49.54%
DP	54.36%	49.86%

The results of this simple horse race are illuminating: adding EWTK variable information to the EDF and the implied ratings gap (IRG) improves forecasting performance, particularly for financial firms. When we add the recent downgrade dummy and the ratings outlook information, we obtain additional increases in the ARs of both corporates and financials, although the performance boost from adding the recent downgrade dummy is smaller than expected. Estimating separate models for specific ratings categories results in further increases in the ARs of the DP model for both corporates and financials, as can be seen by comparing results in the last and penultimate rows of Table 5. These results provide evidence that in most cases, each broad source of forecasting "lift" in the DP model provides valuable information that helps increase forecasting accuracy for downgrade events without duplicating information contained in the other measures. If we try other promising combinations of the variables, such as a model with the IRG and the ratings outlook, we find that these too are dominated by the DP in terms of the accuracy ratio.

4.1.4 PERFORMANCE BY RATING CATEGORY

We now break down the model's performance by rating category. Accuracy ratios attained by the model for each of the broad Moody's Investors Service (MIS) rating categories are show in Figure 5.

Figure 5. Accuracy Ratios by Rating Category: DP Model with Ratings-Based Drivers

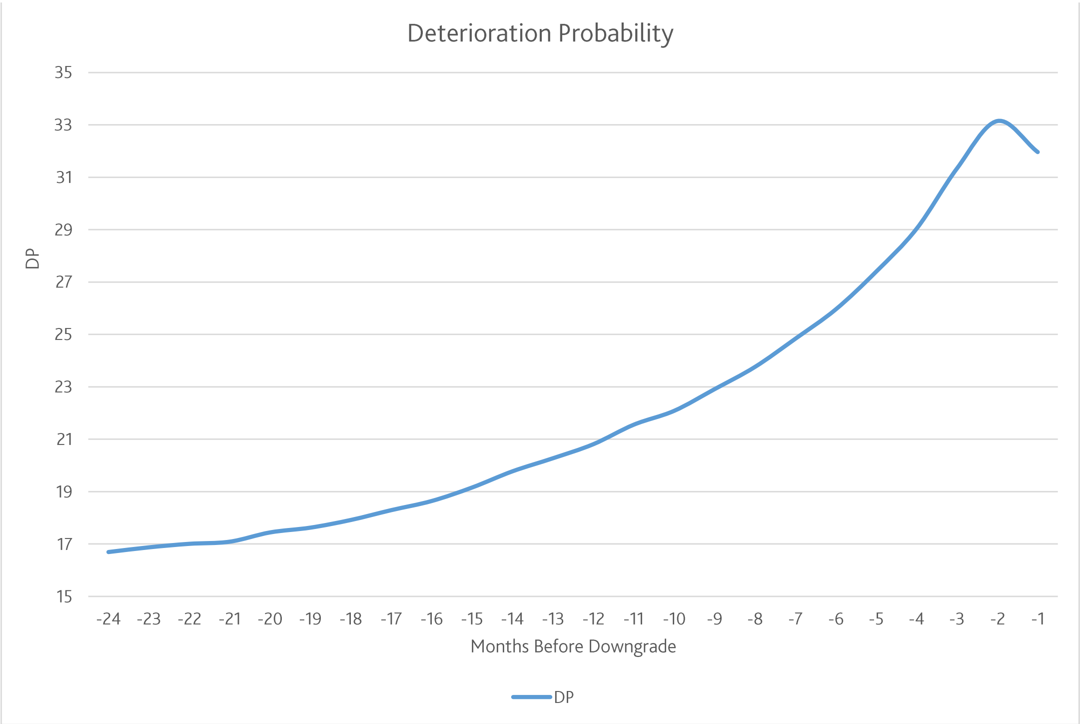


The consistently solid performance of the DP model across the ratings spectrum, with the possible exception of B-rated financials, is noteworthy. Furthermore, the model seems to perform particularly well for Aaa-rated, Ca-rated, and C-rated corporates, as evidenced by accuracy ratios at or above 65% for these rating categories at the ends of the rating spectrum. Estimation of DP model coefficients by rating, with parsimonious specifications used for ratings categories with fewer observations in the case of financials, was key to achieving broadly consistent performance across the ratings spectrum.

4.1.5 DP DYNAMICS: RUN-UP TO DOWNGRADE EVENTS AND PERFORMANCE OVER THE BUSINESS CYCLE

Let us now turn to the dynamics of the DP in the run-up to a downgrade event and through the business cycle. Figure 6 plots the average DP score, across all downgrade events in our sample, during the 24 months prior to the month in which downgrade occurs.

Figure 6. Run-up to Downgrade: Average DP Score During Two Years Preceding a Downgrade Event



From Figure 6, we see that the DP behaves exactly as it should in the two years before a downgrade event occurs: on average it increases smoothly and continuously prior to downgrade, accelerating upwards in the last six months prior to the event. In terms of magnitude, the downgrade probability essentially doubles during the two years preceding a typical downgrade, and the peak in the measure prior to a downgrade occurs on average two months before the event. In other analysis not shown here, we find that the measure falls on average during the two years prior to an upgrade event, which is consistent with the idea that factors that increase the probability of a downgrade should typically be associated with a lower probability of an upgrade, and vice versa.

