Document : [J:\970\Risque de credit\MODÈLES REGLÉMENTAIRES\I. WHLSL\sRRS 8\PD\03\_Documentation\01\_Modèle\New guidelines\DEV\_BRR\_SRRS8\_SOVEREIGN\_2018\_v0.docx

# 

**PD MODEL DEVELOPMENT**

**RETAIL MORTGAGE PORTFOLIO (SRRS4)**

Document : [J:\970\Risque de credit\ MODÈLES REGLÉMENTAIRES\ II. RETAIL\RRS 4\PD\Revision\_2018\03\_Documentation\01\_Modèle\DEV\_PD\_SRRS4\_MORTGAGE\_2018\_V02

# **VERSION HISTORY**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version** | **Date Revised** | **Description of modifications** | **Updated sections** | **Responsible officers** | **Function** |
| V01 | July 31, 2018 | Creation of the document | All | Ablaye Thiaw | Principal Advisor |
| V02 | November 15th. 2018 | Improvement of the document and integration of some validation recommendations | All | Ablaye Thiaw | Principal Advisor |
| V03 | July 9th. 2019 | Including Validation / OSFI comment on consistency in the source of uncertainty and margin of conservatism | Section 8.10  Table 43 | Ablaye Thiaw | Senior Manager |

# **LIST OF TABLES**

[Table 01: Impact summary on the portfolio as of June 30, 2018 14](#_Toc532818019)

[Table 02: Models performance comparison 15](#_Toc532818020)

[Table 03: Validation findings 16](#_Toc532818021)

[Table 04: OSFI findings 19](#_Toc532818022)

[Table 05: Distribution of portfolio for AGF as of June, 2018. 29](#_Toc532818023)

[Table 06: Distribution of portfolio for ALT\_A as of June, 2018 29](#_Toc532818024)

[Table 07: Distribution of portfolio for B2B as of June, 2018. 29](#_Toc532818025)

[Table 08: Distribution of portfolio for Province as of June, 2018. 29](#_Toc532818026)

[Table 09: Distribution of portfolio Insured vs Non-Insured as of June, 2018 30](#_Toc532818027)

[Table 10: Number of defaults and default rate per quarter 31](#_Toc532818028)

[Table 11: Concentration analysis 32](#_Toc532818029)

[Table 12: Mortgage portfolio exposure as of LBC total assets 33](#_Toc532818030)

[Table 13: Model design participants overview 38](#_Toc532818031)

[Table 14: Variable class & WOE 40](#_Toc532818032)

[Table 15: Representativeness of data – Univariate measures 59](#_Toc532818033)

[Table 16: Representativeness of data – Stability Index 60](#_Toc532818034)

[Table 17: Siddiqi rules of thumb for Information Value (page 81) 70](#_Toc532818035)

[Table 18: Preliminary selected variables with their IV 71](#_Toc532818036)

[Table 19: Relationship between Target and selected variables 75](#_Toc532818037)

[Table 20: Pearson Correlation analysis between selected variables. 76](#_Toc532818038)

[Table 21: Ln-Likehood and AIC for distribution selection 77](#_Toc532818039)

[Table 22: Gamma survival model estimation results 80](#_Toc532818040)

[Table 23: Final variable estimates and Weight of Evidence 81](#_Toc532818041)

[Table 24: Variable Inflation (VIF) 82](#_Toc532818042)

[Table 25: BLC’s Master Scale for Retail 83](#_Toc532818043)

[Table 26: Mortgage Long-Run Default Rate 84](#_Toc532818044)

[Table 27: KS rules of thumb values 87](#_Toc532818045)

[Table 28: AUROC rules of thumb values 88](#_Toc532818046)

[Table 29: Statistical measures of performance 88](#_Toc532818047)

[Table 30: Discriminatory result - AUROC 89](#_Toc532818048)

[Table 31: Calibration result – Binomial test 90](#_Toc532818049)

[Table 32: Calibration result – Vasicek test 91](#_Toc532818050)

[Table 33: Calibration result – Blochinger-Leippold test 92](#_Toc532818051)

[Table 34: Stability result – Population Stability Index 93](#_Toc532818052)

[Table 35: Stability result – Maximum grade-level concentration 93](#_Toc532818053)

[Table 36: Final sensitivity analysis 94](#_Toc532818054)

[Table 37: Transition matrix from old to new PD 95](#_Toc532818055)

[Table 38: Default Parameter Value 96](#_Toc532818056)

[Table 39: Benchmarking with Canadian Peers 96](#_Toc532818057)

[Table 43: Sources of uncertainty and associated margins of conservatism 97](#_Toc532818058)

# **LIST OF FIGURES**

[Figure 01: Retail Mortgage Segmentation 27](#_Toc532818059)

[Figure 02: LBC risk rating systems 34](#_Toc532818060)

[Figure 03: Overview of the methodology of survival time calculation 47](#_Toc532818061)

[Figure 04: Survival Time and Censored 49](#_Toc532818062)

[Figure 05: Illustration of datasets layout 50](#_Toc532818063)

[Figure 06: Illustration of seasoning effect 56](#_Toc532818064)

[Figure 07: Canadian GDP – All Industries - Monthly 61](#_Toc532818065)

[Figure 08: Unemployment Rates 61](#_Toc532818066)

[Figure 09: Example of WOE’s Trend 72](#_Toc532818067)

[Figure 10: Cox-Snell residual - Gamma 77](#_Toc532818068)

[Figure 11: Historical Default rate and PDs 84](#_Toc532818069)

[Figure 12: Kolmogorov-Smirnov measure 86](#_Toc532818070)

[Figure 13: ROC curve 87](#_Toc532818071)

[Figure 14: Distribution per risk rating 94](#_Toc532818072)

# **ABBREVIATIONS AND ACRONYMS**

**AFT** Accelerated Failure Time

**AIC** Akaike's Information Criterion

**AIRB** Advanced Internal Rating Based

**AR** Accuracy Ratio

**AUC** Area under ROC Curve

**AUROC** Area Under Receiver Operating Characteristic

**BRR** Borrower Risk Rating

**CMHC**  Canada Mortgage and Housing Corporation

**CRP** Consumer Risk Predictor

**DR** Default Rate

**EAD** Exposure at Default

**EC** Economic capital

**EL** Expected loss

**FRR** Facility Risk Rating

**GDP** Gross Domestic Value

**HELOC** Home Equity Line of Credit

**ICAAP** Internal Capital Adequacy Assessment Process

**IFRS** International Financial Reporting Standards

**IRB** Internal Rating Based

**IV** Information Value

**KS** Kolmogorov-Smirnov

**LBC** Laurentian Bank of Canada

**LGD** Loss Given Default

**OSFI** Office of the Superintendent of Financial Institutions

**PD** Probability of Default

**PIT** Point In Time

**RMAI** Retail Mortgage Acquisition Initiative

**ROC** Receiver Operating Characteristic

**RRS** Risk Rating System

**RWA** Risk-Weighted Assets

**SCHL** Société Canadienne d’Hypothèques et de Logement

**SRRS** Sub Risk Rating System

**TTC** Through-the-cycle

**VIF** Variance Inflation Factors

**WOE** Weight Of Evidence

# **TABLE OF CONTENTS**

[1](#_Toc13556366)

[VERSION HISTORY 2](#_Toc13556367)

[LIST OF TABLES 3](#_Toc13556368)

[LIST OF FIGURES 4](#_Toc13556369)

[ABBREVIATIONS AND ACRONYMS 5](#_Toc13556370)

[TABLE OF CONTENTS 7](#_Toc13556371)

[EXECUTIVE SUMMARY 12](#_Toc13556372)

[1. GENERAL OVERVIEW 14](#_Toc13556373)

[1.1 Background of the model development 15](#_Toc13556374)

[1.2 Main issues of the modeling framework 16](#_Toc13556375)

[1.3 Limitations of the model 16](#_Toc13556376)

[1.4 Comments and findings from the last model validation 16](#_Toc13556377)

[1.5 Modeling objectives 20](#_Toc13556378)

[1.6 Overview of the modeling framework 20](#_Toc13556379)

[1.6.1 Quantitative/qualitative assessment 20](#_Toc13556380)

[1.6.2 Vendor Model or Internal Model 20](#_Toc13556381)

[1.7 Overview of the Structure of entire Model Documentation 20](#_Toc13556382)

[1.8 Illustration of the Modeling Steps 21](#_Toc13556383)

[2. OVERVIEW OF PROCESSES IN PLACE 21](#_Toc13556384)

[2.1 Validation Process 21](#_Toc13556385)

[2.2 Vetting Process 22](#_Toc13556386)

[2.3 Replication Process 22](#_Toc13556387)

[2.4 Monitoring Process 22](#_Toc13556388)

[2.5 Refreshing Process 22](#_Toc13556389)

[2.6 Risk Materiality Assessment 22](#_Toc13556390)

[3 FINAL MODEL(S) 22](#_Toc13556391)

[3.1 Model log of changes 23](#_Toc13556392)

[3.1.1 Changes in Economic Conditions 23](#_Toc13556393)

[3.1.2 Changes in the Modeling Methodology 23](#_Toc13556394)

[3.1.3 Changes in Default Definition 23](#_Toc13556395)

[3.1.4 Changes in Business Practises 25](#_Toc13556396)

[3.2 Scope of the model 26](#_Toc13556397)

[3.2.1 Target portfolio 27](#_Toc13556398)

[3.2.1.1 Industries covered by the model 28](#_Toc13556399)

[3.2.1.2 Products covered by the model 28](#_Toc13556400)

[3.2.1.3 Distribution of obligors 29](#_Toc13556401)

[3.2.2 Concentration analysis 32](#_Toc13556402)

[3.2.3 Business practises 32](#_Toc13556403)

[3.2.4 Materiality of the portfolio 33](#_Toc13556404)

[3.3 Risk Rating System design 33](#_Toc13556405)

[3.3.1 RRS/SRRS 33](#_Toc13556406)

[3.3.2 Time horizon of the parameter 35](#_Toc13556407)

[3.3.3 Rating philosophy 35](#_Toc13556408)

[3.3.4 Rating dimensions 37](#_Toc13556409)

[3.3.5 Rating structure 37](#_Toc13556410)

[3.3.6 Rating Criteria 37](#_Toc13556411)

[3.4 Model Usage 37](#_Toc13556412)

[3.5 Model Design 38](#_Toc13556413)

[3.5.1 Model Design Participants 38](#_Toc13556414)

[3.5.2 Model Design Overview 38](#_Toc13556415)

[3.5.3 Contributions of the business lines 38](#_Toc13556416)

[3.6 Summary of PD assessment process 39](#_Toc13556417)

[4 DEFINITIONS 41](#_Toc13556418)

[4.1 Definition of Risk Type 41](#_Toc13556419)

[4.2 Definition of Default 41](#_Toc13556420)

[4.3 Definition of Loss 43](#_Toc13556421)

[4.4 Definition of Costs 43](#_Toc13556422)

[4.5 Definition of Discounting 43](#_Toc13556423)

[4.6 Definition of Time to Resolution 44](#_Toc13556424)

[4.7 Definition of Re-aging 44](#_Toc13556425)

[4.8 Definition of Overdraft 44](#_Toc13556426)

[4.9 Definition of Cured Account 44](#_Toc13556427)

[4.10 Definition of seasoning 45](#_Toc13556428)

[5 DATA USED IN THE MODELING PROCESS 45](#_Toc13556429)

[5.1 Data Sources 45](#_Toc13556430)

[5.2 Data Refreshing Process 46](#_Toc13556431)

[5.3 Data Processing 46](#_Toc13556432)

[5.3.1 Exclusions and Manipulations 51](#_Toc13556433)

[5.3.2 Treatment of Missing 53](#_Toc13556434)

[5.3.3 Treatment of Outliers 54](#_Toc13556435)

[5.3.4 Treatment of Cured Accounts 55](#_Toc13556436)

[5.3.5 Re-ageing treatment 55](#_Toc13556437)

[5.3.6 Overdraft treatment 55](#_Toc13556438)

[5.3.7 Data Segmentation 55](#_Toc13556439)

[5.3.8 Accounting for Seasoning Effect 56](#_Toc13556440)

[5.3.8.1 Seasoning effect analysis 56](#_Toc13556441)

[5.3.8.2 Seasoning effect treatment 57](#_Toc13556442)

[5.3.7 Sampling 57](#_Toc13556443)

[5.4 Data Pre-treatment 57](#_Toc13556444)

[5.4.1 Data Transformation 58](#_Toc13556445)

[5.4.8.1 Data Linearization 58](#_Toc13556446)

[5.4.8.2 Data Standardization 58](#_Toc13556447)

[5.4.8.3 Data Conversion 58](#_Toc13556448)

[5.4.2 Preliminary Variable Selection 59](#_Toc13556449)

[5.5 Representativeness of Data 59](#_Toc13556450)

[5.6 Historical Data Coverage 60](#_Toc13556451)

[5.6.1 Adverse Economic Conditions & Stress Years 60](#_Toc13556452)

[5.6.2 Sufficiency of Data 62](#_Toc13556453)

[6 METHODOLOGY 62](#_Toc13556454)

[6.1 Methodological Options 63](#_Toc13556455)

[6.2 Selected Modeling Approach 65](#_Toc13556456)

[6.3 Assumptions and Hypothesis 67](#_Toc13556457)

[7 MODEL SPECIFICATION AND ESTIMATION 67](#_Toc13556458)

[7.1 Preliminary Analysis 68](#_Toc13556459)

[7.1.1 Target Variable 68](#_Toc13556460)

[7.1.2 Predictive or Explanatory Variables 68](#_Toc13556461)

[7.1.3 Univariate Analysis 68](#_Toc13556462)

[7.1.4 Multivariate Analysis 73](#_Toc13556463)

[7.1.5 Analysis of the Relationship between Target Variables and the Predictive Variables 74](#_Toc13556464)

[7.1.6 Correlation Analysis between Variables 76](#_Toc13556465)

[7.2 Risk Drivers Selection (Model variables selection) 76](#_Toc13556466)

[7.3 Downturn 79](#_Toc13556467)

[7.3.1 Correlation between Default Rate and LGD/EAD 79](#_Toc13556468)

[7.3.2 Downturn Period Selection 79](#_Toc13556469)

[7.3.3 Downturn Factor 79](#_Toc13556470)

[7.4 Preliminary Results 80](#_Toc13556471)

[7.4.1 Modeling Results 80](#_Toc13556472)

[7.4.2 Modeling Assumptions Violation Check 81](#_Toc13556473)

[7.5 Model parsimony 82](#_Toc13556474)

[7.6 Mapping Mechanism between Model Output and Rating Grade 82](#_Toc13556475)

[8 CALIBRATION AND FINAL RESULTS 83](#_Toc13556476)

[8.1 Calibration Exercise 83](#_Toc13556477)

[8.2 Downturn Parameter 85](#_Toc13556478)

[8.3 Final Model 85](#_Toc13556479)

[8.4 Model Validation Testing 86](#_Toc13556480)

[8.4.1 In-sample and Out-of-sample Performance 86](#_Toc13556481)

[8.4.2 Out-of-time Backtesting 88](#_Toc13556482)

[8.5 Sensitivity Analysis 94](#_Toc13556483)

[8.6 Model Application on the Current Portfolio 94](#_Toc13556484)

[8.7 Default Parameter Value 95](#_Toc13556485)

[8.8 Benchmarking 96](#_Toc13556486)

[8.9 Overrides 96](#_Toc13556487)

[8.10 Sources of Uncertainty and Associated Margin of Conservatism 96](#_Toc13556488)

[REFERENCES 99](#_Toc13556489)

[APPENDIX 100](#_Toc13556490)

[Appendix 01: Files Location and Description 100](#_Toc13556491)

[Appendix 02: Data flow and process chart 101](#_Toc13556492)

[Appendix 03: Useful definitions 101](#_Toc13556493)

[Appendix 04: Data dictionary 107](#_Toc13556494)

[Appendix 05: Model inventory 107](#_Toc13556495)

[Appendix 06: Candidate explanatory variable and associated sources 108](#_Toc13556496)

[Appendix 07: Potential risk drivers according to existing literature 108](#_Toc13556497)

[Appendix 08: Descriptive statistics 110](#_Toc13556498)

[Appendix 09: Missing observations and outliers 110](#_Toc13556499)

[Appendix 10: Univariate analysis on candidate variables 125](#_Toc13556500)

[Appendix 11: Multivariate analysis on candidate variables 130](#_Toc13556501)

[Appendix 12: Distribution of explanatory variables 139](#_Toc13556502)

[Appendix 13: Decision tree (segmentation) 139](#_Toc13556503)

[Appendix 14: Performance testing approach chart 139](#_Toc13556504)

[Appendix 15: Discussion with the business lines 140](#_Toc13556505)

[Appendix 16: List of products in the modeling database 140](#_Toc13556506)

[Appendix 17: Backtesting result for AGF 140](#_Toc13556507)

[Appendix 18: Backtesting result for ALT\_A 141](#_Toc13556508)

[Appendix 19: Backtesting result for B2B 141](#_Toc13556509)

[Appendix 20: Backtesting result without conservatism 142](#_Toc13556510)

# **EXECUTIVE SUMMARY**

Laurentian Bank uses credit risk rating systems to assess and manage retail clients. The Risk Rating system for the mortgage loan portfolio is based on an internal model. The objective of this project is to determine the relationship between the risk drivers for the retail mortgage population to build statistical PD models. The model assigns probability of default to mortgage exposures based on borrower and transaction characteristics as well as historical delinquency. The estimated PD is used for the risk-weighting of assets under the AIRB approach, to estimate expected losses and assess loan provisioning, and in all other areas of portfolio risk management.

It is important to note that for mortgage portfolio, the PD is calculated at product level and the model is used for back-end purposes. For this reason, behavioral variables are considered in the modeling process. The model will be applied to all retail mortgage portfolio except RMAI. The model covers mortgage products excluding the home equity lines of credit, which will be included in PD Heloc's model. The different types of products are the residential mortgage and the personalized mortgage loans. The product coverage for such portfolio includes owner-occupied Residential Mortgages of 1-4 units excluding home equity line of credits (HELOC). It also covers non-owner-occupied residential mortgages of 1-4 units providing the client does not have more than 10 units (excluding principal & secondary residences).

The mortgage portfolio on which this PD model will be applied is composed of retail mortgage loans described in the next table (as of June 2018).

Totaling more than 72,000 accounts (including all tranches) for an exposure of $ 13 billion, the retail mortgage portfolio represents 29% of the Bank's total exposure. The Bank's mortgage portfolio has a high concentration in Quebec with 65% followed by Ontario with 25% of loans. Among these mortgage products, there are conventional loans (53%) and insured loans (47%) either by CMHC or by Genworth.

The mortgage portfolio on which the PD model was developed consists of loans from Laurentian Bank as well as those of B2B, AGF loans that was acquired in 2012 which represent 0.7% of the mortgage portfolio and ALT\_A loans that are alternative financing represent 3% of the mortgage portfolio. The principle of ALT\_A loans is to finance clients who have had problems in the past (credit payments, temporary jobs, etc.) for a risk premium.



Of all the statistical models recognized in the literature that can be used to predict the probability of default, survival analysis has been favored. The reason for this choice are numerous but we can name some: survival analysis is very flexible and can estimate the default probability over any time horizon. This methodology can accommodate censored data of subjects with incomplete information about survival time. Moreover, modelling the time until default provides more statistical power than simply counting the number of defaults. The variables retained in the final model are all recognized in the literature and have a proven impact on the estimate of the probability of default. There are delinquency variables (Maximum number of days of delinquency in the last 6 months and Number of months since the most recent delinquency), characteristic of the borrower (Number of months since the customer is open at BLC, Number of active margin accounts, Average number of months since date opened on all bankcard trades, Number of months since latest activity of revolving trades., Number of inquiries in the last 12 months and Ratio of total balance to credit limit for all bankcard trades) and characteristic of the loan (Number of months before end of amortization).

The main changes carried out by this exercise can be summarized as follows:

**Data used**: Previous model was developed using data from October 2008 to July 2009 and calibrated with data from October 2010 to January 2015. This revision uses data from January 2008 to July 2017 covering a complete economic cycle.

**Segmentation**: Even though the model was developed without any segmentation, the present model is calibrated differently on AGF versus NO AGF exposures. In this proposed review, no segmentation is done. The variable AGF – NO AGF is transformed as a binary variable and tested in the modeling process. However, the binary variable was not selected in the final model.

**Methodological approach**: Present model was constructed with the scorecard approach. This revision uses a survival analysis in the parametric version.

**Impacts of proposed parameters**:

Table 01: Impact summary on the portfolio as of June 30, 2018



# **GENERAL OVERVIEW**

In June 2011, Experian Decision Analytics was contracted by Laurentian Bank of Canada (LBC) to develop a suite of custom behavioral models consisting of five scorecards (including mortgage portfolio) used for predicting the probability of default of retail exposures. The behavioral models rank order retail credit products based on their respective risk of default. The scorecards are used in the risk management of retail portfolios, the estimation of loan losses and for capital calculation.

In June 2013, Laurentian Bank developed PD models based on the behavioral score developed by Experian for use in the AIRB context. Behavioral models calculate scores that are used to segment the population into 17 retail risk ratings. Each retail risk rating is assigned a probability of default based on the average annual 1-year default rate calculated over 23 quarterly cohorts with conservatism factor added to ensure robustness of the PD estimates.

In April 2017, a calibration is done using data from October 2010 to January 2015.

In August 2012, LBC acquired via B2B Bank the AGF mortgage loans. AGF is a trust company. Even though AGF residential mortgage loans are still present in the Bank’s portfolio, this is not a growing portfolio, i.e. sales of new AGF loans are now discontinued.

The decision to redevelop the model was taken to address major issues. The current model had been developed by Experian in 2011. There is no documentation on Experian work. We also do not know the selection criteria for the different variables or the underlying assumptions. During the recent calibration, there was a challenge of data traceability as the current model does not include the data before 2010. Other issues related to consistency in the definition of default or in the choice variables were also raised, either by the regulator or the Validation team. All these points mean that redevelopment was necessary even if the current model has still acceptable performance. Below is a comparison table of the performance of the current and proposed model:

Table 02: Models performance comparison



The comparison was made, on the one hand with the backtesting of the proposed model and the results of the January 2018 backtesting report (the last available). We did not compare the stability of the models because it was not available for the current model. Apart from the binomial test that is right at the limit (the threshold is 80%) for the current model, the rest seems acceptable. However, the proposed model is also very good and can correct the issues listed above.

This document presents the methodology employed for the calculation of Probability of Default for the retail mortgage portfolio. The proposed model use the survival analysis approach based on borrower and transaction characteristics to estimate a probability of default for the retail mortgage portfolio.

## 1.1 Background of the model development

LBC’s Mortgage portfolio falls into the SRRS 4, which includes all the owner-occupied Residential Mortgages with less than 5 units and the 1-4 units non-owner-occupied Residential Mortgages providing that the main client does not have more than 10 total rental units at the Bank. One survival model is developed for all the mortgage portfolio and no segmentation is done.

## 1.2 Main issues of the modeling framework

No issue was found in the modeling framework since the model covers enough date history (2008 to 2017), contains enough defaults and cover a complete economic cycle. The mortgage portfolio is not a low default portfolio.

## 1.3 Limitations of the model

The model does not apply to the RMAI portfolio.

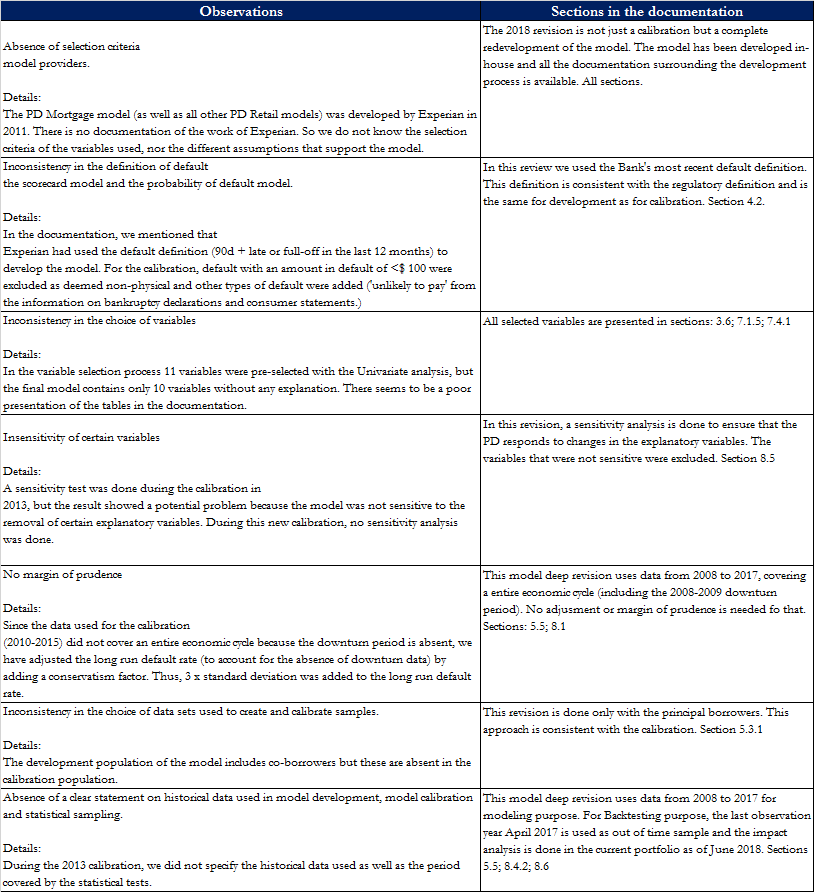
## 1.4 Comments and findings from the last model validation

Some suggestions and recommendations (see table 02) have arisen from the model validation based on the previous versions of the model documentation and associated relevant documents. Where possible, some of those recommendations shall be addressed in the scope of this version of the model documentation with new template.

Table 03: Validation findings

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **N** | **Observation** | **Recommendation** | **Level** | **Section in validation report** | **Section in the present model documentation**  **Where the issue is addressed** | **Action plan** |
| 1 | Data quality and reconciliation | KPMG recommends that an analysis be undertaken to reconcile the data extractions from 2010 and 2011 to ensure that the modeling data is appropriate and complete. Furthermore, enhanced data would ensure that the segmentation criteria be linked and that the datasets may be used for contiguous modeling purposes since modeling data prior to 2010 could not be retrieved. | Preoccupying | Section 4.2.2 Target portfolio and distribution of obligators | Section 3.2.1 Target portfolio | *Segmentation criteria will be refined and will be linked to the datasets for modeling purposes* |
| 2 | Modeling data transformations and consistency | KPMG recommends the alignment of the SAS implementation with respect to the LBC's internal policy for managing co-borrower accounts. The modifications should address the inconsistencies on the various approaches at filtering co-borrowers. Also, KPMG recommends that LBC further documents the approach to manage co-borrower accounts, the foundation for the methodological choices and the implications on the models’ development approach and performance. | Information | Section 6.4.2 Co-borrower accounts management | Section 5.3.1 | *The implementation will be made with respect to the LBC's internal policy for managing co-borrower accounts.*  *Documentation regarding this point will be enhanced.* |
| 3 | Documentation exhaustiveness | KPMG recommends that the Bank revisits its documentation standards to include additional details with regard to the elements previously enumerated. A revised model documentation that contextualizes development decisions and justifications would improve transparency and minimize model risk as more scrutiny would  be required during the modeling process. | Information | All sections | All sections | *A new documentation framework will be developed and will cover all requirements* |
| 4 | Model backtesting results and ongoing model performance: | KPMG recommends that the Bank closely monitors the results of all the tests in order to ensure that the models performance do not deteriorate over time and that the estimation are sufficient conservative  across periods. | Information | Section 7 Performance tests and backtesting | Section 8.4.2 Out-of-time backtesting | *A backtesting framework will be developed and will cover all requirements regarding this point.* |
| 5 | Credit scorecard update and definition of default consistency | KPMG suggests the redevelopment of the behavioral scorecards over the medium term to ensure the adequate performance of the models with respect to the chosen risk factors and their relative weights, in spite of the absence of issues with the actual models in place. We also suggest to conduct comparability analyses over the default definition across time and in accordance with the related changes. | Information | Section 3.1 Validation of the definition of default | Section 3.1.3 Change in default definition | *A log of changes has been documented and includes changes related to recovery policies. Potential impacts of these changes on estimated parameters will be analyzed and documented. Margin of conservatism will be applied if necessary.* |

Table 04: OSFI findings



## 1.5 Modeling objectives

The model aims at building a quantitative tool for predicting the creditworthiness in term of PD for LBC’s mortgage portfolio based on a set of identified borrower and transaction characteristics. For that purpose, we use a survival analysis approach in its parametric version. The estimated PD is used for the risk-weighting of assets under the AIRB approach, to estimate expected losses and assess loan provisioning, and in all other areas of portfolio risk management.

## 1.6 Overview of the modeling framework

Given that we have enough historical data and enough number of defaults, we have opted for purely statistical approach to model the PD for the mortgage portfolio.

### 1.6.1 Quantitative/qualitative assessment

The model entirely uses a quantitative model to build a relationship between survival time and identified borrower and transaction characteristic. As shown in section 8.4, the model shows overall good performance. No qualitative assessment is done.

### 1.6.2 Vendor Model or Internal Model

A vendor model is a model that is developed and calibrated by an external provider, using in large part external data. Capital Adequacy Requirements (CAR) recognize the need for banks to utilize vendor models whenever internal data are insufficient to build a model on their own. Nevertheless, the full complement of validation is still required in order to demonstrate that the acquired model is relevant, representative and appropriate for the portfolio in question.

This 2018 retail mortgage PD model is not a vendor model. It is entirely built by LBC Risk Model Team with internal data.

## 1.7 Overview of the Structure of entire Model Documentation

The following sections give an overview of the process in place and then focus on the main elements regarding the modeling framework and the final model. The rest of the model documentation will be divided into 4 main sections: besides a definition of key concepts for modeling, we describe different sources of data used in the modeling process, present the methodology and detail all of model’s specification and estimation including preliminary results. Finally, we present the final results after calibration. Appendix 01 describes all of SAS programs used for model building.

## 1.8 Illustration of the Modeling Steps

The modeling steps are detailed as follow:

1. Discussions with key stakeholders
2. Literature review
3. Feasibility analysis
4. Target portfolio definition
5. Description of the rating system used
6. Applied process to the model’s construction
7. Criteria used to assess the model’s performance
8. Definition of relevant concepts
9. Data collection and treatment
10. Data pre-treatment and transformations
11. Model parameters development
12. Model valuation
13. Default parameter calculation
14. Impact analysis
15. Model documentation
16. Presentation to the business line(s) impacted by the model and to the credit department
17. Technical implementation document

For a more detailed outline of each steps, please refer to the relevant directive on credit model building.

# **OVERVIEW OF PROCESSES IN PLACE**

This section depicts an overview of all relevant LBC current process related to risk management including risk modeling. This includes validation, vetting, replication, monitoring and risk materiality assessment, each of which is widely described in existing reference documents.

## Validation Process

Refer to the Bank’s reference document pertaining the conceptual validation process of risk models (Model’s conceptual validation document) and associated procedure (Model’s conceptual validation procedure).

## Vetting Process

The roles and responsibilities of internal audit in the scope of vetting process are described in their guideline (Internal guideline on conceptual validation). They are also mentioned in the policy Enterprise-Wide Model Risk Management Policy.

