“Bank Name”

MODEL DEVELOPMENT GROUP

Model Development Report

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

Prepared by:

First Name, Last Name

Title, Model Development (MD)

Reviewed by:

First Name, Last Name

Title, Model Development (MD)

Identification

ID

Model Version Number

Model

File Name

894

3.2

CRE loan balances - Probability of Default model

Model Development

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

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 MODEL METHODOLOGY

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

4.3 Assesment of input data

4.4 Bias and fairness

4.5 Performance monitoring approach

5 MODEL PERFORMANCE

5.1 Performance testing

5.1.1 Data quality

5.1.2 Information value

5.1.3 Model coefficient

5.1.4 Receiver Operating Characteristic curve

5.1.5 KS statistic

5.2 Conclusions and recommendations

5.3 Model limitations

6 REFERENCES

7 APPENDIX

7.1 Final MDR

7.2 Other MD materials

8 ASSESSMENT OF REGULATORY REQUIREMENTS

9 ISSUE LOG

1. Model Performance

Model developers have built v3.2 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 is in Python.

In the current development exercise, model developers have performed the following activities for the selected varibles: Performance testing of the PD model - information value of risk drivers, calculation of model coefficients, data quality checks and Gini.

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 Performance testing

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 Information value

Information Value (IV), calculated at risk driver level, is a measure used to quantify the predictive power of a risk driver. More specifically, how well can a risk driver discriminate between a default and no-default response.

The test returns IV for selected risk drivers listed in the table below.

|  |  |
| --- | --- |
| Feature | Iv total |
| priorfydscr | 0.8183686365 |
| priorfyocc | 0.5424667177 |
| mrfytdocc | 0.409471795 |
| debt\_yield\_p1 | 0.293684185 |
| OLTV | 0.2461692548 |
| Division | 0.235608834 |
| interestonly | 0.0443585828 |

The table above provides the IV of all the variables used in the model, higher IV implies more discriminatory power of the variable between the default and non-default customers.

5.1.3 Model coefficient

The PD model is developed using logistic regression technique with 7 risk drivers (independent variables) in the final model equation. Please find below the list of parameters and their model coefficients.

|  |  |
| --- | --- |
| Parameter | Coef |
| const | -4.4168461324 |
| priorfydscr | -0.7618294827 |
| mrfytdocc | -0.3871917877 |
| debt\_yield\_p1 | -0.1510340457 |
| priorfyocc | -0.4915827429 |
| OLTV | -0.310497772 |
| Division | -0.5789169907 |
| interestonly | -0.3459055608 |

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

5.1.4 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 correspondingly the Gini value of the model is higher than 0.4, indicating that discriminatory power of the model is satisfactory.

5.1.5 KS statistic

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 |

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