Predicting Rating Transitions using Machine Learning

VINAYAK KUMAR, KAY GIESECKE[1], PAUL KAMENSKI, STAS MELNIKOV[2], SATHISH UMAPATHY[3]

Abstract

Using a mixture of advanced data analytics derived from the explosion of information from quant credit markets and machine learning techniques, this paper develops a systematic framework to predict ratings migration in corporate credit. By systematically identifying firms with elevated possibility of downgrade or upgrade, the model supports more proactive portfolio management and better capital allocation. These insights are especially critical for financial institution investors like insurers, who are highly sensitive to credit ratings of securities on balance sheet. The ability to forecast rating transitions consistently and accurately will enhance insurer risk management and bottom-up credit selection, thereby ensuring a more stable and robust financial intermediation process.

[1] Kay Giesecke is Professor of Management Science & Engineering at Stanford University. He is the Founder and Director of Stanford's Advanced Financial Technologies Laboratory, the Director of the Mathematical and Computational Finance Program and a member of the Institute for Computational and Mathematical Engineering.

[2] Stas Melnikov is Head of Quantitative Research and Risk Data Solutions at SAS.

[3] Sathish Umapathy is Portfolio Manager at Pension Insurance Corporation.

The other authors are employees of Man Group.

Introduction

The rapid evolution of quantitative credit promises to revolutionize the asset class in the same way that quant equity changed the shape of equity investment. With the advent of widespread electronic trading has come a corresponding explosion in the availability of associated data. This has led to a rich and fascinating new source of material to feed research projects. We wanted to see whether there was information within this new universe of data that might help us to predict ratings migration with greater accuracy.

Credit ratings play a foundational role in global capital markets. They provide standardized assessments of creditworthiness underpinning investment decisions, regulatory frameworks, and capital allocation. Early prediction of rating transitions could be of benefit to a variety of institutional stakeholders. Active credit managers could generate signals in an early-warning or early-opportunity system for credit rating changes. Insurers, as significant investors in credit-sensitive assets, are heavily exposed to the risk of downgrades from a regulatory capital point of view. By anticipating credit deterioration, predictive models can enhance internal ratings in risk-based regulatory capital regimes such as Solvency UK/II.

This paper develops a systematic framework for predicting ratings changes using a combination of fundamental and market-based signals. Credit rating transition modeling is a challenging research problem due to the relatively limited historical data, non-linear relationships, and multiple key drivers across the time-series and cross-section. To address this, we combine a diverse array of data sources with statistical and machine-learning techniques.

Our work contributes to the existing literature in several ways. First, we model a novel panel of bond-specific and equity-specific features leveraging high-quality measures from industry data providers. Secondly, we contribute to the growing literature on machine learning applied to ratings transition modeling, highlighting substantial and intuitive non-linearities emerging from our model. Finally, we demonstrate a substantial empirical improvement over existing approaches – our model produces AUCs as high as 90% out-of-sample, significantly outperforming existing benchmarks in prior literature.

Literature Review

The existing literature can broadly be divided into two strands depending on the approach to credit rating forecasting taken. The first approach involves modeling the contemporaneous credit rating of an issuer as a function of covariates - the independent variables or predictors used in the model, typically with the aim of predicting credit ratings in a hold-out sample.

So-called ‘static’ credit rating modeling starts from Kaplan and Urwitz (1979), who deployed multiple regression models to explain bond ratings using financial ratios. Early work compared various classification techniques: for instance, Ederington (1985) evaluated linear discriminant analysis, logistic regression, and probit models (which are similar to logistic regression but assume the probabilities follow a cumulative normal distribution rather than a logistic distribution), finding that logistic and probit approaches outperformed linear models in predicting ratings. It is worth noting that static credit rating models can still be considered forecasting, because contemporaneous predictions can be viewed as an empirical view of what the ‘true’ credit rating is of an issuer. Discrepancies between actual credit ratings and model ratings could be expected to converge if the model is accurate.

By the 1990s and early 2000s, ordered categorical models (such as ordered probit) became standard for capturing the ordinal nature of ratings. Studies like Blume, Lim, and MacKinlay (1998) and Amato and Furfine (2004) employed these methods and confirmed that key financial measures such as firm size and leverage were also predictive. Researchers have expanded rating models since by incorporating additional factors beyond financial ratios. Ashbaugh-Skaife, Collins, and LaFond (2006), for example, show that firms with stronger corporate governance practices receive higher credit ratings, highlighting the importance of qualitative factors. In parallel, more advanced techniques from machine learning have been applied to rating prediction: Dutta and Shekhar (1988) demonstrated an early use of neural networks to predict bond ratings, and Huang et al. (2004) introduced support vector machines. More recently, Tavakoli et al. (2025) deployed ML on unstructured or multimodal datasets, employing various combinations of fusion strategies with selected deep-learning models, including convolutional neural networks (CNNs) and variants of recurrent neural networks (RNNs), and pre-trained language models (BERT). Their work indicated text data was more useful than numeric data in predicting credit ratings.

The second approach to credit rating forecasting involves explicitly modeling the transition probabilities between credit ratings, rather than attempting to explain current ratings. In our view, this is a more effective way of forecasting ratings changes, as it isolates features and relationships which correspond directly with transitions.

Early models of rating transition relied on historical transition matrices, assuming constant Markovian probabilities – that is to say that they were memoryless, assuming that the likelihood of transitioning from one credit rating to another depends only on the current rating and not on the past ratings or the path taken to get there. Altman and Kao (1992) documented "rating drift," where downgrades tend to cluster, challenging the memoryless Markov assumption. Empirical evidence shows transition probabilities are time-varying and cyclical (Bangia et al. 2002), worsening in downturns. Discrete-time models like dynamic ordered probit/logit have also gained popularity. Blume, Lim, and MacKinlay (1998) used ordered models to study shifting rating standards. Mizen and Tsoukas (2012) found that including lagged ratings and initial rating improves downgrade prediction, confirming the presence of rating momentum. Feng, Gourieroux, and Jasiak (2008) developed an ordered qualitative model capturing these transitions empirically.

To address hidden regime shifts, researchers introduced Hidden Markov Models to identify the latent credit process. Jarrow, Lando, and Turnbull (1997) proposed a reduced-form Markov intensity model that incorporates both default and rating migration risk into the pricing of credit spreads. Their model allows for arbitrage-free pricing of risky bonds and CDS instruments, and remains widely cited as the basis for dynamic credit term structure models. Nickell, Perraudin, and Varotto (2000) tested the stability of historical rating transition matrices and found strong evidence of time variation, especially in downgrade rates during recessions. This challenges the use of unconditional average matrices in portfolio models and supports time-varying or regime-based transition modeling. Lando and Skødeberg (2002) showed that path and duration dependencies significantly affect migration probabilities. Korolkiewicz and Elliott (2008) modeled rating transitions as noisy signals of unobserved credit states. These models infer latent regimes such as "crisis" or "expansion," aligning predicted transitions with macro cycles. Figlewski, Frydman, and Liang (2012) explicitly modeled how macroeconomic variables like GDP growth, industrial production, and interest rates affect both default probabilities and rating migrations. Their model links firm-level credit events to systemic risk channels, confirming that macro variables significantly improve predictive accuracy beyond firm fundamentals alone.

Finally, modern machine learning has started to emerge as a method of choice. Afonso, Gomes, and Rother (AGR 2011) conducted an extensive comparison of binary classifiers (logit, decision trees, neural nets, SVMs) in predicting upgrades and downgrades. In juxtaposition to our work, AGR focus on predicting the direction of rating transition *conditional* on a transition having already occurred. We believe our formulation more accurately represents the situation of a real-time forecaster, who must determine both the likelihood of a rating transition as well as direction. Our approach differs from the body of existing work by using machine learning techniques, coupled with a unique dataset of firm and bond-specific covariates to develop practical, forward-looking probability forecasts.