## Replication Process

The replication process is the one described in the Bank’s Loan review standards. This is referring to the loan review guideline that is currently under the policies and guidelines of the Bank, available on Intranet. The goal of the Loan review is to be carried out sample analysis to validate the rating generated by “ROMEO” for Commercial portfolios and ensure quality controls for Retail portfolios.

## Monitoring Process

In the scope of monitoring process, the risk solution department regularly produces credit risk reports every month, which shall then be approved by the credit committee. These reports follow every business lines portfolio, limit controls, overdrafts, exposures, authorized amounts, etc.

## Refreshing Process

The refreshing process is the one described in the Credit Risk Quantification Policy. The refreshing in question is the refreshing of the regulatory parameters (PD, LGD, EAD). The requirement is set to a minimum of once a year.

## Risk Materiality Assessment

The risk materiality is assessed in accordance to the criteria described in the Enterprise-Wide Model Risk Management Policy.

# **FINAL MODEL(S)**

The PD mortgage model, an survival analysis approach to built a tool for predicting the probability of default at the product (account) level given identified borrower and transactions characteristics that are reasonably expected to be likely influence credit worthiness. In this section is a depiction of the main elements regarding the modeling framework, including log of changes, scope, usage as well as risk rating system (RRS) and model designs. Finally, we summarize the assessment process with the model.

## Model log of changes

Per the Bank’s policies, all AIRB models should be updated on a regular basis to ensure appropriateness. Such update must account for changes since last model revision. The main changes in this revision are:

### 3.1.1 Changes in Economic Conditions

The previous model was revised in April 2017 with a calibration that the historical data spanned from October 2010 to January 2015. This year’s revision will span from January 2008 to July 2017, which covers the last recession in 2008 and 2009, as well as the recovery period since then.

There is no economic change since the last revision (last 2017 calibration).

### 3.1.2 Changes in the Modeling Methodology

The previous revision used the scorecard technical approach. This revision uses the survival analysis approach in its parametric version.

### 3.1.3 Changes in Default Definition

There has been a change in the default definition over time at the Bank. We noticed that there is no change that is specific to mortgage portfolio. These changes in default are common to the all portfolios in the Bank. Here is the last version of definition of default since the last revision.

**EPO.26 default definition Mat 29th 2017:**

**3-1 Regulatory definition of default**

Debtor default occurs in the context of one or both of the following events:

a. The bank considers the obligor unlikely to pay their credit obligations to the banking group in full, without recourse to actions such as realizing a security (if held).

b. The obligor is more than 90 days past due on any material credit obligation to the banking group. Overdrafts are considered past due once the client has breached the authorized limit, or been advised of a limit lower than current outstanding.

The criteria of this definition are used to identify material default situations and are applicable to all categories of debtors. This enables maintenance of a common understanding and operational uniformity throughout the Bank.

**3-1-1 Exceeding the 90-day threshold**

The use of the 90-day delinquency limit is one of the default definition triggers. Arrears of more than 90 days is a determining criterion for the onset of default, regardless of the rating approach used (Standard or Advanced) in the calculation of regulatory capital.

It should be noted that for Business Services clients, certain nuances apply. The 90-day delinquency limit applies to a repayment delay of a large credit defined as a percentage of the borrower’s total commitment and/or a minimum monetary threshold, determined by the Internal Ratings function together with the Credit function and/or the Bank’s Corporate Risk Committee. Moreover, for the same client, the criteria characterizing arrears vary based on the respective product.

The following factors are used to identify payment in arrears:

o Late payment is equivalent to a failure to pay the principal and interest on time

o Late payment is equivalent to a failure to pay the interest on time

o Late payment is equivalent to a failure to pay the principal on time.

**3-2 Signs indicating unlikeliness to Pay**

To identify the elements indicating unlikeliness to pay, the Bank has a list of indicators pointing to events that might lead to default by a debtor and, consequently, to financial loss for the institution:

o notice of legal mortgage / of sale

o assignment of debt resulting in a significant economic loss

o Negative contact (Investment loans)

o death;

o voluntary deposit

o voluntary bankruptcy

o petition of bankruptcy by the creditor

o proposals to the creditors

o refusal to pay

o forced restructuring

o seizure of assets

o re-aging (failure to comply with the re-aging conditions)

o write-off

o any other situation showing that the borrower will not be able to meet its commitments.

**3-2-1 Situations Unrelated to Regulatory Default**

Certain events may be a precursor to the indicators mentioned above. These events are distinguished from regulatory default because they do not meet the specific criteria of the definition. However, they need to be monitored by a Credit function officer in the event where a credit commitment would not be honoured.

o Technical default (failure to comply with contract clauses, e.g. financial ratios, security not provided)

o Litigation (contested litigation or arbitration).

**3-3 Default of Market and Derivatives products**

The definition of default stated for Business Services and SME clientele is the same as that of market products. The default criteria for this product type are therefore the same.

**3-4 Regulatory default vs. Accounting default**

If a regulatory default is triggered for a respective account, the account also experiences accounting default. As defined by the Basel guidelines; a regulatory default is always triggered before an account is classified as having experienced accounting default. Exceptions must be able to be justified to the Regulatory Authority.

A report showing regulatory default exposures and the default exposures according to the accounting definition of default is automatically generated on a monthly basis. Thereafter, discrepancies between the two definitions of default must be justified by the Credit Risk Management function.

No materiality condition is established for default identification in the retail mortgage portfolio.

### 3.1.4 Changes in Business Practises

* Acquisition of AGF trust (August 2012)

In August 2012, B2B Bank acquired AGF trust and its corresponding data was integrated into the LBC’s databases in October 2013. In terms of risk profile, the exposures from AGF are different from the LBC’s Mortgage loans. Even though the AGF loans are still included in the Bank’s portfolio, it is not a growing portfolio since the sales of new AGF loans are now discontinued.

As the AGF loans are in decline and eventually discontinued, we do not consider it appropriate to segment the Mortgage portfolio between AGF and non-AGF. However, the AGF loans are all included in the modeling process.

As per the description above, the AGF loans have already been integrated into the mortgage portfolio since 2012. These AGF loans have already been considered in the previous PD model and this proposed one. We do not think that it is going to have any impact on this newly PD model application.

* Guideline B-20: Residential Mortgage Underwriting Practices and Procedures (January 2018) (OSFI, 2018)

A “residential mortgage” under this Guideline includes any loan to an individual that is secured by residential property (i.e., one to four unit dwellings).

Since January 1st, 2018, qualification rules for mortgages have been changed. For insured residential mortgage, OSFI expects FRFIs (Federal-Regulated Financial Institutions) to meet mortgage insurers requirements in regard to debt serviceability. For uninsured residential mortgages, FRFIs should contemplate current and future conditions as they consider qualifying rates and make appropriate judgments. At a minimum, the qualifying rate for all uninsured mortgages should be the greater of the contractual mortgage rate plus 2% or the five-year benchmark rate published by the Bank of Canada (OSFI Guideline B-20). This means that it became more difficult to qualify for a mortgage since the qualification is based on a stressed scenario instead of an actual one.

OSFI is requiring lenders to enhance their Loan-To-Value (LTV) measurement and limits so they will be dynamic and responsive to risk. Under the final Guideline, federally regulated financial institutions must establish and adhere to appropriate LTV ratio limits that are reflective of risk and are updated as housing markets and the economic environment evolve.

OSFI is placing restrictions on certain lending arrangements that are designed, or appear designed to circumvent LTV limits. A federally regulated financial institution is prohibited from arranging with another lender a mortgage, or a combination of a mortgage and other lending products, in any form that circumvents the institution’s maximum LTV ratio or other limits in its residential mortgage underwriting policy, or any requirements established by law.

We do not think that these restrictions are going to have an impact on the application of the developed PD model.

## Scope of the model

The PD mortgage model is meant to assess the creditworthiness of LBC’s mortgage portfolio in terms of probability of default measuring the likeliness of occurrence of a default event for the product. The model applies to the LBC’s retail mortgage including B2B and AGF mortgages. The RMAI mortgages are out of scope.

### 3.2.1 Target portfolio

The model is used on the retail residential mortgage portfolio. This portfolio includes all owner-occupied Residential Mortgages with less than 5 units. It also includes 1-4 unit’s non-owner-occupied Residential Mortgages given that the main client does not have more than 10 total rental units at the Bank. The portfolio is defined per the following business rules

Figure 01: Retail Mortgage Segmentation



We look first if the customer is a retail one, if yes, we look if the exposure belongs to the residential mortgage operational system (source system ID\_SYS\_PROD = ‘HYP’). If yes, we then look at whether the loan is in a mortgage kit (Heloc) (IND\_SCI). if it’s not, then we look at the type of mortgage products:

'000112','000114','000115','000211','000212','000213','000214','002009','002010','002011','002012','002013','002014','002015','002016','002017','002018','002019','002020','002021','002022','002023','002024','002025','002026','002027','002028','002029','002030','002031','002032','002033','002034','002035','002036','002037','002038','002039','002040','002041','002042','002043','002044','100001' and finally the number of units that must be smaller than 5. However, we must also look at whether the client is "persistent" per the segmentation tree (A client is deemed persistent when he hits at least 11 dwelling or 2 million in commitment. When labelled persistent, a client remains a Multi-Unit client as long as he detains at least one dwelling). We have no indicator to identify them but persistent customers must be considered in the Multi unit real estate model. Hence, we excluded all the accounts that are in the Multi unit portfolio.

### 3.2.1.1 Industries covered by the model

The retail mortgage is not classified by industry. The NAICS (2012) code is not applicable for retail portfolio.

### 3.2.1.2 Products covered by the model

The model aims at estimating product-level PD within the target portfolio made of eligible clients from mortgage portfolio. The product coverage for such portfolio includes owner-occupied Residential Mortgages of 1-4 units excluding home equity line of credits (HELOC). It also covers non-owner-occupied residential mortgages of 1-4 units providing the client does not have more than 10 units (excluding principal & secondary residences).

The Mortgage Portfolio includes conventional and insured (by Genworth or CMHC) mortgages on residential properties with less than 5 dwellings.

**Conventional**: Uninsured mortgages for which the borrower must provide an initial minimal amount equivalent to 20% of the original property value (25% before June 2007).

**Genworth Insured**: Mortgages insured by Genworth for which the borrower must provide an initial minimal amount equivalent to 5% of the original property value. The mortgage is insured for the total loan amount and for potential unpaid interests. The Canadian government guarantees these loans at up to 90% of the original principal value, were Genworth unable to meet its obligations.

**CMHC Insured**: Mortgages insured by CMHC where the borrower must provide an initial minimal amount equivalent to 5% of the original property value. The insurance covers the total amount of the loan as well as unpaid interests

### 3.2.1.3 Distribution of obligors

These statistics include all the tranches of loans.

Table 05: Distribution of portfolio for AGF as of June, 2018.

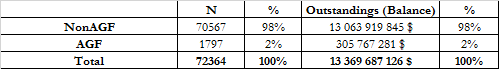
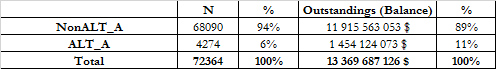


Table 06: Distribution of portfolio for ALT\_A as of June, 2018



They are alternative financing. The Bank finances clients who have had problems in the past (credit payments, temporary jobs, etc.) for a risk premium.

Table 07: Distribution of portfolio for B2B as of June, 2018.

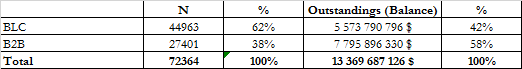


Table 08: Distribution of portfolio for Province as of June, 2018.

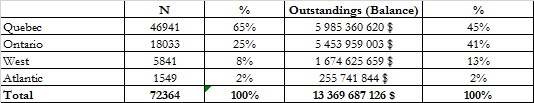


Table 09: Distribution of portfolio Insured vs Non-Insured as of June, 2018

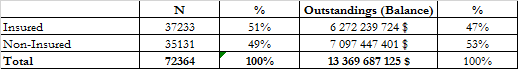
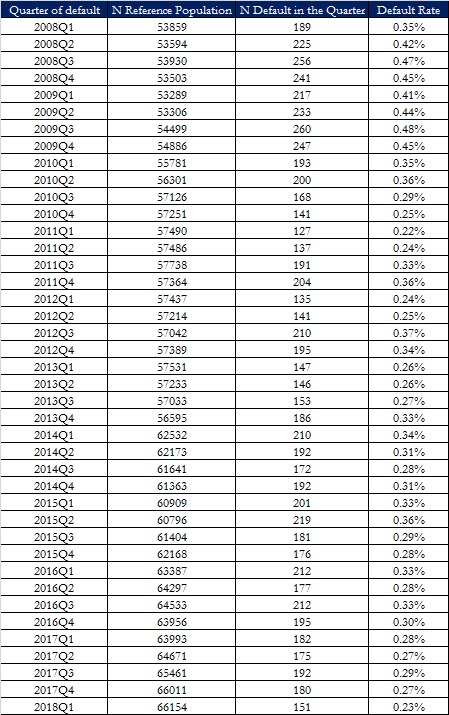


Table 10: Number of defaults and default rate per quarter

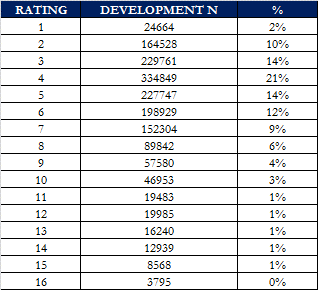


The default rate is calculated by quarter and loan (aggregation of tranches by loan). We cannot say that there is a seasonality effect insofar as there is no systematic and recurring rise in default rates for a given quarter.

### 3.2.2 Concentration analysis

The goal is to ensure that there is no excessive concentration in a rating. We look at the distribution of ratings at the development level. The table below shows that concentration is not an issue since any rating exceeds 25% according to the Backtesting framework.

Table 11: Concentration analysis



### 3.2.3 Business practises

The Laurentian Bank credit policy (section EPH.05) describes the underwriting criteria for Laurentian Bank residential mortgages.

The primary holder’s revenue and total debt ratio is also considered in the adjudication process. Any residential mortgage application that does not meet the criteria is placed into review or declined.

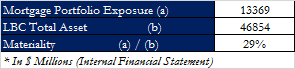
As part of Laurentian Banks credit portfolio risk management, a weekly scoring process assigns each residential mortgage issued to an individual to a risk rating. Internal explanatory variables are updated weekly and credit bureau variables are updated on a quarterly basis.

Exposures in default are assigned a retail risk rating of 17.

### 3.2.4 Materiality of the portfolio

According to the table below, the Bank’s mortgage direct exposure represents 29% of the total 46 billion assets as of June 2018.

Table 12: Mortgage portfolio exposure as of LBC total assets



## Risk Rating System design

As outlined by Basel paragraph 394, “*the term rating system comprises all of the methods, processes, controls, and data collection and IT systems that support the assessment of credit risk, the assignment of internal ratings and the quantification of default and loss estimate*”. The LBC have chosen multiple systems. On one hand, there is a retail Risk Rating System (RRS) covering residential mortgage loans, qualifying revolving retail exposures (credit cards, credit lines and overdrafts), and other retail exposures. On the other hand, there is a non-retail RRS or wholesale RRS, covering corporate, bank and sovereign asset classes. Such classification is consistent with established bank practise. In the scope of PD mortgage modeling framework, not only is it important to define the Bank’s Risk Rating System design, but it is also interesting to describe how the residential mortgage portfolio PD is calculated using the underlying behavior model. This model reflects both borrower and transaction specific variables. A PD is calculated for each residential mortgage exposure based on information from the primary account holder’s credit bureau and internal characteristics. Additionally, specific product level variables are used in the model. The main risk drivers cover delinquency, borrower and transaction specific risks to explain probability of default as required by Basel II.

### 3.3.1 RRS/SRRS

For modeling purpose, it is important to distinguish between a RRS, a SRRS and a risk model. Not only does the RRS comprise credit risk parameters (PD, EAD and LGD), but it also includes all systems that support the assessment of credit risk, the assignment of internal risk ratings, and the quantification of default and loss estimates. Also, banks typically manage their credit-related business in broad business lines or portfolios, each of which may encompass a variety of specific borrower and exposure types. Although specific business line and portfolio delineations can greatly vary across individual banks, key common features that define a business line or portfolio may be related to the nature of the borrower (e.g. governmental, corporate, household, etc.), the nature of the transaction (e.g.: instalment, revolving, mortgage, real estate...), or a combination of both.

The design and features of internal risk rating systems and internal default loss estimation process at the LBC also reflects this broad management approach. According to LBC’s credit policies (EPO.55, May 29th 2017), the Bank’s rating systems are based on the regulatory requirements governing portfolio segmentation of portfolios. These requirements establish asset classes and sub-classes. Multiple rating methodologies/systems may be utilised within each sub-asset class, therefore introducing sub-risk rating systems (SRRS) with associated credit risk models within sub-classes/systems. The criteria used in defining the Bank’s RRS and SRRS were the type of borrower and other borrower risk characteristics, as well as transaction-specific risk characteristics such as product and/or collateral types. This is in line with Basel paragraph 397 requirements on corporate, sovereign and bank RRS design. As highlighted in LBC general credit policies, the following specific classification criteria were also considered in defining RRS/SRRS so as ensure homogeneous and coherent classification of credit risks:

* Harmonization of credit risk parameters with the Bank’s effective operational processes;
* Statistical characteristics of the portfolios:

The conclusion that can be drawn from the statements above is that SRRS and credit risk models are only parts of a RRS. The PD mortgage model uses a survival analysis approach to built a relationship between the PD of retail mortgage exposures and identified creditworthiness determinants such as Bureau Credit or transaction information. This is a predicting tool that shall be used for PD forecasting of any LBC’s retail mortgage except RMAI portfolio.

As shown on the figure below, the Bank has delineated its portfolio into 13 distinct eligible SRRS for AIRB credit risk modeling. The SRRS 4 is the one in the scope of the PD mortgage model. Retail SRRS (1, 2, 3, 5, 12) as well as SRRS 7, 8, 9, 11 and 13 are therefore excluded.

Figure 02: LBC risk rating systems

Wholesale

**SRRS 7 :** Commercal Real Estate

**SRRS 13 :** Specialized lending

**SRRS 11 :** Commercial

Retail

**RRS 4 :** Residential Mortgage

**SRRS 5 :** Investment loan

**SRRS 3 :** VISA

**SRRS2 :** Line of Credit/Overdraft

**SRRS 1 :** RSP/Direct/Indirect

**SME/Retail**

**SRRS 9 :** Commercial

**SRRS 12 :** HELOC

**LBC AIRB Risk Rating Systems**

Sovereign

**SRRS 8 :**

Sovereign

Appendix 14 depicts the decision tree for assigning borrowers to each portfolio including mortgage sub-segment, showing evidence that borrowers’ allocation across portfolios is not cherry-picking. This segmentation is the one automated within the Bank’s platforms for borrower risk rating (ROMEO/SAS).

### 3.3.2 Time horizon of the parameter

The probability of default as an estimate of the likelihood that the default even will occur within a horizon. The time horizon of the PD is of one year. However, the estimation of one-year PD requires the use of historical defaults over a period from 2008 to 2017. Otherwise even though the time horizon is of one-year, a long-term horizon (from 2008 to 2017) have been considered for as long-run PD for calibration purposes. Therefore, capturing the average default experience of mortgage exposures over a mix of economic conditions from 2008 to 2017 including economic downturn conditions (see section 5.6.1) is sufficient to provide a reasonable estimates of mortgage average one-year default rates over the economic cycle.

### 3.3.3 Rating philosophy

The rating philosophy describes the dynamic characteristics of rating systems and refers to how the borrower rating is affected by the choice of economic, business and industry conditions that are considered in the rating process[[1]](#footnote-1). It is necessary to define and clarify the rating philosophy underlying the PD Mortgage model. Rating systems can be classified roughly in two categories depending on the way different systems utilize available obligor-specific and aggregate information[[2]](#footnote-2):

• Point-in-time (PIT) approaches;

• Through-the-cycle (TTC) approaches.

PIT approaches tend to produce ratings that are responsive to changes in current business conditions. Otherwise, PIT systems tend to focus on of the borrower’s current condition and/or most likely condition over the course of a chosen short time horizon, typically one year or less[[3]](#footnote-3). Therefore, PIT ratings will tend to adjust quickly to a changing economic environment: tendency to improve during business expansions as most obligors’ creditworthiness gets better, and to deteriorate during recessions. Under PIT rating systems, obligors are slotted into risk buckets based on the best available information about their current credit quality. Obligors are rapidly transitioned to new buckets as their current credit quality changes. The mapping of the PD to the risk rating is kept constant. A PIT system can be defined by volatile ratings, due to frequent rating migrations, but constant PDs per rating grade.

TTC approaches tend to produce ordinal rankings of obligors that are relatively stable over the business cycle. Thus, they focus on the obligor’s performance at the trough of business cycle or during adverse business conditions. A “risk bucket” or rating grade can be defined as the group of obligors sharing similar PDs. The characteristics of the set of used information to assign an obligor to a risk bucket determine the difference between PIT and TTC rating system. Under TTC rating systems, obligors are assigned to rating grades based on evaluations of their abilities to remain solvent in the trough of business cycle or during severe stress events. Because they place weight on stress conditions rather than current conditions, TTC ratings tend to change less often than PIT ratings and they tend to be more stable over the business cycle. Consequently, TTC ratings will tend to remain more-or-less constant as macroeconomic conditions change over time.

The rating philosophy of this model is point-in-time according to the above definition. Explanatory variables are updated on weekly & quarterly intervals and therefore reflect the most recent information available to evaluate the default probability of the exposure. Clients can therefore be expected to migrate with changes in their credit quality. Under this approach, observed default rates for particular ratings should tend to be more stable over a standard economic cycle (five years).

### 3.3.4 Rating dimensions

As outlined in Basel paragraph 396, a qualifying IRB rating system must have two separate risk dimensions, the risk of the borrower default and the transaction-specific factors. The PD model is perfectly in line with this paragraph since the estimate of the probability of default is based on customer-specific information, credit bureau information, and loan characteristics.

### 3.3.5 Rating structure

The retail rating goes from 1 to 16. The rating 17 is assigned to the defaulted accounts.

### 3.3.6 Rating Criteria

For Retail portfolio, there is no rating criteria as wholesale rating criteria. We must note that the retail mapping table, from PD to ratings does not have the same purposes as the Commercial one. This retail table is only for analysis purposes for which we need a rating class for PDs. Please, see section 7.5.

## Model Usage

The residential mortgage behavioral & PD model is used throughout the Bank for a wide range of activities: credit risk quantification, credit risk management, reporting to senior management, accounting, etc.

The model is used in the following contexts:

• Capital Calculation: The probability of default will be used for capital estimation.

• Stress testing: The PD model is used in stress testing the residential mortgage portfolio. The retail risk ratings generated by the model are tested under various scenarios.

• Reporting: A parameter estimation document is produced annually and authorized by senior management to be used in all activities. Expected loss estimations and credit quality distributions are reported to senior management on a quarterly basis through the credit report. Stress testing results are published once a year and reported to senior management.

• Accounting: The PD model is used in the accounting process for the calculation of the collective allowance with adjustments for IFRS compliance, whose results are recorded in the Bank’s financial statements.

## Model Design

This section portrays an overview of the model design participants, including the heads of Internal Ratings & Credit Risk Management department, and associated analysts in charge of the model building. We also depict the steps surrounding the model design itself.

### 3.5.1 Model Design Participants

The development of the PD mortgage model has therefore involved the following participants from LBC:

Table 13: Model design participants overview

| **Participants (LBC)** | **Role** | **Responsibility** |
| --- | --- | --- |
| Catherine LUSSIER | Assistant Vice President, Internal ratings | * High level supervision |
| Ablaye Thiaw | Principal advisor- Credit Risk Management | * Modeling & Documenting |
| Credit Solution Team |  | * Provide expert opinions on the selection of the final variables and on their integration at the database level |

### 3.5.2 Model Design Overview

The PD mortgage model was developed internally with proprietary data from LBC. These databases contain enough historical information and defaults to develop, calibrate and validate the model.

Several methodological approaches were discussed. Those which are the most used (logistic regression, survival models) or those least used (linear regression, panel models, discriminant analysis, neural networks, decision trees). We opted for a survival approach in its parametric version because it is widely used in the industry for PD modeling, easily adapted for IFRS 9 requirements and easily implementable.

### 3.5.3 Contributions of the business lines

There is no contribution of the business lines to the mortgage DP model development process.

## Summary of PD assessment process

The final estimates from model development are the following:

The PD mortgage was modeled within the framework of a gamma survival model (GSM). The 1-year PD is therefore evaluated as the probability of survival for 365 days.

For GAMMA distribution:

where is the cumulative distribution function for the gamma distribution with parameters, having:

(is the shape)

( is the scale)

b =

t = 365 for PD estimation

= calibration factor

Here is the PD mortgage final model:

*λ* = 10.81

+0.4637 \* WOE\_NB\_MOIS\_ECHEANCE\_AMORT

+0.8381 \* WOE\_NB\_JR\_DELQ\_MAX\_6MOIS

+0.4688 \* WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN

+0.4657 \* WOE\_BC\_NB\_INQ\_LAST12

+0.3573 \* WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC

+0.2837 \* WOE\_BC\_AVG\_MOS\_BC\_TRD

+0.368 \* WOE\_BC\_MOS\_SNC\_RCNT\_DELQ

+0.2321 \* WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT

+0.13 \* WOE\_PASF\_NB\_PRDT\_MC\_ACTIF

The scale parameter, *σ*, was fitted to the following value:

*σ* = 2.2761

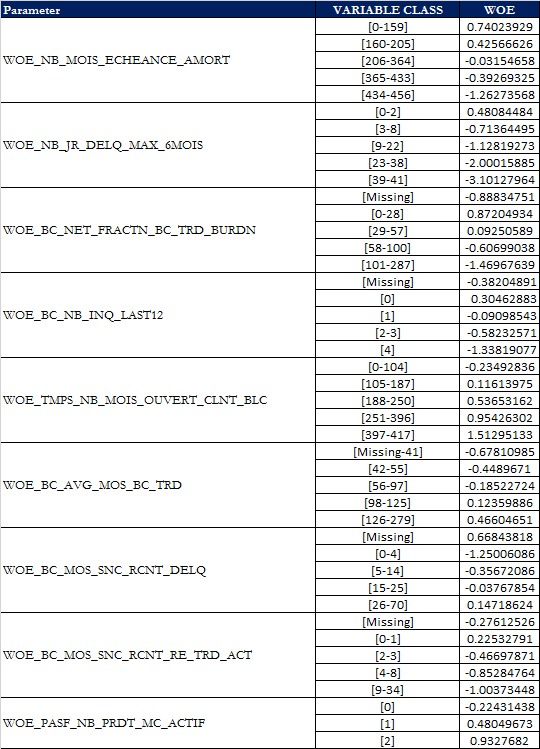
The shape parameter, *δ*, was fitted to the following value:

*δ* = -0.2376

The calibration factor, , was fitted to the following value:

= 0.33398

Table 14: Variable class & WOE



**Final Model variable description:**

**WOE\_NB\_MOIS\_ECHEANCE\_AMORT**: Number of months before end of amortization.

**WOE\_NB\_JR\_DELQ\_MAX\_6MOIS**: Maximum number of days of delinquency in the last 6 months.

**WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN**: Ratio of total balance to HC/CL for all bankcard trades.

**WOE\_BC\_NB\_INQ\_LAST12**: Number of inquiries in the last 12 months.

**WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC**: Number of months since the customer is open at BLC.

**WOE\_BC\_AVG\_MOS\_BC\_TRD**: Average number of months since date opened on all bankcard trades.

**WOE\_BC\_MOS\_SNC\_RCNT\_DELQ**: Number of months since the most recent delinquency.

**WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT**: Number of months since latest activity of revolving trades.

**WOE\_PASF\_NB\_PRDT\_MC\_ACTIF**: Number of active margin accounts.

# **DEFINITIONS**

The following sections define all other relevant concepts in PD modeling as per LBC’s policies, including risk type, re-aging, overdrafts and cured accounts. First of all, we properly define the type of risk in the scope of the PD mortgage model.

## Definition of Risk Type

The PD mortgage model aim at assessing the creditworthiness of mortgage loans given predetermined risk drivers. The mortgage PD is at account level and borrowers in the scope of the model are exclusively LBC’s customers and its subsidiaries like B2B Bank. The risk assessment methodology developed for mortgage PD modeling is mainly concerned with mortgage default risk. So, mortgage risk can be defined as the risk that the account defaults on principal or interest. Assessment of mortgage risk therefore focuses on both the inability to repay as well as the unwillingness to repay.

## Definition of Default

If we want to meet the regulatory requirements, we must make sure the following used definition of default is Basel-compliant, including proper treatment of cured accounts. This definition of default is also the same as the LGD and EAD parameters.

**Definition used in the Model: LBC’s default definition (EPO.26 – General credit policies 2017 version)**

According to LBC, a debtor default occurs in the context of one or both following events:

* The bank considers the obligor unlikely to pay their credit obligations to the banking group in full, without recourse to actions such as realizing a security (if held);
* The obligor is more than 90 days past due on any material credit obligation to the banking group. Overdrafts are considered past due once the client has breached the authorized limit, or been advised of a limit lower than current outstanding.

The criteria of this definition are used to identify material default situations and are applicable to all categories of debtors.

More precisely, Regulatory default may be triggered because of the following indicators which will be detailed herein. These indicators help to target real defaults and to safeguard the Bank against them. These indicators are:

* An arrear of 90 days or more on principal and interest;
* Unlikeliness to pay;
* The Bank sells the debt at a loss (e.g. 80%-90% of the commitment);
* The Bank requests the payment of a revocable debt for example, to the debtor and the guarantor, without receiving payment at final payment date;
* An impaired loan status is assigned / IFRS 9 provisioning process;
* The Bank takes a provision or performs a write-off due to a credit risk deterioration;
* The Bank petitions for a client’s bankruptcy.

The definition of default used for the mortgage PD model also reflects the operational default definition used in the collection and recovery process and the calculation of credit provision.

No other materiality threshold is established for default identification as one can see in certain documentations (a threshold of $100).

**Reconciliation to the Reference Definition**

OSFI defines default in guideline A-1 of paragraph 452: “A default is considered to have occurred regarding a particular obligor when either or both of the two following events have taken place.

* The bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realising security (if held).
* The obligor is past due more than 90 days on any material credit obligation to the banking group. Overdrafts will be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than current outstanding”.

In paragraph 453, OSFI defines the elements to be taken as indications of unlikeliness to pay as follows:

* The bank puts the credit obligation on non-accrued status.
* The bank makes a charge-off or account-specific provision resulting from a significant perceived decline in credit quality after the bank taking on the exposure.
* The bank sells the credit obligation at a material credit-related economic loss.
* The bank consents to a distressed restructuring of the credit obligation where this is likely to result in a diminished financial obligation caused by the material forgiveness, or postponement, of principal, interest or (where relevant) fees.
* The bank has filed for the obligor’s bankruptcy or a similar order in respect of the obligor’s credit obligation to the banking group.”

By comparing paragraphs 452 and 453 and the definition of default of the Bank, we see that the Bank is in the same line with the regulator. Therefore, we can say that the definition of default used in the model is Basel compliant.

Meanwhile, the definition of default used for the mortgage PD model is consistent with the one used for mortgage LGD modeling.

## Definition of Loss

Definition of loss is only applicable to LGD model.

## Definition of Costs

Definition of costs is only applicable to LGD model.

## Definition of Discounting

Definition of discounting is only applicable to LGD models.

## Definition of Time to Resolution

Definition of time to resolution is only applicable to LGD models.

