Using Alternate Data and Machine Learning to Improve Timeliness in Default Prediction

Abstract

Credit ratings are expected to be a summary measure of the probability of default of a given company/instrument. Rating agencies publish the likelihood of default by rating category and over next 1,2 and 3 years. If creditworthiness of a company goes down owing to poor financial performance, in theory the credit rating is revised downwards. One would expect that a company that defaults has undergone several rating downgrades along the way. However, in reality, most of the defaulting companies continue to enjoy credit rating anywhere between AA to BB nearly a month before a default (and often till the day before the default).

In this paper, we approach the problem of default forecasting through the lens of timeliness of a revision in forecast of default. Our hypothesis is that using higher frequency alternate data (in addition to conventional fundamental data from companies) combined with machine learning techniques can improve the timeliness of default forecast revision. As detailed below, we conclude that our approach is indeed able to improve this measure of timeliness when compared to the conventional credit rating revisions.

Background

Conventional credit rating process relies predominantly on the analysis of company-specific data, as received from the company itself. This approach has a serious shortcoming in the context of timeliness of default forecasting revision. For one, companies publish even quarterly results with a lag of 1-2 months. Secondly, there are a number of ‘financial shenanigans’ possible that allow companies to window dress their financials so that they look less problematic than they really are. This leads to the common occurrence of companies continuing to be rated A, BBB or BB almost up to the day of their default. While it is not reasonable to expect credit research to forecast each default with perfect foresight, it is a matter of concern that the downgrades in credit ratings of companies is neither frequent enough nor adequate enough as their financial health deteriorates.

We attempt to remedy the problem of timeliness by using higher frequency data as well as advanced machine learning models. Our target is to arrive at a method that suggests significant enough increase in measures of probability of default of companies as they approach default.

Data used

We carried out our analysis on the universe of listed companies in India. We used publicly available data on company financials as published by the companies each quarter and year. The credit ratings of companies (and the changes therein) are taken from public announcements by the companies.

We also used higher frequency alternate data about the companies, their sectors as well as about the state of broader economy. The most common example of alternate data in context of credit is the equity market data. Being updated daily, equity prices are a very useful gauge of future prospects of companies and sectors. Details of other alternate data are proprietary.

Methodology

We studied all listed companies that were rated between 2009 and 2018 (any rating at all). Next, we considered the subset of companies that defaulted on their debt obligations. In this defaulting subset, we tracked the rating of each company n months before the date of default where n ranged from 12 months to 1 month. This created the base-case of credit rating change on the way to default.

To calculate our estimates of probability of default, we used features drawn from fundamental data published by companies as well as various alternate datasets mentioned above. These features were pre-processed to normalize, winsorize and stationarize them.

For machine learning models, we tried several alternatives including logistic regression, neural networks, support vector machines and random forests. Of these logistic regression and neural networks looked promising. We created an ensemble model of these that created a forecasted of probability of default for each company for each week between 2009 and 2018.

For the defaulting company subset, we tracked the values of probability of default as predicted by our model – 12 months to 1 month before the default. We also tracked the implicit probability of default as per the credit rating of the company. Both these values were averaged for the full universe for each value of lag i.e. 12 months to 1 month.

Results

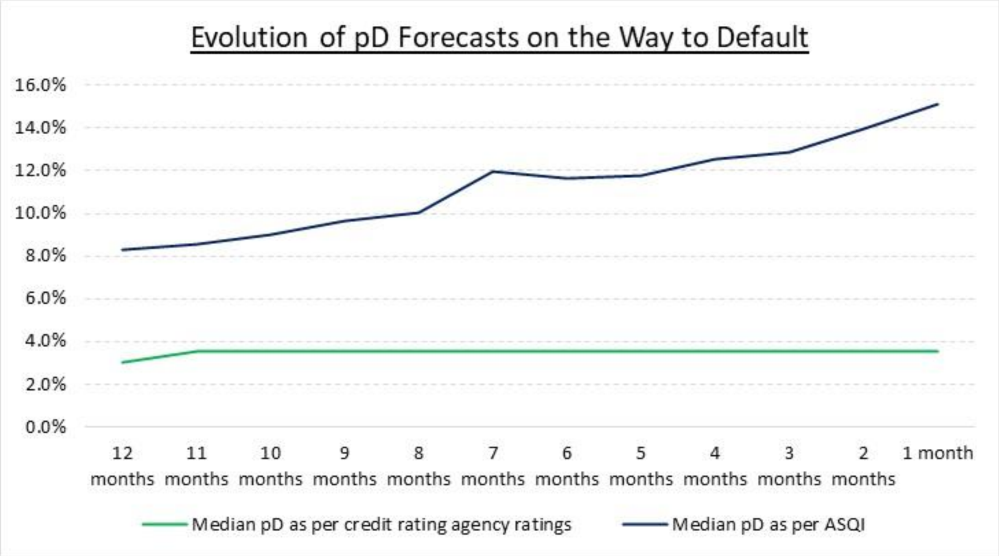
Following table shows the results of our analysis. (pD = probability of default).