Data

We analyze credit rating transitions between 2001 and 2024, focusing on corporate bonds covered by the IUC0 IceBofA Index[4]. IUC0 tracks over 12,000 issues and 2000 issuers, focusing primarily on corporate bonds denominated in USD. Our database covers 8,792 unique bond issuers and over 507,527 issuer-month samples.

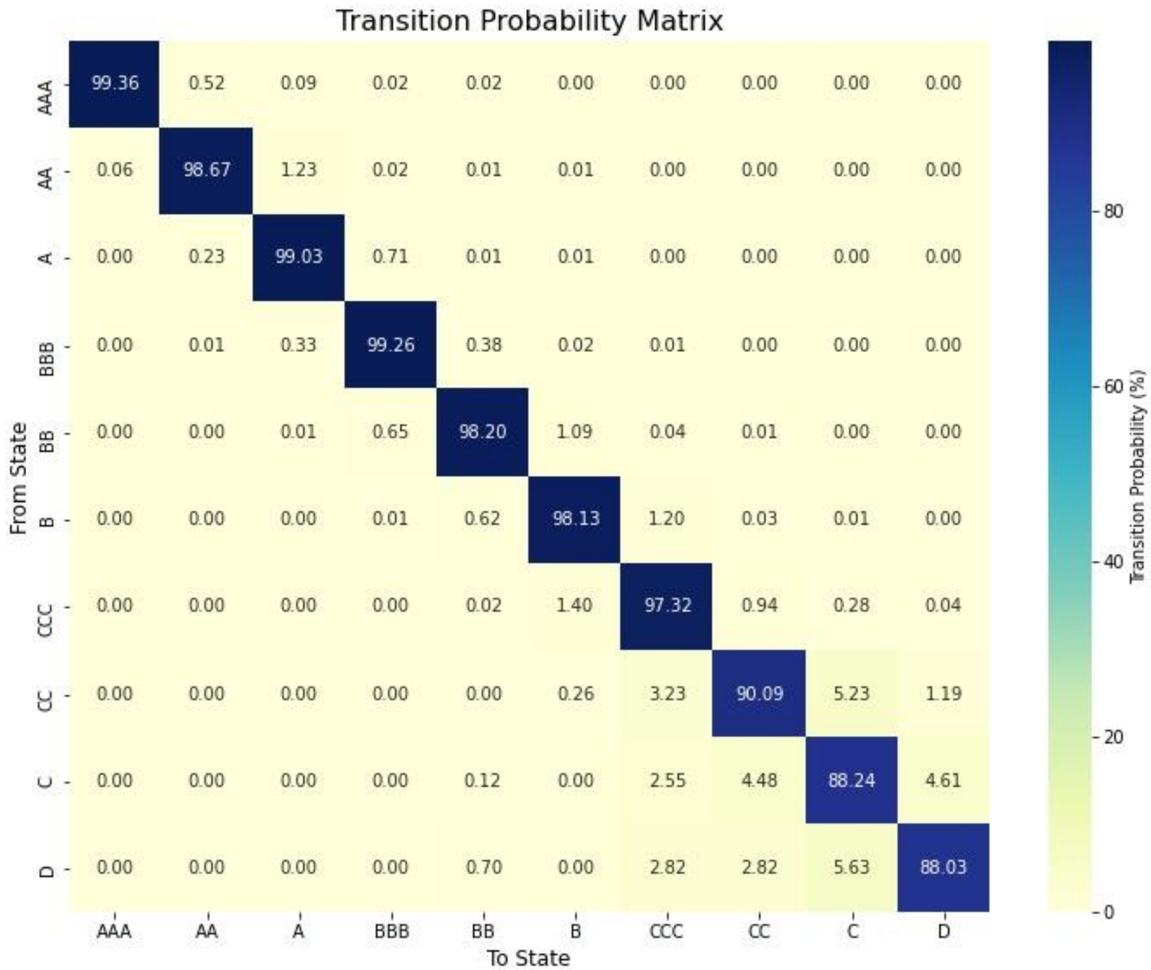
We aggregate issue-level ratings to obtain a single issuer-level rating.[5] We focus on *bucket-level transitions*, subsequently aggregating ratings to ten buckets {AAA, AA, A, ..., CC, C, D}, where D represents default. Transitions are modeled at a monthly frequency.

The figure below shows the empirical transition probability matrix obtained from our sample. A cell gives the likelihood (in percent) of a transition from a given bucket (row) to another bucket (column) over one month. Thus, the values in a row sum to one. These likelihoods are averages across issuers and time. They provide model-free estimates of transition probabilities for an issuer with a given rating.

---

[4] Source: ICE Data Indices, Man Group. Relevant ICE Data Index is a product of ICE Data Indices, LLC and is used with permission. ICE® is a registered trademark of ICE Data Indices, LLC or its affiliates. The index data referenced herein is the property of ICE Data Indices, LLC, its affiliates (“ICE Data”) and/or its third party suppliers and has been licensed for use by Man Group. ICE Data and its Third Party Suppliers accept no liability in connection with the use of such index data or marks.

[5] Face-value weighting balances stability with liquidity considerations in determining aggregate issuer ratings.



We highlight several observations. Across all rating buckets, the most likely outcome is no change; the probabilities in the diagonal of the matrix are the largest in every row. For example, the probability of an AA-rated issuer maintaining its rating over the next month is 98.67%. Upgrades and downgrades to adjacent buckets (e.g. from AA to A) are less likely for IG rated issuers than for HY issuers. Multi-bucket transitions (e.g. from AA to BBB) are exceedingly rare, with frequencies being typically (much) lower than 10 basis points. The only exceptions are HY issuers rated CCC and below, for which multi-bucket transitions are in the order of a few percent. Transitions to default by firms rated CCC and above are very rare.

To reduce the dimensionality of the ratings prediction problem, we focus on four primary transitions from any given rating bucket: No change, upgrade, downgrade, and default. Default is treated as an absorbing state: once an issuer enters that state we stop tracking it. Upgrades, downgrades, and defaults cover multi-bucket transitions along with the more common single-bucket transitions. Modelled probabilities for these four transitions can be used to accurately approximate the full 10x10 bucket-to-bucket transition probability matrix.

Following this labeling approach, we obtain 499,978 issuer-month samples of no change, 4,583 downgrade samples, 2,867 upgrade samples, and just 99 default samples. The transition sample is highly unbalanced: the vast majority of the samples deliver the same outcome (no change). The default count is lower than in alternative issuer universes over the same period (e.g. Moody's rated bonds). This is due to selection bias: the IUC0 Index representing our universe focuses on performant issues subject to index eligibility criteria – i.e. they have already cleared several hurdles in order to be included within the index. The selection bias reflects the institutional reality for market participants who use these benchmarks to obtain accurate pricing and security characteristics. As a result, we focus less on the default probability component of our classification system and primarily examine upgrades and downgrades.

Our goal is to predict conditional transition probabilities over a future period given a vector of features. We construct 48 variables representing several different categories: (1) Equity-related variables such as momentum from sources including Barra and Man; (2) bond variables such as yield and option-adjusted spread from ICE, (3) ratings variables such as the current issuer-level rating and the split-rated-ness of the issuer, (4) one-year default probabilities from SAS/Kamakura and (5) macro-economic variables such as interest rates from Bloomberg. Table 1 in the Appendix provides a complete list of these variables.

All features except the macroeconomic variables are issuer-specific – that is to say, that we seek to distinguish between systemic and idiosyncratic factors. The modelled dependence of conditional transition probabilities on variables common to all firms captures the correlation of transitions. For example, low GDP growth tends to increase the probability of downgrade across the board, implying downgrades correlated across the issuer universe.