Figure 7. DP over the Business Cycle, Corporates and Financials, 2004-2016

Corporates Financials

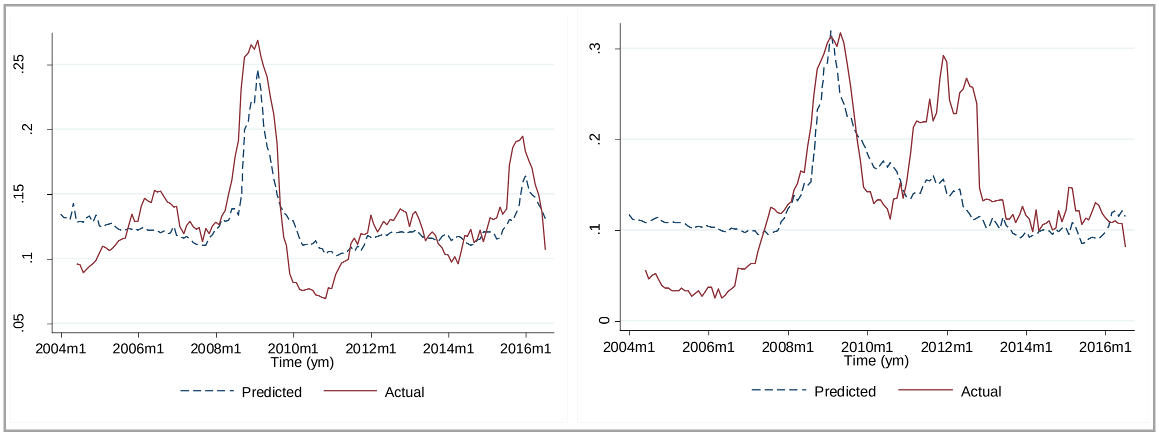


Figure 7 plots the median DPs for corporates (left panel) and financials (right panel) over the business cycle. In an given month, the "actual" variable shown in the plots is the average across firms of the dependent variable from our logit models. Recall that our dependent variable is set equal to one at time *t* when the firm experiences a downgrade during any month from t+1 to t+12, inclusive, and set to zero otherwise. Since peaks in raw monthly downgrade frequencies tend to lead peaks in the dependent variable by five months on average, we shift the "actual" line in the plot forward by five months to account for this difference.

Median downgrade probabilities spike during high stress economic periods such as the recent financial crisis, as expected. The peak median DP for financial firms during the financial crisis is around six percentage points greater than the peak median DP for corporates during that period. We find this result plausible in light of the central role played by the financial sector in the Great Recession. Actual downgrade rates seems to under- and over-shoot the predicted downgrade rates pre- and post-crisis, respectively, although this behavior is more pronounced for financials than corporates, and may be related to a post-crisis change in rating methodology for financial firms.

Out-of-sample validation

As an out-of-sample test of model performance, we consider how the model would have fared during the recent bear market for oil that lasted from approximately August 2014 until December 2016, which is the latest date in the estimation dataset. We estimate the coefficients of each of the models in Table 5 using data from a training period, and then assess model performance during a testing, or out-of-sample period that begins in the month following the end of the training period and extends until December 2016. For robustness, we consider three separate end dates for the training period: July 2013, July 2014, and July 2015.

Table 6 reports accuracy ratios for each of the models considered in Table 5 for the population of corporate firms, and Table 7 reports the same information for financials. Each data column of Tables 6 and 7 displays model ARs obtained using a different choice of the training sample end date.

Table 6. Out-of-sample Accuracy Ratios (ARs) for Selected Models: Corporate Firms

Model Inputs	Training Sample End Dates		
	July 2013	July 2014	July 2015
EDF	33.72%	34.76%	39.62%
IR Gap	44.74%	47.28%	50.98%
EDF + IR Gap	45.98%	48.44%	52.26%
EDF + EWTK Measures + IR Gap	46.92%	50.16%	55.00%
EDF + EWTK Measures + IR Gap + Recent Downgrade	45.70%	48.82%	54.02%
EDF + EWTK Measures + IR Gap + Recent Downgrade + Outlook	54.16%	55.72%	61.56%
DP	56.56%	57.48%	62.06%

Table 7. Out-of-sample Accuracy Ratios (ARs) for Selected Models: Financial Firms

Model Inputs	Training Sample End Dates		
	July 2013	July 2014	July 2015
EDF	15.88%	13.04%	23.54%
IR Gap	13.34%	12.60%	23.70%
EDF + IR Gap	13.00%	12.54%	23.26%
EDF + EWTK Measures + IR Gap	17.44%	13.72%	19.40%
EDF + EWTK Measures + IR Gap + Recent Downgrade	19.24%	17.34%	29.78%
EDF + EWTK Measures + IR Gap + Recent Downgrade + Outlook	41.84%	41.54%	42.90%
DP	42.92%	44.70%	47.88%

We find that the DP model outperforms all of the competitor models out-of-sample for each of the three different choices for the end date of the training period. Overall, the results suggest that the DP performs quite well out-of-sample compared to variety of plausible alternative models. In fact, the out-of-sample AR for corporates is higher in each of the exercises shown in Table 6 than in the in-sample results reported in Table 5.

Additional Out-of-Sample Exercises

In addition to the above out-of-sample exercises, we performed a number of additional tests to evaluate the robustness of the model. The results from these exercises are shown in Table 8 below.

Table 8. Additional Out-of-Sample Exercises: Accuracy Ratios

Out-of-Sample Test	Corporates	Financials
Constant Coefficients	58.28%	52.06%
Three-Month Re-Estimation	57.00%	53.00%
Extended Dataset	62.96%	57.48%
Capped/Floored	58.28%	52.06%

In the first exercise, we estimated the model up to the month marking the end of the first half of the sample (June 2010), and then used the coefficients estimated at that time to produce DP outputs from July 2010 to September 2016. The AR obtained out-of-sample for this exercise was 58.28% for corporates and 52.06% for financials.