## Definition of Re-aging

According to the bank’s policy (EPO.26, May 29th 2017), re-aging is intended to prevent exposure default due to a retail borrower’s unfortunate and temporary financial situation. The exposure in question is between 1-90 days’ delinquent, yet not in default. When a re-aging is performed, a current status is reassigned to the exposure without collecting the total amount (principal, interest, fees) stipulated under the loan covenant.

Re-aging is a manual process carried out by the officer whereby the Bank must communicate directly with the borrower, who needs to demonstrate:

* Their willingness to fulfill their commitment; and
* Their capacity to fully reimburse their loan

Re-aging does not necessarily mean default. Defaults happen only when the re-aging conditions are not respected. In this case, the is going to captured in the model.

## Definition of Overdraft

Overdraft does not exist in the mortgage portfolio.

## Definition of Cured Account

A cured account has met the qualification for being in default, but has finally been repaid in full with no possibility of loss. The account goes back to the current portfolio and the bank suffers no loss. In the scope of mortgage PD modeling, an account that emerges from default gets back in the current portfolio.

Cured accounts are used in the modeling process. These will be present at a specific “observation window” if and only if, on the date of the observation window, the account was already cured. If, for example, an account is cured in August 2011, it will not appear at the July 2011 observation window, but will appear in the October 2011 window and the ones following.

## Definition of seasoning

Banks are encouraged to capture seasoning effects in their risk models. Seasoning effect is the tendency for default hazard rates to increase during the early months in the life of a loan and tend to flatten out or decline after. In other words, from a probability of default perspective, the seasoning effect refers to the effect of time on the likelihood of default. Seasoning occurs when the risk of default for the loans in the portfolio/pool is not uniform over time. For instance, long-term retail exposures characterised by seasoning effects are loans that experience a peak in default rate several years after origination. For instance, a product where the default probability is much greater if the loan is in the second year of its lifecycle or relatively low default rates in their first year, rising default rates in the next few years, and declining default rates for the remainder of their terms. Especially for retail exposures with long effective maturities such as credit cards and mortgage, the behaviour of the default rate throughout the loan age or “seasoning” is well-known to play a role in predicting defaults. Seasoning effect is also observed in the LGD.

# **DATA USED IN THE MODELING PROCESS**

This section summarizes the sources of the data for Mortgage PD modeling. We also discuss the issues regarding the data refreshing, pre-processing and processing, representativeness of the data and the historical data coverage.

## Data Sources

The Mortgage PD model was entirely developed with internal data. Three sources have been used in the process of building the development population:

***DMC & DTM*** (Credit Datamart): Theses sources contain relevant data covering the loan’s life, on a monthly basis. They were used for the following steps:

* Selection of the mortgage portfolio (ID\_SYS\_PROD, TP\_CLIENT, CD\_PRODUIT)
* Creation of quarterly cohort.
* Provide information of the credit bureau and other internal variables.
* Selection of input variables.

***EDC*** (Corporate Database): This table contains all historical data of the Bank. Data are classified according to a start date and an end date. It was used for the following steps:

* Selection of the mortgage portfolio (IND\_SCI, NB\_LOGEMENT, NB\_MAGASIN)
* Provide information of mortgage credit and other internal variables.
* Selection of input variables.

***LBC’s Default Database***: creation of the table “Unification” that serve to identify defaulted accounts. It was used for the following step:

* Identification of defaulted accounts.

## Data Refreshing Process

As we said earlier, the data are updated every month. This monthly update is very practical since it allows us to have more latitude in the choice of the quarterly cohort. However, credit bureau information is updated every 3 months.

## Data Processing

***Historical quarterly data***

The starting table is the Credit Datamart tables (DTM), a monthly table that allows us to choose our quarterly cohorts. Then, to get other information that we needed, either to identify the mortgage portfolio, or to identify potential variables for modeling purposes, we join the Corporate Database tables (EDC) making sure that each date of cohort (DTM\_DT\_CLND) is between the start date and the end date of the EDC tables.

This historical table is also used to identify defaults in the mortgage portfolio. We make a join with the default table by EDC\_ID\_DAFF (account number) and all accounts that match will be treated as mortgage defaults. We do this because for defaulted accounts from radiation, information on the product in default is not available.

***Default data***

The default table is created to meet the Bank's regulatory requirements for identifying defaulted borrowers / accounts. This table of default includes the 90 days defaulted account, unlikely to pay and bankruptcies. No materiality amount is set for the mortgage defaults.

The default table is made with an entry date and an exit date of default. An account may be in default several times in a year (the account is in default, the amount past due is paid and the account is cured, but may be in default again 90 days after).

For modeling purposes, with a survival approach, the default is counted only once and the first occurrence is considered. However, for the calibration of the model where a default rate per cohort (year) is calculated, the default is counted once within a cohort, but if the account has failed in another cohort, then it will be considered new but in the cohort where it appeared.

For model development, 5647 defaults were used.

**Measure of time**

In the development process, an “overlapping windows data structure” was used to estimate the PD parameter. One of the advantages of this method is to increase the number of observation windows used in the modeling process, as the same accounts may figure in several windows. Another advantage of this method, not least, is to take rapidly into account all relevant changes of the client behavior. The same observation date is used for all observations during the performance period associated to that window. The figure below aims to illustrate how this process operates:

Figure 03: Overview of the methodology of survival time calculation



**Case A**: The loan is open before the first observation date and is current at the analysis end date. The covariates (number days of delinquency, number of months since open date, etc.) are measured at the observation date. This loan has never been in default situation; time is measured as number of days between end date of analysis and first observation date and is censored. The loan will be present in all windows.

**Case B**: The loan is open before the first observation date but was closed before the end date of analysis. This loan has never been in default situation; time is measured as number of days between closed date and first observation date and is censored. The loan will be present in all windows between first observation date and closing date.

**Case C**: The loan is open after the first observation date and is current at the analysis end date. Time is measured as number of days between open and end analysis dates and is censored. The loan will be present in all windows between open and closing date.

**Case D**: The loan is open after the first observation date but was closed before the end of analysis date. This loan has never been in default situation; time is measured as number of days between open date and the closing dates and is censored. The loan will be present in all windows between open date and closing date.

**Case E**: The loan is open before the first observation date and was in default before the end date of analysis. Time is measured as number of days between default date and first observation date and is not censored. The loan will be present in all windows between first observation date and default date.

**Case F**: The loan is open before the first observation date and was in default 90 days and then the account is “cured”. Time is measured as the number of days between the first default date and first observation date and is not censored. The loan will be present only in the windows between first observation date and the first default date.

We have a total of 35 windows (January 2008 to July 2016) for the PD mortgage model. An account may sometimes fall into several spans and, a single default accordingly. For instance, defaults that occurred in July 2010 are accounted for in all windows before that date. Accounts that do not default are presented in each window since their opening.

In the final database, all time dependent covariates were measured every 3 months for a total of 35 windows. For example, for account “A”, we may have up to 35 measures of covariates; for accounts “C”, “D” or “E” we will have less than 35 measures, depending of opening and closing/default date.

In order to have a window, we have to have 12 months of data before the window date. For example, for the first window of January 2008, we take the current portfolio as of January 2008, but we observe the behavior of these clients on the January 2007-January 2008 period (for example, maximum delinquency over the last 12 months). The advantage that we have with the way that LBC’s databases were built is the fact that at the observation date we can have the variables 3, 6, 12 months before (for example NB\_JR\_DELQ\_MAX\_12MOIS). After that, we look at which account (that were in the current portfolio on January 31st, 2008) have defaulted after January 31st, 2008.

The figure below shows how we define the survival time and process to the censoring:

Figure 04: Survival Time and Censored



Time was measured as follows:

i. In the case of no default and no closure of the account before the end of study (EoS – July 31st, 2017), survival time would be the number of days between the EoS and the window date (Wdw).

ii. In the case of no default, but closure of the account (Cls), survival time would be the number of days between Cls and Wdw.

iii. Finally, in the case of default (Dft), survival time would be the number of days between Dft and Wdw.

All information available throughout the life cycle of accounts was used, therefore, all relevant available information was used in the modelling process.

Figure 05: Illustration of datasets layout

The final dataset (“Treat.Variable\_Treatment\_Mortgage”) is constructed as follows (illustration with 2 observations dates)



As illustrated by the above figure, each line of the “Treat.Variable\_Treatment\_Mortgage.sas7dat” dataset contains the characteristics (variables such as number days of delinquency, credit bureau information, etc.) of an observation at a given observation date.

**Censoring**

Censoring come in many forms and occurs for many different reasons. The most basic distinction is between left censoring and right censoring. An observation on a variable *T* is right censored if all we know about *T* is that it is greater than some value *c*. In survival analysis, *T* is typically the time of occurrence for some event, and cases are right censored because observation is terminated before the even occurs.

Symmetrically, left censored occurs when all we know about an observation on a variable *T* is that it is less than some value. In the context of survival analysis, left censored is most likely to occur when you begin observing a sample at a time when some of the individuals may have already experienced the even.

In the literature, right censoring is far more common than left censoring, and most computer programs for survival analysis do not allow for left censored data.

For mortgage PD model, the right censoring is use that’s why we count the default on time and its first occurrence. This method is the simplest and most common. The other kinds of censoring require some potentially dubious assumptions.

### 5.3.1 Exclusions and Manipulations

The following exclusions and manipulations apply to the models:

* **90 days past-due at observation data**

In treating the >90 days overdue defaults, the decision was made to exclude from certain windows any account that was, at the time of the observation, in default, even though it could cure within the year. On the other hand, had it been cured at observation, it would have been included in the database. For example: If a >90days default occurs December 2010 and account becomes “cured” in February 2011, it will not be present in the January 2011 window as it is not in current portfolio at that date, but will be present in the April 2011 window because it is in the current portfolio. This is made to avoid calculating a survival time which would have included both a “curing-time” and a real “survival-time”.

* **Closed accounts**

All accounts at the observation date with status “closed” are excluded from the development sample. An account is closed because it’s transferred to the other institution, or the loan has come to term, or the account is in default and the guarantee is sold, etc. Excluding closed accounts excludes write-off or bankrupt accounts. When there is a write-off, the status of the account is changed in the databases and becomes closed even if on the other side the recovery process continues. We can also check it with the variable RADT\_MNT\_RADIE which must always be missing.

* **Inactive accounts**

All accounts at the observation date with status “inactive” are excluded from the development sample. An account is inactive because it’s opened but not disbursed yet. The outstanding is zero for the Mortgage.

* **Co-borrowers**

The initial database contains all the borrowers including the co-borrowers. For modeling purposes, we excluded co-borrowers. This is consistent with how defaults are assigned once the account is in default. Indeed, in the Default Management System (DMS), for Retail portfolio, it is the product that defaults and when the product that is in default belongs to two customers (principal and co-borrower), the default is attached to the main customer (the principal). To be consistent, we will then take the information of the main borrower as part of the modeling.

The same treatment is applied to the calibration and back testing exercise.

* **Other manipulations and treatments**

We can have in the mortgage portfolio several accounts or sub-accounts (tranches) that belong to the same mortgage. When a tranche is in default, all the other tranches under the same loan will be in default. In order to identify and regroup all the tranches belonging to the same loan we use ID\_GRNT as an indicator for the loan. The PD that we develop is a PD by loan and not a PD by tranche.

Sometimes tranches belonging to the same loan have different characteristics, such as different interest rates, number of delinquency days, and so on. In this case, choices must be made to aggregate at the loan level. For example, for interest rates, we would take the maximum that seems to be the most conservative choice. For delinquency, we take the maximum number of days of delinquency. In a situation where we have two tranches, one the purpose of the loan is "refinancing", the other is missing, so we assume that the loan is "refinancing".

Other manipulations have been made on some variables. For example, the internal delinquency variables that come from the EDC databases. When these variables (for illustration NB\_JR\_DELQ\_ACT) are missing, it is not a true missing but a zero.

By the way, in the EDC data warehouse, there is the delinquency table (EDC\_CRDT\_DELQC). At each update, a join of all accounts is made with this delinquency table. If there is a match it is that the account has a delinquency otherwise the system returns a missing value (which means that there is no delinquency). So, all internal delinquency variables with a missing value are set to 0.

The credit bureau variables may contain particular values that should be considered as missing and should not be included in treatment of outliers in the next section. Specifically, this reasoning applies to variables that should be strictly positive and have values of (-1). In addition, some variables contain (999) or (99999) values that are clearly missing values and do not represent a valid data entry.

### 5.3.2 Treatment of Missing

There is no commonly accepted recipe for treating missing values. Nevertheless, some methods are often used depending on the nature of the dataset, the model, and the event that is modeled. We present some of these methods below:

1. **Case wise deletion**

The method consists of deleting every observation for which a value is missing. Given the big size of missing values for the bureau credit information, this method does not appear well suited for our dataset.

1. **Mean substitution**

The method consists of replacing missing values with the mean of the variable. The main advantage of such a method is that it allows every observation in the dataset to be used. Furthermore, it ensures that the method has no effect on the central tendency, avoiding any bias. However, by substituting the missing values with the mean of the variable, it artificially reduces the standard deviation of the variables, making the regression coefficients more significant than they should be.

1. **Regression substitution**

The method consists of fitting a regression that predicts the missing value based on the value of other variables in the dataset. The imputed values being calculated as a linear combination of other included variables causes the variance to be underestimated, with the same consequences as in the mean substitution method.

1. **Including missing values as it is**

The method consists of treating missing values as a separate attribute, grouped and used in the model as an input. The model can then be allowed to assign weights to this attribute. In some cases, this assigned weight may be close to the “neutral” or mean value, but in cases where the weigh is closer to another attribute, it may shed light on the exact nature of the missing value (See Naeem Siddiqi: Developing and Implementing Intelligent Credit Scoring).

For PD mortgage model, we chose the method that include missing value as it is for several reasons. First, given the big size of the number of missing values in our dataset, we couldn’t afford a case wise deletion. Second, the methods a), b) and c) assume that missing data holds no value, that no further information can be gleaned from analyzing the missing data. This is not necessary true, missing values may be part of a trend, linked to other characteristics, or indicative of bad performance. Missing value are not usually random. For example, those who are new at their work may be more likely to leave the “Years at Employment” field blank on an application form. If characteristics or records with missing values are excluded, none of these insights can be made. Therefore, it is recommended that missing data be included in the analysis, and be assigned weight in the model. This method recognizes that missing data holds some information value, and that there is business benefit in including such data in our analysis.

### 5.3.3 Treatment of Outliers

An outlier is defined as an observation that is abnormally distant from the rest of the distribution. Outliers must be assessed carefully since regression models are sensitive to these values that can bias the coefficients estimates. An observation could be far in the tails of a distribution for two reasons: it could reflect an extreme event or it could be due to a data entry error. In the first case, even if the observation is extreme, it contains useful information, particularly if we try to model creditworthiness deterioration. In the second case, the observation clearly must be corrected.

There are commonly two approaches to deal with outliers. The first approach is data trimming, i.e. deleting the outliers from the dataset. However, this method may delete some valuable information about particular accounts. The second approach is data winsorization, i.e. setting the value of outliers to a certain cut-off level. Therefore, we can limit the range of standard deviation and preserve the business benefits of the data at the same time. In this work, we apply the second approach and we compute the z-score to identify the variables containing potential outliers.

The z-score is computed as follows |z-score| = | | > 3.29, where is the sample average, and std is the standard deviation. The cut-off level is set to 3.29, which is associated to a 0.001 p-value of the standard normal distribution. The z-score is influenced by extreme values, especially if the skewness is important. Therefore, the percentage of outliers changes less than with other methods as Tukey or Median. It is important to note that the main goal is to establish a cut-off for continuous variables only if it is considered necessary to avoid eliminating the real data values.

We use the SAS Standard Procedure to standardize continuous variables to a standard normal distribution. Z-scores are calculated and values exceeding ±3.29 are considered outliers in order to correct the lower and upper bounds of variables containing extreme values. For each potential variable, we perform a special analysis of the outliers in order to set a lower or upper bound to each of the outliers. If an outlier is identified using the z-score method, the less restrictive, the cut-off is calculated as the 99 percentiles. All of these boundaries will be used at the implementation of the model.

### 5.3.4 Treatment of Cured Accounts

No special treatment is done for cured accounts in PD modeling. Even if the account is cured, it is counted as default. For the calibration, if an account is cured and falls in default again, the default will be considered again if it appears in a different performance window.

### 5.3.5 Re-ageing treatment

The use of re-aging should be performed in a prudent manner to not impact the performance and the portfolio delinquency status. Re-aging does not automatically lead to a default. A certain number of re-ageing is allowed by the Bank for accounts that meet the conditions. A re-ageing account will be in default if the account does not respect the terms of the arrangement. To ensure compliance, a "re-aging" indicator is included in the Default Management System (SGD). It is therefore possible to measure and monitor re-aged loan performance.

### 5.3.6 Overdraft treatment

There is no overdraft treatment in the mortgage PD model because overdraft is not allowed to the mortgage portfolio.

### 5.3.7 Data Segmentation

Except the segmentation to determine the mortgage portfolio, no other segmentation was performed. There is no segmentation between AGF & Not AGF as it was done in the last revision:

* Even though AGF residential mortgage loans are still present in the Bank’s portfolio, this is not a growing portfolio, i.e. sales of new AGF loans are now discontinued. They represent 2% of the mortgage portfolio as of June 30, 2018 and they are decreasing.
* The variable AGF – NO AGF is transformed as a binary variable and tested in the modeling process. The variable is not retained because it was not significant.
* Even if there is no segmentation, the AGF effect is captured in the other variables, as can be seen in the average PDs in the portfolio as of June 30, 2018. Average PD of AGF accounts (2.23%) is 82% higher than the average PD of non-AGF (1.22%).
* Another way to ensure that the model applies to AGF accounts is to do the Backtesting exercise on AGF accounts. Indeed, the model is well calibrated for AGF accounts as can be seen in the appendix 17.

The same consideration was done for ALT\_A and B2B. The variables are transformed as a binary variable and tested in the modeling process. They are not retained because they were not significant.

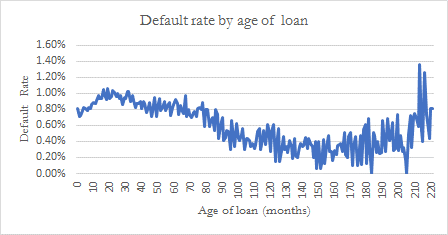
### Accounting for Seasoning Effect

#### Seasoning effect analysis

This section only applies to PD estimation of retail exposures (Basel paragraph 467) especially those displaying a material seasoning pattern, meaning that the exposures have relatively low default rates in their first year, rising default rates in the next few years, and declining default rates for the remainder of their terms. Consequently, in such cases, there is a material relationship between the time since origination of exposures and the long-run (average one-year) default rate of the portfolio. Otherwise, loan age (time on book) plays a role in predicting defaults. Given the one-year IRB horizon, the regulatory framework therefore requires the Bank to adjust its PD estimates upward for anticipated seasoning effects.

For mortgage PD, this seasoning effect described above can be analyzed by computing the relationship between the age of the account and the default rate.

Figure 06: Illustration of seasoning effect



Even if it is not clearly defined towards the end (from the 210th month), we can see a trend that reflects a seasoning effect. with defaults increasing at the beginning of the loan, stabilize a little and then drop.

#### Seasoning effect treatment

We do not have the variable age of loan in the final model (the variable was tested but not selected in the final model). Nonetheless, we have a similar variable in the model that better capture the impact of age on the loan which is: BC\_AVG\_BC\_MOS\_TRD. We do not need an addition margin of conservatism.

### 5.3.7 Sampling

We use the SAS Procedure PROC SURVEYSELECT with stratified sampling in order to split our data in a development and test samples.

PROC SURVEYSELECT provides methods for both equal probability sampling and probability proportional to size (PPS) sampling. In equal probability sampling, each unit in the sampling frame, or in a stratum, has the same probability of being selected for the sample. In PPS sampling, a unit’s selection probability is proportional to its size measure. For details about probability sampling methods, see Lohr (2009), Kish (1965, 1987), Kalton (1983), and Cochran (1977). We use the simple random sampling without replacement.

PROC SURVEYSELECT can perform stratified sampling by selecting samples independently within strata, which are nonoverlapping subgroups of the survey population. Stratification controls the distribution of the sample size in the strata. It is widely used in practice toward meeting a variety of survey objectives. For example, with stratification we can ensure adequate sample sizes for subgroups of interest, including small subgroups, or you can use stratification toward improving the precision of the overall estimates. The data are stratified by Date\_obs, Province, AGF, ALT\_A and B2B.

The method was to separate our population in two parts:

* In order to leave an out-of-sample data for testing the models, a sampling process occurred; 80% of the portfolio was selected for development (1608167 observations)
* The remaining 20% was considered as our out-of-sample data on which we test the performance of the model (401711 observations).

## Data Pre-treatment

Data pre-treatment or pre-processing is an often neglected but important step in a modeling process. Knowing that the quality of data affects the modeling result and in order to improve quality, the initial dataset should be pre-processed so as to improve efficiency and ease the modeling process. Data pre-treatment mainly deals with the transformation of the initial dataset, which facilitates a preliminary selection of risk factors.

### 5.4.1 Data Transformation

The use of different data sources sometimes result in information having various formats, units and scales. Especially for models using advanced statistical techniques, some data transformation is therefore required to improve the quality of modeling data and safely perform a preliminary data selection. Data transformation is undertaken to provide an alternative representation of the data, that it is hoped will lead to a better (more predictive) model than would result from using the data in its original form. Data transformation can involve: linearization, standardisation and conversion.

#### Data Linearization

Linearization transformations are applied so that the relationships between the explanatory variables and the target variable are (approximately) linear. Having linear relationships is important for methods such as linear regression and logistic regression. If the relationships in the data are highly non-linear then poor models will result using these methods. In our modelling framework, we have performed WOE of some factors to capture the linear and logical relationship with the survival time.

#### Data Standardization

Data standardization is very important when dealing with parameters of different units and scales. Such transformation will transform the data set to have zero mean and unit variance. For example, if one predictor variable takes values in the range 10,000 to 1,000,000 and another takes values in the range 0.01 to 1, then the parameter coefficients (the model weights) will be very different, even if the two variables contribute equally to the model. This is not an issue for all model construction techniques, but as a rule, it is good practice to transform interval variables so that they all take values that lie on the same scale.

Apart from the standardization of the data used in the detection and processing of outliers, no other was necessary in the processing of the data for PD mortgage.

#### Data Conversion

Some learning algorithms such as ordinary Least square can only handle numeric features, while some others can only handle nominal features. Only few can handle both. Data features therefore have to be converted to satisfy the requirement of learning algorithms, for example numeric to nominal, nominal to numeric or nominal to binary transformations. No data conversion was performed for PD mortgage modeling.

### 5.4.2 Preliminary Variable Selection

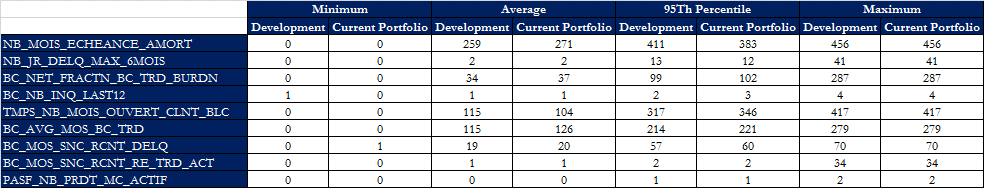
No factor exclusion was considered at this stage of data processing.

## Representativeness of Data

The data used in the modeling process consists solely of internal LBC data and represents the mortgage loans portfolio. A history covering the period from January 2008 to July 2017 was used to ensure large history coverage. The model was then later tested on the most recent population available (dated June 2018 for impact analysis) to ensure a out-of-time testing.

However, to find out differences between development and current population, a comparison for principal risk drivers was carried out, the results are presented in following table:

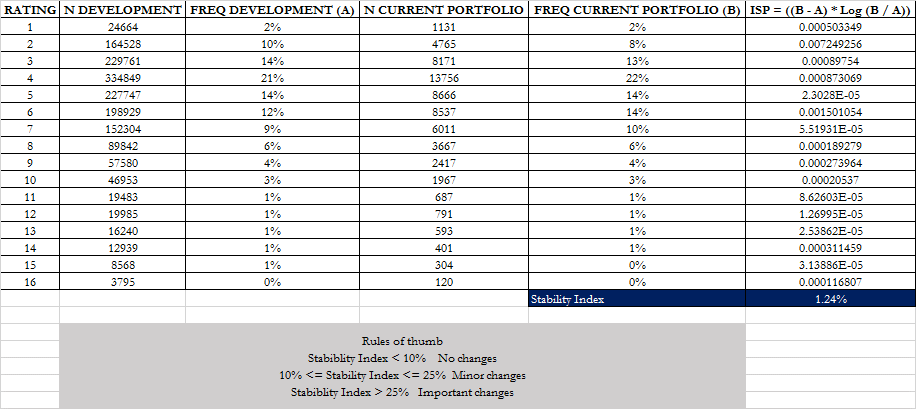
Table 15: Representativeness of data – Univariate measures



For the variables presented in the tables above, we do not observe significant differences between their distributions. Based on the results of our “key variables” comparisons between the development population and the portfolio as of June 30, 2018, we conclude that the development dataset is appropriate to develop a PD model that will be applied to the current portfolio.

Another way to prove that the representativeness is not an issue, the stability index from development and current population as of June 2018, was calculated:

Table 16: Representativeness of data – Stability Index



Since the index is less than 10%, according to the rule of thumb we conclude that there is no change in two populations and the data used for estimation is comparable to the bank’s current exposures. (refer to Backtesting framework)

## Historical Data Coverage

It is important to examine the data coverage because Basel II framework specifies adequate data period coverage for accurate and robust estimates. Otherwise, banks’ internal assessments of the performance of their own rating systems must be based on long data histories, covering a range of economic conditions, and ideally one or more complete business cycles: at least five years of historical data are needed for mortgage PD model (paragraphs 466). Besides, the model operational processes must be sensitive to adverse economic conditions.

We used 35 windows to build the model. Snapshots of January, April, July and October of each year were taken, starting with January 2008 and ending with July 2017 (then end of analysis). This period cover both an economic downturn (2008-2009) as well as the following recovery period, making us compliant with the previously mentioned Basel II requirements.

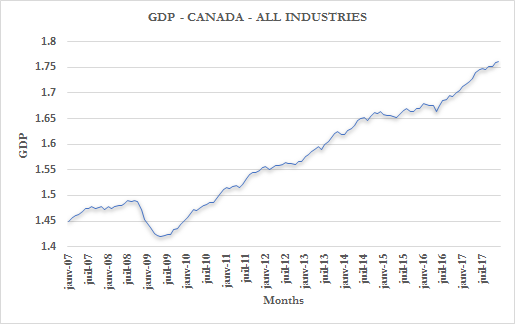
### 5.6.1 Adverse Economic Conditions & Stress Years

An economic downturn (Downturn) can be interpreted in many ways. A definition, called a "technical recession" by economists, is at least two consecutive quarters of negative growth in real GDP, corresponding to the period from contraction to economic trough. The economic slowdown is usually accompanied by an increase in the unemployment rate. Within the bank, the increase in credit risk can be observed on default rates, drawn amounts and losses.

Gross domestic product (GDP) is the total unduplicated value of the goods and services produced in the economic territory of a country or region during a given period (source: Statcan).

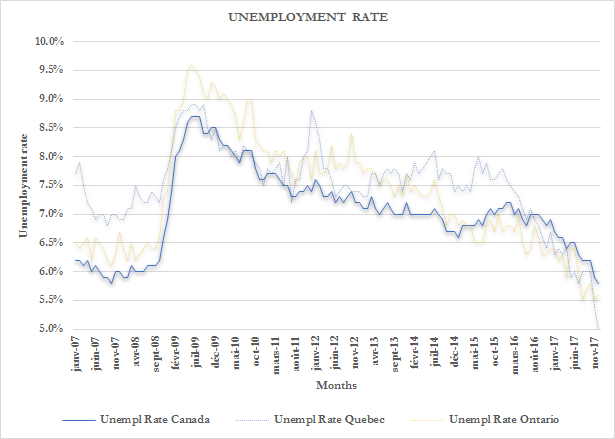
Generally, PD estimated vary according to the economic context. It is essential to determine whether the data used for the PD development has been gathered in periods of economic prosperity or of economic depression. The methodology employed to determine in which parts of the economic cycle the development data was gathered is to look at the Canadian GDP growth rate and unemployment rate.

Figure 07: Canadian GDP – All Industries - Monthly



The decline in activity, which corresponds to the fall in GDP, can be observed from the middle of 2008 until the middle of 2009. This same observation can also be made when we analyze the trend in the rate of unemployment for Canada as a whole, but also for Quebec and Ontario in the graph below:

Figure 08: Unemployment Rates



As shown in the graphs above[[4]](#footnote-4), the model development considers the last recession period from mid-2008 to mid-2009.

### 5.6.2 Sufficiency of Data

The 5-year period coverage condition recommended by Basel paragraph 466 for the mortgage PD modeling for data sufficiency are met since the data are from 2008 to 2017, meaning that the data used for model building is sufficient.

# **METHODOLOGY**

In this section, we present an overview of the different methodological options for PD modelling and we explain the rationale for the chosen approach as well as the underlying assumptions and hypotheses.

## Methodological Options

In modelling credit risk, three different options generally arise: a purely statistical approach, a judgmental approach and a hybrid one.

1. **Purely statistical approach**

This approach is always preferred when it is possible to put together a relatively important default database with all relevant information on potential risk drivers. Below is the list of the available approaches. The logistic regression and the survival analysis remain the most commonly used models.

* **Logit and Probit models**

These models estimate the probability of occurrence of default given a set of predictor variables that could be numeric or categorical. The dependent variable is binary and indicates whether the individual defaulted. The results can be interpreted directly as default probabilities and the individual significance and performance of each variable could be easily verified. The choice between a Logit and a Probit model boils down to choosing a normal or a logistic distribution. The two functions have a similar form with the latter having fatter tails. Note however that the differences in the results are often negligible. The parameters are generally estimated using the maximum likelihood technique.

* **Survival models**

Unlike classical models that focus on whether a specific borrower is likely to default, survival models predict the time at which the borrower may default. This class of models is very popular in credit risk modelling and has been widely applied in the recent years. These models are also able to effectively handle censored observations, for which information about the survival time is incomplete, such as individuals who withdraw from the study or do not experience the event before the study ends.

Three different types of survival models are available: parametric, semi-parametric and non-parametric.

1- The parametric approach assumes that the survival time follows a certain known probability distribution. Some commonly used distributions are: Exponential, Weibull, Gamma, Log-logistic and Log-normal. The parameters are estimated using maximum likelihood.

2- The semi-parametric approach widely known as Cox proportional hazards model assumes a nonlinear relationship between the predictors and the hazard function but makes no assumption about the underlying distribution of the survival function. In addition, the hazard ratio of two subjects depends only on values of their covariates and does not depend on time. Under this approach, the parameters are estimated using the partial likelihood whilst the baseline hazard needs to be calibrated independently.

3- The non-parametric approach relies mainly on the Kaplan-Meier estimator and do not make any distributional assumptions. This method is mainly used to estimate and graph survival probabilities and also make simple comparisons.

* **Linear regression**

This classical model assumes a linear relationship between the predictors and the default variable. The parameters are generally estimated using the ordinary least squares technique. However, there is no specific range for the output, which makes the interpretation and conversion to default probabilities complex.