Features are time-stamped and matched with the transition samples. The variables are updated at different frequencies (e.g. daily, monthly); we use the most recent values available at the beginning of a month. Specifically, for issuer i, we denote by X_{it} the vector of feature variables evaluated at the beginning of month t.

About 50% of the samples have missing feature values. There are some broad missingness patterns that we can identify. For example, private (non-listed) firms lack the complete set of equity-related variables (e.g. momentum). We treat missing values by including in X_{it}: a missingness dummy as well as a numeric variable measuring the percentage of missing values for a sample. A missing feature value is imputed by replacing it with the crosssectional median of the observed feature values at the relevant period. With the missing dummy and missing fraction variables, our model can discern samples with imputed feature values from samples with fully populated values. This approach enables us to harness all 500,000+ transition samples for model construction and testing.

Transition Probability Model

The observed data consist of issuer-month transition-feature pairs$(\mathrm{Y}_{it}, \mathrm{X}_{it})$ where the "label" Y_{it} represents one of the four possible transitions we track (no-change, upgrade, downgrade, default) observed for issuer i during month t, and X_{it} is the associated feature vector. For convenience, the transitions are enumerated such that $\mathrm{Y}_{it} \in \{1,2,3,4\}$. We will use a subset of the observed data to estimate the conditional probabilities $P_t(Y_{it} = k)$ for transitions k=1,...,4 during month t, given all information available at the beginning of t. To this end, we assume that

$$\mathrm{P}_t(\mathrm{Y}_{it} = \mathrm{k}) \ = \ \mathrm{f}_\theta\left(\mathrm{k}, \mathrm{X}_{\{\mathrm{it}\}}\right)$$

Where $\mathrm{f}_\theta$is a transition function (model) to be specified, θ is a vector of parameters to be estimated, and k=1,...,4. Note we do not report statistics on the default class as the sampling is too small to conduct tests.

f_θ does not depend on time or issuer; it represents the transition behavior across issuers and time. The feature vector X_{it}creates an issuer- and time-dependent transition probability.

Although our time horizon is a single month, we can generate transition probability estimates for any multi-month horizon by stacking single-month predictions. For example, the probability of no rating change over two months is the product of the conditional probability of no change over the next month and the conditional probability of no change over the second month. The first probability can be directly evaluated. To evaluate the second probability, we need to evolve the covariate X_{it} forward to the beginning of the second month. The simplest approach would be to freeze X_{it} at the most recent value. Another approach is to formulate hypothetical scenarios for selected elements of X_{it}, such as key macro-economic variables. We leave the exploration of long-horizon ratings prediction for future research.

We consider alternative specifications of the transition function f_θ, including a linear logistic regression baseline and machine learning (ML) classifiers such as a random forest (an ensemble learning method that combines multiple decision trees) and gradient boosting (another ensemble method that builds models sequentially, optimizing performance by reducing errors from previous iterations). We also consider a naïve benchmark leveraging the in-sample transition matrix to construct out-of-sample forecasts solely according to ratings class.

The ML models have the capacity to capture complex non-linearities, including interactions between variables, in the transition data. This, in turn, can enable more accurate transition forecasts. We choose to focus on random forests and gradient boosting because they are easier to train and do not require the normalization of the feature variables when compared to alternative techniques such as neural networks.

Model Training

A model f_θ is trained by minimizing a cross-entropy error objective (a loss function that measures the difference between the model's predicted probabilities and the actual observed outcomes) over the parameter set θ given a set of training pairs (Y_{it}, X_{it}). The optimization can entail a regularization term to guard against overfitting the training data at the expense of generalizability (that is, reduced out-of-sample predictive accuracy).

The training set is often partitioned into two subsets, the smaller of which (the validation set) is used to tune any hyper-parameters associated with f_θ or the optimization (training) method. A hyper-parameter can, for example, specify the complexity of an ML model or the weight of a regularization term.

Unless noted otherwise, we use pairs (Y_{it}, X_{it}) observed up to 12/2014 (about 50% of the entire sample) for model training and validation, and pairs between 1/2015 and 12/2024 for out-of-sample testing of model predictive performance.

One issue we face is the severe imbalance of the transition data set: the vast majority of the samples represents no-change transitions. This renders model training challenging, especially for the minority outcomes (upgrades, downgrades, defaults). A standard approach to mitigate this issue is under-sampling the majority class and/or over-sampling the minority classes for training. This approach is equivalent to re-weighting the samples in the objective function. It introduces a bias into predicted probabilities that we will need to remove later.

Model Evaluation – Out-of-sample predictive performance

A fitted model $\hat{f}_{\hat{\theta}}$, where $\hat{\theta}$ minimize the training objective, delivers predicted one-month ahead transition probabilities for an issuer i with feature vector X_{it} via (*). We measure the accuracy of these predicted probabilities on the test set using two approaches.

The first is based on the standard tool of Receiver Operator Characteristic (ROC) curves, and more specifically, the Area Under the Curve (AUC). The AUC measures a model's ability to correctly *rank firms* according to the predicted likelihood of a transition. For example, for downgrade, the AUC is the probability that a model forecasts a higher likelihood of downgrade for a randomly chosen downgrade sample than a randomly chosen upgrade/unchanged/default sample. An AUC of 1 represents a crystal ball model and an AUC of ½ a useless coin flip model. Note that the under/over-sampling of the training samples does not affect the AUC.

Class	Transition Matrix AUC	Logistic Regression AUC	Random Forest AUC	Gradient Boosting AUC
Unchanged	0.63	0.70	0.72	0.81
Upgrade	0.56	0.65	0.70	0.88
Downgrade	0.54	0.77	0.80	0.90

Exhibit 1: Model AUCs

We document substantial AUCs for our feature set. The non-linear versions of the model generate exceptional results over our sample, with AUCs in the high-80% over the OOS period. In addition to examining ROC AUCs, we also examine the precision-recall tradeoff of our model. Below, we plot the precision-recall curve and ROC curve for the gradient boosting model, focusing on the downgrade class for illustration.