Time before default	Median pD as per credit ratings	Median pD as per ASQI
12 months	3.0%	8.3%
11 months	3.5%	8.6%
10 months	3.5%	9.0%
9 months	3.5%	9.7%
8 months	3.5%	10.0%
7 months	3.5%	11.9%
6 months	3.5%	11.6%
5 months	3.5%	11.8%
4 months	3.5%	12.5%
3 months	3.5%	12.8%
2 months	3.5%	13.9%
1 month	3.5%	15.1%
Taken for ~500 defaults between 2009 and 2018		

As is clear from the table, the implicit probability of default in conventional credit rating hardly changed over time as the companies approached default. On the other hand, the probability of default as measured by ASQI's approach went up steadily as date of default approached. Another noteworthy point is that the starting point – 12 months prior to default – for ASQI's approach is already about 3 times higher than that of conventional credit ratings. Hence even the steady state likelihood of default prediction is superior in this approach.

For a typical user, a high value of pD which keeps rising further to cross 10% nearly 8 months prior to default should be an adequate red flag for a given company being analyzed.

Following diagram shows the above data graphically.



In the following table, we show the same results as above but converted into the form of conventional credit ratings – by using the rating-to-pD estimated published by rating agencies.

Time before default	Median credit rating as per credit rating	Equivalent rating for ASQI's pD estimate
12 months	BB+	B
11 months	BB	B
10 months	BB	B
9 months	BB	B-
8 months	BB	B-
7 months	BB-	B-
6 months	BB-	B-
5 months	BB-	B-
4 months	BB-	C
3 months	BB-	C
2 months	BB-	C
1 month	BB-	C
Taken for ~500 defaults between 2009 and 2018		

Conclusions

We conclude that use of alternate data and advanced machine learning models significantly improves the timeliness of revision in default probability estimates. This has the potential to improve the credit risk underwriting as well as monitoring by lenders. It can also improve the handling of counterparty risk assessment by companies for their receivables. Lastly it can help companies track the financial health of their key suppliers.

To know more about how to access the real-time updates produced by ASQI Systems, please write to us at asqi.systems@asqi.in

DISCLAIMER:

The information in this white paper are for the personal use of the authorised recipient(s). Under no circumstances, this information should be distributed publicly or reproduced without our prior permission.

The information provided in our publications is collated from publicly available data, which may be available freely e.g. annual reports, websites and media releases or which may be purchased by us from data providers/aggregators. We make every effort to validate the accuracy of the information included in the publications. We, to the fullest extent permitted by law, make no warranty (express, implied or otherwise), or assume any legal liability or responsibility for the accuracy, completeness, reliability and usefulness of this information. In many instances, the opinion expressed in the publications is our current opinion based on the current market trends and is subject to change from time to time.

The report provided is a statement of opinion and not a statement of fact. It contains forward looking statements. The white paper is not a guarantee of any particular outcome and you may therefore not use any information, in full or in part as the sole basis for any decision you take. It should not be construed as recommendation for lending to or avoiding lending to a given company, as well as investing in or not investing in any instruments issued by a company. An independent analysis and due diligence should be done before making any decision to invest/lend or not lend to the specific company. Opinion, projections and estimates are subject to change without notice.

The information on this document does not constitute any offer, recommendation or solicitation to any person to enter into any transaction or adopt any trading or investment strategy, nor does it constitute any prediction of likely future movement in rates or prices or any representation that any such future movements will not exceed those shown in any illustration. The information in our publications should be interpreted as indicative guidance only. We assert that any business or investment decisions should not be made solely based on the information presented in our publications. We recommend that the users should take proper / relevant professional advice before making any such decision.

ASQI Systems Private Limited is not a credit rating agency or an investment advisor, and is not purporting to provide you with investment, legal or tax advice. We, nor any of our affiliates, directors, employees or other representatives, accept any liability and will not be liable for any loss or damage arising directly or indirectly (including special, incidental or consequential loss or damage) from your use of this document, howsoever arising, and including any loss, damage or expense arising from, but not limited to, any defect, error, imperfection, fault, mistake or inaccuracy with this document, its contents or associated services, or due to any unavailability of the document or any thereof or due to any contents or associated services.