Second, we ran a variation of the same exercise in which DP estimates are produced beginning in July 2010, but model coefficients were reestimated every three months. The ARs obtained from this exercise were 57% for corporates and 53% for financials, respectively. Together, the similarity of the "fixed estimation" and "periodic reestimation" results indicates that frequent reestimation of results may not be crucial for achieving good model performance across the business cycle.

Third, we extended our research dataset, which comprised data from the period January 2004 until September 2016, to include an additional 14 months of data from the beginning of our production dataset. This extended the dataset up to November 2017. We then used model estimates computed using the in-sample period ending September 2016 to generate DP metrics during the "production out-of-sample" period from October 2016 until November 2017. The resulting accuracy ratios of the DP metric during the production out-of-sample period were 62.96% for corporates and 57.48% for financials, respectively.

Fourth and finally, we examined implications for a version of the DP metric that imposes a cap of 70% and a floor of 1% on the DP value in production, analogous to the manner in which our EDF metric is reported with a cap of 50% for corporates, or 35% for financials, and a floor of 0.01% for both. We computed the 70% cap value by grouping downgrade events and associated DP metrics and using the approximate value of the DP corresponding to the 2.5th percentile of the quantity $1 - DP$. Similarly, we determined the 1% floor by grouping non-downgrade events and associated DP metrics and using the approximate value of the DP corresponding to approximately to the 2.5th percentile of the DP in this case. To evaluate the effects of imposing said cap and floor on the DP, we reran the exercise mentioned above in which we estimated the model up to the month marking the end of the first half of the sample (June 2010), and then used the coefficients estimated at that time to produce DP outputs from July 2010 to September 2016. The ARs obtained for the DP metric with cap and floor imposed on it were identical to the ones obtained for the unrestricted DP: 58.28% for corporates and 52.06% for financials.

We now present three useful case studies that illustrate the value of DP as a predictor of deterioration in a firm's credit quality.

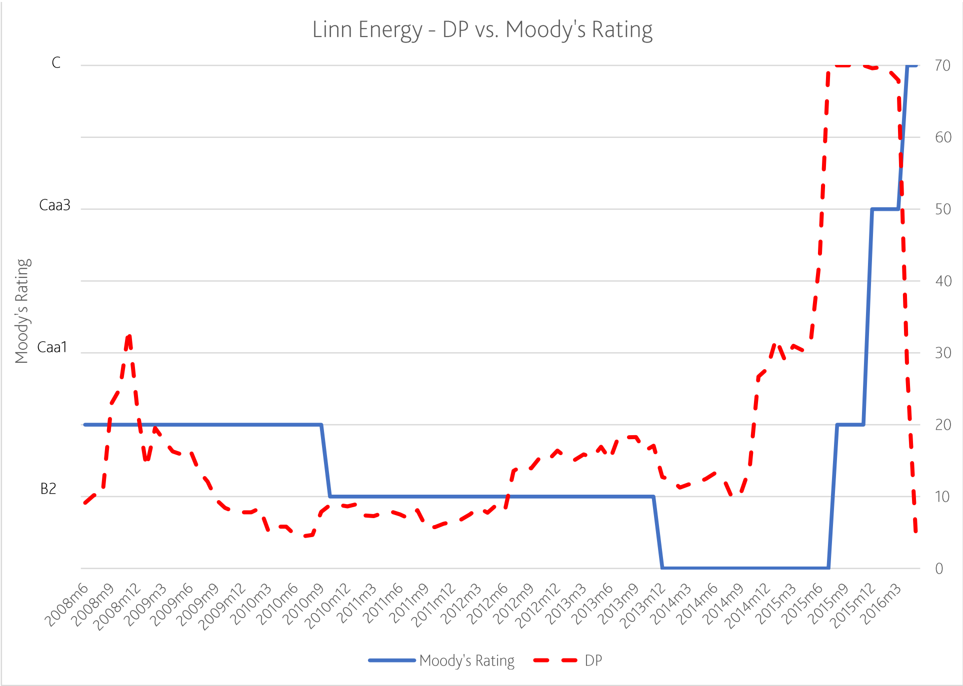
4.1.6 CASE STUDIES OF THE DP MODEL IN PRACTICE

The first case study involves Linn Energy, a U.S. oil and gas company that ultimately defaulted, and the second case study concerns Abengoa, S.A., a diversified Spanish telecom company with exposures to the energy, environment, and transportation sectors. The third case study involves a Chinese coal mining firm, Yanzhou Coal, which transitioned from being unrated to rated during the sample.

Linn Energy

Linn Energy, founded in 2003, is an American oil and gas company located in Houston, Texas[8]. Linn has reserves throughout East Texas, Northern Louisiana, and Oklahoma, according to the company's website.[9] Unfortunately for many oil firms, crude oil prices have only recently risen from historically low levels, with West Texas Intermediate prices averaging around $50 a barrel over the past year, not too far from the $30 nadir reached in February 2016, and far from the peak of $145 reached in July of 2008[10]. Oversupply, driven by fracking, has been the main driver of lower oil prices, sometimes interrupted by slight price increases when OPEC nations promised to cut back production.

Figure 8. DP and Rating versus Time for Linn Energy



Linn Energy filed for Chapter 11 Bankruptcy in May 2016[11] following a series of downgrades that began in 2015, with the latest coming in March 2016. The common threads linking these downgrades are a persistently high cost of capital and weak asset coverage of debt[12].