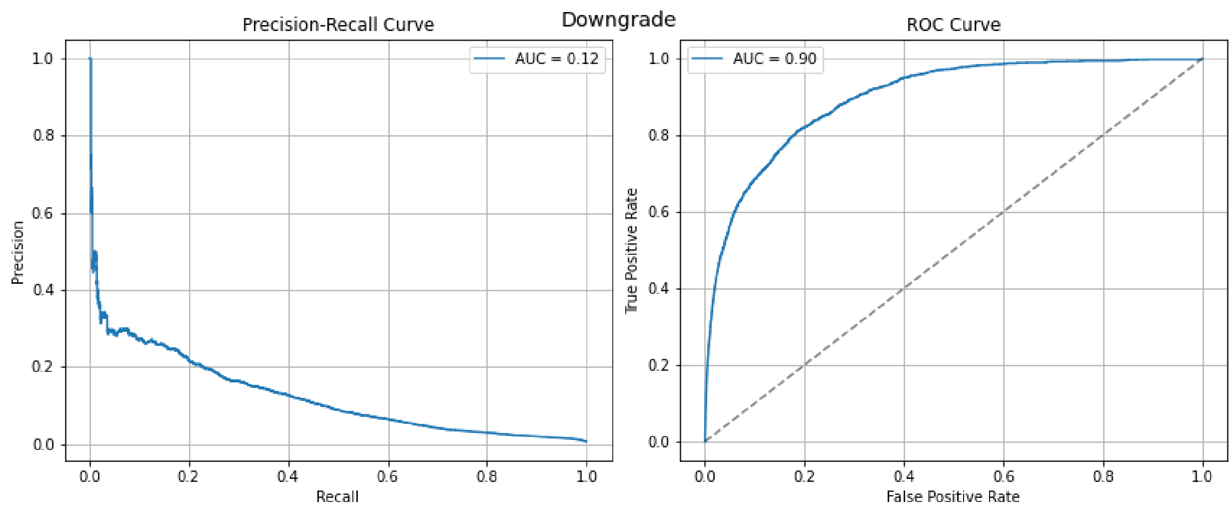


Exhibit 2: Precision-Recall and ROC Curve for Gradient Boosting Model

As a reminder, our panel is significantly imbalanced with 0.6% of the full sample representing upgrades and 0.9% of the sample representing downgrades. In the out-of-sample period, the proportions are even smaller at 0.7% and 0.5% for downgrades and upgrades, respectively. The baseline PR-AUC associated with a random classifier corresponds to the positive-class proportion of the dataset. The OOS downgrade PR-AUC of 12% thus represents a 20x improvement over random classification, which is substantial in light of the class imbalance. Besides PR-AUC, we also seek high precision at a reasonable level of recall for risk management applications. In this respect, our model achieves ~20% precision at 20% recall, representing a 28-fold improvement over random classification. For upgrades, we achieve a precision of ~5% at 20% re-call, which is 10X improvement over random classifier. As an additional benchmark, the PR-AUC for a naïve transition matrix generates 0.6% PR-AUC for upgrades and 0.8% PR-AUC for downgrades, which are only slightly above the baseline proportions of downgrades and upgrades.

Our downgrade prediction results are particularly strong, while our upgrade performance is also robust. This pattern reflects real-world behavior: downgrades tend to be triggered by more abrupt and observable financial deterioration, resulting in clearer model signals. Upgrades, on the other hand, occur more slowly and are subject to greater conservatism by rating agencies. In our sample, we observe twice as many downgrades as upgrades, leading to a higher signal-to-noise ratio for downgrade detection. As such, the relative difference in performance is consistent with prior findings in the literature.

Our second approach to measuring out-of-sample predictive performance entails comparing forecasted transition probabilities with realized transition rates. The predicted probabilities are biased due to the under/over-sampling of the samples for model training. We need to remove this bias by calibrating the predicted probabilities to actual transition rates. To perform this calibration, we take pre-2015 as the in-sample period and construct 10 equally distributed bins mapping predicted probabilities to the average transition rate within each bin. The bin thresholds and values are then used out-of-sample to compute aggregate forecasted transition rates over time, plotted below for downgrades.

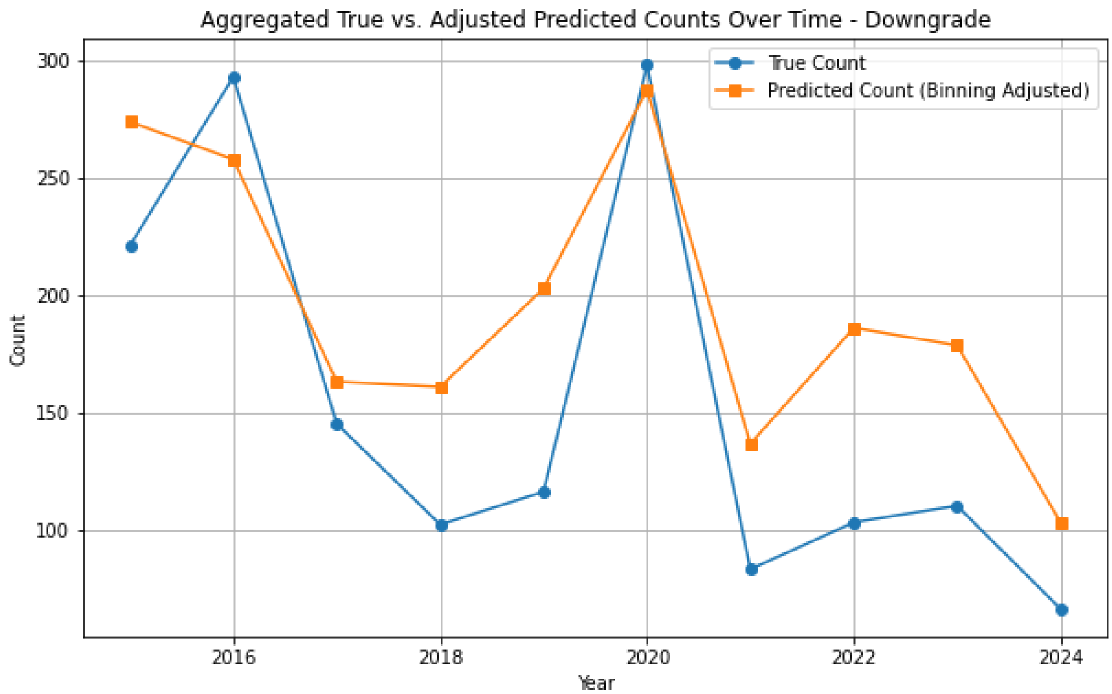


Exhibit 3: Forecasted vs. Actual Aggregate Counts for Downgrades

Note this procedure can be generalized by refining the calibration function $g(\hat{f}, X_{it})$ to better match physical transition rates. Our current approach can be viewed as having a naïve, non-parametric calibration which could easily be extended by incorporating macro variables or other covariates. It is encouraging that we see close aggregate predicted counts with a simple procedure.

Robustness Analysis

We perform several robustness checks:

1. To test for potential leakage from the feature variables causing look-ahead bias, we construct models with lagged feature values (3- and 6-month lags). The out-of-sample AUCs drop a few percent for each transition but not excessively. There is no evidence of leakage.
2. We investigate whether a global model trained on the full-sample is more optimal than models trained within separate sub-universes. As mentioned above, about half of the sample has missing feature values. We partition the data according to missingness patterns and train alternative models on these partitions. Zooming into appropriate partitions of the test set, we compare sub-model performance with full model performance over the training sample. We find the full model performs at least as well as the alternative sub-models and, in some cases, significantly better. This implies the ability of ML models to learn patterns in the transition data even if some features are missing.
3. Finally, we have also constructed and tested models on different train/test partitions of the data set. The main empirical results are qualitatively similar to those specified here.

The AUC results as a function of monthly lag are tabulated in the appendix. We are happy to provide the other results upon request.

Hidden Insights

First – we highlight feature importances derived from our XGBoost model. Feature importance in XGBoost quantifies the relative contribution of each predictor to the model's predictive performance. This measures the improvement in the model's loss function from splits involving a given feature, aggregated across all trees.

Feature	Importance
OAS	0.19
Yield-to-Maturity	0.17
1-year KDP	0.08
9-month Equity Momentum	0.07
Barra Momentum	0.07
Issuer Rating	0.06
12 Month Equity Momentum	0.06
6 Month Equity Momentum	0.04
Barra Spec Risk	0.03
Issue Market Value	0.03
Issuer Market Value (Bond)	0.02
3 Month Equity Momentum	0.02
Split-rating	0.02
Barra Volatility	0.02
Barra Residual Volatility	0.02

Exhibit 4: Feature Importances

The top two features are bond-market features – the option-adjusted spread and yield-to-maturity. The dominance of these features is a well-known result consistent with prior understanding: the bond market tends to price ratings changes well before they occur, implying market-based indicators of credit risk should be highly useful. Another notable cluster of features is equity market-related, as we see the full term-structure of equity momentum across horizons rank high in importance. Lastly, in terms of fundamental indicators of credit risk, we see the SAS Kamakura default probability and current issuer-level rating as critical covariates.

Notably, these features drive rating transitions in a non-linear way; the empirical AUC results indicate non-linearity adds substantial value on top of a logistic-linear specification with the same underlying data. To further explore the structure of the non-linearity, we examine one and two-dimensional partial dependence plots as a useful way of visualizing and interpreting the models.