* **Discriminant analysis**

Discriminant analysis is a widely-known classification technique usually applied to corporate bankruptcies. The goal is to find the discriminant function and separate groups according to specific characteristics by maximizing the differences between the groups and minimizing the differences within individuals. The main drawback of this approach is the assumption of multivariate normality, which is typically not applicable for variables used in credit scoring.

* **Panel models**

This type of models is used when we have observations of the borrower characteristics for more than one period. We can also integrate other time-varying variables into the model to enhance its performance. However, the assumption of independent observations is violated because of the additional time dimension. The correlation between the observations of the same individual makes the use of traditional estimators inappropriate and requires using advanced econometric techniques.

* **Neural networks**

Neural networks belong to the class of non-parametric methods inspired by the functioning of biological nervous systems. They consist of several connected nodes that generate a certain output based on specific inputs. They mainly differ from classical models by their blackbox nature and the non-linear relation between variables. Neural networks are free of any distributional assumptions and can quickly integrate new information. However, interpreting the output and converting it to default probabilities is not straightforward.

* **Decision trees**

Decision trees are another non-parametric statistical option that can easily handle nonlinear relationships between the output and the predictive variables. The sample is divided into groups according to their characteristics by using a set of if-then split conditions. Different algorithms and splitting criteria could be used, namely the Classification and Regression Trees algorithms (C&RT) and the Chi-square Automatic Interaction Detector (CHAID). Decision trees are particularly efficient when the predictive variables are known to be interactive and their number is limited. However, to compute default probabilities, additional steps are required since the output is risk categories and not a continuous variable.

1. **Purely expert or judgmental approach**

This approach is adopted when it is not possible to put together a significant default database. In this case, experts with deep knowledge of the risk framework help in selecting relevant risk factors and map the final score to a PD.

1. **Hybrid approach**

This approach is adopted when the available dataset is not important enough to entirely rely on a purely statistical approach. In this case, elements of both of the above approaches are combined to develop the model

## Selected Modeling Approach

Given data availability, we opt for a statistical approach for PD modelling. Specifically, we will apply the survival analysis since it is very flexible and can estimate the default probability over any time horizon. In addition, it presents several advantages over logistic regression and other statistical methods. In fact, we are able to use more information in model building than other conventional models. This methodology can accommodate censored data of subjects with incomplete information about survival time. Moreover, modelling the time until default provides more statistical power than simply counting the number of defaults. Unlike logistic regression, time-varying variables could be easily integrated, particularly macroeconomic factors, which can enhance the model predictive performance.

Several papers in the financial literature support the use of this type of models and provide evidence of its performance. Banasik, Crook and Thomas (1999) find that survival models and logistic regression are competitive and that in some cases the former are more predictive. Belloti and Crook (2007) also show that survival analysis methods are more competitive for default prediction in comparison with logistic regression. Tong, Mues and Thomas (2010) show that the calibration performance for the survival approaches is superior to logistic regression for intermediate time intervals and useful for fixed 12-month time horizon estimates. Similar conclusions were also reached by Stepanova and Thomas (2000) and Laitinen (2005) who confirm the previous findings.

In our context, we will apply a survival analysis based on a parametric approach. This approach is particularly useful when we want to investigate the effects of other variables on the survival time and carry out formal tests on the form of the survival function. Before estimating the parameters, we must specify a plausible distribution for the random variable T and choose a functional form to associate the explanatory variables with this distribution. Thus, we will be able to derive a concise equation for the survival time and the hazard rate. The most commonly used underlying distributions are: Exponential, Weibull, Gamma, Log-logistic and Log-normal. The choice is not arbitrary and will be based on statistical tests to determine the best fit. On the other hand, parametric models can accommodate left, right and interval censored data whilst proportional hazard models can only handle right censored data. Another advantage of parametric models over semi-parametric Cox models is that they predict survival times since the distribution is already specified and extrapolation is possible. We can distinguish two types of predictions: point prediction (a single value) and interval prediction.

Next, we define the main functions used in the context of survival analysis. Let T be a non-negative random variable representing the time until the occurrence of the event. The distribution of T is described either by its cumulative distribution function or by the survival function , which gives the probability of surviving up to time t. We also define the hazard function or the instantaneous rate of occurrence of default given that the individual has survived up to the specified time.

.

The numerator of the above expression is the conditional probability that the event will occur in the interval given that it has not occurred before. By dividing this quantity by the width of the interval and taking the limit to zero, we obtain an instantaneous rate of occurrence.

is a non-increasing function with the following boundary condition: , since the event is sure not to have occurred at time 0. As time goes to infinity, the survival curve goes to 0. The survival time could also be expressed as a function of the hazard rate:

.

All the development process was done under a Gamma distribution (which was found to be the best fit for both the mortgage PD model at step 2).

A short technical review about Gamma distribution is as follows: The *SAS Lifereg procedure* was used for modeling. The CDF of the time to default *t* with parameters  takes on the following form:



Where:

 ε ~ N (0, 1)

The value of λ will then be obtained by the SAS estimation at x=0 ().

## Assumptions and Hypothesis

Survival analysis does not have predefined assumptions regarding the distribution of predictors. However, the assumptions of Accelerated Failure Time model can be summarized in the following points:

**H1**: Correct specification of underlying survival time distribution.

**H2**: The regressors are not linearly dependent (no perfect multicollinearity).

**H3**: The number N of observations is higher than the number K of parameters to be estimated.

# **MODEL SPECIFICATION AND ESTIMATION**

In this section, we highlight the specification of the survival analysis framework for mortgage PD modeling. As part of the modeling framework, we should first perform some preliminary analysis in which we discuss about the target variable considered when representing the risk rating and, through univariate as well as multivariate analysis, identify potential explanatory variable and analyse relationship with target. Then, we should select most relevant risk drivers among candidates determine in which ways they could potentially affect the creditworthiness of mortgage portfolio.

## Preliminary Analysis

The preliminary step is to review the literature on the different methodologies used to develop PD. Univariate analyzes are also done on different variables to study their nature.

Before analyzing how these factors might influence the mortgage PD in the scope of available data, we should document the definition of the target variable.

### 7.1.1 Target Variable

A survival model is used to analyze time-to-event historical data and to generate estimates, referred to as survival curves, that show how the probability of the event occurring changes over time. In many life situations, as time progresses, certain events are more likely to occur. The survival models help decision makers to form better estimates than guessing about the expected timing of certain events. The estimates take into account the impact of other variables, referred to as independent, predictor variables or covariates, on the expected timing of the event to occur.

Analogous to a linear regression analysis, a survival analysis typically examines the relationship of the survival variable (the time until the event) and the predictor variables (the covariates). The event of interest is frequently referred to as a hazard. The analysis specifies a linear-like function for the event called the hazard function. For mortgage PD model the target variable is time before default.

Note that the final model output will be then transformed into probability of default (PD).

### 7.1.2 Predictive or Explanatory Variables

The explanatory variables in our model are factors that are reasonably expected to be likely influence mortgage credit worthiness. Our first step on the path to understanding the dataset is to look at each variable, one at a time, through univariate analysis. After that, relevant candidates are all analyzed together during multivariate analysis in order to retain the best subset of risk drivers likely to evaluate mortgage creditworthiness.

### 7.1.3 Univariate Analysis

Initial characteristic analysis involves two main tasks. The first step is to assess the strength of each characteristic individually as a predictor of performance. This is also known as univariate screening, and is done to screen out weak or illogical characteristics. The strongest characteristics are then grouped. This applies to attributes in both continuous and discrete characteristics, and it is done for an obvious reason. Grouping offers some advantages (Siddiqi):

* It offers an easier way to deal with outliers with interval variables and rare classes.
* Grouping makes it easy to understand relationships, and therefore gain far more knowledge of the portfolio. A chart displaying the relationship between attributes of a characteristic and performance is a much powerful tool than a simple variable strength statistic. It allows users to explain the nature of this relationship, in addition to the strength of the relationship.
* Nonlinear dependencies can be modeled with linear models.
* It allows unprecedented control over the development process by shaping the group, one shapes the final composition of the model.
* The process of grouping characteristics allows the user to develop insights into the behavior of risk predictors and increases knowledge of the portfolio, which can help in developing better strategies for portfolio.

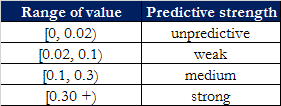
Once the strongest characteristics are grouped and ranked, variable selection is done. At the end of initial characteristic analysis, we will have a set of strong, grouped characteristics, preferably representing independent information types, for use in the survival analysis step.

Weight of evidence (WOE) measures the likelihood that a modality of a grouped variable is predictive of default or not. The overall predictive power of the variable is measured using the Information Value (IV). The IV is widely used measure in the industry, and different practitioners have different rules of thumb regarding what constitutes weak or strong characteristics.

Let consider a class variables with modalities . Also, let (resp. ) the frequence of the kth modality within the non-default (resp. default) borrowers. Then the information value of this variable is defined by:

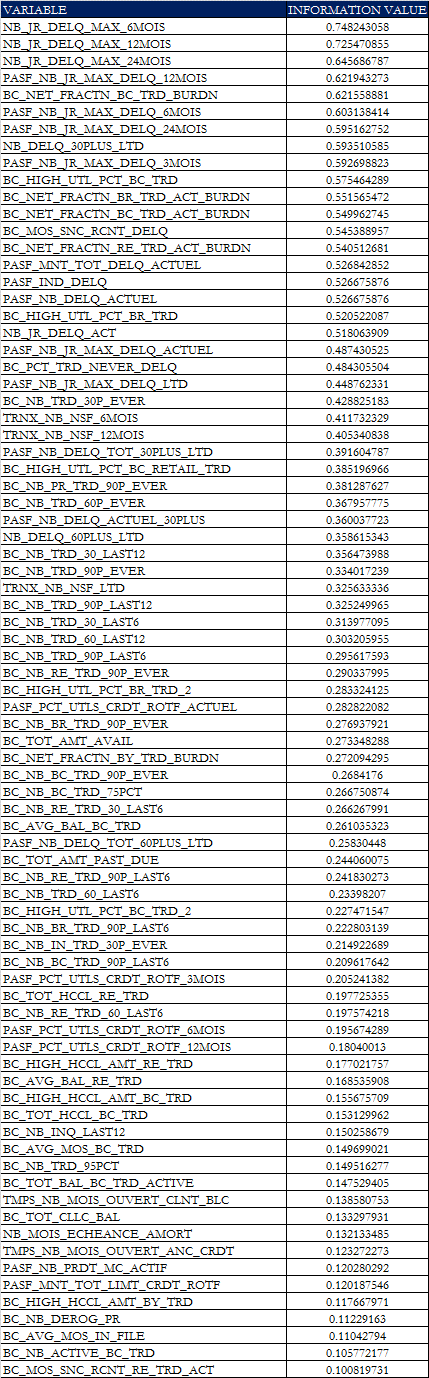
Siddiqi reported the rules of thumb to assess the predictive power of a variable using the Information Value. The table below reports this rules:

Table 17: Siddiqi rules of thumb for Information Value (page 81)



For the Mortgage loans PD model, there are 80 variables out of more than 206 that are initially selected because they have an IV higher than 0.1. For continuous variables, we would like to compute the IV for a different number of bins (4 to 10 for example) and consider the number of bins giving the maximum IV. But we are running our model in SAS 9.3 so we cannot use the SAS Procedure PROC HPBIN. We compute the IV by selecting a bin of 5 that is often used in the literature and because computing IV with several bins will take hours to run. For categorical variables, the number of bins is the number of level of the variable. Then, we perform a correlation analysis on the 80 continuous variables and exclude those that are highly correlated. Given that we have enough potential variables on which to build a model, we have set the correlation threshold at 50%. We consider two independent variables correlated if the correlation coefficient is high than 50% (See Bart Bassens: Credit risk modeling using SAS) and the p-value of the hypothesis test less than 1%. From a set of highly correlated variable, we keep the one with the highest information value. Table 18 presents the list of selected variables with their Information Value.

Table 18: Preliminary selected variables with their IV



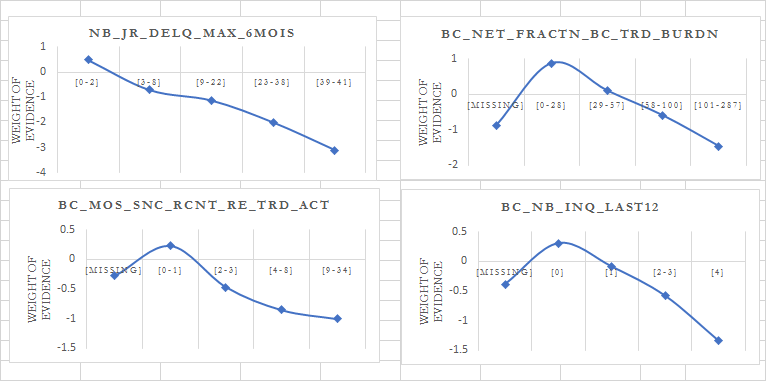
Once the variables are preliminarily sectioned with the IV, we will now transform the continuous variables into binning with the WOE. We chose bins of 5 to make sure we have enough accounts in each bin and more bin may be less convenient for implementation. The best bin is the one that maximizes the IV (See Mamdouh Refaat, Credit Risk Scorecards: Development and Implementation Using SAS, page 284 to 405)).

The WOE measures the strength of each attribute, or grouped attributes, in separating good and bad accounts. It’s a measure of the difference between the proportion of goods and bads in each attribute. The WOE is based on the log of odds calculation:

WOE = Ln (Distribution of good / Distribution of bad)

The statistical strength, measured in terms of WOE and IV, is, howevernot the only factor in choosing a characteristic for further analysis, or designating it as a strong predictor. The grouped attribute strengths must also be in a logical order, and make operational sense. In the graphs below we present a set WOE:

Figure 09: Example of WOE’s Trend



For NB\_JR\_DELQ\_MAX\_6\_MOIS: we can see that the groupings in this characteristic have a linear relationship with WOE; that is, they denote a linear and logical relationship between the attributes in these variables and the time survival. This confirm business experience in the Bank. An increase in the number of days of delinquency causes a decrease in the survival time.

For BC\_NET\_FRACTN\_BC\_TRD\_BURDN, we can see, apart from “missing”, the other groupings in this characteristic have a linear relationship with WOE; that is, they denote a linear and logical relationship between the attributes in these variables and the time survival. This confirm that the increase in the utilization rate causes a decrease in the survival time.

Other examples of WOE will be presented in the appendix 10.

Other variable selection techniques exist. For example, the model is developed using non-grouped characteristics, statistics to evaluate predictive strength include R-square and Chi-square. Both these methods used goodness-of-fit criteria to evaluate characteristics. The R-squared technique uses a stepwise selection method that rejects characteristics that do not meet incremental R-square increase cut-offs. A typical cut-off for stepwise R-squared is 0.005. Chi-square operates in a similar fashion, with a minimum typical cut-off value of 0.5. The cut-offs can be increased if too many characteristics are retained in the model.

Even though these techniques could have done the work of variable selection, we chose the method of IV and WOE for the reasons described above

### 7.1.4 Multivariate Analysis

The order in which variables were selected to be tested/added was chosen based on a partial correlation method. At each iteration, the potential variables were ranked in function of their partial correlation with the dependent variable (censored), given the variables that were already in. the model (through the use of the SAS’ Proc Corr Procedure and the Partial Option)[[5]](#footnote-5).

The partial correlation measures the mutual connection between two variables y and xj when others variables (x1, x2, x3 ...) are held constant with respect to the two variables involved y and xj (unlike the previous case). The partial correlation coefficient is very useful in multiple regression, where it "... directly assesses the proportion of the unexplained variation of y that becomes explained by the addition of the variable xj. "In the case where we have an explanatory variable x1 and a variable kept constant x2, the coefficient of partial correlation is calculated as follows (Scherrer, B. 1984. Biostatistique: equation 18-50, p. 704):

The best candidate was then chosen, and a Survival Model was estimated through the use of *SAS’ Lifereg Procedure*. The Model’s output was then observed, as well as its AIC, Log-Likelihood and KS statistic. In order to be retained in a model, any candidate must meet the following criteria:

* Make economic sense;
* Do not exhibit multicollinearity problems based on the correlation matrix and the Variance Inflation Factor (VIF);
* Be statistically significant (p-value below 0.001)[[6]](#footnote-6),
* Produce coefficients with the expected sign.
* Diminish the AIC or augment the KS[[7]](#footnote-7).

If any variable chosen by the automatic-selection described above did not fulfill these criteria, it was manually excluded from the model, and the process was re-iterated.

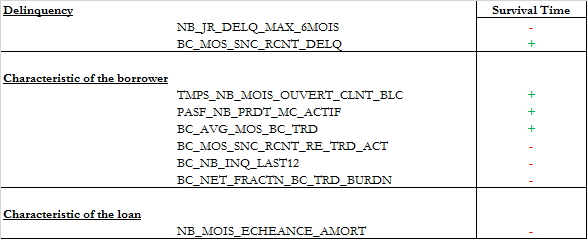
After choosing a certain set of variables which minimized the AIC or maximized the KS, transformations of certain variables were tested. Finally, reductions of the number of variables were tested, to ensure “parsimony” in the model.

The final model was then chosen based on the above criteria were fulfilled and results were satisfactory.

### 7.1.5 Analysis of the Relationship between Target Variables and the Predictive Variables

The relationship between target and explanatory variables is displayed below:

Table 19: Relationship between Target and selected variables



NB\_JR\_DELQ\_MAX\_6MOIS: Number of days of maximum delinquency in the last 6 months. The variable is transformed to Weight of Evidence. As expected, it is negatively related to the survival time.

BC\_MOS\_SNC\_RCNT\_DELQ: Number of months since the most recent delinquency. The variable is transformed to Weight of Evidence. As expected, it is positively related to the survival time.

TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC: Time with the bank. Number of months since opening the client file at the BLC. The variable is transformed to Weight of Evidence. As expected, it is positively related to the survival time.

PASF\_NB\_PRDT\_MC\_ACTIF: Number of active margin accounts. The variable is transformed to Weight of Evidence. As expected, it is positively related to the survival time.

BC\_AVG\_MOS\_BC\_TRD: Average number of months since date opened on all bankcard trades. The variable is transformed to Weight of Evidence. As expected, it is positively related to the survival time.

BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT: Number of months since latest activity of revolving trades. The variable is transformed to Weight of Evidence. As expected, it is negatively related to the survival time.

BC\_NB\_INQ\_LAST12: Number of inquiries in last 12 months. The variable is transformed to Weight of Evidence. As expected, it is negatively related to the survival time.

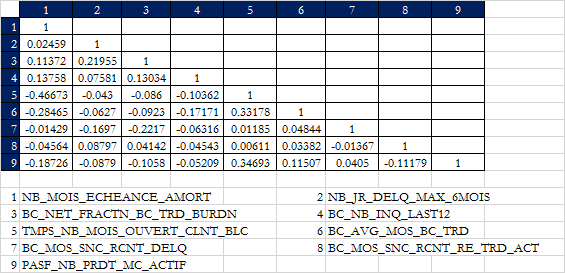
BC\_NET\_FRACTN\_BC\_TRD\_BURDN: Ratio of total balance to HC/CL for all bankcard trades. The variable is transformed to Weight of Evidence. As expected, it is negatively related to the survival time.

NB\_MOIS\_ECHEANCE\_AMORT: Number of months remaining until the end of depreciation. The variable is transformed to Weight of Evidence. As expected, it is negatively related to the survival time.

### 7.1.6 Correlation Analysis between Variables

Higher correlation among variables could lead to non-convergence of the optimization routines, or could lead to a higher standard error of the estimates. Thus, we check for multicollinearity among selected variables:

Table 20: Pearson Correlation analysis between selected variables.

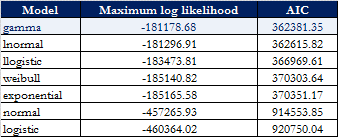


The result of the Pearson correlation matrix already shows a weak correlation (less than 50%) between the dependent variables. We will confirm it later with the variable inflation analysis (VIF).

## Risk Drivers Selection (Model variables selection)

The first step followed in the modeling process is selecting a baseline distribution to start the iterative process of variable selection. In order to do that, all variables are put in a LIFEREG regression and, using the AIC criterion, the initial distribution is selected. The AIC value for the 7 distributions available in SAS is as follows:

Table 21: Ln-Likehood and AIC for distribution selection



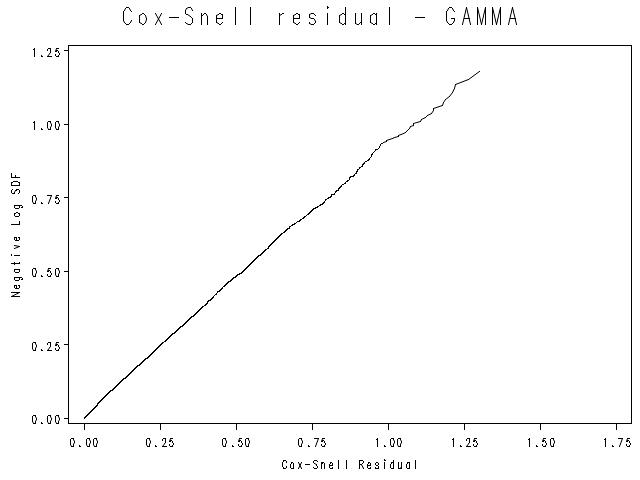
The log-likelihood value can be used to select which parametric model provides the best fit. The Akaike Information Criteria (AIC) provides a score for assessing fit. An approximation of the AIC score is:

AIC

Where k is the number of parameters in the model, and ln(L) is the log likelihood. As a rule of thumb, a smaller AIC value represents a better fit.

The AIC criterion suggests that data could be better fitted using Gamma model. Iterative process will start using the suggested distribution. Cox-Snells residuals analysis confirms the choice.

Figure 10: Cox-Snell residual - Gamma



It is common practice to use Cox-Snell residuals to check for overall goodness of fit in survival models. We evaluate the presumed relation of unit exponentially distributed residuals for a good model fit and evaluate under some violations of the model. This is done graphically with the usual graphs of Cox-Snell residual. It is observed that residuals from a correctly fitted model follow unit exponential distribution. It is also not uncommon to see some slight jumps occurring at the extremities of the graph.

Selecting risk drivers has been done using partial correlation as described in section 7.1.4. Before presenting the selected model, we describe the method used to choose a survival time distribution. The procedure to select the survival time distribution can be summarized as follows:

**Step 1**: Identify the risk factors which are significant and make economic sense in both a univariate and a multivariate model.

**Step 2**: After identifying the risk factors, compare Akaike’s Information Criterion (AIC) between the available survival functions: Gamma, log-logistics, log-normal, weibull and exponential.[[8]](#footnote-8)

**Step 3**: Select the best model. (Through iterations and selection of variables by partial correlation).

**Step 4**: Once the model is chosen, compare AIC of the model using the available survival functions. Based on AIC and Cox-Snell residuals, select the best survival function.

## Downturn

The regulatory framework requires the economic or market conditions underlying the modeling data to be relevant with current and foreseeable conditions (Basel paragraph 450). On one hand, if only data covering a period of economic prosperity is available, elements ought to be added to the estimates to account for a downturn effect. On the other hand, if the data covers a complete economic cycle, the estimates should be reliable. In the scope mortgage PD modeling, the data covers full economic cycle (see section 5.6.1). Therefore, no additional analysis should be made in order to account for downturn effect. Otherwise, section 7.3.1 below does not apply to the model.

### 7.3.1 Correlation between Default Rate and LGD/EAD

This is not applicable for PD model.

### 7.3.2 Downturn Period Selection

It can be noticed from the analysis in section 5.6.1 that the period covering mid 2008 to mid 2009 can be considered as a downturn period for mortgage portfolio given the decrease in GDP growth and the Unemployment rate in Canada.

### 7.3.3 Downturn Factor

Since the modeling data cover full economic cycle, there will be no need to assess for a downturn factor capturing the downturn effect.

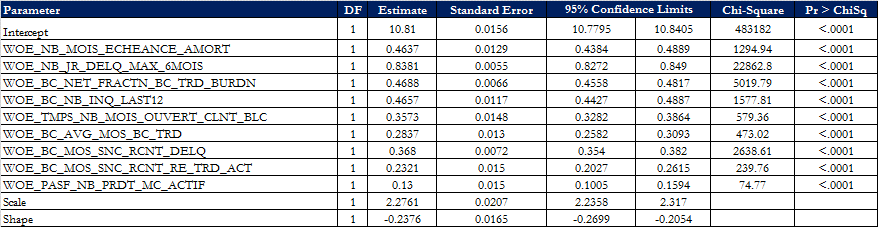
## Preliminary Results

In this section, we display preliminary estimates, i.e. modeling results before calibration. We also check for the violation of the model assumptions as depicted in section 6.3.

### 7.4.1 Modeling Results

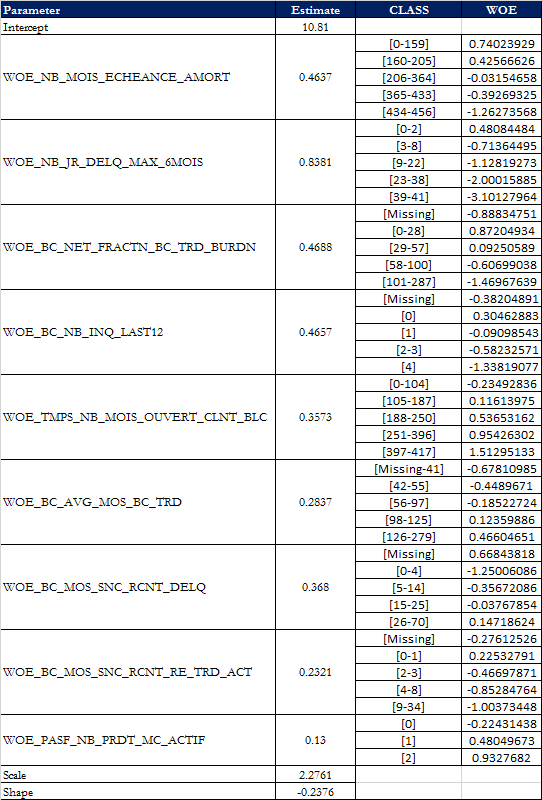
The following tables summarize the results for the fitted Gamma model, the “Parameter” column is the variable in final equation; “Estimate” is the coefficient value, the “Standard Error” is the standard error for estimation of coefficient, “Wald Chi-Square” column is the estimation of value for the Chi-Square test and “Pr > ChiSq” is the the p-value of the Wald chi-square statistic with respect to a chi-square distribution with one degree of freedom.

Table 22: Gamma survival model estimation results



The multivariable analysis results with a list of parsimonious variables that are predictive of the time to default. The table above presents the output of the Gamma distribution is a standard output of a regression, except that we must put another table that indicates the class interval to specify the subdivision of class variables.

Table 23: Final variable estimates and Weight of Evidence

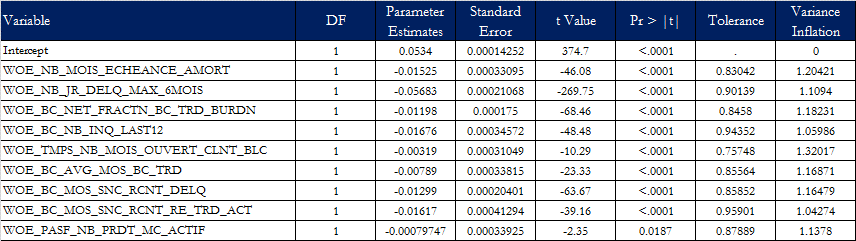


### 7.4.2 Modeling Assumptions Violation Check

H1: Correct specification of underlying survival time distribution: This issue has been treated with the AIC criteria and Cox-Snell residual plots.

H2: The regressors are not linearly dependent (no perfect multicollinearity): Regression methods are sensitive to extremely high correlations among predictor variables. This issue has been treated with collinearity diagnostics incorporated in REG procedure in SAS. For the final model, the VIF criterion (less than 10) is verified.

Table 24: Variable Inflation (VIF)



H3: The number N of observations is higher than the number K of parameters to be estimated: Test based on maximum likelihood methods could have problems when there are too few cases relative to the number of predictor variables, nevertheless, this is not a problem for the Mortgage portfolio as there were more than 1 million observations.

## Model parsimony

The model of PD mortgage is a purely statistical model, developed over a fairly broad period covering a complete economic cycle. We have enough historical data and enough defaults to develop a model that meets the performance requirements (please see section 8.4.2 out-of-time backtesting). Therefore, we did not have to make a trade-off between performance and model complexity.

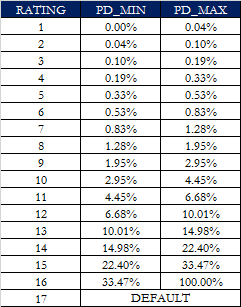
## Mapping Mechanism between Model Output and Rating Grade

The mortgage portfolio PD model use the master scale from LBC’s risk rating system. Once the model has calculated and generated the PDs, they are mapped to the master table in order to assign a rating grade. The master table was developed by the risk modeling team. It is based on the PD’s value interval and numerical values that increase in a one point assigned rating grade differentiator where lower numerical values correspond to the lower PD’s and higher numerical values correspond to the higher PD’s.

The modeling team did not consider it useful to change the mapping table for Retail models. Retail PD models do not come out of risk ratings like commercial PD models. So, no update is done on the ratings of Retail except for the needs of analysis as for the Backtesting models. The notion of risk rating does not really exist for Retail in the same way that we talk about BRR (Borrower Risk Rating) for the Commercial.

The following table summarizes the LBC retail risk rating scale:

Table 25: BLC’s Master Scale for Retail



# **CALIBRATION AND FINAL RESULTS**

To assess the quality (or validity) of a model, we must determine if the model is correctly calibrated. Correct calibration means that the probabilities of default computed at the beginning of a period are “compatible” with realized defaults during the period or historical realized defaults.

## Calibration Exercise

Note that the model is calibrated to make sure that the total average PD is as large as the historical default rate. As mentioned earlier, a first calibration of the model is made by running a “proc lifereg” with the selected variables on the development samples which automatically calibrates the model on the development data historical default rate.

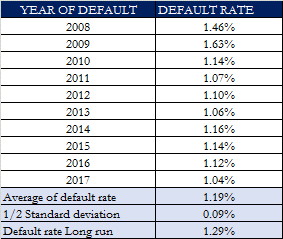
There is also another element that must be considered in the calibration process. Indeed, for the development of the model the default was counted once in its existence and we took the first occurrence. But in the calibration process where we calculate the annual default rate, we count the default every defaulted year if this is the case.

We compute the yearly default rate from 2008 to 2017. The second column of the table below gives the default times series for mortgage portfolio.

Usually, we add a standard deviation to address the uncertainty that is related to the fluctuation of default rates calculated by year. However, since the model successfully passes the backtesting exercise (See appendix 18) without the addition of the standard deviation or other adjustment factor, we add a half standard deviation.

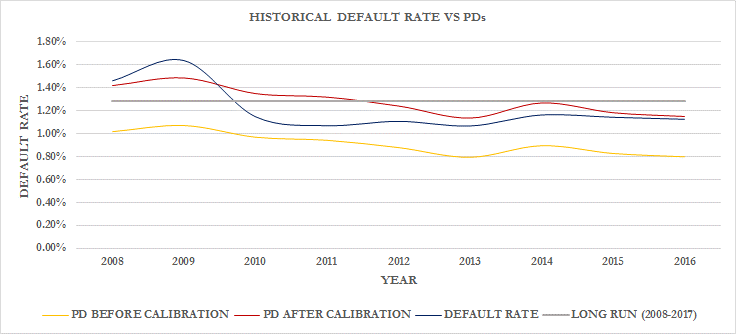
The average default rate equals 1.19%. The half standard deviation of 0.09% is added to Default Rate (DR) as a conservative factor to cover for general uncertainties related to the used of models. Finally, the Long Run Default Rate was set at 1.29%.