Figure 8 depicts Linn Energy's senior unsecured or equivalent Moody's rating (as produced by the Senior Unsecured Rating Algorithm) and the DP over time. During the 2008-2013 period, Linn was upgraded twice, from B3 to B1. The DP remained low during that time. In Q3 of 2014, the DP

---

[8] Jeansonne, Clay. "LINN Energy Founder Michael C. Linn Retiring", NASDAQ Global NewsWire, 01 Dec 2011

[9] LINN Energy Company Website, http://www.linnenergy.com/operations/

[10] U.S. Energy Information Administration (2017), "Crude Oil Prices: West Texas Intermediate", Retrieved Aug. 4, 2017 from EIA database.

[11] French, Gretchen and Steven Wood, "Moody's downgrades Linn Energy's PDR to D-PD on bankruptcy filing", Moody's Investors Service, 12 May 2016.

[12] French, Gretchen and Steven Wood, "Moody's Downgrades Linn Energy's Corporate Family Rating to Ca", Moody's Investors Service, 16 Mar 2016.

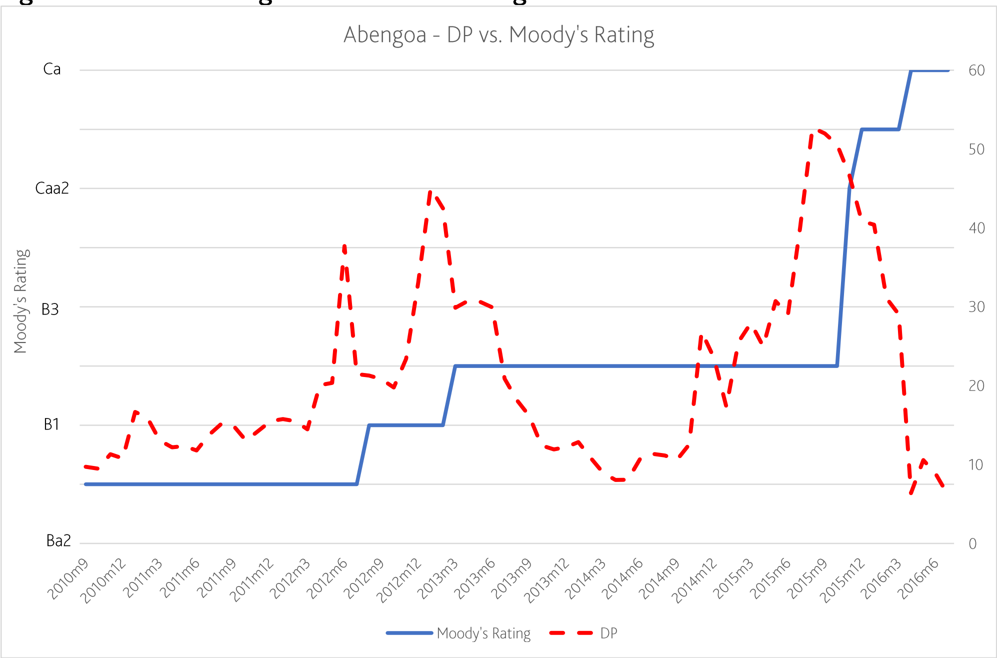
began a rapid ascent to 70, corresponding to a 70% downgrade probability, which it attained soon before Linn Energy was downgraded two notches from B1 to B3 around July of 2015. The DP remained around 70 as Linn was downgraded again to Caa3, and then fell substantially just before the firm's default, at which point its rating was withdrawn.

Linn Energy currently does not have a Moody's rating, but has recently completed a financial restructuring.

Abengoa S.A.

Abengoa, S.A. is a telecommunications company based in Seville, Spain, and was founded in 1941. Its label as a telecommunications firm notwithstanding, Abengoa is fairly well diversified, providing solutions in the energy, environment, and transportation sectors. The firm maintains operations across the globe, including in North America, South America, and the rest of Europe[13]. In 2010, Moody's Investors Service initially assigned Abengoa a Ba3 (stable) rating, citing well-established operations, a sound market position, and a diversified portfolio of services in various sectors.[14]

Figure 9. DP and Rating versus Time for Abengoa S.A.



After a downgrade in July 2012 (to B1, stable), Abengoa was again downgraded to B2 (stable) in March 2013. In both events, ratings analysts cited "slow deleveraging", a weak Spanish macroeconomic climate (home to a quarter of the firm's revenue-generating activities), and need for consistent regulatory support for its solar energy projects. The firm's liquidity profile also caused concerns for ratings analysts, who indicated that a worsening liquidity profile could lead to further downgrades.

Figure 9 plots Abengoa's Moody's rating and DP measure versus time during the period from September 2010 until September 2016. As is apparent in the Figure, the DP measure of downgrade risk spiked two months before each of the aforementioned downgrades occurred. Further, following a stable period for the firm's rating in which the DP metric temporarily reached a nadir of just over 12% in Q2 of 2014, the measure rose markedly during the period from Q3 of 2014 to Q3 of 2015, reaching a peak near 55% around August 2015. In November 2015, Moody's Investors Service

---

[13] Bloomberg, L.P., "Company Overview of Abengoa, S.A.", 17 Nov 2017.

[14] Bodard, Eric de and Wolfgang Draack, "Moody's assigns Ba3 CFR to Abengoa (Spain); outlook stable", Moody's Investors Service, 13 Sep 2010

further downgraded the firm not once, but twice: first, from B2 to B3 (outlook stable), and then from B3 to Caa2 (outlook negative). [15] The firm was downgraded to Caa3 in December 2015.