One-dimensional partial dependence plots (1D PDP) show the marginal effect of a single feature on our predicted transition probability. Two-dimensional partial dependence plots (2D PDP) visualize the marginal effect of two features on a model's predicted outcome, averaging over the distribution of all other features. The resulting surface highlights interactions and non-linearities by plotting predicted values across a grid of joint values for the selected feature pair. These tools aid in interpreting models such as gradient-boosted trees by illustrating how combinations of features influence model output.

Below, we plot 1D PDPs for Barra momentum and the SAS/Kamakura default probability (KDP). The patterns are consistent with intuition – downgrades tend to be more likely for issuers with higher default probability and with negative equity momentum. Additionally, univariate analysis alone indicates some substantial non-linearity in the relationships; downgrade probability is generally monotonic with both covariates, but the kinks and plateaus in the curve are learned by the model.

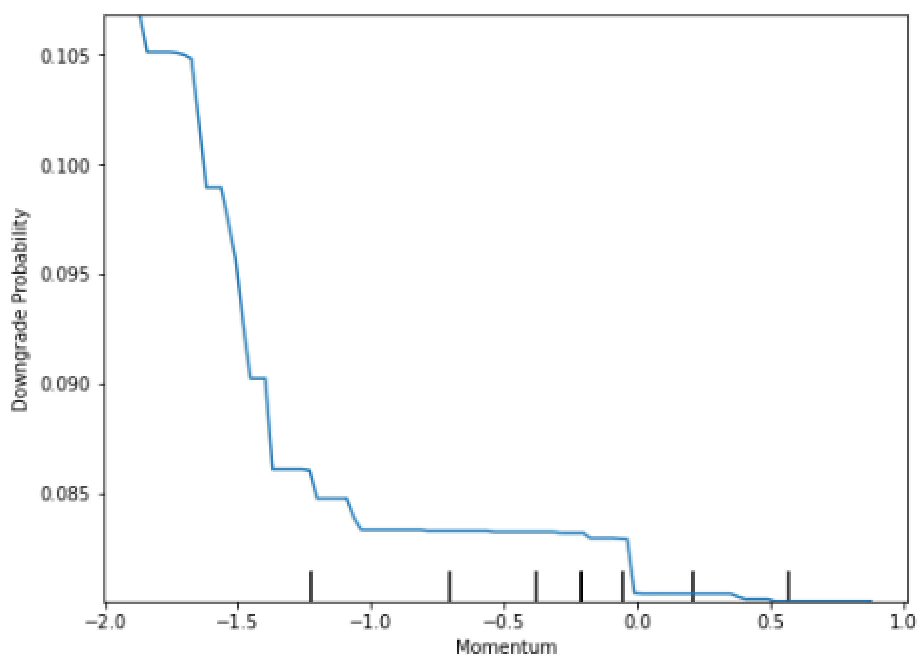


Exhibit 5: Momentum 1D PDP:

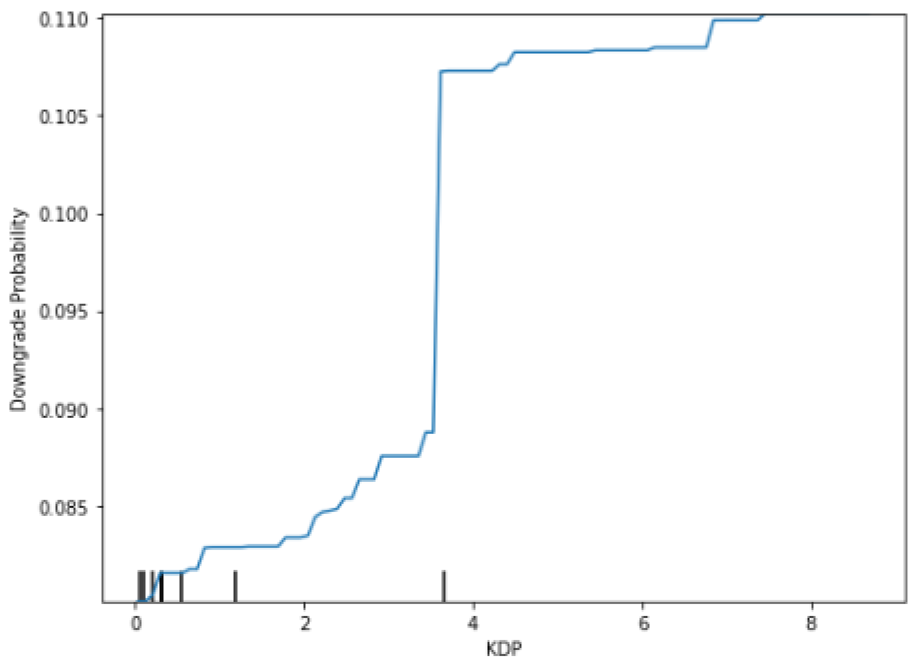


Exhibit 6: KDP 1D PDP

To deepen the analysis, we plot 2D PDPs of our downgrade probability model jointly against Kamakura default probability and momentum indicators. Consistent with the univariate interactions, the downgrade probability tends to increase with increasing KDP and with negative momentum. There is a substantial non-linearity detected in the interaction of the two covariates - the probability surface is steepest at the smallest and largest values of KDP and momentum, while flattening out in the middle region of the distribution.

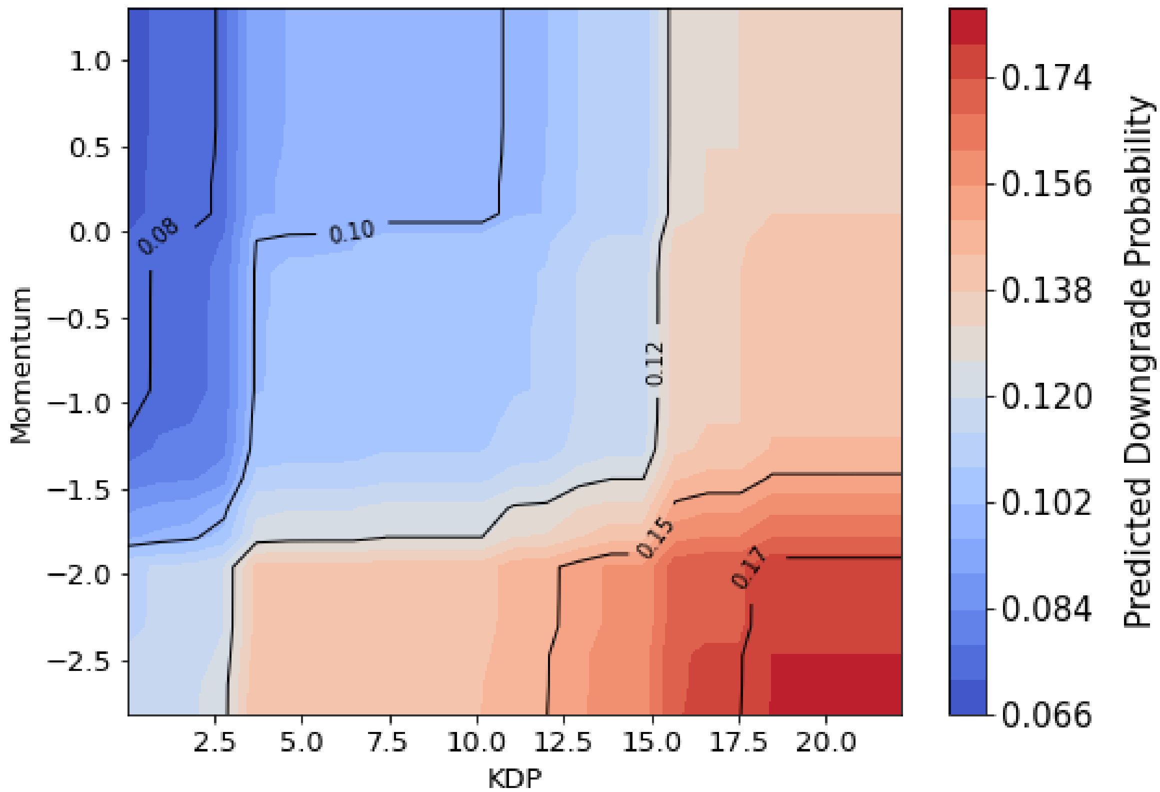


Exhibit 7: Momentum-KDP 2D PDP

We also examine the 2D PDP of KDP against the OAS of the issuer. Another intuitive pattern emerges – on a univariate basis, the bonds with widest spreads as priced by credit markets demonstrate the largest downgrade risk. However, the KDP adds orthogonal information – for any fixed OAS, a higher KDP drives a larger downgrade probability, indicating the value-add of KDPs above and beyond what is visible in market spreads.