Table 26: Mortgage Long-Run Default Rate



Once the level of calibration is defined, the adjustment is made by successive try-error method to make sure that the predicted mean PD is close to the chosen calibration level.

Figure 11: Historical Default rate and PDs



## Downturn Parameter

No downturn parameter is needed for PD model since the PD model cover a complete economic cycle including downturn period.

## Final Model

As mentioned above, the goal of calibration process is to bring the estimated PD to the observed default rate on the overall population. Here is the PD mortgage final model:

*λ* = 10.81

+0.4637 \* WOE\_NB\_MOIS\_ECHEANCE\_AMORT

+0.8381 \* WOE\_NB\_JR\_DELQ\_MAX\_6MOIS

+0.4688 \* WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN

+0.4657 \* WOE\_BC\_NB\_INQ\_LAST12

+0.3573 \* WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC

+0.2837 \* WOE\_BC\_AVG\_MOS\_BC\_TRD

+0.368 \* WOE\_BC\_MOS\_SNC\_RCNT\_DELQ

+0.2321 \* WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT

+0.13 \* WOE\_PASF\_NB\_PRDT\_MC\_ACTIF

The scale parameter, *σ*, was fitted to the following value:

*σ* = 2.2761

The shape parameter, *δ*, was fitted to the following value:

*δ* = -0.2376

The calibration factor, , was fitted to the following value:

= 0.33398

PD is then calculated using the relation described earlier in section 6.2.

## Model Validation Testing

The development of methodologies for validating internal rating systems is an important regulatory issue. A common approach to validate rating systems is the comparison of estimated parameters with empirical ex-post realizations. For that purpose, statistical testing is performed on new, yet unseen data (in-sample, out-of-sample or out-of-time), for a systematic detection of biases in model results. Most of performance testing methods proposed in the literature focus on the discriminatory power of PD estimates. Among the best-known are the Kolmogorov Smirnov test (KS), the Accuracy Ratio (AR) and the area under the ROC (Receiver Operating Characteristics) curve. More advanced methods aim at directly assessing the calibration quality, i.e., the accuracy and reliability of PD estimates, namely binomial, Hosmer-Lemeshow and Spiegelhalter tests. For mortgage PD model, we will compare the KS and the ROC (AUROC) in the development and validation sample.

### 8.4.1 In-sample and Out-of-sample Performance

KS is measured by the maximum distance between cumulative distribution of defaults and cumulative distribution of non-defaults (D in the graphic). Generally speaking, for back-end models, we expected K-S ranges from 0.5 to 0.7.

Figure 12: Kolmogorov-Smirnov measure

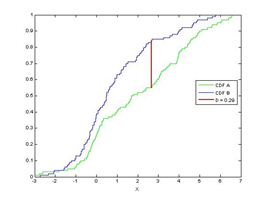
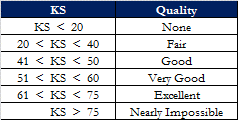


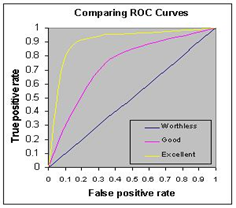
Table 27: KS rules of thumb values



The ROC curve helps to assess the accuracy of a model graphically. AUROC summarizes the information of the ROC curve into an indicator on a scale of 0% to 100%.

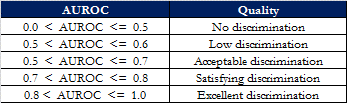
A model with maximal discriminatory power ranks the observations from the riskier borrower to the less risky one, allowing one to draw any two observations from the population and be 100% certain that the one deemed to be riskier by the model is, in fact, riskier. By contrast, a model with no discriminatory power, one we would expect to be wrong half of the time, would have an AUROC of 50%.

Figure 13: ROC curve



According to general industry standards, the AUROC statistics relating to the qualitative assessment standards are the following:

Table 28: AUROC rules of thumb values



The statistical measures of the performance for the mortgage PD model is as follow:

Table 29: Statistical measures of performance



Following the rules of thumb, the above table shows that the model has a good performance.

### 8.4.2 Out-of-time Backtesting

The main goals of model backtesting are:

* Determine how well the model performs on a out-of-time sample;
* Ensure that a model has not been overfit and its out-of-time performance is reliable;
* Confirm that the modelling approach, not just an individual model, is robust through time and credit cycles.

The Bank’s risk management framework includes a regular (annual) backtesting process. Backtesting implies testing several distinct aspects of a model:

* Discriminatory power – the model’s inner ability to classify borrowers or facilities according to their risk level.
* Calibration – the accuracy of the model output and its consistency with actual figures.
* Stability – the consistency between the development sample and the actual portfolio.

For Backtesting purpose, the last observation year (April 30st, 2017) is used as out of time sample.

1. **Discriminatory power**

The discriminatory power of a PD model refers to its ability to identify risky borrowers and rank them accordingly. Discriminatory power will be backtested by comparing the ranking with actual default observations, using the Area Under the Receiver Operating Characteristics curve (AUROC).

The rationale behind discriminatory power tests is that a model with a high discriminatory power must accurately classify defaults in the worst ratings and non-defaults in the best ratings.

In the Backtesting framework, the granularity of the PDs is not used but a mean or median PD per rating. We chose to use the median PD to assign the PD by rating because the median is more representative and allows for better management of extreme cases compared to the mean.

Table 30: Discriminatory result - AUROC



According to the rule of thumb of AUROC described above, the discriminatory test is a success.

1. **Calibration**

Calibration refers to the process of assigning a forward-looking PD to each individual rating. It is a crucial step in backtesting because it estimates the adequacy of the PD parameter, which has a direct impact on regulatory capital calculation.

A model may have a high discriminatory power, but that does not necessarily mean that the probability of default is accurate in predicting the realized default rate. Therefore, calibration backtesting ensures that the difference between realized and predicted default rates is not statistically significant.

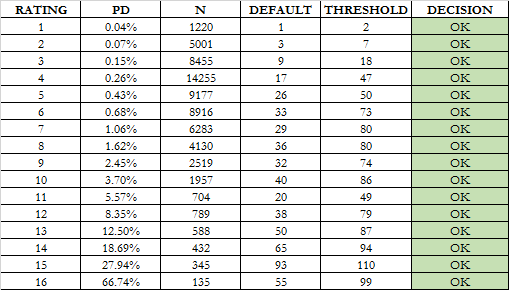
Calibration backtesting is explicitly required by the OSFI in paragraph 334 “banks must regularly compare realized default rates with estimated PDs for each grade and be able to demonstrate that the realized default rates are within the expected range for that grade.”

In order to do so, the following analyses will be performed: Binomial test, Vasicek test and Blochinger-Leippold test.

* **Binomial test (without correlation)**

The binomial test is a rating-level test which helps to assess whether or not the realized default rate is significantly higher than the estimated PD. The test result is expressed in terms of a default threshold (the number of defaults over which we can consider that realized default rate is underestimated) or a p-value expressing the significance of the difference between the estimated and realized PD. The test will be successful if at least 80% of ratings pass the test at a 5% confidence level.

Table 31: Calibration result – Binomial test



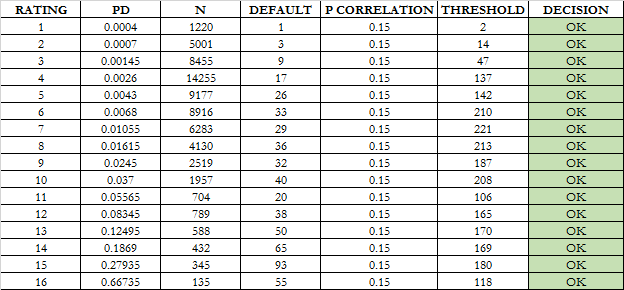
The Binomial test without correlation is a success since at least 80% (100% in our case) of ratings pass the test at a 5% confidence level.

* **Vasicek test**

The main hypothesis of the binomial test is that there is no correlation between defaults. Consequently, the threshold is lower than it would be in the presence of correlation. This assumption leads to a test statistic that is too conservative and more frequent recalibration than necessary. The literature has developed several tests to overcome this issue, including the Vasicek test.

The Vasicek test is an adaptation of the binomial test incorporating the impact of the correlation. If the binomial test is passed, then there is no need to perform the binomial test with correlation. However, if the test fails, the Vasicek test may help to assess whether or not the difference between the PD and the realized default rate is due to the default correlation. The test will be successful if at least 80% of ratings pass the test at a 5% confidence level.

Table 32: Calibration result – Vasicek test



The Vasicek test correlation is a success since at least 80% (100% in our case) of ratings pass the test at a 5% confidence level.

* **Blochinger – Leippold test**

One shortfall of the binomial test (or the Vasicek test) is that it is performed at the grade level. In certain situations, it may be difficult to judge whether or not the calibration is adequate as some ratings may pass the test and others may fail, leading to a mixed result. This is why some tests are focussed on testing the whole PD scale at once, such as the Hosmer-Lemeshow test or the Blochinger Leippold (“BL”), which are both based on the χ2 distribution. Unlike the Hosmer-Lemeshow, the BL test has the ability to take into account the impact of correlation.

The BL test is a “goodness of fit test” as it measures if the probability of defaults estimated corresponds to the realized defaults (see reference [4]). The BL is based on two components:

* the level of the PD curve, which refers to the accuracy of estimated probabilities compared to the rate of default per risk class; and
* the shape of the PD curve, which is measured by comparing the theoretical AUROC with the empirical area. Unlike other models which mainly focus on the level of the curve, the BL test also measures the shape of the model which can be distorted.

These two components (level and shape) can be combined in a single statistic that is asymptotically χ 2 distributed.

The test will be successful if the BL is lower than the critical value χ 2 at a 5% confidence level.

Table 33: Calibration result – Blochinger-Leippold test



The Blochinger-Leippold test is also a success.

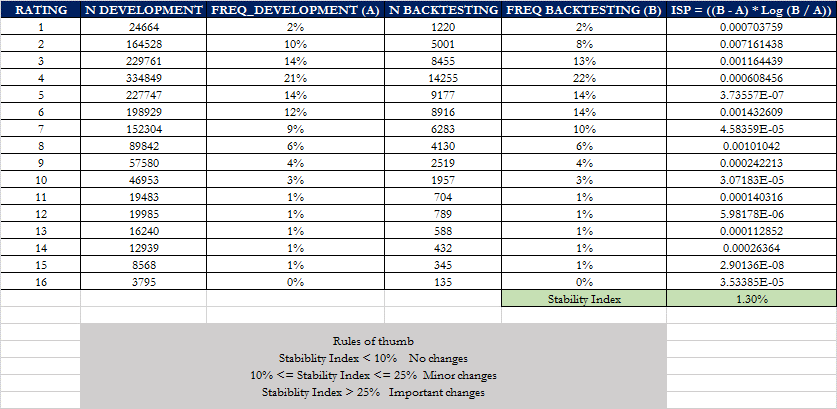
1. **Stability**

Two tests are computed to analyse the stability of the model: Population Stability Index and Maximum grade-level concentration.

* **Population Stability Index (PSI)**

The population stability index (PSI) compares the population used to create the model (development sample) with the current population (backtesting sample). This test helps to assess if the model, which has been developed on a sample with given characteristics, is still adequate given the change in the sample. In order to compute the PSI, frequencies of each ranking are compared individually as a score by class and then calculated as a whole.

Table 34: Stability result – Population Stability Index



According to the rules of thumb, there no changes in the population since the development.

* **Maximum grade-level concentration**

The maximum grade-level concentration represents the maximum observed portfolio percentage within a single risk rating. The LBC has set the maximum threshold at 25%, which is lower than the threshold suggested in the preliminary draft and in the first draft of the Basel II document.

However, since it is possible to observe a higher concentration than 25% for certain types of portfolio when adequately justified, we also compare the concentration with the concentration at model development.

The test will be successful if the maximum grade-level concentration is less than 25% or less than 1.1 times the maximum concentration at model development.

Table 35: Stability result – Maximum grade-level concentration

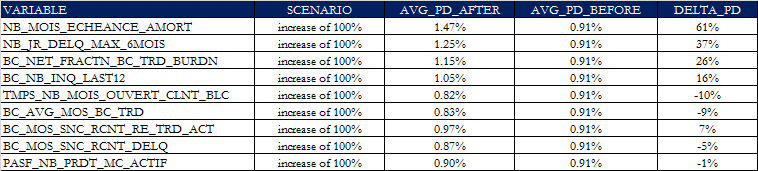


## Sensitivity Analysis

The main objective of the sensitivity analysis is to demonstrate that the model is sensitive to movements in the explanatory variables in the way that sufficiently large movements imply variations of the predicted outcome, as it would be reasonably observed in reality. A sensitivity analysis is conducted variable by variable.

We want to know for example, what will happen if the variables increase abruptly from single to double? To do this, we will take a hypothesis. We want to observe the impact on PD variable by variable. For example, if the number of days of delinquency increases by 100%, all things being equal, what will be the impact on PD?

Table 36: Final sensitivity analysis



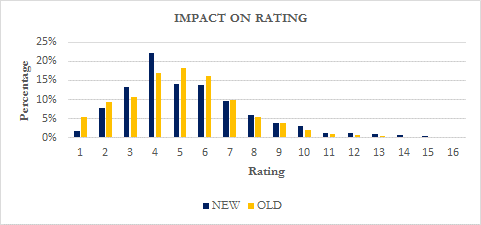
The PD is sensitive to the variation of all variables at various degrees. Of all the variables of the models, it is the increase of NB\_MOIS\_ECHEANCE\_AMORT which would impact most the PD.

## Model Application on the Current Portfolio

It is interesting to make estimations and examine the impact of our model on the population to which the PD model must be applied, i.e. the mortgage portfolio. At the time an impact analysis was performed as of June 2018.

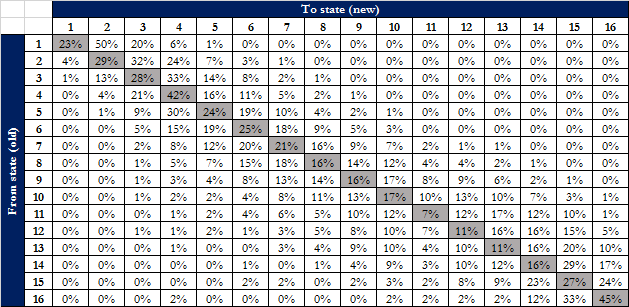
In the previous model, PD is done by tranche. However, in the operationalization and implementation, the maximum PD of the tranches is taken as the PD of the loan. As a result, in the impact analysis, we took the maximum PD of the tranches for the previous model in order to compare it with the PD of the proposed model.

Figure 14: Distribution per risk rating



The risk rating distribution shows that the new model is more concentrated in the investment grades (1, 2, 3, 4) than the old one.

Table 37: Transition matrix from old to new PD



## Default Parameter Value

If, for some operational reasons, there is missing information in the systems, and it is impossible to calculate the PD using the models presented in this document, default parameters will be assigned. The default PD value that will be given is the target long-run default rate. The table below presents the default PD value:

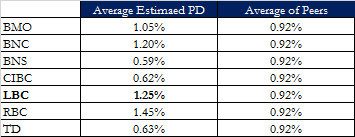
Table 38: Default Parameter Value



## Benchmarking

To assess the adequacy of our model estimates, we benchmarked our portfolio PD with the other banks corresponding portfolio PD. The following table represents the PD divulgated by other Canadian banks for residential mortgage in their Supplementary Regulatory Capital information in 2018Q1(available on the web). We compare with the average PD in the current portfolio as of June, 2018.

Table 39: Benchmarking with Canadian Peers



The comparison with the peers should always be done sparingly since we never know exactly how these PD are obtained. However, this may give some idea of the Bank's position relative to peers. Except RBC, LBC is above the peers in terms of average PD.

The Peers parameters that is presented there is from January 2018 while the BLC one is from June 2018. But that gives just an order of magnitude because we do not have all the information on how the PD of the peers presented in the reports of the various banks is calculated.

## Overrides

Overrides are not allowed for retail portfolio.

## Sources of Uncertainty and Associated Margin of Conservatism

All possible sources of uncertainty arising from model development, as well as and associated margin of conservatism, should be described in the table below:

Table 43: Sources of uncertainty and associated margins of conservatism

|  |  |
| --- | --- |
| **Source of Uncertainty** | **Conservatism** |
| New approach | This is not an issue since survival model is not a new approach since it is a widely-used approach in the literature. Section 6.2. |
| Reproducibility issues: Lack of clarity in several model variables | Any reproducibility issues are noted in the model process. Internal variables are used and are defined in the model in accordance with the bank definition. Sections 7.1.2; 7.1.5. |
| Omission of key risk drivers | All the pertinent internal variables are used including the borrowers, the loans and bureau of credit information. Sections 7.1.2; 7.1.5. |
| Invalidity of an hypothesis | The hypotheses are those of the survival analysis, no other hypothesis is emitted. Sections 6.3; 7.4.2. |
| Inadequation of the default definition | The definition of default is that used by the Bank and is consistent with the OSFI definition. Section 4.2. |
| Absence of default data | Defaulted accounts from 2008 to 2018 are used in the model process. Section 3.2.1.3. |
| Numerous expert judgments: Lack of documentation in regards to the exchanges with the experts | There are no expert judgments in the mortgage PD model. |
| Counterintuitive hypotheses | The hypotheses are those of the survival analysis, no other hypothesis is emitted. Sections 6.3; 7.4.2. |
| Numerous hypotheses | The hypotheses are those of the survival analysis, no other hypothesis is emitted. Sections 6.3; 7.4.2. |
| Out-of-sample estimation | This is not an issue since a validation test is done with out-of-sample data. Section 8.4. |
| Sign of coefficient | This is not an issue since all the signs of coefficient are good and coherent with what expected. Section 7.1.5. |
| Inadequately justified adjustment | No specific adjustment is used in this model. |
| Inadequate downturn adjustment | This is not an issue since the data comprises periods of economic slowdown as well as periods of economic prosperity. Otherwise, the economic conditions underlying the data are relevant to current and foreseeable conditions. Section 5.6.1; 7.3. |
| Sensitivity of parameter to shocks on explaining variables | This is not an issue since a sensibility analysis is done on the selected variables. Those which are not sensitive are removed from the model. Section 8.5. |
| Sample is representative of portfolio | This point is not an issue for the mortgage loans PD model as our development dataset was entirely built from internal sources. Further analysis in section 5.5 has shown that the distribution of the main drivers across the development and current populations were very similar. Section 5.5. |
| Data integrity | Some data integrity was raised during the model development:   * More than one principal client hold same ID\_GRNT. * A significant jump in noted in the default rate after 210 months for the mortgage accounts. Must be explained. * Issue in some variables like “nb\_mois\_depuis\_debour”.   However, a margin of conservatism is added at the model and will be enough to address theses issues. |
| Incoherence between treatments on external Vs. internal data | This is not an issue since no external data is used for the mortgage PD model. Section 1.6.2. |
| Insufficient number of defaults/observations | This point is not an issue for the mortgage loans PD model as our dataset covers 10 years of data (2008-2017), which covers both expansion periods and slowdown periods. The mortgage PD model contains thousands of defaults. Section 3.2.1.3. |
| Insufficient historical data coverage | This point is not an issue for the mortgage loans PD model as our dataset covers 10 years of data (2008-2017), which covers both expansion periods and slowdown periods. Section 5.6. |
| Inadequate use of model | The mortgage PD model is used to calculate the probability of default for mortgage portfolio of the LBC. In addition to this, it is used for capital calculation, stress-testing, reporting, etc. Section 3.4. |
| Operational treatment is not adequately aligned with model | This is not an issue. |
| Inadequate algorithm or program | This is not an issue since the model used algorithms from SAS that are replicable. Section 3.6. |

# **REFERENCES**

1. **Mamdouh Refaat**, Credit Risk Scorecards: Development and Implementation Using SAS, P 283 - 405
2. **Naeem Siddiqi**, Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring, Chapter 6, P 73 - 87
3. **Paul D. Allison**, Survival Analysis Using SAS: A Pratical Guide - Second Edition, Chapter 1 to 4.
4. **Bart Baesens**, Credit Risk Modeling Using SAS – SAS Institute
5. **Scherrer, B.** 1984. Biostatistique: equation 18-50, p. 704

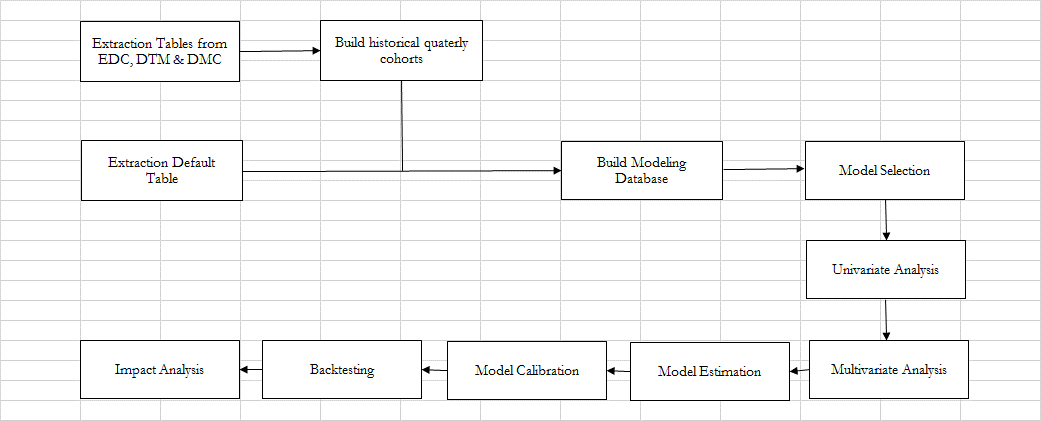
# **APPENDIX**

### Appendix 01: Files Location and Description

J:\970\Risque de credit\MODÈLES REGLÉMENTAIRES\II. RETAIL\RRS 4\PD\REVISION\_2018