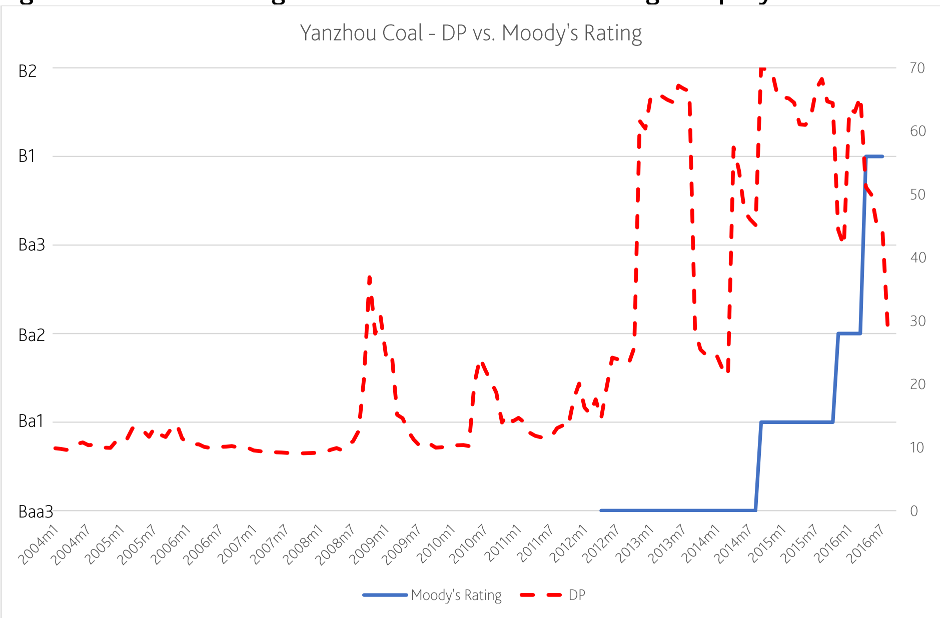
The firm's inability to attract and retain investors proved to be central to its demise, as the downgrade to Caa2 was announced following its failure to secure a desperately needed EUR 250 million in equity. The firm also filed a pre-insolvency claim which, even if it was not considered a default, reflected a perpetually worsening liquidity profile. Abengoa formally began bankruptcy proceedings in March 2016.[16]

For each of Abengoa's downgrade events, the DP measure provided at least two months of fairly clear early warning that a downgrade was likely. Even in the nearly three-year period in which Abengoa maintained its rating, the model is able to detect the firm's increasing credit risk, as evidenced by the DP's peak at a level just shy of 55% before the series of downgrades that led to Abengoa's default. Note that the sharp drop in the measure near the end of the sample period does not signal a false negative *per se*. Rather, when firms reach the lower end of the rating scale distribution, they are more likely to have their rating withdrawn due to a default rather than to receive another downgrade. At the time of writing, Abengoa appears to have made progress towards completing its financial restructuring plan.

Yanzhou Coal Mining Company

Yanzhou Coal Mining Company is a Chinese coal and energy company. Founded in 1997, the company operates in and serves the mainland China region. It has market presence in most of China and some parts of northeastern Asia, as well as operations in Australia.

Figure 10. DP and Rating vs. time for Yanzhou Coal Mining Company



Chinese coal companies face two main challenges. One is the fact that coal prices have been well below their mid-2008 peak of about $140/short ton for some time. In particular, since the financial crisis, coal prices have ranged from $40-80 a short ton, with the maximum being a little over $80/short ton in late 2011[17]. The second challenge is that the market for coal in China is oversaturated. Even with Yanzhou controlling a sizable market share, it has had to scale back production, especially right before the winter

[15] Heck, Matthias and Matthias Hellstern, "Moody's downgrades Abengoa to Caa2; outlook negative", Moody's Investors Service, 26 Nov 2015.

[16] Heck, Matthias, Matthias Hellstern, and Wen Li, "Insights Into Abengoa's Default", Moody's Investors Service, 17 Mar 2016.

[17] Quandl Financial, Economic, & Alternative Data (2017), "US Coal Prices by Region", Retrieved 3 Aug 2017 from Quandl database.

months, when coal is the most valuable[18]. Adding to these problems, an Australian subsidiary, Yancoal Australia, has operated at a loss since 2013.[19]

At the time the company received its initial Moody's long-term (LT) corporate family rating (Baa3, stable) in May, 2012, analysts praised the company's optimal market position, compliance with safety and environmental regulation, and its close relationship with the local government.[20] The DP and senior unsecured or equivalent Moody's rating over time for Yanzhou are shown in Figure 15. There had been a couple of short-lived spikes in the DP, in parts of 2008 and 2010, prior to the conferral of the rating. However, the DP began to rise significantly and persistently shortly after the rating was published, from around 14% in May 2012 to just under 70% by mid-2013. This rise is due largely to the fact that, once the firm was rated, the modeled DP began to take advantage of information embodied in the implied ratings gap (IRG) variable, which is a valuable predictor of downgrades.

As the DP approached its upper bound of 70%, the LT corporate family rating of the firm was subsequently downgraded several times: to Ba1 (September 2014), Ba2 (November 2015), and finally to B1 in April 2016. The DP sat at a still high but somewhat modest level of around 30% as of September 2016. As of March 2018, Yanzhou holds a B2, stable rating from Moody's, and has not defaulted.