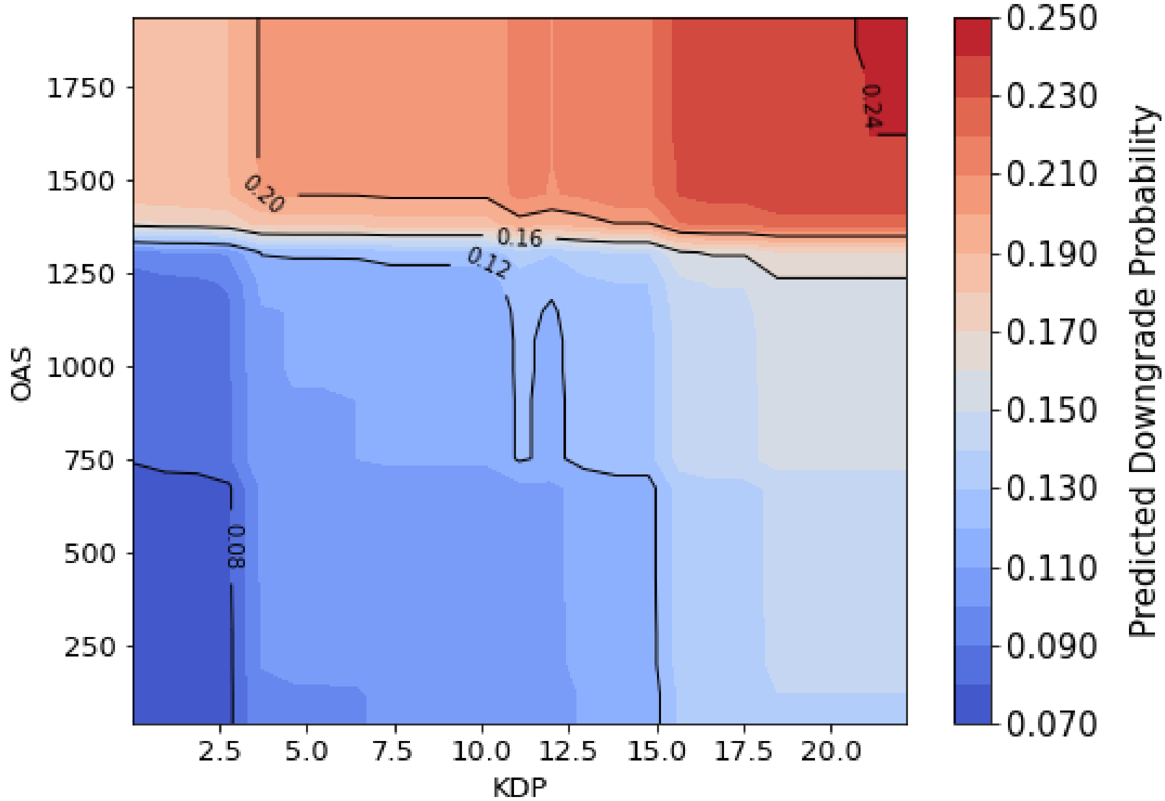


Exhibit 8: OAS-KDP 2D PDP

Beyond these graphics, we also include OAS-Momentum and Rating-KDP 2D PDP plots in the appendix. The interpretation is qualitatively similar – high downgrade probability firms have wide spreads and negative equity momentum. High downgrade probability firms also tend to exhibit the highest default probability within their rating class.

Conclusion

This study presents a practical framework for forecasting credit rating transitions using machine learning. By systematically identifying firms with elevated probabilities of downgrade or upgrade, the model supports more proactive portfolio management and better capital allocation. These insights are especially critical for financial institution investors like insurers, who are highly sensitive to credit ratings of securities on balance sheet. The ability to forecast rating transitions consistently and accurately will enhance insurer risk management and bottom-up credit selection. The aggregate impact of such improvements should support a more stable financial intermediation process. More generally, we expect the framework to be useful to any institution looking to enhance monitoring tools, manage tail risks, and refine bottom-up credit selection.

We also highlight novel non-linear interactions between bond and equity-market features in forecasting ratings transition. One area of future work involves exploring and enhancing long-horizon forecasts, possibly by integrating macroeconomic forecasts into the framework. Another area could involve integrating additional data sources, including the vast array of alternative datasets used in systematic investing processes. Lastly, further work should certainly focus on integrating such models into portfolio management systems or capital allocation frameworks.

The credit markets have historically been slower than equity markets to adopt technology. It is incumbent on us as market participants to use the full range of tools available to us to make the right decisions for our clients. We believe that the future of credit is only going in one direction: increasing sophistication, technological innovation, and systematization. Work such as this study will help the most evolved and forward-looking credit managers and allocators take proactive and informed decisions.

Appendix

Feature	Category	Description
Missing proportion	Missingness	Missingness indicator defined as the proportion of features that are missing pre-imputation in the sample.
BASPCAAA Index	Macro	US Corporate AAA 10 Yr Spread from Bloomberg.
VIX Index	Macro	Chicago Board Options Exchange Volatility Index from Bloomberg.
SPX Index	Macro	S&P 500 Index from Bloomberg
INJCJC Index	Macro	US Initial Jobless Claims Seasonally-adjusted from Bloomberg
CPI Index	Macro	US CPI Urban Consumers YoY NSA
CL1 Comdty	Macro	Generic 1st Crude Oil, WTI (NYM)
CO1 Comdty	Macro	Generic 1st Crude Oil, Brent (ICE)
USGG10YR Index	Macro	US Generic Govt 10 Yr from Bloomberg
USYC5Y30 Index	Macro	5s30s spread from Bloomberg
USYC2Y10 Index	Macro	2s10s spread from Bloomberg
USURTOT Index	Macro	US Unemployment Rate, Seasonally-adjusted from Bloomberg
USYC3M10 Index	Macro	3m10y spread from Bloomberg
1-year KDP	Fundamental Credit Risk	SAS Kamakura 1-year forward default probability of the issuer.
Issuer Rating	Fundamental Credit Risk	Issuer-aggregated credit rating.
Split-rating	Fundamental Credit Risk	Dummy variable defined as 1 if there is disagreement between S&P/Moody's/Fitch on credit rating.
Log Change KDP	Fundamental Credit Risk	Max{0, Log(KDP_{t}/KDP_{t-12})
9-month Equity Momentum	Equity Market Indicator	Point-to-point 9-month price return.
12 Month Equity Momentum	Equity Market Indicator	Point-to-point 12-month equity price return.
6 Month Equity Momentum	Equity Market Indicator	Point-to-point 6-month equity price return.
3 Month Equity Momentum	Equity Market Indicator	Point-to-point 3-month equity price return
1 Month Equity Momentum	Equity Market Indicator	Point-to-point 1-month equity price return.
Barra Earnings Variability	Equity Market Indicator	Proprietary MSCI Barra factor. See MSCI documentation for definition.[6]
OAS	Bond Market Indicator	Option-adjusted spread of the representative bond from ICEBofA.
Yield-to-Maturity	Bond Market Indicator	Yield-to-maturity of the representative bond from ICEBofA.
Issue Market Value	Bond Market Indicator	Issue market value of the representative bond from ICEBofA.
Issuer Market Value (Bond)	Bond Market Indicator	Summed market-value for all bonds within the issuer.
Spread Duration	Bond Market Indicator	Spread duration of the representative bond from ICEBofA.
Barra Momentum	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Spec Risk	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Volatility	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Residual Volatility	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Pbeta	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Gbeta	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Size	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Book-to-price	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Liquidity	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Long-term reversal	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra beta	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra earnings yield	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Dividend Yield	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Profitability	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Growth	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra leverage	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Investment Quality	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Mid-cap	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.
Barra Earnings Quality	Barra Risk Factor	Proprietary MSCI Barra factor. See MSCI documentation for definition.