### Appendix 02: Data flow and process chart



### Appendix 03: Useful definitions

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | VARIABLE NAME | DEFINITION | | NB\_MOIS\_ECHEANCE\_AMORT | Temps restant jusqu'à la fin de l'amortissement | | NB\_MOIS\_DEPUIS\_DEBOUR | Mois depuis déboursé. Nombre de mois depuis le déboursé du crédi | | ACTF\_MNT\_TOT\_ACTIF | Total des soldes des placements non clos et des comptes d'épargne non clos. | | ACTF\_MNT\_TOT\_EPRG | Total des soldes des comptes d'épargne non clos | | ACTF\_MNT\_TOT\_PLACEMENT | Total des valeurs des placements non clos. | | MNT\_AUTORISE | Limite ou montant autorisé du produit de crédit. | | MNT\_SOLDE | Solde du produit de crédit. | | NB\_JR\_DELQ\_ACT | Nombre de jours de délinquance actuelle du produit de crédit. | | NB\_JR\_DELQ\_MAX\_6MOIS | Nombre de jours maximum observés lors des 6 derniers mois pour un dossier d'affaire. | | NB\_JR\_DELQ\_MAX\_12MOIS | Nombre de jours maximum observés lors des 12 derniers mois pour un dossier d'affaire. | | HDEM\_NB\_ACCEPT\_3MOIS | # acceptations dernier 3 mois | | HDEM\_NB\_ACCEPT\_6MOIS | # acceptations dernier 6 mois | | HDEM\_NB\_ACCEPT\_12MOIS | # acceptations dernier 12 mois | | HDEM\_NB\_DMND\_3MOIS | # demandes dernier 3 mois | | HDEM\_NB\_DMND\_6MOIS | # demandes dernier 6 mois | | HDEM\_NB\_DMND\_12MOIS | # demandes dernier 12 mois | | HDEM\_NB\_MOIS\_DERN\_DMND | Mois depuis dernière demande | | PASF\_MNT\_TOT\_DELQ\_ACTUEL | Total des soldes des produits de crédit délinquants dont le statut est non clos. | | PASF\_MNT\_TOT\_SLD\_CRDT\_VERS | Somme des soldes des produits actifs de crédit à versements. | | PASF\_MNT\_TOT\_SLD\_HYPT | Somme des soldes des hypothèques actives | | PASF\_NB\_DELQ\_ACTUEL | Nombre de crédits non clos actuellement délinquant | | PASF\_NB\_JR\_MAX\_DELQ\_ACTUEL | Délinquance max 6 derniers mois | | PASF\_NB\_DELQ\_ACTUEL\_30PLUS | # délinquant actuel 30 jpd et + | | PASF\_NB\_DELQ\_ACTUEL\_60PLUS | # délinquant actuel 60 jpd et + | | PASF\_NB\_JR\_MAX\_DELQ\_3MOIS | Délinquance max 3 derniers mois | | PASF\_NB\_JR\_MAX\_DELQ\_6MOIS | Délinquance max 6 derniers mois | | PASF\_NB\_JR\_MAX\_DELQ\_12MOIS | Délinquance max 12 derniers mois | | PASF\_NB\_PRDT\_ACTIF | Nombre de produits de crédit qui sont actifs | | PASF\_NB\_PRDT\_CRDT\_VERS\_ACTIF | Nombre de produits actifs de crédit à versements | | PASF\_NB\_PRDT\_HYPT\_ACTIF | Nombre d'hypothèques actives. | | TMPS\_MOY\_MOIS\_CRDT\_BLC | Temp moyen avec crédit BLC | | TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC | Temps plus vieux crédit | | TRNX\_NB\_CRDT\_6MOIS | Nombre total de transactions au crédit durant les 180 jours précédant la date d'extraction des données pour l'ensemble des dossiers d'affaire d'un client. | | TRNX\_NB\_CRDT\_12MOIS | Nombre total de transactions au débit durant les 365 jours précédant la date d'extraction des données pour l'ensemble des dossiers d'affaire d'un client. | | TRNX\_NB\_NSF\_6MOIS | # NSF 6 mois | | TRNX\_NB\_NSF\_12MOIS | # NSF 12 mois | | BC\_AVG\_BAL\_BC\_TRD | Average balance of all open bankcard trades | | BC\_AVG\_BAL\_RE\_TRD | Average balance of all revolving trades | | BC\_AVG\_BAL\_RE\_TRD\_LAST3 | Average revolving balance over past 3 months | | BC\_AVG\_MOS\_BC\_TRD | Bankcard: Average number of months since date opened on  all bankcard trades. Count months difference between date opened and today date & divide by total number of bankcard trades. | | BC\_AVG\_MOS\_IN\_FILE | Average number of months since date opened on all trades. Count months difference between date opened and today date & divide by total number of trades. | | BC\_CD\_ORIG\_BUR\_CRDT | Identifie le bureau de crédit ayant fourni l'information. | | BC\_DT\_DERN\_BC | Ce champ contient la date de dernier chargement des informations bureau de crédit. | | BC\_FRAUD\_SYS\_CD | Fraud system code | | BC\_FRAUD\_VICTIM | Fraud victim indicator. | | BC\_HIGH\_HCCL\_AMT\_BC\_TRD | Bankcard: Highest credit limit on active bankcard trades. | | BC\_HIGH\_HCCL\_AMT\_BC\_TRD\_EVER | Highest amount of bankcard trade HC/CL, ever. | | BC\_HIGH\_HCCL\_AMT\_BY\_TRD | Highest amount of bank line of credit trade HC/CL, active. | | BC\_HIGH\_HCCL\_AMT\_RE\_TRD | Highest amount of revolving trade HC/CL, active. | | BC\_HIGH\_UTL\_BAL\_BC\_TRD | Balance corresponding to highest utilization. Bankcard | | BC\_HIGH\_UTL\_BAL\_BC\_TRD\_2 | Balance corresponding to highest utilization. Bankcard | | BC\_HIGH\_UTL\_BAL\_BR\_TRD | Balance corresponding to highest utilization. Bank revolving. | | BC\_HIGH\_UTL\_BAL\_BR\_TRD\_2 | Balance corresponding to second highest utilization. Bank revolving. | | BC\_HIGH\_UTL\_PCT\_BC\_RETAIL\_TRD | Highest credit line utilization on a retail tradeline. | | BC\_HIGH\_UTL\_PCT\_BC\_TRD | Highest utilization on bankcard trades. | | BC\_HIGH\_UTL\_PCT\_BC\_TRD\_2 | Second highest utilization. Bankcard. | | BC\_HIGH\_UTL\_PCT\_BR\_TRD | Highest Utilization on bank revolving trades. | | BC\_HIGH\_UTL\_PCT\_BR\_TRD\_2 | Second Highest Utilization. Bank revolving. | | BC\_MOS\_ON\_FILE | Months on file. | | BC\_MOS\_SNC\_OLD\_TRD\_OPN | Months since oldest trade opened. | | BC\_MOS\_SNC\_RCNT\_30\_DELQ | Number of months since last reported 30 day delinquency. | | BC\_MOS\_SNC\_RCNT\_60\_DELQ | Number of months since last reported 60 day delinquency. | | BC\_MOS\_SNC\_RCNT\_90P\_DELQ | Number of months since last reported 90 day delinquency. | | BC\_MOS\_SNC\_RCNT\_BC\_TRD\_ACT | # of months since last activity on bankcard trades. | | BC\_MOS\_SNC\_RCNT\_BC\_TRD\_OPN | Months since most recent bankcard trade opened. | | BC\_MOS\_SNC\_RCNT\_BKRP\_ITEM | # of months since most recent item in bankruptcy segment. | | BC\_MOS\_SNC\_RCNT\_BY\_TRD\_OPN | Months since most recent bank line of credit trade opened. | | BC\_MOS\_SNC\_RCNT\_CLLC | Months since most recent collection. | | BC\_MOS\_SNC\_RCNT\_DELQ | Months since most recent delinquency. | | BC\_MOS\_SNC\_RCNT\_DEROG\_PR | Months since most recent derogatory public record. | | BC\_MOS\_SNC\_RCNT\_IN\_TRD\_OPN | Months since most recent installment trade opened. All Installment | | BC\_MOS\_SNC\_RCNT\_OF\_TRD\_DELQ | Months since most recent other finance trade delinquency. Other Finance: | | BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT | # of months since latest activity of revolving trades. All revolving. | | BC\_MOS\_SNC\_RCNT\_RR\_TRD\_DELQ | Number of months since the most recent delinquency (rating of 2 or more) on retail revolving trades. | | BC\_MOS\_SNC\_RCNT\_RR\_TRD\_OPN | Number since most recent retail revolving trade opened. | | BC\_MOS\_SNC\_RCNT\_RT\_TRD\_OPN | Months since most recent retail trade opened. | | BC\_MOS\_SNC\_RCNT\_TRD\_OPN | Months since most recent trade opened. | | BC\_NB\_ACTIVE\_AI\_TRD | Number of active auto installment trades. | | BC\_NB\_ACTIVE\_BC\_TRD | Number of active bankcard trades. | | BC\_NB\_ACTIVE\_BI\_TRD | Number of active bank installment trades. | | BC\_NB\_ACTIVE\_BR\_TRD | Number of active bank revolving trades. | | BC\_NB\_ACTIVE\_BY\_TRD | Number of active bank line of credit trades. | | BC\_NB\_ACTIVE\_FI\_TRD | Number of active finance installment trades. | | BC\_NB\_ACTIVE\_FR\_TRD | Number of active finance revolving trades. | | BC\_NB\_ACTIVE\_IN\_TRD | Number of active installment trades. | | BC\_NB\_ACTIVE\_OF\_TRD | Number of active other finance trades. | | BC\_NB\_ACTIVE\_PF\_TRD | Number of active personal finance trades. | | BC\_NB\_ACTIVE\_RE\_TRD | Number of active revolving trades. | | BC\_NB\_ACTIVE\_RR\_TRD | Number of active retail revolving trades. | | BC\_NB\_ACTIVE\_RT\_TRD | Number of active retail trades. | | BC\_NB\_ACTIVE\_SD\_TRD | Number of active deferred student loans. | | BC\_NB\_ACTIVE\_SL\_TRD | Number of active student loans. | | BC\_NB\_ACTIVE\_TRD | Number of actives trades. | | BC\_NB\_ACTIVE\_TRD\_PAST\_DUE | Number of active trades with amount past due. | | BC\_NB\_BC\_TRD | Number of bankcard trades. | | BC\_NB\_BC\_TRD\_75PCT | Number of active bankcard trades with utilization >=75%. | | BC\_NB\_BC\_TRD\_90P\_EVER | Number of bankcard trades ever 90+ or more days past due. | | BC\_NB\_BC\_TRD\_90P\_LAST6 | Number of bankcard trades 90+ days past due in the last 6 months. B | | BC\_NB\_BC\_TRD\_LAST12\_BAL0P | Number of bankcard trades opened in past 12 months with balance > 0. | | BC\_NB\_BC\_TRD\_OPN | Number of open bankcard trades. Bankcard: Number of open bankcard trades. | | BC\_NB\_BC\_TRD\_OPN\_LAST3 | Number of bankcard trades opened in past 3 months. | | BC\_NB\_BC\_TRD\_OPN\_LAST12 | Number of bankcard trades opened in past 12 months | | BC\_NB\_BI\_TRD\_90P\_LAST6 | Number of bank installment trades 90+ past due in last 6 months. | | BC\_NB\_BI\_TRD\_OPN\_LAST12 | Number of bank installment trades opened in past 12 months. | | BC\_NB\_BR\_TRD | Number of bank revolving trades. | | BC\_NB\_BR\_TRD\_90P\_EVER | Number of bank revolving trades ever 90+ or more days past due. | | BC\_NB\_BR\_TRD\_90P\_LAST6 | Number of bank revolving trades 90+ or more days past due in the last 6 months. | | BC\_NB\_BR\_TRD\_LAST12\_BAL0P | Number of bank revolving trades opened in past 12 months with balance >0. | | BC\_NB\_BR\_TRD\_OPN\_LAST12 | Number of bank revolving trades opened in past 12 months. | | BC\_NB\_BY\_TRD\_BAL0P | Number of bank line of credit trades with balance > 0. | | BC\_NB\_CLLC\_LAST12\_BAL0P\_AMT250 | # of variables Collections within 12 months, balance >0, original amount>$250. | | BC\_NB\_DEROG\_PR | Number of derogatory public records. | | BC\_NB\_FIN\_INQ\_LAST6 | Number of finance inquiries within 6 months. | | BC\_NB\_INQ\_LAST2 | Number of inquiries in last 2 months. | | BC\_NB\_INQ\_LAST6 | Number of inquiries in last 6 months. | | BC\_NB\_INQ\_LAST12 | Number of inquiries in last 12 months. | | BC\_NB\_INQ\_LAST6\_EX\_LAST7 | Number of inquiries within last 6 months, excluding last 7 days. | | BC\_NB\_IN\_TRD | Number of installment trades. | | BC\_NB\_IN\_TRD\_30P\_EVER | Number of installment trades ever 30 or more days past due. | | BC\_NB\_IN\_TRD\_90P\_LAST6 | Number of installment trades 90+ days past due in last 6 months. | | BC\_NB\_IN\_TRD\_BAL0P | Number of active installment trades with balance >0. | | BC\_NB\_IN\_TRD\_OPN\_LAST6 | Number of installment trades opened in past 6 months. | | BC\_NB\_MT\_INQ\_LAST6 | # of mortgage inquiries within past 6 months. | | BC\_NB\_OF\_TRD | Number of other finance trades. | | BC\_NB\_OF\_TRD\_OPN\_LAST6 | Number of other finance trades opened in past 6 months. | | BC\_NB\_PF\_TRD | Number of personal finance trades. | | BC\_NB\_PF\_TRD\_OPN\_LAST6 | Number of personal finance trades opened in past 6 months. | | BC\_NB\_PR\_TRD\_90P\_EVER | # of variables derog trades and public records. | | BC\_NB\_RE\_TRD | Number of revolving trades. | | BC\_NB\_RE\_TRD\_30\_LAST6 | Number of revolving trades 30 days past due in last 6 months. | | BC\_NB\_RE\_TRD\_60\_LAST6 | Number of revolving trades 60 days past due in last 6 months. | | BC\_NB\_RE\_TRD\_90P\_EVER | Number of revolving trades ever 90+ or more days past due. | | BC\_NB\_RE\_TRD\_90P\_LAST6 | Number of revolving trades 90+ days past due in last 6 months. | | BC\_NB\_RE\_TRD\_LAST12\_BAL0P | Number of revolving trades opened in past 12 months with balance >0. | | BC\_NB\_RE\_TRD\_OPN\_LAST6 | Number of revolving trades opened in past 6 months. All revolving | | BC\_NB\_RE\_TRD\_OPN\_LAST12 | Number of revolving trades opened in past 12 months. All revolving | | BC\_NB\_RR\_TRD | Number of retail revolving trades. Retail Revolving | | BC\_NB\_RR\_TRD\_90P\_EVER | Number of retail revolving trades ever 90+ or more days past due. | | BC\_NB\_RR\_TRD\_90P\_LAST6 | Number of retail revolving trades 90+ days past due in last 6 months. | | BC\_NB\_RR\_TRD\_OPN | Number of open retail revolving trades. Retail revolving | | BC\_NB\_RR\_TRD\_OPN\_LAST6 | Number of retail revolving trades opened in past 6 months. Retail revolving | | BC\_NB\_RT\_TRD\_OPN | Number of open retail trades. Retail | | BC\_NB\_SATS | Number of satisfactory trades. | | BC\_NB\_SKIP\_LOCATE\_NOT | # of variables skip/locate notices on file. | | BC\_NB\_TRD | Number of trades. All industries: Total number of trades on credit file. | | BC\_NB\_TRD\_30P\_EVER | Number of trades ever 30 or more days past due. All industries | | BC\_NB\_TRD\_30\_EVER | Number of trades worse ever 30 days past due. | | BC\_NB\_TRD\_30\_LAST6 | Number of trades 30 days past due in the last 6 months. All industries | | BC\_NB\_TRD\_30\_LAST12 | Number of trades 30 days past due in the last 12 months. All industries: | | BC\_NB\_TRD\_60P\_EVER | Number of trades ever 60 or more days past due. | | BC\_NB\_TRD\_60\_EVER | Number of trades worse ever 60 days past due. | | BC\_NB\_TRD\_60\_LAST6 | Number of trades 60 days past due in the last 6 months. All industries | | BC\_NB\_TRD\_60\_LAST12 | Number of trades 60 days past due in the last 12 months. All industries | | BC\_NB\_TRD\_90P\_EVER | Number of trades ever 90 or more days past due. All industries | | BC\_NB\_TRD\_90P\_LAST6 | Number of trades 90+ past due in the last 6 months. All industries | | BC\_NB\_TRD\_90P\_LAST12 | Number of trades 90+ past due in the last 12 months. All industries | | BC\_NB\_TRD\_95PCT | # of active trades with utilization > % 95. All industries. | | BC\_NB\_TRD\_LAST12\_BAL0P | Number of trades opened in past 12 months with a balance >0. All industries. | | BC\_NB\_TRD\_OPN | Number of open trades. All industries: | | BC\_NB\_TRD\_OPN\_LAST3 | Number of trades opened in past 3 months. All industries | | BC\_NB\_TRD\_OPN\_LAST6 | Number of trades opened in past 6 months. All industries | | BC\_NB\_TRD\_OPN\_LAST12 | Number of trades opened in past 12 months. All industries: | | BC\_NB\_TRD\_OPN\_LAST24 | Number of trades opened in past 24 months. All industries: | | BC\_NET\_FRACTN\_BC\_TRD\_ACT\_BURDN | Utilization percent of active bankcard trades. | | BC\_NET\_FRACTN\_BC\_TRD\_BURDN | Ratio of total balance to HC/CL for all bankcard trades. | | BC\_NET\_FRACTN\_BR\_TRD\_ACT\_BURDN | Utilization percent of active bank revolving trades. | | BC\_NET\_FRACTN\_BY\_TRD\_BURDN | Ratio of total balance to HC/CL for all bank line of credit trades. | | BC\_NET\_FRACTN\_OPN\_RE\_TRD\_BURDN | Ratio of total balance to HC/CL for all open revolving trades | | BC\_NET\_FRACTN\_RE\_TRD\_ACT\_BURDN | Utilization percent of active revolving trades. | | BC\_NET\_FRACTN\_RR\_TRD\_ACT\_BURDN | Utilization percent of active retail revolving trades. | | BC\_NET\_FRACTN\_RR\_TRD\_BURDN | Ratio of total balance to HC/CL for all retail revolving trades. | | BC\_NET\_FRACTN\_RT\_TRD\_BURDN | Ratio of total balance to HC/CL for all retail trades. | | BC\_PCT\_TRD\_NEVER\_DELQ | Percent of trades never delinquent. | | BC\_REJECT\_REASON\_CD | Edit Reject Reason code. Reason why the data inquiry has been rejected. | | BC\_TOT\_AMT\_AVAIL | Available credit not utilized. All industries. | | BC\_TOT\_AMT\_PAST\_DUE | Total amount now past due. All industries: | | BC\_TOT\_BAL\_BC\_TRD\_ACTIVE | Total balance of active bankcard trades. | | BC\_TOT\_BAL\_BI\_TRD | Total balance of all bank installment trades. Bank Installment | | BC\_TOT\_BAL\_BR\_TRD | Total balance of all bank revolving trades. Bank revolving | | BC\_TOT\_BAL\_BY\_TRD | Total balance of all bank line of credit trades. Bank line of credit | | BC\_TOT\_BAL\_IN\_TRD | Total balance of all installment trades. All Installment | | BC\_TOT\_BAL\_IN\_TRD\_ACTIVE | Total balance of active installment trades. All Installment | | BC\_TOT\_BAL\_MT\_TRD | Total balance of mortgage trade. All industries. | | BC\_TOT\_BAL\_OF\_TRD | Total balance of all other finance trades. | | BC\_TOT\_BAL\_PF\_TRD | Total balance of all personal finance trades. Personal Finance | | BC\_TOT\_BAL\_RE\_TRD | Total balance of all revolving trades. All revolving: | | BC\_TOT\_BAL\_RT\_TRD | Total balance of all retail trades. Retail: Total balance of all retail trades | | BC\_TOT\_BAL\_TRD | Total Balance of all trades. All industries | | BC\_TOT\_BAL\_TRD\_EXCL\_MT | Total balance of all trades excluding mortgage. All industries | | BC\_TOT\_CLLC\_BAL | Total balance due to collections. | | BC\_TOT\_HCCL\_BC\_TRD | Total high credit/credit limit on active bankcard trade lines. | | BC\_TOT\_HCCL\_RE\_TRD | Total high credit - credit limit on active revolving trades. | | BC\_TOT\_HCCL\_TRD | Total high credit/credit limit. All industries: Total high credit/credit limit | | BC\_TOT\_MM\_COMMIT | Total Monthly commitment - variation of AT60. All industries. | | BC\_TOT\_MM\_PMNTS | Total monthly payments. All industries: Total amount of payments on all trades. | | BC\_TOT\_MM\_PMNTS\_IN\_TRD | Total monthly installment trade payments. All Installment | | BC\_TOT\_SEC\_BAL | Total secured balance. | | BC\_TOT\_SEC\_CL | Total secured credit limit. | | BC\_WORST\_BC\_RATING | Bankcard worst current rating | | BC\_WORST\_BI\_RATING | Bank installment worst credit rating. Bank Installment. | | BC\_WORST\_BR\_RATING | Bank revolving worst credit rating. Bank revolving. | | BC\_WORST\_IN\_RATING | Worst installment rating. All Installment. | | BC\_WORST\_RE\_RATING | Worst revolving rating. All revolving. | | BC\_WORST\_RR\_RATING | Retail revolving worst credit rating. Retail revolving. | |

### Appendix 04: Data dictionary

Please see Appendix 03 above.

### Appendix 05: Model inventory

N/A

### Appendix 06: Candidate explanatory variable and associated sources

Please see sections 7.1.3 & 7.1.4 of the documentation.

### Appendix 07: Potential risk drivers according to existing literature

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Factor** | **Effect on Default** | **Sources** |
| **Macro-economic** | Unemployment Rate | (+) | Belloti, T and J. Crook, 2009, Credit scoring with macroeconomic  variables using survivala analysis. |
| House Price volatility | (+) | Quercia, Roberto G., Anthony Pennington-Cross, and Chao Yue Tian,  2011, Mortgage default risk and local unemployment |
| Personal loan interest | (+) | Crook, Jonathan, and John Banasik, 2012, Forcasting and explaining  aggregate consumer credit delinquency behavior |
| Personal disposable income | (-) | Crook, Jonathan, and John Banasik, 2012, Forcasting and explaining  aggregate consumer credit delinquency behavior |
| Stock market index | (-) week significance | Belloti, T and J. Crook, 2009, Credit scoring with macroeconomic variables  using survivala analysis. |
| **Loan** | Loan to value ratio | (+) LTV's both at origination and at current time have a positive effect  on default; the positive effect kicks in after LTV exceeds certain threshold | Campbell, John Y, and Joao F. Cocco, 2014, A model of mortgage default,  working paper. |
| Loan rate volatility | (+) | Quercia, Roberto G., Anthony Pennington-Cross, and Chao Yue Tian, 2011,  Mortgage default risk and local unemployment |
| Loan size | (-) | Divino, Jose Angelo, Edna Souza Lima, and jaime Orrillo, 2013, Interest rate  and default in unsecured loan markets. |
| Age of mortgage | (\*) default display a rise-then fall pattern as mortgage age. | Campbell, Tim S., and J. Kimball Dietrich, 1983, The determinants of default  on insured conventional residential mortgage loans. |
| Mortgage term | (\*) Some find longer term mortgages default  more; others find the opposite or the insignificance of mortgage term. | Herzog, john P., and James S. Earley, 1970, The major determinants of differential  mortgage quality. |
| **Borrower** | Income | (-) High income relates to lower default, but  the effect is weaker or reversed for Graduated Payment Mortgages | Garriga, Carlos, and Don E. Schlagenhauf, 2010 Home equity, foreclosures, and bail-out  programs during the subprime crisis. |
| Credit quality | (-) | Quercia, Roberto G., Anthony Pennington-Cross, and Chao Yue Tian, 2011, Mortgage  default risk and local unemployment |
| Occupation | (\*) this may be category of occupation,  duration in current job, or other related measures. | Cunningham, Donald F., and Charles A. Capone, Jr., 1990, The relative termination  experience of adjustable to fixed-rate mortgages. |
| Age of borrower | (\*) a fall-rise fall features across borrower age cohort; borrower age is at the time when loan status is observed instead of at loan origination. | Cunningham, Donald F., and Charles A. Capone, Jr., 1990, The relative termination  experience of adjustable to fixed-rate mortgages. |
| Marital status | (\) | Herzog, john P., and James S. Earley, 1970, The major determinants of differential  mortgage quality. |
| **Property** | Property condition | (\*) | Campbell, Tim S., and J. Kimball Dietrich, 1983, The determinants of default  on insured conventional residential mortgage loans. |
| region | (\*) | Herzog, john P., and James S. Earley, 1970, The major determinants of differential  mortgage quality. |
| Neighborhood | (\*) | Mian, Atif, and Amir Sufi, 2009, The consequences of mortgage credit expansion:  evidence from the U.S. mortgage default crisis. |

(+): positive relationship between explanatory variable and default; higher probability of default is associated with higher value of the variable.

(-): negative relationship between explanatory variable and default; higher probability of default is associated with lower value of the variable.

(\*): significant relationship between explanatory variable and default, but the pattern may be non-monotonic, non-linear or inconsistent among studies, or the relationship does not have a positive or negative interpretation.

(\): insignificant relationship between explanatory variable and default.

### Appendix 08: Descriptive statistics

Please see section 3.2.1.3 of the documentation.

### Appendix 09: Missing observations and outliers

List 1 after clean up

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N Miss** | **%Missing** | **Minimum** | **Mean** | **Median** | **Maximum** |
| MNT\_AUTORISE | 2009878 | 0 | 0% | 2105.58 | 170615.94 | 150000 | 3854260 |
| MNT\_SOLDE | 2009878 | 0 | 0% | 0.01 | 152425.48 | 135200.4 | 2980086.08 |
| NB\_MOIS\_ECHEANCE\_AMORT | 2009878 | 0 | 0% | 0 | 259.2217572 | 271 | 600 |
| PASF\_NB\_PRDT\_ACTIF | 2009878 | 0 | 0% | 1 | 2.5741602 | 2 | 31 |
| PASF\_NB\_PRDT\_CRDT\_ROTF\_ACTIF | 2009878 | 0 | 0% | 0 | 1.0938331 | 1 | 15 |
| PASF\_NB\_PRDT\_CRDT\_VERS\_ACTIF | 2009878 | 0 | 0% | 1 | 1.4803272 | 1 | 30 |
| PASF\_NB\_PRDT\_HYPT\_ACTIF | 2009878 | 0 | 0% | 1 | 1.3997178 | 1 | 30 |
| PASF\_NB\_PRDT\_MC\_ACTIF | 2009878 | 0 | 0% | 0 | 0.4240849 | 0 | 9 |
| PASF\_NB\_PRDT\_OUV | 2009878 | 0 | 0% | 1 | 2.5946545 | 2 | 487 |
| PASF\_NB\_PRDT\_PRET\_ACTIF | 2009878 | 0 | 0% | 0 | 0.0806094 | 0 | 8 |
| PASF\_NB\_PRDT\_PSD | 2009878 | 0 | 0% | 0 | 0.2001126 | 0 | 8 |
| PASF\_NB\_PRDT\_TOTAL | 2009878 | 0 | 0% | 1 | 3.7304329 | 3 | 2203 |
| PASF\_NB\_PRDT\_VISA\_ACTIF | 2009878 | 0 | 0% | 0 | 0.4696355 | 0 | 5 |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_12MOIS | 1385463 | 624415 | 31% | 0 | 1.9650829 | 0.2223899 | 7418.03 |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_3MOIS | 1364595 | 645283 | 32% | 0 | 2.3760728 | 0.2148334 | 11165.08 |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_6MOIS | 1372208 | 637670 | 32% | 0 | 2.2622356 | 0.2177187 | 11055.38 |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_ACTUEL | 1223038 | 786840 | 39% | 0 | 0.430493 | 0.2872711 | 11237.59 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCH\_12M | 47974 | 1961904 | 98% | 0 | 0.576653 | 0.6585138 | 17.098889 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCH\_3M | 41251 | 1968627 | 98% | 0 | 0.5704937 | 0.6663172 | 17.098889 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCH\_6M | 43506 | 1966372 | 98% | 0 | 0.5722584 | 0.6638214 | 17.098889 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCH\_ACT | 39096 | 1970782 | 98% | 0 | 0.5656607 | 0.6689708 | 12.0869271 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCP\_12M | 554879 | 1454999 | 72% | 0 | 2.3413044 | 0.3742112 | 2106.11 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCP\_3M | 541731 | 1468147 | 73% | 0 | 2.9037298 | 0.370619 | 8421.43 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCP\_6M | 546242 | 1463636 | 73% | 0 | 2.6926292 | 0.3707702 | 4211.22 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_MCP\_ACT | 535701 | 1474177 | 73% | 0 | 0.4304499 | 0.3664435 | 6.2641 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_PSD\_12M | 297269 | 1712609 | 85% | 0 | 7.4303285 | 0.0037375 | 1735.4 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_PSD\_3M | 283170 | 1726708 | 86% | 0 | 7.9827679 | 0 | 1735.4 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_PSD\_6M | 287884 | 1721994 | 86% | 0 | 7.8206276 | 0 | 1735.4 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_PSD\_ACT | 278516 | 1731362 | 86% | 0 | 7.6650483 | 0 | 1935.48 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_VISA\_12M | 731603 | 1278275 | 64% | 0 | 0.3848275 | 0.2285506 | 3273.54 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_VISA\_3M | 695643 | 1314235 | 65% | 0 | 0.3901146 | 0.2281368 | 3273.54 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_VISA\_6M | 708582 | 1301296 | 65% | 0 | 0.3862875 | 0.2279963 | 3273.54 |
| PASF\_PCT\_UTLS\_CR\_ROTF\_VISA\_ACT | 681360 | 1328518 | 66% | 0 | 0.389747 | 0.217295 | 4128.86 |
| TMPS\_MOY\_MOIS\_CRDT\_BLC | 2009878 | 0 | 0% | 0 | 49.8332361 | 37 | 398 |
| TMPS\_NB\_MOIS\_OUVERT\_ANC\_CRDT | 2009878 | 0 | 0% | 0 | 81.9441648 | 60 | 713 |
| TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC | 2009878 | 0 | 0% | 0 | 115.3157291 | 82 | 1398 |
| TMPS\_NB\_MOIS\_OUVERT\_DERN\_CRDT | 2009878 | 0 | 0% | 0 | 34.2669968 | 23 | 398 |

List 2 after clean up

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N Miss** | **%Missing** | **Minimum** | **Mean** | **Median** | **Maximum** |
| NB\_DELQ\_30PLUS\_LTD | 2009878 | 0 | 0% | 0 | 0.121894 | 0 | 30 |
| NB\_DELQ\_60PLUS\_LTD | 2009878 | 0 | 0% | 0 | 0.0175931 | 0 | 9 |
| NB\_DELQ\_90PLUS\_LTD | 2009878 | 0 | 0% | 0 | 0.0013483 | 0 | 4 |
| NB\_JR\_DELQ\_ACT | 2009878 | 0 | 0% | 0 | 0.4743775 | 0 | 89 |
| NB\_JR\_DELQ\_MAX\_12MOIS | 2009878 | 0 | 0% | 0 | 2.5752548 | 0 | 773 |
| NB\_JR\_DELQ\_MAX\_24MOIS | 2009878 | 0 | 0% | 0 | 3.4871918 | 0 | 773 |
| NB\_JR\_DELQ\_MAX\_6MOIS | 2009878 | 0 | 0% | 0 | 1.8705946 | 0 | 773 |
| PASF\_NB\_DELQ\_ACTUEL | 2009878 | 0 | 0% | 0 | 0.0848559 | 0 | 29 |
| PASF\_NB\_DELQ\_ACTUEL\_30PLUS | 2009878 | 0 | 0% | 0 | 0.0135491 | 0 | 11 |
| PASF\_NB\_DELQ\_ACTUEL\_60PLUS | 2009878 | 0 | 0% | 0 | 0.0034136 | 0 | 9 |
| PASF\_NB\_DELQ\_ACTUEL\_90PLUS | 2009878 | 0 | 0% | 0 | 0.0009712 | 0 | 9 |
| PASF\_NB\_DELQ\_TOT\_30PLUS\_LTD | 2009878 | 0 | 0% | 0 | 0.2718309 | 0 | 33 |
| PASF\_NB\_DELQ\_TOT\_60PLUS\_LTD | 2009878 | 0 | 0% | 0 | 0.06824 | 0 | 13 |
| PASF\_NB\_DELQ\_TOT\_90PLUS\_LTD | 2009878 | 0 | 0% | 0 | 0.0246328 | 0 | 11 |
| PASF\_NB\_JR\_MAX\_DELQ\_12MOIS | 2009878 | 0 | 0% | 0 | 6.8680606 | 0 | 2085 |
| PASF\_NB\_JR\_MAX\_DELQ\_24MOIS | 2009878 | 0 | 0% | 0 | 9.6430271 | 0 | 2085 |
| PASF\_NB\_JR\_MAX\_DELQ\_3MOIS | 2009878 | 0 | 0% | 0 | 3.5880994 | 0 | 1824 |
| PASF\_NB\_JR\_MAX\_DELQ\_6MOIS | 2009878 | 0 | 0% | 0 | 4.9307724 | 0 | 1824 |
| PASF\_NB\_JR\_MAX\_DELQ\_ACTUEL | 2009878 | 0 | 0% | 0 | 1.1367964 | 0 | 1741 |
| PASF\_NB\_JR\_MAX\_DELQ\_LTD | 2009878 | 0 | 0% | 0 | 16.498794 | 2 | 2085 |
| TRNX\_NB\_CRDT\_12MOIS | 1451790 | 558088 | 28% | 0 | 83.7369936 | 57 | 2440 |
| TRNX\_NB\_CRDT\_6MOIS | 1451790 | 558088 | 28% | 0 | 41.8041101 | 28 | 1294 |
| TRNX\_NB\_DEBT\_12MOIS | 1451790 | 558088 | 28% | 0 | 316.469041 | 184 | 3810 |
| TRNX\_NB\_DEBT\_6MOIS | 1451790 | 558088 | 28% | 0 | 157.559393 | 92 | 2360 |
| TRNX\_NB\_NSF\_12MOIS | 1451790 | 558088 | 28% | 0 | 0.7612396 | 0 | 153 |
| TRNX\_NB\_NSF\_6MOIS | 1451790 | 558088 | 28% | 0 | 0.3956943 | 0 | 100 |
| TRNX\_NB\_NSF\_LTD | 1451790 | 558088 | 28% | 0 | 3.8489127 | 0 | 447 |
| PASF\_MNT\_TOT\_DELQ\_ACTUEL | 2009878 | 0 | 0% | 0 | 9071.11 | 0 | 5309671.54 |
| PASF\_MNT\_TOT\_LIMT\_CRDT\_ROTF | 2009878 | 0 | 0% | 0 | 11138 | 3000 | 2101500 |
| PASF\_MNT\_TOT\_LIMT\_PSD | 2009878 | 0 | 0% | 0 | 179.553903 | 0 | 9000 |
| PASF\_MNT\_TOT\_RADIE | 2009878 | 0 | 0% | 0 | 101.395211 | 0 | 140362.32 |
| PASF\_MNT\_TOT\_SLD\_CRDT\_ROTF | 2009878 | 0 | 0% | 0 | 4966.47 | 0 | 1974874.99 |
| PASF\_MNT\_TOT\_SLD\_CRDT\_VERS | 2009878 | 0 | 0% | 0.01 | 186506.37 | 145676.4 | 11228978.7 |
| PASF\_MNT\_TOT\_SLD\_HYPT | 2009878 | 0 | 0% | 0.01 | 185588.66 | 144849.6 | 11228978.7 |
| PASF\_MNT\_TOT\_SLD\_MC | 2009878 | 0 | 0% | 0 | 4006.39 | 0 | 1935628.67 |
| PASF\_MNT\_TOT\_SLD\_PRDT | 2009878 | 0 | 0% | 0 | 191440.87 | 149454.6 | 11228978.7 |
| PASF\_MNT\_TOT\_SLD\_PRET | 2009878 | 0 | 0% | 0 | 917.710302 | 0 | 2183064.19 |
| PASF\_MNT\_TOT\_SLD\_PSD | 2009878 | 0 | 0% | 0 | 34.0426084 | 0 | 12036.53 |
| PASF\_MNT\_TOT\_SLD\_VISA | 2009878 | 0 | 0% | 0 | 930.302703 | 0 | 234841.9 |
| AGE\_MOIS\_PRET | 2009878 | 0 | 0% | 0 | 49.1111053 | 35 | 397 |