4.1.7 THE DISTRIBUTION OF DOWNGRADE MAGNITUDES

In addition to providing a downgrade event probability in the form of the DP, we include information on the magnitude distribution for ratings downgrades in order to facilitate calculations that require information on downgrade severity.

Figure 11. Empirical Frequencies of Downgrade Magnitude, All Ratings

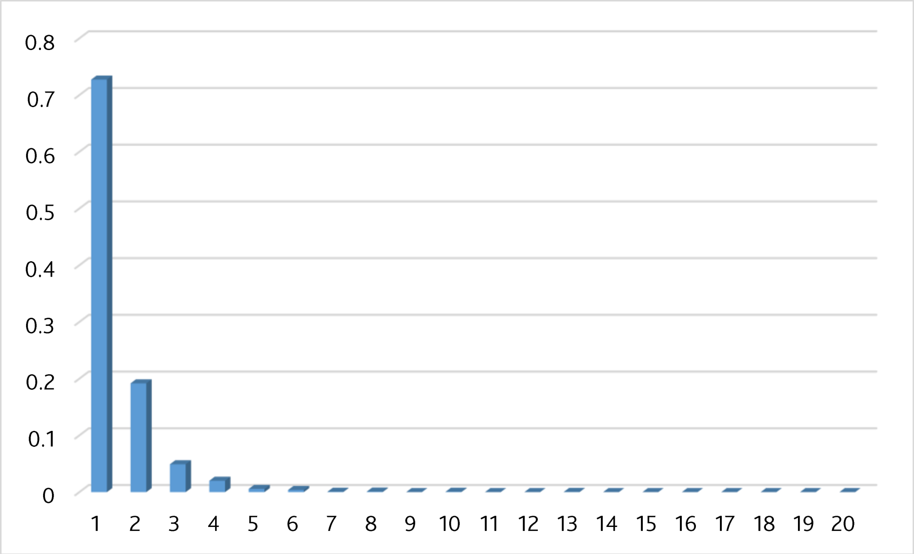


Figure 11 plots the empirical frequencies with which downgrades, measured by number of ratings notches, occur. As the quantitative rating scale ranges from 1 (Aaa) to 21 (C), the range for the number of possible downgrade notches goes from 1 to 20.

The downgrade notch distribution in Figure 11 is heavily skewed to the right, with the mode and median equal to 1 and the mean equal to 1.4. The largest downgrade in-sample for Top 90 public firms during the period from 2004-2016 was 13 notches. For 1 to 6 notch downgrades, the

[18] Murray, Lisa, "China's coal industry wrestles with overcapacity, pollution, and a restive public", Australian Financial Review, 27 Nov 2016.

[19] "Yanzhou: Yancoal Australia still not breaking even", Mining Acquisitions, 3 May 2017, Retrieved Aug. 2, 2017 from https://miningacquisitions.com/2017/05/03/yanzhou-yancoal-australia-still-not-breaking-even/.

[20] Chung, Ivan and Gary Lau, "Moody's assigns first-time Baa3 ratings to Yanzhou Coal", Moody's Investor's Service, 27 Apr 2012.

empirical frequencies are 72.72%, 19.17%, 4.91%, 2.00%, 0.56%, and 0.38%, respectively, with downgrades of more than six notches comprising only 0.25% of the sample.

In Table 9, we break down the frequencies of downgrade notches by initial (pre-downgrade) rating category. One-notch downgrades are the most common for all initial ratings categories. Interestingly, the ratings with the lowest probability of one-notch downgrades are Baa3 and Caa2. Baa3, the lowest possible investment grade rating, has a one-notch downgrade probability of 57%, and Caa2 has a one-notch downgrade probability of 47%. Ratings at these levels have a greater probability of being downgraded two or more notches conditional on the event that a downgrade occurs. Note that the probability of a downgrade occurring in the first place is a separate issue, and this probability — as shown in Section 2 — differs by initial rating as well.

Table 9. Empirical Frequencies of Downgrade Magnitude by Initial Agency Credit Rating

	Aaa	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa1	Caa2	Caa3	Ca
1	**0.647**	**0.719**	**0.800**	**0.686**	**0.768**	**0.738**	**0.791**	**0.828**	**0.805**	**0.570**	**0.682**	**0.720**	**0.712**	**0.740**	**0.728**	**0.671**	**0.720**	**0.471**	**0.706**	**1.000**
2	**0.235**	**0.228**	**0.122**	**0.265**	**0.181**	**0.194**	**0.150**	**0.118**	**0.152**	**0.279**	**0.221**	**0.173**	**0.193**	**0.212**	**0.193**	**0.203**	**0.140**	**0.368**	**0.294**	0.000
3	0.000	**0.018**	**0.044**	**0.025**	**0.019**	**0.055**	**0.043**	**0.029**	**0.014**	**0.088**	**0.051**	**0.053**	**0.069**	**0.043**	**0.045**	**0.077**	**0.086**	**0.162**	0.000	0.000
4	0.000	**0.035**	**0.022**	**0.008**	**0.019**	**0.008**	**0.008**	**0.014**	**0.020**	**0.037**	**0.021**	**0.036**	**0.017**	**0.005**	**0.020**	**0.042**	**0.054**	0.000	0.000	0.000
5	**0.059**	0.000	0.000	**0.008**	**0.007**	0.000	0.000	**0.007**	**0.003**	**0.015**	**0.015**	**0.004**	**0.004**	0.000	**0.010**	**0.007**	0.000	0.000	0.000	0.000
6	**0.059**	0.000	0.000	**0.008**	0.000	0.000	**0.004**	**0.004**	**0.003**	**0.004**	**0.010**	**0.009**	**0.004**	0.000	**0.005**	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.007**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	**0.004**	0.000	**0.003**	0.000	0.000	**0.004**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	**0.011**	0.000	0.000	**0.004**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	**0.007**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

