“Bank Name”

MODEL VALIDATION MODEL RISK MANAGEMENT

Annual Validation Report

CRE loan balances - Probability of Default Model (Model ID 894 Version 3.2)

Prepared by:

First Name, Last Name

Title, Model Validation (MV)

Reviewed by:

First Name, Last Name

Title, Model Validation (MV)

Identification

ID

Model Version Number

Model

File Name

894

3.2

CRE loan balances - Probability of Default model

Annual Validation

Table of Contents

1 EXECUTIVE SUMMARY

2 INTRODUCTION

2.1 Model source

2.2 Model composition

2.3 Model version

2.4 Model usage and product definition

2.5 Validation coverage

3 MODEL DESCRIPTION

3.1 Model overview

3.2 Theoretical background

3.3 Significant assumptions and limitations

3.4 Model input

3.5 Model output

4 MV METHODOLOGY ASSESSMENT

4.1 Literature review

4.2 Conceptual soundness

4.2.1 Model Framework and Calculation Method

4.2.2 Alternate Methodologies

4.2.3 Model Segmentation

4.2.4 Model Driver/Variable Selection

4.2.5 COVID-19 impact assessment

4.2.6 Review of any outstanding regulatory or IA findings, if applicable

4.3 Assesment of input data

4.4 Bias and fairness

4.5 Assesment of significant assumptions and limitations

4.6 Assesment of source of uncertainty and margin of conservatism

4.7 Assesment of the documentation

4.8 Assesment of performance monitoring approach

4.9 Vendor model contingency plan (if applicable, otherwise delete)

5 MODEL PERFORMANCE ASSESSMENT

5.1 PD model replication

5.1.1 Data quality

5.1.2 Univariate analysis

5.1.3 Model coefficient comparison

5.2 PD model stress test

5.2.1 Kolmogorov–Smirnov test

5.2.2 Variance Inflation Factor

5.2.3 Receiver Operating Characteristic curve

5.2.4 Normal test

5.3 Conclusions and recommendations

5.4 Model limitations

5.5 Model risk assessment

6 REFERENCES

7 APPENDIX

7.1 Final MDR

7.2 Other MV materials

8 ASSESSMENT OF REGULATORY REQUIREMENTS

9 ISSUE LOG

1. Model Performance Assessment

MV performed an annual validation of the CRE loan balances PD model. The model is developed using logistic regression technique with 7 independent variables. Data was sourced from TREPP and was split into training and testing data by random sampling approach. The development of the model was in Python, however the validation is performed in R.

In the current validation exercise, MV has performed the following activities for the selective varibles: 1.) Replication of the PD model - univariate analysis of risk drivers, comparison of independently replicated model coefficients with values as provided by development team 2.) Validation testing of the PD model - KS test, ROC curve, VIF and Normal test have been performed by the MV

The definations of the selected variables are listed below:

priorfydscr:Preceding Fiscal Year DSCR (NOI). A ratio of net operating income (NOI) to debt service for the most recent fiscal year end statement available as reported by the servicer.

mrfytdocc: Most Recent Occupancy Rate. The most recent available percentage of rentable space occupied. Should be derived from a rent roll or other document indicating occupancy consistent with most recent documentation.

debt\_yield\_p1: Debt yield of preceding financial year.

priorfyocc:Preceding Fiscal Year Occupancy Rate. A percentage of rentable space occupied as of the most recent fiscal year end operating statement available. Should be derived from a rent roll or other document indicating occupancy, and in most cases should be within 45 days of the most recent fiscal year end financial statement.

OLTV:LTV at the time of origination.

Division: Division created from the states of US.

interestonly:Interest Only (Y/P/N). If loan is interest only for life, then Y; if loan has more than one interest only period but is not fully interest only then P; else N.

The details of the above are provided in the sections below.