List 3 after clean up

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N Miss** | **%Missing** | **Minimum** | **Mean** | **Median** | **Maximum** |
| BC\_NB\_IN\_TRD\_30P\_EVER | 1986694 | 23184 | 1% | 0 | 0.0917776 | 0 | 9 |
| BC\_NB\_IN\_TRD\_90P\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0100906 | 0 | 8 |
| BC\_NB\_IN\_TRD\_BAL0P | 1986694 | 23184 | 1% | 0 | 0.8574149 | 1 | 23 |
| BC\_NB\_IN\_TRD\_OPN\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.1368142 | 0 | 20 |
| BC\_NB\_MT\_INQ\_LAST6 | 1566996 | 442882 | 22% | 0 | 0 | 0 | 0 |
| BC\_NB\_OF\_TRD | 1986694 | 23184 | 1% | 0 | 1.659027 | 1 | 32 |
| BC\_NB\_OF\_TRD\_OPN\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0360629 | 0 | 6 |
| BC\_NB\_PF\_TRD | 1986694 | 23184 | 1% | 0 | 1.1195227 | 0 | 58 |
| BC\_NB\_PF\_TRD\_OPN\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.011796 | 0 | 5 |
| BC\_NB\_PR\_TRD\_90P\_EVER | 1986694 | 23184 | 1% | 0 | 0.4334543 | 0 | 24 |
| BC\_NB\_RE\_TRD | 1986694 | 23184 | 1% | 0 | 11.0005924 | 10 | 102 |
| BC\_NB\_RE\_TRD\_30\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.1274323 | 0 | 18 |
| BC\_NB\_RE\_TRD\_60\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0340742 | 0 | 18 |
| BC\_NB\_RE\_TRD\_90P\_EVER | 1986694 | 23184 | 1% | 0 | 0.1918479 | 0 | 21 |
| BC\_NB\_RE\_TRD\_90P\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0556996 | 0 | 19 |
| BC\_NB\_RE\_TRD\_LAST12\_BAL0P | 1566996 | 442882 | 22% | 0 | 0.270674 | 0 | 11 |
| BC\_NB\_RE\_TRD\_OPN\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.240796 | 0 | 23 |
| BC\_NB\_RE\_TRD\_OPN\_LAST12 | 1566996 | 442882 | 22% | 0 | 0.4938679 | 0 | 23 |
| BC\_NB\_RR\_TRD | 1986694 | 23184 | 1% | 0 | 2.3336684 | 2 | 32 |
| BC\_NB\_RR\_TRD\_90P\_EVER | 1986694 | 23184 | 1% | 0 | 0.0319324 | 0 | 7 |
| BC\_NB\_RR\_TRD\_90P\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.006682 | 0 | 5 |
| BC\_NB\_RR\_TRD\_OPN | 1986694 | 23184 | 1% | 0 | 1.1366028 | 1 | 22 |
| BC\_NB\_RR\_TRD\_OPN\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0252837 | 0 | 4 |
| BC\_NB\_RT\_TRD\_OPN | 1986694 | 23184 | 1% | 0 | 1.1447103 | 1 | 22 |
| BC\_NB\_SATS | 1986694 | 23184 | 1% | 0 | 17.2065245 | 16 | 117 |
| BC\_NB\_SKIP\_LOCATE\_NOT | 1986694 | 23184 | 1% | 0 | 0 | 0 | 0 |
| BC\_NB\_TRD | 1986694 | 23184 | 1% | 0 | 18.0790645 | 16 | 140 |
| BC\_NB\_TRD\_30P\_EVER | 1986694 | 23184 | 1% | 0 | 1.0468069 | 0 | 25 |
| BC\_NB\_TRD\_30\_EVER | 126173 | 1883705 | 94% | 0 | 0.6606723 | 0 | 13 |
| BC\_NB\_TRD\_30\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.1565943 | 0 | 18 |
| BC\_NB\_TRD\_30\_LAST12 | 1986694 | 23184 | 1% | 0 | 0.2735988 | 0 | 22 |
| BC\_NB\_TRD\_60P\_EVER | 1986694 | 23184 | 1% | 0 | 0.3999886 | 0 | 24 |
| BC\_NB\_TRD\_60\_EVER | 126173 | 1883705 | 94% | 0 | 0.172739 | 0 | 7 |
| BC\_NB\_TRD\_60\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0434601 | 0 | 18 |
| BC\_NB\_TRD\_60\_LAST12 | 1986694 | 23184 | 1% | 0 | 0.0805997 | 0 | 20 |
| BC\_NB\_TRD\_90P\_EVER | 1986694 | 23184 | 1% | 0 | 0.251165 | 0 | 24 |
| BC\_NB\_TRD\_90P\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0760872 | 0 | 21 |
| BC\_NB\_TRD\_90P\_LAST12 | 1986694 | 23184 | 1% | 0 | 0.1015994 | 0 | 23 |
| BC\_NB\_TRD\_95PCT | 1566996 | 442882 | 22% | 0 | 0.7591213 | 0 | 20 |
| BC\_NB\_TRD\_LAST12\_BAL0P | 1566996 | 442882 | 22% | 0 | 0.5660582 | 0 | 25 |
| BC\_NB\_TRD\_OPN | 1986694 | 23184 | 1% | 0 | 7.3305637 | 7 | 52 |
| BC\_NB\_TRD\_OPN\_LAST3 | 1986694 | 23184 | 1% | 0 | 0.1821483 | 0 | 43 |
| BC\_NB\_TRD\_OPN\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.3998507 | 0 | 43 |
| BC\_NB\_TRD\_OPN\_LAST12 | 1986694 | 23184 | 1% | 0 | 0.851349 | 1 | 43 |
| BC\_NB\_TRD\_OPN\_LAST24 | 1986694 | 23184 | 1% | 0 | 1.7925448 | 1 | 43 |
| BC\_NET\_FRACTN\_BC\_TRD\_ACT\_BURDN | 1799728 | 210150 | 10% | 0 | 35.374945 | 22 | 814 |
| BC\_NET\_FRACTN\_BC\_TRD\_BURDN | 1896388 | 113490 | 6% | 0 | 43.5107098 | 17 | 59028 |
| BC\_NET\_FRACTN\_BR\_TRD\_ACT\_BURDN | 1858001 | 151877 | 8% | 0 | 41.4267554 | 34 | 2364 |
| BC\_NET\_FRACTN\_BY\_TRD\_BURDN | 1296470 | 713408 | 35% | 0 | 44.5462355 | 34 | 95400 |
| BC\_NET\_FRACTN\_OPN\_RE\_TRD\_BURDN | 123314 | 1886564 | 94% | 0 | 56.3618973 | 28 | 91630 |
| BC\_NET\_FRACTN\_RE\_TRD\_ACT\_BURDN | 1873276 | 136602 | 7% | 0 | 40.2938355 | 33 | 1364 |
| BC\_NET\_FRACTN\_RR\_TRD\_ACT\_BURDN | 498886 | 1510992 | 75% | 0 | 13.8690883 | 1 | 681 |
| BC\_NET\_FRACTN\_RR\_TRD\_BURDN | 1163225 | 846653 | 42% | 0 | 17.7234112 | 0 | 93570 |
| BC\_NET\_FRACTN\_RT\_TRD\_BURDN | 1168688 | 841190 | 42% | 0 | 18.0546356 | 0 | 93570 |
| BC\_PCT\_TRD\_NEVER\_DELQ | 1986694 | 23184 | 1% | 0 | 92.5650659 | 100 | 100 |
| BC\_TOT\_AMT\_AVAIL | 1566996 | 442882 | 22% | 0 | 37776.05 | 23665 | 7459411 |
| BC\_TOT\_AMT\_PAST\_DUE | 1986694 | 23184 | 1% | 0 | 32.01864 | 0 | 309638 |
| BC\_TOT\_BAL\_BC\_TRD\_ACTIVE | 1986694 | 23184 | 1% | 0 | 4957.74 | 1717 | 451505 |
| BC\_TOT\_BAL\_BI\_TRD | 1986694 | 23184 | 1% | 0 | 8587.97 | 0 | 7795170 |
| BC\_TOT\_BAL\_BR\_TRD | 1986694 | 23184 | 1% | 0 | 17821.52 | 5261 | 3171255 |
| BC\_TOT\_BAL\_BY\_TRD | 1986694 | 23184 | 1% | 0 | 12081.45 | 0 | 3168056 |
| BC\_TOT\_BAL\_IN\_TRD | 1986694 | 23184 | 1% | 0 | 12497.48 | 2249 | 7877432 |
| BC\_TOT\_BAL\_IN\_TRD\_ACTIVE | 1986694 | 23184 | 1% | 0 | 11788.21 | 1709 | 7865959 |
| BC\_TOT\_BAL\_MT\_TRD | 1566996 | 442882 | 22% | 0 | 439.055526 | 0 | 3384981 |
| BC\_TOT\_BAL\_OF\_TRD | 1986694 | 23184 | 1% | 0 | 3778.33 | 0 | 578594 |
| BC\_TOT\_BAL\_PF\_TRD | 1986694 | 23184 | 1% | 0 | 229.302385 | 0 | 134592 |
| BC\_TOT\_BAL\_RE\_TRD | 1986694 | 23184 | 1% | 0 | 18164.51 | 5551 | 3171255 |
| BC\_TOT\_BAL\_RT\_TRD | 1986694 | 23184 | 1% | 0 | 247.777099 | 0 | 96758 |
| BC\_TOT\_BAL\_TRD | 1986694 | 23184 | 1% | 0 | 31229.09 | 15922 | 7984269 |
| BC\_TOT\_BAL\_TRD\_EXCL\_MT | 1566996 | 442882 | 22% | 0 | 35720.64 | 17753 | 8363154 |
| BC\_TOT\_CLLC\_BAL | 1986685 | 23193 | 1% | 0 | 67.5740281 | 0 | 217798 |
| BC\_TOT\_HCCL\_BC\_TRD | 1986694 | 23184 | 1% | 0 | 19728.53 | 13400 | 488000 |
| BC\_TOT\_HCCL\_RE\_TRD | 1986694 | 23184 | 1% | 0 | 47021.11 | 26510 | 4841430 |
| BC\_TOT\_HCCL\_TRD | 1986694 | 23184 | 1% | 0 | 65994 | 43349 | 8860945 |
| BC\_TOT\_MM\_COMMIT | 1566996 | 442882 | 22% | 0 | 582.411022 | 361 | 306149 |
| BC\_TOT\_MM\_PMNTS | 1986694 | 23184 | 1% | 0 | 494.334952 | 337 | 306149 |
| BC\_TOT\_MM\_PMNTS\_IN\_TRD | 1986694 | 23184 | 1% | 0 | 284.040923 | 0 | 267407 |
| BC\_TOT\_SEC\_BAL | 1566996 | 442882 | 22% | 0 | 2206.98 | 0 | 1935629 |
| BC\_TOT\_SEC\_CL | 1566996 | 442882 | 22% | 0 | 3213.68 | 0 | 2000000 |

List 4 after clean up

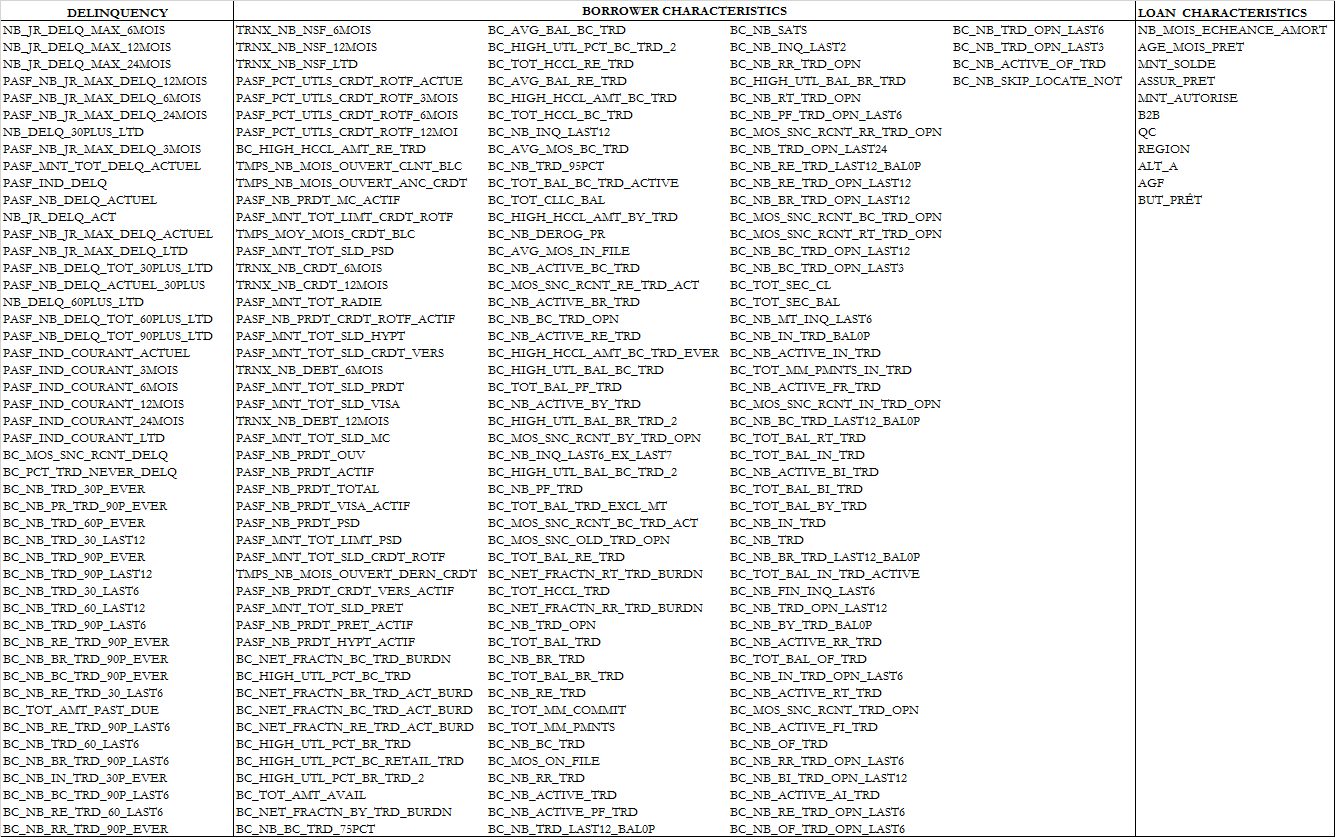
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N Miss** | **%Missing** | **Minimum** | **Mean** | **Median** | **Maximum** |
| BC\_AVG\_BAL\_BC\_TRD | 1986694 | 23184 | 1% | 0 | 1916.14 | 737 | 233468 |
| BC\_AVG\_BAL\_RE\_TRD | 1965624 | 44254 | 2% | 0 | 1630.88 | 615 | 365930 |
| BC\_AVG\_BAL\_RE\_TRD\_LAST3 | 421740 | 1588138 | 79% | 0 | 1280.81 | 64 | 378363 |
| BC\_AVG\_MOS\_BC\_TRD | 1986694 | 23184 | 1% | 0 | 115.160896 | 111 | 544 |
| BC\_AVG\_MOS\_IN\_FILE | 1986694 | 23184 | 1% | 0 | 110.245148 | 111 | 489 |
| BC\_HIGH\_HCCL\_AMT\_BC\_TRD | 1986694 | 23184 | 1% | 0 | 9555.37 | 7500 | 450000 |
| BC\_HIGH\_HCCL\_AMT\_BC\_TRD\_EVER | 1986694 | 23184 | 1% | 0 | 11375.09 | 10000 | 450000 |
| BC\_HIGH\_HCCL\_AMT\_BY\_TRD | 1566996 | 442882 | 22% | 0 | 18465.02 | 5000 | 4000000 |
| BC\_HIGH\_HCCL\_AMT\_RE\_TRD | 1566996 | 442882 | 22% | 0 | 22853.61 | 11400 | 4000000 |
| BC\_HIGH\_UTL\_BAL\_BC\_TRD | 1566996 | 442882 | 22% | 0 | 3273.43 | 1444 | 99510 |
| BC\_HIGH\_UTL\_BAL\_BC\_TRD\_2 | 1566996 | 442882 | 22% | 0 | 1166.6 | 0 | 99219 |
| BC\_HIGH\_UTL\_BAL\_BR\_TRD | 1566963 | 442915 | 22% | 0 | 6901.15 | 2874 | 99998 |
| BC\_HIGH\_UTL\_BAL\_BR\_TRD\_2 | 1566993 | 442885 | 22% | 0 | 3195.02 | 348 | 99997 |
| BC\_HIGH\_UTL\_PCT\_BC\_RETAIL\_TRD | 1986694 | 23184 | 1% | 0 | 33.2031737 | 15 | 814 |
| BC\_HIGH\_UTL\_PCT\_BC\_TRD | 1849237 | 160641 | 8% | 0 | 45.528837 | 36 | 13100 |
| BC\_HIGH\_UTL\_PCT\_BC\_TRD\_2 | 1003892 | 1005986 | 50% | 0 | 24.5498669 | 5 | 251 |
| BC\_HIGH\_UTL\_PCT\_BR\_TRD | 1890813 | 119065 | 6% | 0 | 57.4379682 | 63 | 95400 |
| BC\_HIGH\_UTL\_PCT\_BR\_TRD\_2 | 1239400 | 770478 | 38% | 0 | 35.3776061 | 19 | 251 |
| BC\_MOS\_ON\_FILE | 1974005 | 35873 | 2% | 0 | 237.505913 | 234 | 966 |
| BC\_MOS\_SNC\_OLD\_TRD\_OPN | 1970382 | 39496 | 2% | 0 | 231.313077 | 228 | 962 |
| BC\_MOS\_SNC\_RCNT\_30\_DELQ | 645924 | 1363954 | 68% | 1 | 20.965883 | 15 | 75 |
| BC\_MOS\_SNC\_RCNT\_60\_DELQ | 283672 | 1726206 | 86% | 1 | 26.1456506 | 22 | 73 |
| BC\_MOS\_SNC\_RCNT\_90P\_DELQ | 194799 | 1815079 | 90% | 1 | 27.8459027 | 25 | 141 |
| BC\_MOS\_SNC\_RCNT\_BC\_TRD\_ACT | 1949742 | 60136 | 3% | 0 | 2.9899058 | 1 | 242 |
| BC\_MOS\_SNC\_RCNT\_BC\_TRD\_OPN | 1949742 | 60136 | 3% | 0 | 43.5830971 | 28 | 544 |
| BC\_MOS\_SNC\_RCNT\_BKRP\_ITEM | 43592 | 1966286 | 98% | 0 | 41.1311709 | 35 | 243 |
| BC\_MOS\_SNC\_RCNT\_BY\_TRD\_OPN | 1455215 | 554663 | 28% | 0 | 71.8295297 | 59 | 582 |
| BC\_MOS\_SNC\_RCNT\_CLLC | 67777 | 1942101 | 97% | 0 | 24.715759 | 21 | 147 |
| BC\_MOS\_SNC\_RCNT\_DELQ | 834618 | 1175260 | 58% | 1 | 19.4886044 | 13 | 151 |
| BC\_MOS\_SNC\_RCNT\_DEROG\_PR | 58545 | 1951333 | 97% | 0 | 38.1214109 | 32 | 243 |
| BC\_MOS\_SNC\_RCNT\_IN\_TRD\_OPN | 1828537 | 181341 | 9% | 0 | 41.7009259 | 27 | 441 |
| BC\_MOS\_SNC\_RCNT\_OF\_TRD\_DELQ | 91745 | 1918133 | 95% | 1 | 33.8424764 | 33 | 151 |
| BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT | 1550973 | 458905 | 23% | 0 | 1.730092 | 1 | 239 |
| BC\_MOS\_SNC\_RCNT\_RR\_TRD\_DELQ | 216164 | 1793714 | 89% | 1 | 27.3492302 | 25 | 107 |
| BC\_MOS\_SNC\_RCNT\_RR\_TRD\_OPN | 1531517 | 478361 | 24% | 0 | 100.098147 | 84 | 637 |
| BC\_MOS\_SNC\_RCNT\_RT\_TRD\_OPN | 1540333 | 469545 | 23% | 0 | 99.3874448 | 83 | 637 |
| BC\_MOS\_SNC\_RCNT\_TRD\_OPN | 1968881 | 40997 | 2% | 0 | 19.3606323 | 12 | 489 |
| BC\_NB\_ACTIVE\_AI\_TRD | 1986694 | 23184 | 1% | 0 | 0.2397964 | 0 | 8 |
| BC\_NB\_ACTIVE\_BC\_TRD | 1986694 | 23184 | 1% | 0 | 2.244929 | 2 | 25 |
| BC\_NB\_ACTIVE\_BI\_TRD | 1986694 | 23184 | 1% | 0 | 0.5945475 | 0 | 23 |
| BC\_NB\_ACTIVE\_BR\_TRD | 1986694 | 23184 | 1% | 0 | 3.0413793 | 3 | 30 |
| BC\_NB\_ACTIVE\_BY\_TRD | 1986694 | 23184 | 1% | 0 | 0.7945043 | 1 | 16 |
| BC\_NB\_ACTIVE\_FI\_TRD | 1986694 | 23184 | 1% | 0 | 0.2582129 | 0 | 8 |
| BC\_NB\_ACTIVE\_FR\_TRD | 1986694 | 23184 | 1% | 0 | 0.1011354 | 0 | 6 |
| BC\_NB\_ACTIVE\_IN\_TRD | 1986694 | 23184 | 1% | 0 | 0.8574486 | 1 | 23 |
| BC\_NB\_ACTIVE\_OF\_TRD | 1986694 | 23184 | 1% | 0 | 0.2781359 | 0 | 8 |
| BC\_NB\_ACTIVE\_PF\_TRD | 1986694 | 23184 | 1% | 0 | 0.0851138 | 0 | 6 |
| BC\_NB\_ACTIVE\_RE\_TRD | 1986694 | 23184 | 1% | 0 | 3.4459278 | 3 | 34 |
| BC\_NB\_ACTIVE\_RR\_TRD | 1986694 | 23184 | 1% | 0 | 0.3001489 | 0 | 6 |
| BC\_NB\_ACTIVE\_RT\_TRD | 1986694 | 23184 | 1% | 0 | 0.3048411 | 0 | 6 |
| BC\_NB\_ACTIVE\_SD\_TRD | 1986694 | 23184 | 1% | 0 | 0.0010087 | 0 | 5 |
| BC\_NB\_ACTIVE\_SL\_TRD | 1986694 | 23184 | 1% | 0 | 0.0107062 | 0 | 5 |
| BC\_NB\_ACTIVE\_TRD | 1986694 | 23184 | 1% | 0 | 4.6155216 | 4 | 46 |
| BC\_NB\_ACTIVE\_TRD\_PAST\_DUE | 126173 | 1883705 | 94% | 0 | 0.0552258 | 0 | 12 |
| BC\_NB\_BC\_TRD | 1986694 | 23184 | 1% | 0 | 5.9672713 | 5 | 73 |
| BC\_NB\_BC\_TRD\_75PCT | 1986694 | 23184 | 1% | 0 | 0.4408434 | 0 | 20 |
| BC\_NB\_BC\_TRD\_90P\_EVER | 1986694 | 23184 | 1% | 0 | 0.1275803 | 0 | 15 |
| BC\_NB\_BC\_TRD\_90P\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0414573 | 0 | 15 |
| BC\_NB\_BC\_TRD\_LAST12\_BAL0P | 1986694 | 23184 | 1% | 0 | 0.1795052 | 0 | 11 |
| BC\_NB\_BC\_TRD\_OPN | 1986694 | 23184 | 1% | 0 | 3.1572134 | 3 | 33 |
| BC\_NB\_BC\_TRD\_OPN\_LAST3 | 1566996 | 442882 | 22% | 0 | 0.0759625 | 0 | 23 |
| BC\_NB\_BC\_TRD\_OPN\_LAST12 | 1566996 | 442882 | 22% | 0 | 0.3351719 | 0 | 23 |
| BC\_NB\_BI\_TRD\_90P\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0071325 | 0 | 7 |
| BC\_NB\_BI\_TRD\_OPN\_LAST12 | 1986694 | 23184 | 1% | 0 | 0.2216043 | 0 | 21 |
| BC\_NB\_BR\_TRD | 1986694 | 23184 | 1% | 0 | 7.5258379 | 7 | 83 |
| BC\_NB\_BR\_TRD\_90P\_EVER | 1986694 | 23184 | 1% | 0 | 0.1485055 | 0 | 19 |
| BC\_NB\_BR\_TRD\_90P\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0470571 | 0 | 18 |
| BC\_NB\_BR\_TRD\_LAST12\_BAL0P | 1986694 | 23184 | 1% | 0 | 0.2359578 | 0 | 11 |
| BC\_NB\_BR\_TRD\_OPN\_LAST12 | 1566996 | 442882 | 22% | 0 | 0.4289303 | 0 | 23 |
| BC\_NB\_BY\_TRD\_BAL0P | 1986694 | 23184 | 1% | 0 | 0.615656 | 0 | 15 |
| BC\_NB\_CLLC\_LAST12\_BAL0P\_AMT250 | 1986694 | 23184 | 1% | 0 | 0.0081875 | 0 | 9 |
| BC\_NB\_DEROG\_PR | 1986694 | 23184 | 1% | 0 | 0.0333136 | 0 | 8 |
| BC\_NB\_FIN\_INQ\_LAST6 | 1986694 | 23184 | 1% | 0 | 0.0127614 | 0 | 8 |
| BC\_NB\_INQ\_LAST2 | 1986694 | 23184 | 1% | 0 | 0.0788657 | 0 | 13 |
| BC\_NB\_INQ\_LAST6 | 126173 | 1883705 | 94% | 0 | 0.3480539 | 0 | 17 |
| BC\_NB\_INQ\_LAST12 | 1986694 | 23184 | 1% | 0 | 0.589964 | 0 | 25 |
| BC\_NB\_INQ\_LAST6\_EX\_LAST7 | 1986694 | 23184 | 1% | 0 | 0.2722578 | 0 | 18 |
| BC\_NB\_IN\_TRD | 1986694 | 23184 | 1% | 0 | 6.3924238 | 5 | 89 |

Outliers

|  |  |  |
| --- | --- | --- |
| **VARIABLE** | **UPPER** | **% HIGHER THAN UPPER** |
| AGE\_MOIS\_PRET | 221 | 0.98% |
| BC\_AVG\_BAL\_BC\_TRD | 15262 | 0.99% |
| BC\_AVG\_BAL\_RE\_TRD | 16670 | 0.98% |
| BC\_AVG\_MOS\_BC\_TRD | 279 | 0.98% |
| BC\_AVG\_MOS\_IN\_FILE | 219 | 0.98% |
| BC\_HIGH\_HCCL\_AMT\_BC\_TRD | 37000 | 0.97% |
| BC\_HIGH\_HCCL\_AMT\_BC\_TRD\_EVER | 40000 | 0.96% |
| BC\_HIGH\_HCCL\_AMT\_BY\_TRD | 260000 | 0.77% |
| BC\_HIGH\_HCCL\_AMT\_RE\_TRD | 260000 | 0.77% |
| BC\_HIGH\_UTL\_BAL\_BC\_TRD | 23069 | 0.78% |
| BC\_HIGH\_UTL\_BAL\_BC\_TRD\_2 | 15449 | 0.78% |
| BC\_HIGH\_UTL\_BAL\_BR\_TRD | 67143 | 0.78% |
| BC\_HIGH\_UTL\_BAL\_BR\_TRD\_2 | 38075 | 0.78% |
| BC\_HIGH\_UTL\_PCT\_BC\_RETAIL\_TRD | 108 | 0.98% |
| BC\_HIGH\_UTL\_PCT\_BC\_TRD | 115 | 0.87% |
| BC\_HIGH\_UTL\_PCT\_BC\_TRD\_2 | 101 | 0.41% |
| BC\_HIGH\_UTL\_PCT\_BR\_TRD | 115 | 0.89% |
| BC\_HIGH\_UTL\_PCT\_BR\_TRD\_2 | 101 | 0.43% |
| BC\_MOS\_ON\_FILE | 497 | 0.97% |
| BC\_MOS\_SNC\_OLD\_TRD\_OPN | 500 | 0.98% |
| BC\_MOS\_SNC\_RCNT\_BC\_TRD\_ACT | 64 | 0.96% |
| BC\_MOS\_SNC\_RCNT\_BC\_TRD\_OPN | 236 | 0.96% |
| BC\_MOS\_SNC\_RCNT\_BY\_TRD\_OPN | 253 | 0.71% |
| BC\_MOS\_SNC\_RCNT\_DELQ | 70 | 0.30% |
| BC\_MOS\_SNC\_RCNT\_IN\_TRD\_OPN | 199 | 0.91% |
| BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT | 34 | 0.77% |
| BC\_MOS\_SNC\_RCNT\_RR\_TRD\_OPN | 356 | 0.76% |
| BC\_MOS\_SNC\_RCNT\_RT\_TRD\_OPN | 356 | 0.76% |
| BC\_MOS\_SNC\_RCNT\_TRD\_OPN | 107 | 0.96% |
| BC\_NB\_ACTIVE\_AI\_TRD | 2 | 0.18% |
| BC\_NB\_ACTIVE\_BC\_TRD | 8 | 0.57% |
| BC\_NB\_ACTIVE\_BI\_TRD | 3 | 0.96% |
| BC\_NB\_ACTIVE\_BR\_TRD | 9 | 0.91% |
| BC\_NB\_ACTIVE\_BY\_TRD | 4 | 0.40% |
| BC\_NB\_ACTIVE\_FI\_TRD | 2 | 0.23% |
| BC\_NB\_ACTIVE\_FR\_TRD | 2 | 0.18% |
| BC\_NB\_ACTIVE\_IN\_TRD | 4 | 0.56% |
| BC\_NB\_ACTIVE\_OF\_TRD | 2 | 0.35% |
| BC\_NB\_ACTIVE\_PF\_TRD | 2 | 0.14% |
| BC\_NB\_ACTIVE\_RE\_TRD | 11 | 0.59% |
| BC\_NB\_ACTIVE\_RR\_TRD | 2 | 0.75% |
| BC\_NB\_ACTIVE\_RT\_TRD | 2 | 0.76% |
| BC\_NB\_ACTIVE\_TRD | 13 | 0.71% |
| BC\_NB\_BC\_TRD | 19 | 0.98% |
| BC\_NB\_BC\_TRD\_75PCT | 4 | 0.49% |
| BC\_NB\_BC\_TRD\_90P\_EVER | 3 | 0.65% |
| BC\_NB\_BC\_TRD\_90P\_LAST6 | 1 | 0.86% |
| BC\_NB\_BC\_TRD\_LAST12\_BAL0P | 2 | 0.21% |
| BC\_NB\_BC\_TRD\_OPN | 10 | 0.90% |
| BC\_NB\_BC\_TRD\_OPN\_LAST3 | 1 | 0.36% |
| BC\_NB\_BC\_TRD\_OPN\_LAST12 | 3 | 0.16% |
| BC\_NB\_BI\_TRD\_OPN\_LAST12 | 2 | 0.47% |
| BC\_NB\_BR\_TRD | 23 | 0.93% |
| BC\_NB\_BR\_TRD\_90P\_EVER | 3 | 0.92% |
| BC\_NB\_BR\_TRD\_90P\_LAST6 | 1 | 0.98% |
| BC\_NB\_BR\_TRD\_LAST12\_BAL0P | 2 | 0.47% |
| BC\_NB\_BR\_TRD\_OPN\_LAST12 | 3 | 0.37% |
| BC\_NB\_BY\_TRD\_BAL0P | 3 | 0.72% |
| BC\_NB\_DEROG\_PR | 1 | 0.30% |
| BC\_NB\_FIN\_INQ\_LAST6 | 1 | 0.13% |
| BC\_NB\_INQ\_LAST2 | 1 | 0.75% |
| BC\_NB\_INQ\_LAST12 | 4 | 0.75% |
| BC\_NB\_INQ\_LAST6\_EX\_LAST7 | 3 | 0.34% |
| BC\_NB\_IN\_TRD | 25 | 0.86% |
| BC\_NB\_IN\_TRD\_30P\_EVER | 2 | 0.37% |
| BC\_NB\_IN\_TRD\_BAL0P | 4 | 0.56% |
| BC\_NB\_IN\_TRD\_OPN\_LAST6 | 2 | 0.13% |
| BC\_NB\_OF\_TRD | 8 | 0.83% |
| BC\_NB\_OF\_TRD\_OPN\_LAST6 | 1 | 0.10% |
| BC\_NB\_PF\_TRD | 10 | 0.97% |
| BC\_NB\_PF\_TRD\_OPN\_LAST6 | 1 | 0.08% |
| BC\_NB\_PR\_TRD\_90P\_EVER | 6 | 0.79% |
| BC\_NB\_RE\_TRD | 33 | 0.85% |
| BC\_NB\_RE\_TRD\_30\_LAST6 | 2 | 0.61% |
| BC\_NB\_RE\_TRD\_60\_LAST6 | 1 | 0.49% |
| BC\_NB\_RE\_TRD\_90P\_EVER | 4 | 0.72% |
| BC\_NB\_RE\_TRD\_90P\_LAST6 | 2 | 0.54% |
| BC\_NB\_RE\_TRD\_LAST12\_BAL0P | 2 | 0.53% |
| BC\_NB\_RE\_TRD\_OPN\_LAST6 | 2 | 0.66% |
| BC\_NB\_RE\_TRD\_OPN\_LAST12 | 3 | 0.57% |
| BC\_NB\_RR\_TRD | 10 | 0.70% |
| BC\_NB\_RR\_TRD\_90P\_EVER | 1 | 0.30% |
| BC\_NB\_RR\_TRD\_OPN | 6 | 0.61% |
| BC\_NB\_RR\_TRD\_OPN\_LAST6 | 1 | 0.07% |
| BC\_NB\_RT\_TRD\_OPN | 6 | 0.62% |
| BC\_NB\_SATS | 48 | 0.92% |
| BC\_NB\_TRD | 50 | 0.94% |
| BC\_NB\_TRD\_30P\_EVER | 7 | 0.78% |
| BC\_NB\_TRD\_30\_LAST6 | 2 | 0.86% |
| BC\_NB\_TRD\_30\_LAST12 | 3 | 0.82% |
| BC\_NB\_TRD\_60P\_EVER | 5 | 0.90% |
| BC\_NB\_TRD\_60\_LAST6 | 1 | 0.64% |
| BC\_NB\_TRD\_60\_LAST12 | 2 | 0.50% |
| BC\_NB\_TRD\_90P\_EVER | 5 | 0.70% |
| BC\_NB\_TRD\_90P\_LAST6 | 2 | 0.77% |
| BC\_NB\_TRD\_90P\_LAST12 | 3 | 0.61% |
| BC\_NB\_TRD\_95PCT | 4 | 0.73% |
| BC\_NB\_TRD\_LAST12\_BAL0P | 3 | 0.74% |
| BC\_NB\_TRD\_OPN | 20 | 0.98% |
| BC\_NB\_TRD\_OPN\_LAST3 | 2 | 0.42% |
| BC\_NB\_TRD\_OPN\_LAST6 | 3 | 0.51% |
| BC\_NB\_TRD\_OPN\_LAST12 | 5 | 0.39% |
| BC\_NB\_TRD\_OPN\_LAST24 | 7 | 0.89% |
| BC\_NET\_FRACTN\_BC\_TRD\_ACT\_BURDN | 106 | 0.83% |
| BC\_NET\_FRACTN\_BC\_TRD\_BURDN | 287 | 0.94% |
| BC\_NET\_FRACTN\_BR\_TRD\_ACT\_BURDN | 104 | 0.78% |
| BC\_NET\_FRACTN\_BY\_TRD\_BURDN | 100 | 0.47% |
| BC\_NET\_FRACTN\_RE\_TRD\_ACT\_BURDN | 103 | 0.79% |
| BC\_NET\_FRACTN\_RR\_TRD\_BURDN | 99 | 0.54% |
| BC\_NET\_FRACTN\_RT\_TRD\_BURDN | 99 | 0.56% |
| BC\_TOT\_AMT\_AVAIL | 260711 | 0.78% |
| BC\_TOT\_AMT\_PAST\_DUE | 544 | 0.99% |
| BC\_TOT\_BAL\_BC\_TRD\_ACTIVE | 43182 | 0.99% |
| BC\_TOT\_BAL\_BI\_TRD | 76149 | 0.99% |
| BC\_TOT\_BAL\_BR\_TRD | 219651 | 0.99% |
| BC\_TOT\_BAL\_BY\_TRD | 204921 | 0.99% |
| BC\_TOT\_BAL\_IN\_TRD | 89089 | 0.99% |
| BC\_TOT\_BAL\_IN\_TRD\_ACTIVE | 84419 | 0.99% |
| BC\_TOT\_BAL\_OF\_TRD | 42995 | 0.99% |
| BC\_TOT\_BAL\_PF\_TRD | 7514 | 0.99% |
| BC\_TOT\_BAL\_RE\_TRD | 220735 | 0.99% |
| BC\_TOT\_BAL\_RT\_TRD | 5253 | 0.99% |
| BC\_TOT\_BAL\_TRD | 284207 | 0.99% |
| BC\_TOT\_BAL\_TRD\_EXCL\_MT | 324937 | 0.78% |
| BC\_TOT\_CLLC\_BAL | 786 | 0.99% |
| BC\_TOT\_HCCL\_BC\_TRD | 103300 | 0.99% |
| BC\_TOT\_HCCL\_RE\_TRD | 386200 | 0.99% |
| BC\_TOT\_HCCL\_TRD | 460808 | 0.99% |
| BC\_TOT\_MM\_COMMIT | 4168 | 0.78% |
| BC\_TOT\_MM\_PMNTS | 2598 | 0.99% |
| BC\_TOT\_MM\_PMNTS\_IN\_TRD | 1729 | 0.99% |
| BC\_TOT\_SEC\_BAL | 39642 | 0.78% |
| BC\_TOT\_SEC\_CL | 50500 | 0.77% |
| MNT\_AUTORISE | 537276.5 | 1.00% |
| MNT\_SOLDE | 513291.88 | 1.00% |
| NB\_DELQ\_30PLUS\_LTD | 3 | 0.90% |
| NB\_DELQ\_60PLUS\_LTD | 1 | 0.32% |
| NB\_JR\_DELQ\_ACT | 18 | 1.00% |
| NB\_JR\_DELQ\_MAX\_12MOIS | 49 | 0.99% |
| NB\_JR\_DELQ\_MAX\_24MOIS | 56 | 0.96% |
| NB\_JR\_DELQ\_MAX\_6MOIS | 41 | 0.98% |
| NB\_MOIS\_ECHEANCE\_AMORT | 456 | 0.97% |
| PASF\_MNT\_TOT\_DELQ\_ACTUEL | 233923.12 | 1.00% |
| PASF\_MNT\_TOT\_LIMT\_CRDT\_ROTF | 128000 | 1.00% |
| PASF\_MNT\_TOT\_LIMT\_PSD | 1500 | 0.82% |
| PASF\_MNT\_TOT\_RADIE | 751.85 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_CRDT\_ROTF | 70938.53 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_CRDT\_VERS | 924700.1 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_HYPT | 921795.62 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_MC | 68015.73 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_PRDT | 954466.66 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_PRET | 21515.6 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_PSD | 1045.41 | 1.00% |
| PASF\_MNT\_TOT\_SLD\_VISA | 10493.16 | 1.00% |
| PASF\_NB\_DELQ\_ACTUEL | 2 | 0.37% |
| PASF\_NB\_DELQ\_ACTUEL\_30PLUS | 1 | 0.12% |
| PASF\_NB\_DELQ\_TOT\_30PLUS\_LTD | 3 | 0.62% |
| PASF\_NB\_DELQ\_TOT\_60PLUS\_LTD | 1 | 0.83% |
| PASF\_NB\_DELQ\_TOT\_90PLUS\_LTD | 1 | 0.19% |
| PASF\_NB\_JR\_MAX\_DELQ\_12MOIS | 73 | 0.95% |
| PASF\_NB\_JR\_MAX\_DELQ\_24MOIS | 86 | 0.97% |
| PASF\_NB\_JR\_MAX\_DELQ\_3MOIS | 57 | 0.97% |
| PASF\_NB\_JR\_MAX\_DELQ\_6MOIS | 64 | 0.96% |
| PASF\_NB\_JR\_MAX\_DELQ\_ACTUEL | 31 | 0.92% |
| PASF\_NB\_JR\_MAX\_DELQ\_LTD | 138 | 0.98% |
| PASF\_NB\_PRDT\_ACTIF | 9 | 0.72% |
| PASF\_NB\_PRDT\_CRDT\_ROTF\_ACTIF | 4 | 0.44% |
| PASF\_NB\_PRDT\_CRDT\_VERS\_ACTIF | 6 | 0.83% |
| PASF\_NB\_PRDT\_HYPT\_ACTIF | 6 | 0.75% |
| PASF\_NB\_PRDT\_MC\_ACTIF | 2 | 0.46% |
| PASF\_NB\_PRDT\_OUV | 9 | 0.75% |
| PASF\_NB\_PRDT\_PRET\_ACTIF | 1 | 0.89% |
| PASF\_NB\_PRDT\_PSD | 2 | 0.11% |
| PASF\_NB\_PRDT\_TOTAL | 16 | 0.95% |
| PASF\_NB\_PRDT\_VISA\_ACTIF | 2 | 0.11% |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_12MOIS | 1.011804631 | 0.69% |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_3MOIS | 1.039424242 | 0.68% |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_6MOIS | 1.026844706 | 0.68% |
| PASF\_PCT\_UTLS\_CRDT\_ROTF\_ACTUEL | 1.04225 | 0.61% |
| TMPS\_MOY\_MOIS\_CRDT\_BLC | 210 | 0.98% |
| TMPS\_NB\_MOIS\_OUVERT\_ANC\_CRDT | 303 | 0.99% |
| TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC | 417 | 0.99% |
| TMPS\_NB\_MOIS\_OUVERT\_DERN\_CRDT | 194 | 1.00% |
| TRNX\_NB\_CRDT\_12MOIS | 416 | 0.72% |
| TRNX\_NB\_CRDT\_6MOIS | 209 | 0.71% |
| TRNX\_NB\_DEBT\_12MOIS | 1455 | 0.72% |
| TRNX\_NB\_DEBT\_6MOIS | 730 | 0.72% |
| TRNX\_NB\_NSF\_12MOIS | 13 | 0.62% |
| TRNX\_NB\_NSF\_6MOIS | 7 | 0.63% |
| TRNX\_NB\_NSF\_LTD | 57 | 0.70% |