5 Conclusion

This paper presents and validates a model for the probability of credit deterioration within a 12-month horizon. The model applies to publicly traded firms. For rated public firms, the model's output is equal to the probability of a downgrade of the Moody's agency rating in the next twelve months. For unrated public firms, the model's output should be interpreted as the probability of experiencing a downgrade-like credit quality event over 12 months. The latter output is computed as the downgrade probability of a rated firm with the same subset of observable common characteristics. Validation results indicate that on average, the Deterioration Probability measure rises sharply during the two years prior to a downgrade event. The model performs well out-of-sample during the recent bear market for oil, and is robust to alternative choices of the out-of-sample period. Our case studies indicate how the measure would have provided effective early warning of downgrades for one US and two international firms, which subsequently defaulted, declared bankruptcy, or experienced multiple downgrades. To complement our downgrade probability model, we provide statistics on the distribution of downgrade magnitudes, as measured by number of notches, to aid computations of potential downside exposure linked to deteriorations in credit quality.

6 Appendix

Table A1. Correlation matrix*

	Downgrade 12-month	EDF	Slope	TrEX	Rel EDF	IndMedΔ	IR Gap	Recent DG	STA	POS	NEG
Downgrade 12-month	1										
EDF	0.16	1									
Slope	-0.12	-0.88	1								
TrEX	0.18	0.63	-0.39	1							
Relative EDF	0.05	0.39	-0.35	0.24	1						
Ind Med Δ	0.10	0.14	-0.19	0.08	-0.01	1					
IRG	-0.21	-0.20	0.06	-0.28	-0.07	-0.04	1				
Recent DG	0.11	0.22	-0.18	0.20	0.06	0.07	-0.10	1			
STA	-0.19	-0.13	0.09	-0.14	-0.04	-0.02	0.07	-0.14	1		
POS	-0.07	-0.03	0.03	-0.04	-0.01	-0.03	0.08	-0.08	-0.39	1	
NEG	0.17	0.15	-0.10	0.16	0.04	0.05	-0.15	0.22	-0.55	-0.11	1

*Table A1 reports the Pearson correlation coefficients between the dependent variable and the independent variables used in the DP model.

Table A2. Variance Inflation Factors*

Variable	VIF	1/VIF
EDF	8.38	0.1194
Slope	5.57	0.1794
TrEX	2.25	0.4435
STA	1.29	0.7750
Recent Downgrade	1.21	0.8235
NEG	1.20	0.8299
Relative EDF	1.20	0.8308
IR Gap	1.14	0.8751
Industry Median Growth	1.08	0.9239
POS	1.04	0.9576

*Table A2 reports the variance inflation factors (VIFs) for each of the independent variables in the DP model.

Table A3. Coefficients for Corporates – By Rating*

	Coefficients/*Standard Errors*						
Variable	Aa	A	Baa	Ba	B	Caa	Ca-C
EDF	-0.359**	0.013	0.081***	0.091***	0.112***	0.156**	0.018
	(0.156)	*(0.045)*	*(0.026)*	*(0.023)*	*(0.023)*	*(0.076)*	*(0.025)*
TrEx	0.830	-0.665	0.992***	0.846***	0.606***	0.845***	--
	(1.208)	*(0.638)*	*(0.121)*	*(0.076)*	*(0.062)*	*(0.170)*	--
Slope	-1.125***	-0.518***	-0.345***	-0.007	-0.033***	-0.003	-0.022
	(0.193)	*(0.067)*	*(0.030)*	*(0.017)*	*(0.011)*	*(0.014)*	*(0.036)*
Relative EDF	0.187***	-0.011	-0.010*	0.0009	-0.003***	0.001	0.0004
	(0.032)	*(0.034)*	*(0.006)*	*(0.004)*	*(0.000)*	*(0.001)*	*(0.003)*
IndMedEDF Growth	0.001***	0.0009***	0.0006***	0.0008***	0.0005***	0.0002	0.001*
	(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
TrEx x EDF	-0.14	-0.328***	-0.295***	-0.083***	-0.095***	-0.121	--
	(0.153)	*(0.118)*	*(0.031)*	*(0.024)*	*(0.022)*	*(0.076)*	--
IR Gap	-0.048***	-0.151***	-0.211***	-0.154***	-0.139***	-0.159***	-0.537***
	(0.014)	*(0.006)*	*(0.006)*	*(0.007)*	*(0.008)*	*(0.023)*	*(0.200)*
Recent Downgrade	0.378***	0.270***	0.071*	0.153***	0.340***	0.132	2.433***
	(0.110)	*(0.060)*	*(0.043)*	*(0.045)*	*(0.045)*	*(0.082)*	*(0.630)*
STA	-1.78***	-1.573***	-1.708***	-0.889***	-0.526***	-0.107	1.787***
	(0.091)	*(0.043)*	*(0.035)*	*(0.042)*	*(0.046)*	*(0.107)*	*(0.587)*
POS	--	-2.949***	-2.217***	-1.632***	-1.021***	-0.326	--
	--	*(0.183)*	*(0.103)*	*(0.085)*	*(0.087)*	*(0.211)*	--
NEG	-0.415***	-0.289***	-0.128***	-0.041	0.002	0.110	1.974***
	(0.109)	*(0.050)*	(0.039)	*(0.050)*	*(0.055)*	*(0.106)*	*(0.506)*
Cons	-0.230***	-0.907***	-1.186***	-1.379***	-1.729***	-2.541***	-7.166***
	(0.084)	*(0.039)*	*(0.031)*	*(0.039)*	*(0.047)*	*(0.172)*	*(0.896)*
N	7,163	37,601	69,479	32,972	26,394	5,296	732
LR X^2	678.11***	3,320.10***	7,536.98***	3,458.40***	2,886.96***	890.06***	95.99***
Pseudo R^2	0.1033	0.1228	0.1693	0.1218	0.1244	0.1621	0.2490
AR	41.72%	50.60%	59.24%	48.50%	48.48%	54.80%	72.04%

*Table A3 reports the model coefficients for the DP model estimated by rating category for corporate firms rated between Aa and Ca-C.