Exhibit 9: Model Variables

Lag	Class	Logistic Regression AUC	Random Forest AUC	Gradient Boosting AUC
1	Unchanged	0.685	0.722	0.789
1	Upgrade	0.630	0.698	0.806
1	Downgrade	0.760	0.789	0.838
3	Unchanged	0.679	0.714	0.758
3	Upgrade	0.624	0.694	0.790
3	Downgrade	0.756	0.780	0.796
6	Unchanged	0.668	0.707	0.739
6	Upgrade	0.633	0.701	0.764
6	Downgrade	0.689	0.764	0.792

Exhibit 10: AUCs vs. Lag

Threshold	Upgrade Precision	Upgrade Recall	Upgrade TP	Upgrade FP	Upgrade FN	Downgrade Precision	Downgrade Recall	Downgrade TP	Downgrade FP	Downgrade FN
0.1	0.02	0.67	725	30756	360	0.04	0.72	1105	28087	432
0.15	0.03	0.5	541	16120	544	0.06	0.63	963	15492	574
0.2	0.04	0.37	399	9557	686	0.08	0.54	837	10149	700

Exhibit 11: Precision-Recall Statistics

[6] Barra, LLC's analytics and data (www.msci.com) were used in the preparation of this report. Copyright 2025 Barra, LLC. All Rights Reserved

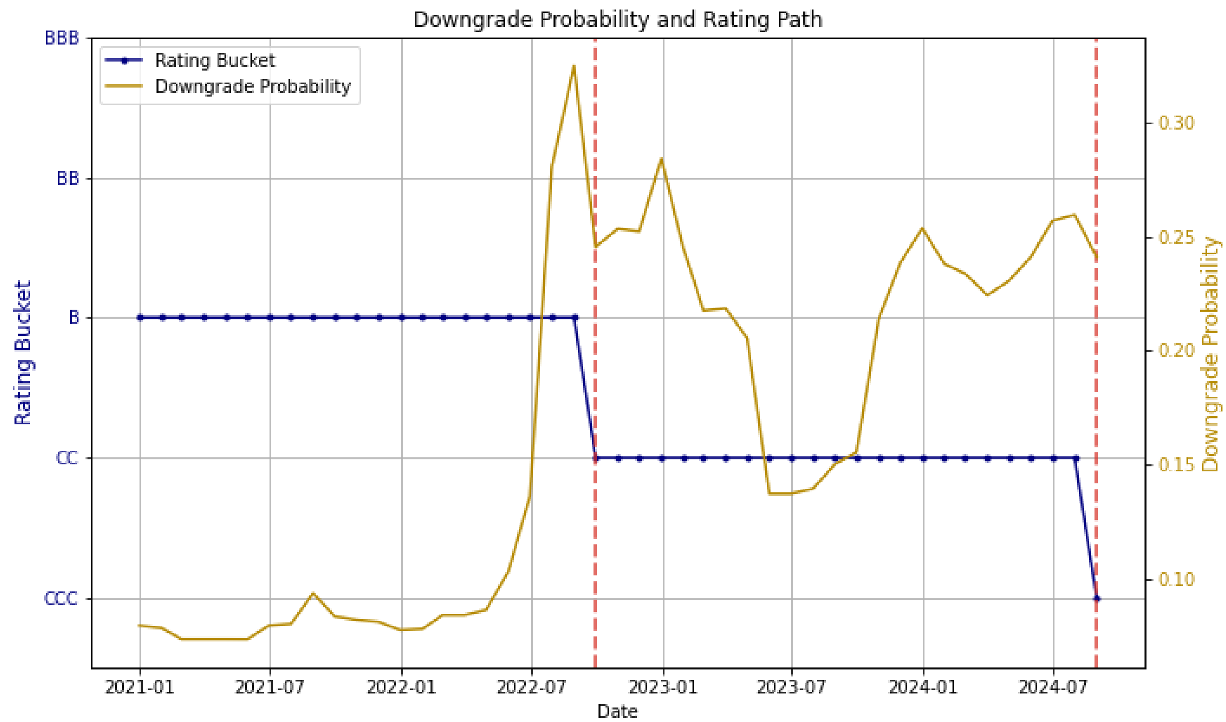


Exhibit 12: XGB downgrade probability and ratings trajectory for mid-sized issuer in financial services