5.1 PD model replication

5.1.1 Data quality

In data quality analysis, we focussed on assessing whether the data used in input of the model (e.g., data from systems and databases) are reliable. Following tests are performed: 1) Missing values 2) Descriptive statistics

Missing values

|  |  |
| --- | --- |
| Feature | Missing |
| mrfytdocc | 42.87% |
| priorfyocc | 25.28% |
| priorfydscr | 21.82% |
| debt\_yield\_p1 | 17.51% |
| OLTV | 3.4% |

Descriptive statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Metrics | Priorfydscr | Mrfytdocc | Debt yield p1 | Priorfyocc | Oltv | Division | Interestonly | Bad flag final v3 |
| count | 72687 | 53116 | 76691 | 69471 | 89812 | 92974 | 92974 | 92974 |
| unique | N/A | N/A | N/A | N/A | N/A | 10 | 3 | N/A |
| top | N/A | N/A | N/A | N/A | N/A | South-Atlantic | N | N/A |
| freq | N/A | N/A | N/A | N/A | N/A | 19311 | 85294 | N/A |
| mean | 1.8395755298 | 93.4809228426 | 12.5337343705 | 93.754880108 | 66.4377253596 | N/A | N/A | 0.011960333 |
| std | 1.8170684715 | 9.0892893303 | 13.8366159547 | 9.2084476055 | 17.2802191395 | N/A | N/A | 0.1087079139 |
| min | -2.8302 | 0.78 | -26.38654378 | 1 | 0.9 | N/A | N/A | 0 |
| 25% | 1.2978 | 91.9175 | 6.8698379192 | 92 | 62.8 | N/A | N/A | 0 |
| 50% | 1.53 | 95.1 | 10.461944328 | 96 | 72.59 | N/A | N/A | 0 |
| 75% | 1.86 | 100 | 14.0404936145 | 100 | 77.6 | N/A | N/A | 0 |
| max | 63.97 | 100 | 408.3590591 | 100 | 96.24 | N/A | N/A | 1 |

5.1.2 Univariate analysis

In a univariate analysis, we set up a multitude of logistic regression models where there is only one explanatory variable (the risk driver) and the response variable (the default). Among those models, the ones that describes best the response variable can indicate the most significant explanatory variables, which can then be used to perform a multivariate analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Coef | P z | Gini coefficient |
| priorfydscr | -0.881726911 | 0 | 0.452 |
| mrfytdocc | -1.2194863213 | 0 | 0.321 |
| debt\_yield\_p1 | -0.8919738737 | 0 | 0.281 |
| priorfyocc | -1.1645963421 | 0 | 0.383 |
| OLTV | -0.7754851517 | 0 | 0.217 |
| Division | -0.8479998614 | 0 | 0.238 |
| interestonly | -0.9347473414 | 8.31589183e-158 | 0.057 |

The model coefficients, p-values and Gini coefficients of each of the chosen explanatory variables indicate that these parameters explain the reponse variable quite well.

5.1.3 Model coefficient comparison

The PD model is independently replicated by the validation team to ensure correctness of the implementation. In order to assess whether the replicated PD model yields the same output as the original model, the risk drivers (independent parameters) and their model coefficients are compared.

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Coef development | Coef validation | Coef diff |
| const | -4.4168461324 | -4.4168461324 | 0 |
| priorfydscr | -0.7618294827 | -0.7618294827 | -2.220446049e-16 |
| mrfytdocc | -0.3871917877 | -0.3871917877 | -9.992007222e-16 |
| debt\_yield\_p1 | -0.1510340457 | -0.1510340457 | 0 |
| priorfyocc | -0.4915827429 | -0.4915827429 | -5.551115123e-17 |
| OLTV | -0.310497772 | -0.310497772 | -8.881784197e-16 |
| Division | -0.5789169907 | -0.5789169907 | 0 |
| interestonly | -0.3459055608 | -0.3459055608 | 0 |

The differences between the results of the independent model replication and the results of the development team, both for the model coefficients and their standard deviation, indicate the correctness of the implementation of the PD model. In case these differences are not equal to zero, further investigation is needed into why this would be the case (e.g., implementation issues, specifications of the underlying data set, parameterization of the PD model, etc).

5.1.3.2 Scoring

Based on the forecasted probability of default, each facility is assiged to a rating. Below the distribution of the ratings for the train and test dataset.