7 Glossary of Terms

DP – Deterioration Probability; a model driven estimate of the probability of a downgrade event for rated firms, and the probability of a similar deterioration of credit quality for unrated firms

Dummy variable – a variable that only takes on values of zero or one; used to represent the presence/absence of a certain attribute; also known as a binary variable or indicator variable.

EDF™ – Expected Default Frequency; model driven estimate of the likelihood that a particular firm will default in a certain time frame (typically one year).

EWTK – Early Warning Toolkit; a set of EDF-based metrics designed to help users detect problem names in their portfolios. Also used in the construction of the Deterioration Probability model.

IRG – the Implied Rating Gap; computed as a firm's Moody's rating minus its Market Implied Rating, where ratings are converted to numerical values on a scale ranging from 1 to 21 for the purpose of calculation.

MIR – Market Implied Rating; a market-driven measure of a firm's credit risk, which is mapped to a Moody's rating. The three types of MIR model are EDF-Implied Ratings, CDS-Implied Ratings, and Bond-Implied Ratings.

MIS – Moody's Investors Service; one of the primary U.S. rating agencies.

Multicollinearity – in a regression model, when the values of a covariate can be largely explained by the remaining covariates.

Nonlinear – not linear; in context, not constantly increasing or decreasing based on changes in model inputs.

Out-of-sample testing – a method of evaluating the performance of a model by estimating model coefficients on a portion of the data and evaluating it on the remaining part of the data

Univariate model – a model that has only one explanatory variable

Variance Inflation Factor – a quantitative measure of the severity of multicollinearity in a regression model; it is defined as $1/(1-R^2)$, where the R^2 comes from the regression of the variable in question on the remaining explanatory variables; values of 10 and above are often regarded as indicating the presence of a high level of multicollinearity.

8 References

Bloomberg, L.P., "Company Overview of Abengoa, S.A.", 17 Nov 2017.

Bodard, Eric de and Wolfgang Draack, "Moody's assigns Ba3 CFR to Abengoa (Spain); outlook stable", Moody's Investor's Service, 13 Sep 2010.

Chung, Ivan and Gary Lau, "Moody's assigns first-time Baa3 ratings to Yanzhou Coal", Moody's Investor's Service, 27 Apr 2012.

Ferry, Danielle and Irina Baron. "Using EDF Measures to Identify At-Risk Names – A Monitoring & Early Warning Toolkit", Moody's Analytics White Paper, 2016.

French, Gretchen and Steven Wood, "Moody's Downgrades Linn Energy's Corporate Family Rating to Ca", Moody's investors Service, 16 Mar 2016.

French, Gretchen and Steven Wood, "Moody's downgrades Linn Energy's PDR to D-PD on bankruptcy filing", Moody's Investors Service, 12 May 2016.

Hamilton, David and Richard Cantor. "Rating Transitions and Defaults Conditional on Watchlist, Outlook and Rating History", Moody's Investor Service, February 2004.

Heck, Matthias and Matthias Hellstern, "Moody's downgrades Abengoa to Caa2; outlook negative", Moody's Investors Service, 26 Nov 2016.

Heck, Matthias, Matthias Hellstern, and Li, W., "Insights Into Abengoa's Default", Moody's Investor's Service, 17 Mar 2016.

Jeansonne, Clay. "LINN Energy Founder Michael C. Linn Retiring", NASDAQ Global NewsWire, 01 Dec 2011.

Kim, Franklin and David Munves. Moody's Analytics Viewpoints: "Moody's Market Implied Ratings: Description, Methodology, and Analytical Applications", July 2016.

LINN Energy Company Website, http://www.linnenergy.com/operations/

Metz, Albert and Richard Cantor. *Introducing Moody's Credit Transition Model*. Moody's Investor Service, 2007.

Murray, Lisa, "China's coal industry wrestles with overcapacity, pollution, and a restive public", Australian Financial Review, 27 Nov 2016.

Nazeran, Pooya and Douglas Dwyer (2015), "Credit Risk Modeling of Public Firms: EDF9", Modeling Methodology, Moody's Analytics.

Quandl Financial, Economic, & Alternative Data (2017), "US Coal Prices by Region", Retrieved 3 Aug 2017 from Quandl database.

U.S. Energy Information Administration (2017), "Crude Oil Prices: West Texas Intermediate", Retrieved Aug. 4, 2017 from EIA database.

Wang, Yikai, Min Ding, Jason Pan and Samuel W. Malone (2017). "Credit Transition Model 2017 Update: Methodology and Performance Review," Moody's Analytics.

"Yanzhou: Yancoal Australia still not breaking even", Mining Acquisitions, 3 May 2017, Retrieved Aug. 2, 2017 from https://miningacquisitions.com/2017/05/03/yanzhou-yancoal-australia-still-not-breaking-even/.