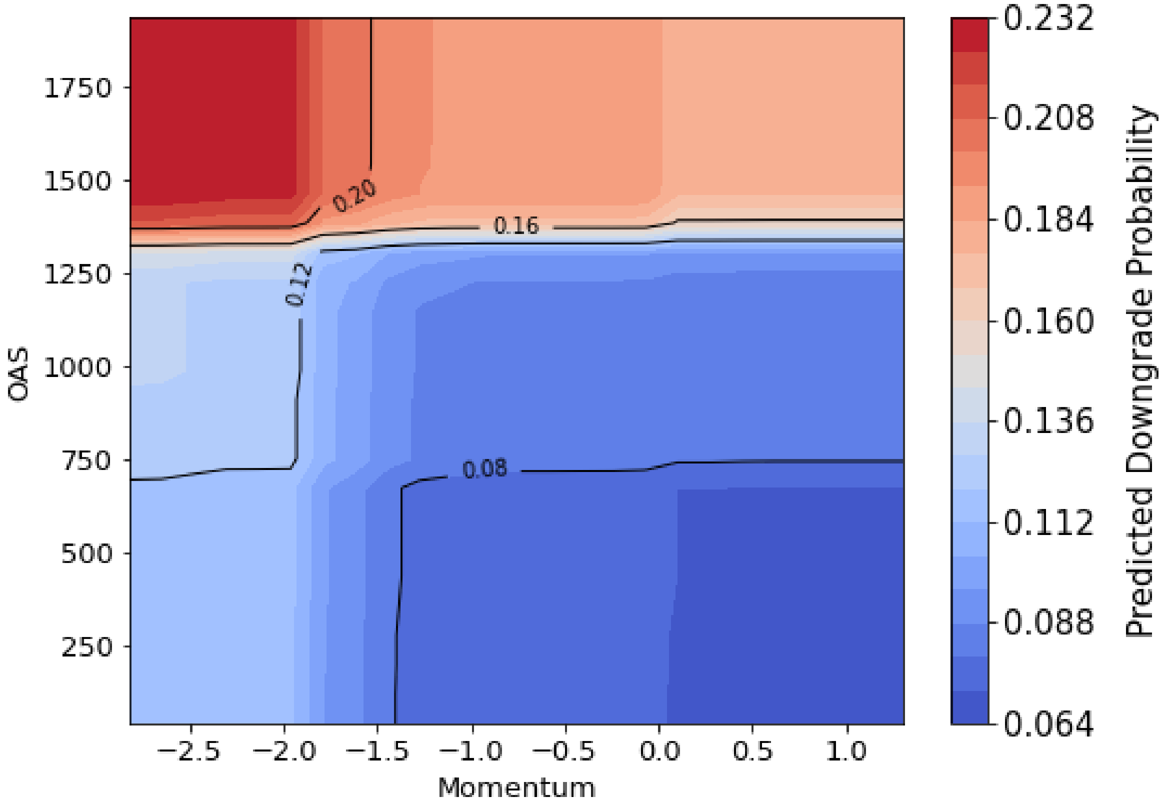


Exhibit 13: OAS-Momentum 2D PDP

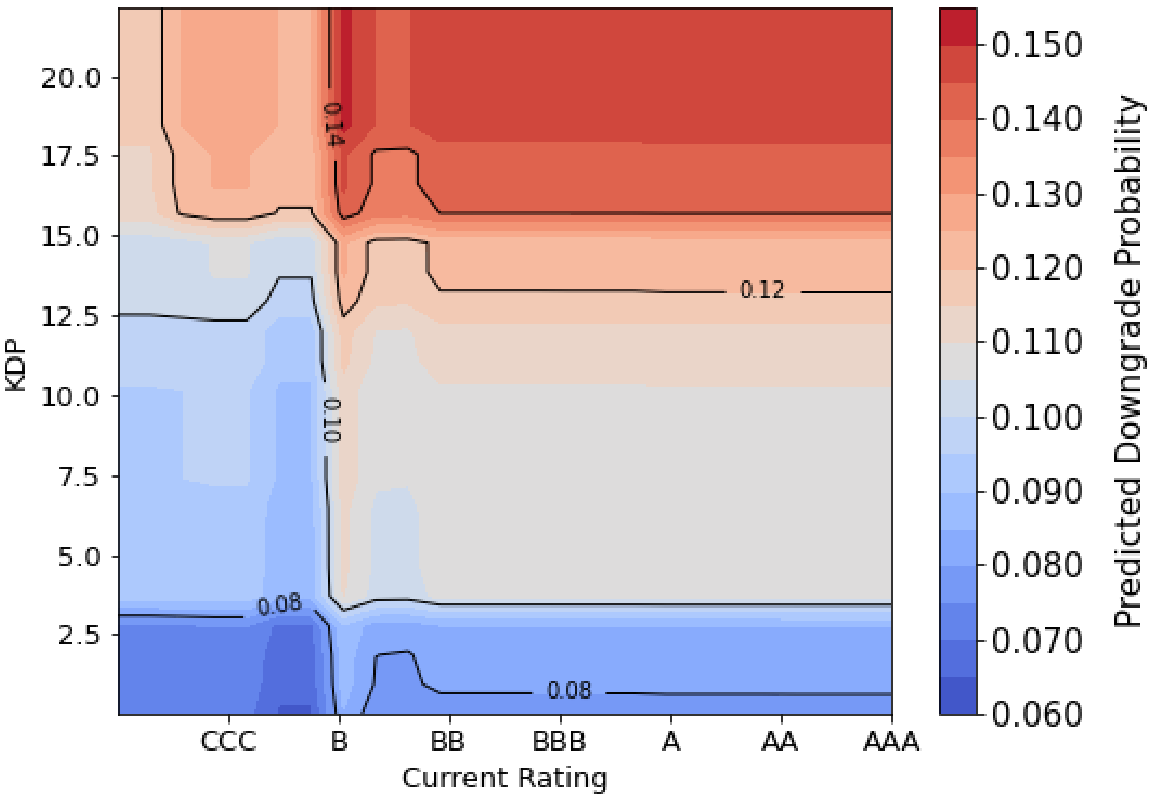


Exhibit 14: KDP-Rating 2D PDP

References

Afonso, António, Pedro Gomes, and Philipp Rother. "An Empirical Evaluation of the Performance of Binary Classifiers in the Prediction of Credit Rating Changes." Journal of Forecasting 30, no. 6 (2011): 510–531.

Altman, Edward I., and Duen-Li Kao. "The Implications of Corporate Bond Rating Drift." Financial Analysts Journal 48, no. 3 (1992): 64–75.

Bangia, Ananth, Francis X. Diebold, André Kronimus, Christian Schagen, and Til Schuermann. "Ratings Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing." Journal of Banking & Finance 26, no. 2–3 (2002): 445–474.

Bayraktar, Erhan, and Hao Xing. "A Unified Framework for Pricing Credit and Equity Derivatives." Mathematical Finance 17, no. 3 (2007): 345–375.

Becker, Bo, and Todd Milbourn. "How Did Increased Competition Affect Credit Ratings?" Journal of Financial Economics 101, no. 3 (2011): 493–514.

Blume, Marshall E., Felix Lim, and A. Craig MacKinlay. "The Declining Credit Quality of U.S. Corporate Debt: Myth or Reality?" Journal of Finance 53, no. 4 (1998): 1389–1413.

Bolton, Patrick, Xavier Freixas, and Joel Shapiro. "The Credit Ratings Game." Journal of Finance 67, no. 1 (2012): 85–111.

Brakatsoulas, Sotiris, and Jiri Kukacka. "Credit Rating Downgrade Risk on Equity Returns." SSRN Electronic Journal (2021).

Feng, Dai, Christian Gouriéroux, and Jolanta Jasiak. "The Ordered Qualitative Model for Credit Rating Transitions." Journal of Empirical Finance 15, no. 1 (2008): 111–130.

Figlewski, Stephen, Halina Frydman, and Wenyuan Liang. “Modeling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions.” International Review of Economics & Finance 21, no. 1 (2012): 87–105.

Gagliardini, Patrick, and Christian Gouriéroux. “Stochastic Migration Models with Application to Corporate Risk.” Journal of Financial Econometrics 3, no. 2 (2005): 188–226.

Hilscher, Jens, and Mungo Wilson. “Credit Ratings and Credit Risk: Is One Measure Enough?” Management Science 63, no. 10 (2017): 3414–3437.

Jarrow, Robert A., David Lando, and Stuart M. Turnbull. “A Markov Model for the Term Structure of Credit Risk Spreads.” Review of Financial Studies 10, no. 2 (1997): 481–523.

Kaplan, Robert S., and Gabriel Urwitz. “Statistical Models of Bond Ratings: A Methodological Inquiry.” Journal of Business 52, no. 2 (1979): 231–261.

Koopman, Siem Jan, André Lucas, and Annerose Monteiro. “The Multi-State Latent Factor Intensity Model for Credit Rating Transitions.” Journal of Econometrics 142, no. 1 (2008): 399–424.

Korolkiewicz, Marek, and Robert J. Elliott. “A Hidden Markov Model of Credit Quality.” Journal of Economic Dynamics and Control 32, no. 12 (2008): 3807–3819.

Lando, David, and Torben M. Skødeberg. “Analyzing Rating Transitions and Rating Drift with Continuous Observations.” Journal of Banking & Finance 26, no. 2–3 (2002): 423–444.

Mizen, Paul, and Serafeim Tsoukas. “Forecasting US Bond Default Ratings Allowing for Previous and Initial State Dependence in an Ordered Probit Model.” International Journal of Forecasting 28, no. 1 (2012): 273–287.

Nickell, Pamela, William Perraudin, and Simone Varotto. “Stability of Rating Transitions.” Journal of Banking & Finance 24, no. 1–2 (2000): 203–227.

Pinches, George E., and Kent A. Mingo. “A Multivariate Analysis of Industrial Bond Ratings.” Journal of Finance 28, no. 1 (1973): 1–18.

Yang, Bohui. “Modeling Path-Dependent State Transitions by a Recurrent Neural Network.” Big Data and Information Analytics 7, no. 1 (2022): 1–12.

Yue, Huifang, Sheng Xia, and Han Liu. “Multi-Task Envisioning Transformer-Based Autoencoder for Corporate Credit Rating Migration Early Prediction.” Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2022): 3183–3193