Train dataset distribution

Test dataset distribution

5.2 PD model stress test

5.2.1 Kolmogorov–Smirnov Test

The KS statistic is used to measure the discriminatory power of the PD model. It is defined as the maximum difference between the cumulative percentage of good samples (i.e., non-defaulters) and the cumulative percentage of bad samples (i.e., defaulters). A higher KS value implies a good fit of the model.

The test returns a KS value.

|  |  |
| --- | --- |
| Ks statistic train | Ks statistic test |
| 0.4219 | 0.3811 |

For training sample, the KS value is 0.4219, which is higher than the threshold of 0.4, indicating that discriminatory power of the model is good. For testing sample, the KS value is 0.3811, which is lower than the threshold of 0.4, indicating that discriminatory power of the model is not sufficient.

5.2.2 Variance Inflation Factor

This test is to assess the multicollinearity among the risk drivers. The variance inflation factor (VIF) quantifies the severity of multicollinearity in an regression analysis caused by correlation between multiple ‘independent variables’ in a model (i.e., the risk drivers). It is a measure of how much the variance of an estimated regression coefficient is inflated because of multicollinearity in the model. High levels of multicollinearity results in unstable parameter estimates and renders the model statistically invalid.

The test returns VIF value for each variable listed in the table below.

|  |  |
| --- | --- |
| Variable | Vif |
| priorfydscr\_transformed | 1.3352 |
| mrfytdocc\_transformed | 1.6708 |
| debt\_yield\_p1\_transformed | 1.364 |
| priorfyocc\_transformed | 1.6813 |
| OLTV\_transformed | 1.2259 |
| Division\_transformed | 1.0953 |
| interestonly\_transformed | 1.045 |

As shown in the above table, all variables have a VIF lower than 4, indicating the model does not suffer from multicollinearity.

5.2.3 Receiver Operating Characteristic curve

Receiver Operating Characteristic (ROC) is a curve in a unit square which is used to assess the accuracy of the diagnostic (e.g. rating) system. The area under curve (AUC) is the area under the ROC curve. Higher AUC means higher discriminatory power of model.

|  |  |
| --- | --- |
| Area under curve | Gini coefficient |
| 0.7743 | 0.5487 |

As shown in the above table, the AUC value is higher than 0.7 and the Gini value of the model is higher than 0.4, indicating that discriminatory power of the model is good.

5.2.4 Normal Test

The aim of the test is to evaluate the adequacy of observed default rates comparing to predicted PD. The test is performed on the level of rating grade. It is applied under the assumption that the mean default rate does not vary too much over time and that default events in different years are independent. The normal test is motivated by the Central Limit Theorem and is based on a normal approximation of the distribution of the time-averaged default rates. With high enough (>30) number (n) of observations which are independent, Point in Time (PIT) observed default rate (ODR) is expected to be normally distributed. The algorithm of the test is to calculate time-weighted (or simple arithmetic) average of all PiT ODRs, calculate standard error of this statistic and derive confidence interval for it.

The test returns a table of p values and a brief interpretation of the test result for each rating grade.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rating class | Predicted pd upper boundary | Default rate | Normal test p value | Normal test result |
| 1 | 0.1% | 0.13% | 0.1063 | Acceptable |
| 2 | 0.26% | 0.22% | 0.1439 | Acceptable |
| 3 | 0.42% | 0.42% | 0.1088 | Acceptable |
| 4 | 0.56% | 0.53% | 0.1982 | Acceptable |
| 5 | 0.71% | 0.7% | 0.2359 | Acceptable |
| 6 | 0.93% | 0.91% | 0.2364 | Acceptable |
| 7 | 1.12% | 0.7% | 0.2665 | Acceptable |
| 8 | 1.41% | 1.1% | 0.4014 | Acceptable |
| 9 | 2.56% | 1.91% | 0.4329 | Acceptable |
| 10 | 15% | 5.37% | 0.0627 | Acceptable |

As shown in the above table, the p-value of Normal test for all segments is higher than 0.05, which means the null hypothesis cannot be rejected and there is evidence that the predicted PD does not deviate from the long run average of observed values. Therefore, the accuracy of the model is adequate.