### Appendix 10: Univariate analysis on candidate variables

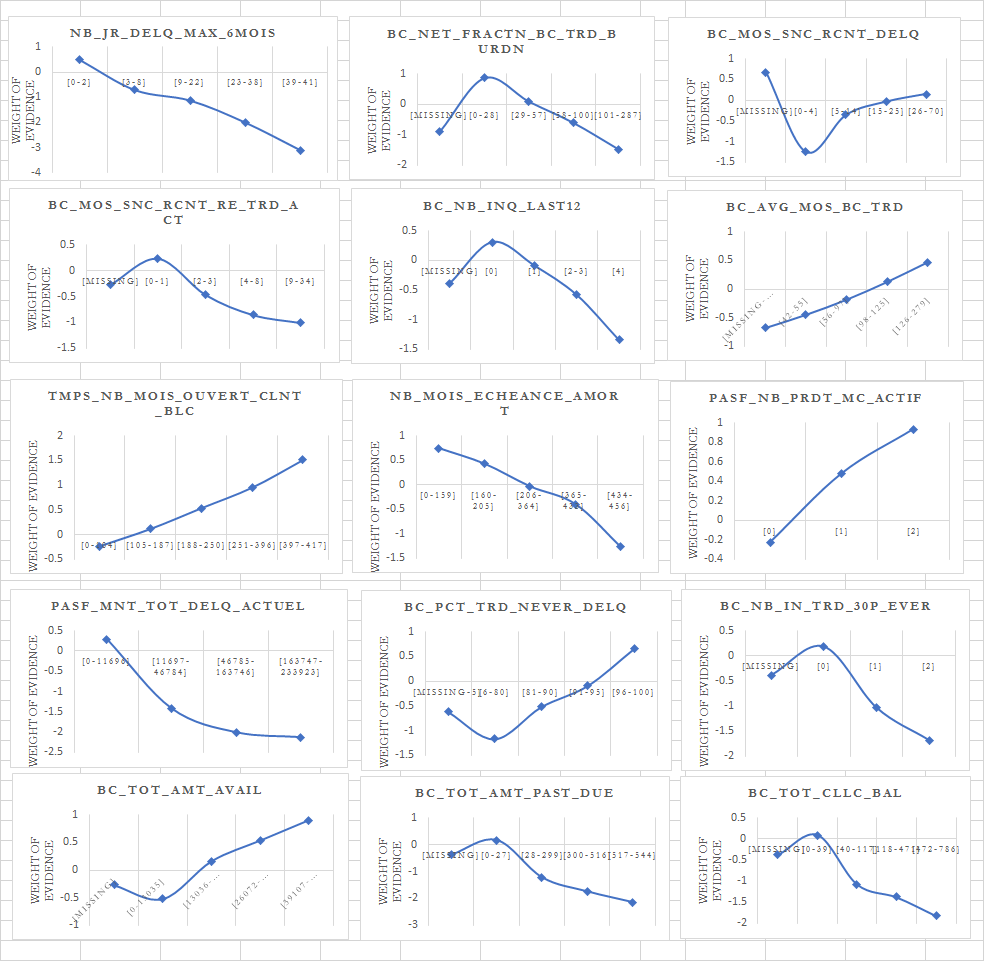
All tested variables



Information Value of tested variables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **VARIABLE** | **INFORMATION VALUE** | | NB\_JR\_DELQ\_MAX\_6MOIS | 0.748243058 | | NB\_JR\_DELQ\_MAX\_12MOIS | 0.725470855 | | NB\_JR\_DELQ\_MAX\_24MOIS | 0.645686787 | | PASF\_NB\_JR\_MAX\_DELQ\_12MOIS | 0.621943273 | | BC\_NET\_FRACTN\_BC\_TRD\_BURDN | 0.621558881 | | PASF\_NB\_JR\_MAX\_DELQ\_6MOIS | 0.603138414 | | PASF\_NB\_JR\_MAX\_DELQ\_24MOIS | 0.595162752 | | NB\_DELQ\_30PLUS\_LTD | 0.593510585 | | PASF\_NB\_JR\_MAX\_DELQ\_3MOIS | 0.592698823 | | BC\_HIGH\_UTL\_PCT\_BC\_TRD | 0.575464289 | | BC\_NET\_FRACTN\_BR\_TRD\_ACT\_BURD | 0.551565472 | | BC\_NET\_FRACTN\_BC\_TRD\_ACT\_BURD | 0.549962745 | | BC\_MOS\_SNC\_RCNT\_DELQ | 0.545388957 | | BC\_NET\_FRACTN\_RE\_TRD\_ACT\_BURD | 0.540512681 | | PASF\_MNT\_TOT\_DELQ\_ACTUEL | 0.526842852 | | PASF\_IND\_DELQ | 0.526675876 | | PASF\_NB\_DELQ\_ACTUEL | 0.526675876 | | BC\_HIGH\_UTL\_PCT\_BR\_TRD | 0.520522087 | | NB\_JR\_DELQ\_ACT | 0.518063909 | | PASF\_NB\_JR\_MAX\_DELQ\_ACTUEL | 0.487430525 | | BC\_PCT\_TRD\_NEVER\_DELQ | 0.484305504 | | PASF\_NB\_JR\_MAX\_DELQ\_LTD | 0.448762331 | | BC\_NB\_TRD\_30P\_EVER | 0.428825183 | | TRNX\_NB\_NSF\_6MOIS | 0.411732329 | | TRNX\_NB\_NSF\_12MOIS | 0.405340838 | | PASF\_NB\_DELQ\_TOT\_30PLUS\_LTD | 0.391604787 | | BC\_HIGH\_UTL\_PCT\_BC\_RETAIL\_TRD | 0.385196966 | | BC\_NB\_PR\_TRD\_90P\_EVER | 0.381287627 | | BC\_NB\_TRD\_60P\_EVER | 0.367957775 | | PASF\_NB\_DELQ\_ACTUEL\_30PLUS | 0.360037723 | | NB\_DELQ\_60PLUS\_LTD | 0.358615343 | | BC\_NB\_TRD\_30\_LAST12 | 0.356473988 | | BC\_NB\_TRD\_90P\_EVER | 0.334017239 | | TRNX\_NB\_NSF\_LTD | 0.325633336 | | BC\_NB\_TRD\_90P\_LAST12 | 0.325249965 | | BC\_NB\_TRD\_30\_LAST6 | 0.313977095 | | BC\_NB\_TRD\_60\_LAST12 | 0.303205955 | | BC\_NB\_TRD\_90P\_LAST6 | 0.295617593 | | BC\_NB\_RE\_TRD\_90P\_EVER | 0.290337995 | | BC\_HIGH\_UTL\_PCT\_BR\_TRD\_2 | 0.283324125 | | PASF\_PCT\_UTLS\_CRDT\_ROTF\_ACTUE | 0.282822082 | | BC\_NB\_BR\_TRD\_90P\_EVER | 0.276937921 | | BC\_TOT\_AMT\_AVAIL | 0.273348288 | | BC\_NET\_FRACTN\_BY\_TRD\_BURDN | 0.272094295 | | BC\_NB\_BC\_TRD\_90P\_EVER | 0.2684176 | | BC\_NB\_BC\_TRD\_75PCT | 0.266750874 | | BC\_NB\_RE\_TRD\_30\_LAST6 | 0.266267991 | | BC\_AVG\_BAL\_BC\_TRD | 0.261035323 | | PASF\_NB\_DELQ\_TOT\_60PLUS\_LTD | 0.25830448 | | BC\_TOT\_AMT\_PAST\_DUE | 0.244060075 | | BC\_NB\_RE\_TRD\_90P\_LAST6 | 0.241830273 | | BC\_NB\_TRD\_60\_LAST6 | 0.23398207 | | BC\_HIGH\_UTL\_PCT\_BC\_TRD\_2 | 0.227471547 | | BC\_NB\_BR\_TRD\_90P\_LAST6 | 0.222803139 | | BC\_NB\_IN\_TRD\_30P\_EVER | 0.214922689 | | BC\_NB\_BC\_TRD\_90P\_LAST6 | 0.209617642 | | PASF\_PCT\_UTLS\_CRDT\_ROTF\_3MOIS | 0.205241382 | | BC\_TOT\_HCCL\_RE\_TRD | 0.197725355 | | BC\_NB\_RE\_TRD\_60\_LAST6 | 0.197574218 | | PASF\_PCT\_UTLS\_CRDT\_ROTF\_6MOIS | 0.195674289 | | PASF\_PCT\_UTLS\_CRDT\_ROTF\_12MOI | 0.18040013 | | BC\_HIGH\_HCCL\_AMT\_RE\_TRD | 0.177021757 | | BC\_AVG\_BAL\_RE\_TRD | 0.168535908 | | BC\_HIGH\_HCCL\_AMT\_BC\_TRD | 0.155675709 | | BC\_TOT\_HCCL\_BC\_TRD | 0.153129962 | | BC\_NB\_INQ\_LAST12 | 0.150258679 | | BC\_AVG\_MOS\_BC\_TRD | 0.149699021 | | BC\_NB\_TRD\_95PCT | 0.149516277 | | BC\_TOT\_BAL\_BC\_TRD\_ACTIVE | 0.147529405 | | TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC | 0.138580753 | | BC\_TOT\_CLLC\_BAL | 0.133297931 | | NB\_MOIS\_ECHEANCE\_AMORT | 0.132133485 | | TMPS\_NB\_MOIS\_OUVERT\_ANC\_CRDT | 0.123272273 | | PASF\_NB\_PRDT\_MC\_ACTIF | 0.120280292 | | PASF\_MNT\_TOT\_LIMT\_CRDT\_ROTF | 0.120187546 | | BC\_HIGH\_HCCL\_AMT\_BY\_TRD | 0.117667971 | | BC\_NB\_DEROG\_PR | 0.11229163 | | BC\_AVG\_MOS\_IN\_FILE | 0.11042794 | | BC\_NB\_ACTIVE\_BC\_TRD | 0.105772177 | | BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT | 0.100819731 | | BC\_NB\_ACTIVE\_BR\_TRD | 0.09895045 | | BC\_NB\_BC\_TRD\_OPN | 0.09671553 | | TMPS\_MOY\_MOIS\_CRDT\_BLC | 0.096077313 | | BC\_NB\_ACTIVE\_RE\_TRD | 0.094058605 | | BC\_HIGH\_HCCL\_AMT\_BC\_TRD\_EVER | 0.092218934 | | BC\_HIGH\_UTL\_BAL\_BC\_TRD | 0.089469563 | | BC\_TOT\_BAL\_PF\_TRD | 0.083504606 | | BC\_NB\_ACTIVE\_BY\_TRD | 0.082968471 | | BC\_HIGH\_UTL\_BAL\_BR\_TRD\_2 | 0.082737437 | | BC\_MOS\_SNC\_RCNT\_BY\_TRD\_OPN | 0.08270792 | | BC\_NB\_INQ\_LAST6\_EX\_LAST7 | 0.082031816 | | BC\_HIGH\_UTL\_BAL\_BC\_TRD\_2 | 0.081836833 | | BC\_NB\_PF\_TRD | 0.081596063 | | AGE\_MOIS\_PRET | 0.081476978 | | BC\_TOT\_BAL\_TRD\_EXCL\_MT | 0.081057672 | | PASF\_MNT\_TOT\_SLD\_PSD | 0.08027992 | | TRNX\_NB\_CRDT\_6MOIS | 0.080012903 | | BC\_MOS\_SNC\_RCNT\_BC\_TRD\_ACT | 0.078224823 | | TRNX\_NB\_CRDT\_12MOIS | 0.078089117 | | BC\_MOS\_SNC\_OLD\_TRD\_OPN | 0.076067584 | | BC\_TOT\_BAL\_RE\_TRD | 0.075872672 | | BC\_NET\_FRACTN\_RT\_TRD\_BURDN | 0.073774325 | | BC\_TOT\_HCCL\_TRD | 0.072012975 | | BC\_NET\_FRACTN\_RR\_TRD\_BURDN | 0.071350999 | | BC\_NB\_TRD\_OPN | 0.065142877 | | BC\_TOT\_BAL\_TRD | 0.065038759 | | BC\_NB\_BR\_TRD | 0.063849504 | | BC\_TOT\_BAL\_BR\_TRD | 0.062775948 | | PASF\_MNT\_TOT\_RADIE | 0.057816608 | | BC\_NB\_RE\_TRD | 0.052866065 | | BC\_NB\_RR\_TRD\_90P\_EVER | 0.052404468 | | PASF\_NB\_DELQ\_TOT\_90PLUS\_LTD | 0.05121152 | | BC\_TOT\_MM\_COMMIT | 0.049404346 | | MNT\_SOLDE | 0.047549795 | | BC\_TOT\_MM\_PMNTS | 0.045310484 | | PASF\_NB\_PRDT\_CRDT\_ROTF\_ACTIF | 0.043534518 | | PASF\_MNT\_TOT\_SLD\_HYPT | 0.04351204 | | BC\_NB\_BC\_TRD | 0.043365028 | | BC\_MOS\_ON\_FILE | 0.043066933 | | BC\_NB\_RR\_TRD | 0.04270892 | | PASF\_IND\_COURANT\_ACTUEL | 0.042485031 | | PASF\_MNT\_TOT\_SLD\_CRDT\_VERS | 0.042433547 | | BC\_NB\_ACTIVE\_TRD | 0.041323457 | | BC\_NB\_ACTIVE\_PF\_TRD | 0.04018781 | | TRNX\_NB\_DEBT\_6MOIS | 0.038912745 | | BC\_NB\_TRD\_LAST12\_BAL0P | 0.038522891 | | PASF\_MNT\_TOT\_SLD\_PRDT | 0.037965994 | | PASF\_MNT\_TOT\_SLD\_VISA | 0.036577207 | | TRNX\_NB\_DEBT\_12MOIS | 0.036131566 | | BC\_NB\_SATS | 0.03595602 | | BC\_NB\_INQ\_LAST2 | 0.035598874 | | BC\_NB\_RR\_TRD\_OPN | 0.034821769 | | BC\_HIGH\_UTL\_BAL\_BR\_TRD | 0.033001157 | | PASF\_MNT\_TOT\_SLD\_MC | 0.03283156 | | BC\_NB\_RT\_TRD\_OPN | 0.031749368 | | BC\_NB\_PF\_TRD\_OPN\_LAST6 | 0.031684721 | | PASF\_NB\_PRDT\_OUV | 0.031167737 | | PASF\_NB\_PRDT\_ACTIF | 0.030499959 | | BC\_MOS\_SNC\_RCNT\_RR\_TRD\_OPN | 0.030011558 | | BC\_NB\_TRD\_OPN\_LAST24 | 0.02865555 | | BC\_NB\_RE\_TRD\_LAST12\_BAL0P | 0.028062161 | | BC\_NB\_RE\_TRD\_OPN\_LAST12 | 0.0275031 | | BC\_NB\_BR\_TRD\_OPN\_LAST12 | 0.026410564 | | BC\_MOS\_SNC\_RCNT\_BC\_TRD\_OPN | 0.02597868 | | BC\_MOS\_SNC\_RCNT\_RT\_TRD\_OPN | 0.025399129 | | BC\_NB\_BC\_TRD\_OPN\_LAST12 | 0.024543473 | | BC\_NB\_BC\_TRD\_OPN\_LAST3 | 0.024390399 | | BC\_TOT\_SEC\_CL | 0.024328623 | | BC\_TOT\_SEC\_BAL | 0.024320166 | | BC\_NB\_MT\_INQ\_LAST6 | 0.024320121 | | PASF\_NB\_PRDT\_TOTAL | 0.024192299 | | PASF\_NB\_PRDT\_VISA\_ACTIF | 0.020478263 | | BC\_NB\_IN\_TRD\_BAL0P | 0.019860593 | | BC\_NB\_ACTIVE\_IN\_TRD | 0.0198522 | | BC\_TOT\_MM\_PMNTS\_IN\_TRD | 0.019624467 | | ASSUR\_PRET | 0.017998316 | | BC\_NB\_ACTIVE\_FR\_TRD | 0.017870238 | | BC\_MOS\_SNC\_RCNT\_IN\_TRD\_OPN | 0.017830468 | | BC\_NB\_BC\_TRD\_LAST12\_BAL0P | 0.015574542 | | BC\_TOT\_BAL\_RT\_TRD | 0.014928402 | | BC\_TOT\_BAL\_IN\_TRD | 0.014748919 | | MNT\_AUTORISE | 0.013615027 | | BC\_NB\_ACTIVE\_BI\_TRD | 0.013562702 | | BC\_TOT\_BAL\_BI\_TRD | 0.013458466 | | BC\_TOT\_BAL\_BY\_TRD | 0.013226491 | | BC\_NB\_IN\_TRD | 0.012290957 | | BC\_NB\_TRD | 0.011874968 | | BC\_NB\_BR\_TRD\_LAST12\_BAL0P | 0.011852072 | | PASF\_NB\_PRDT\_PSD | 0.01170153 | | PASF\_IND\_COURANT\_3MOIS | 0.011690803 | | BC\_TOT\_BAL\_IN\_TRD\_ACTIVE | 0.011217673 | | BC\_NB\_FIN\_INQ\_LAST6 | 0.011079443 | | BC\_NB\_TRD\_OPN\_LAST12 | 0.010834621 | | PASF\_IND\_COURANT\_6MOIS | 0.010602361 | | PASF\_MNT\_TOT\_LIMT\_PSD | 0.010451925 | | PASF\_MNT\_TOT\_SLD\_CRDT\_ROTF | 0.009937097 | | B2B | 0.009795643 | | BC\_NB\_BY\_TRD\_BAL0P | 0.009677301 | | PASF\_IND\_COURANT\_12MOIS | 0.009173694 | | BC\_NB\_ACTIVE\_RR\_TRD | 0.008628271 | | BC\_TOT\_BAL\_OF\_TRD | 0.007436711 | | PASF\_IND\_COURANT\_24MOIS | 0.007323512 | | BC\_NB\_IN\_TRD\_OPN\_LAST6 | 0.007084486 | | BC\_NB\_ACTIVE\_RT\_TRD | 0.006427172 | | BC\_MOS\_SNC\_RCNT\_TRD\_OPN | 0.0062186 | | QC | 0.006094254 | | Region | 0.006089767 | | BC\_NB\_ACTIVE\_FI\_TRD | 0.005752146 | | TMPS\_NB\_MOIS\_OUVERT\_DERN\_CRDT | 0.005063056 | | PASF\_IND\_COURANT\_LTD | 0.004593058 | | BC\_NB\_OF\_TRD | 0.004552315 | | PASF\_NB\_PRDT\_HYPT\_ACTIF | 0.004500819 | | BC\_NB\_RR\_TRD\_OPN\_LAST6 | 0.003917525 | | ALT\_A | 0.002961833 | | PASF\_NB\_PRDT\_CRDT\_VERS\_ACTIF | 0.00291796 | | BC\_NB\_BI\_TRD\_OPN\_LAST12 | 0.002814699 | | BC\_NB\_ACTIVE\_AI\_TRD | 0.002651649 | | BC\_NB\_RE\_TRD\_OPN\_LAST6 | 2.47E-03 | | BC\_NB\_OF\_TRD\_OPN\_LAST6 | 0.002389365 | | BC\_NB\_TRD\_OPN\_LAST6 | 0.002290983 | | BC\_NB\_TRD\_OPN\_LAST3 | 0.002269408 | | BC\_NB\_ACTIVE\_OF\_TRD | 0.002152768 | | PASF\_MNT\_TOT\_SLD\_PRET | 0.002116001 | | PASF\_NB\_PRDT\_PRET\_ACTIF | 0.002101506 | | BC\_NB\_SKIP\_LOCATE\_NOT | 0.002054405 | | AGF | 2.10E-06 | |

**WOE of all variables**

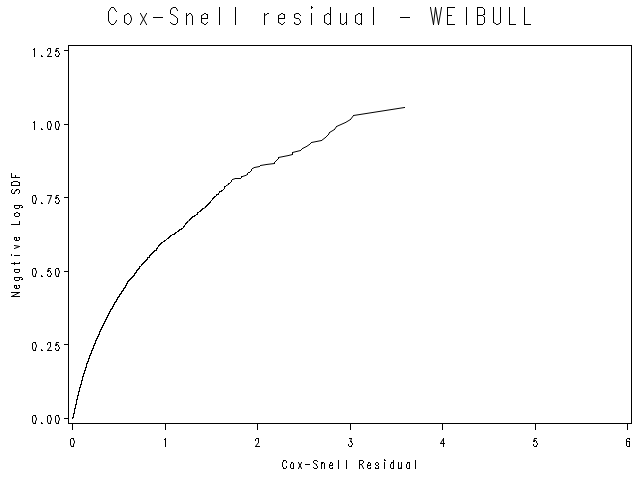


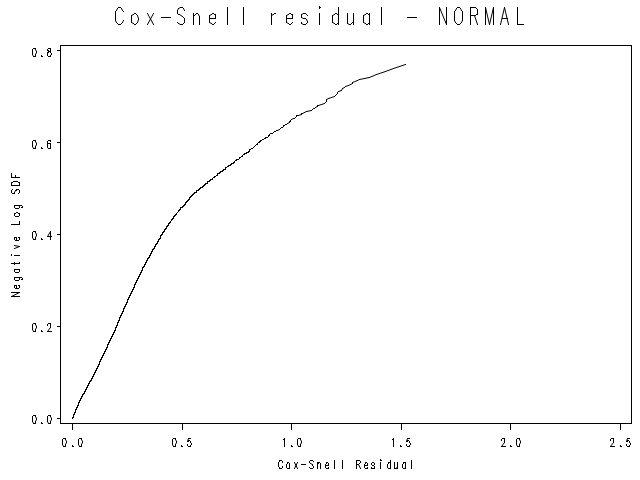
### Appendix 11: Multivariate analysis on candidate variables

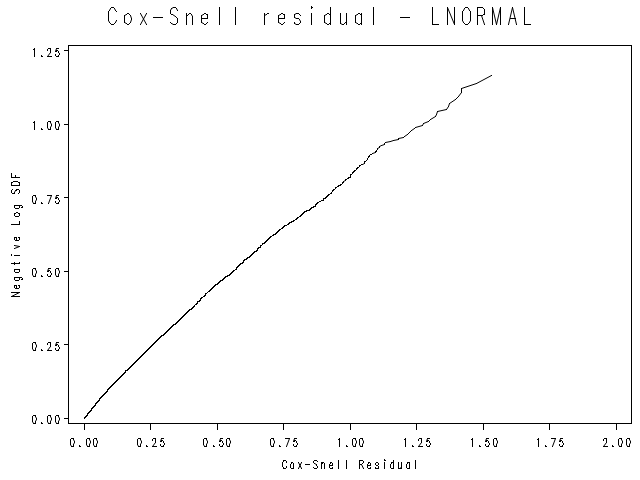
**Model Iteration**

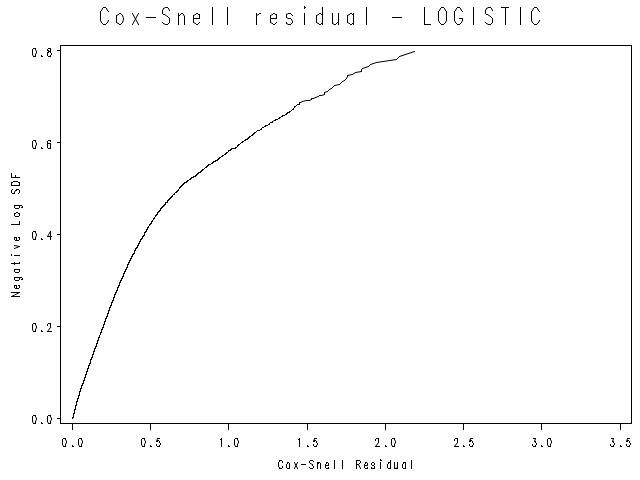
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **StdErr** | **LowerCL** | **UpperCL** | **ChiSq** | **ProbChiSq** | **iteration** | **KS** | **variable name** |
| var004 | 1 | 0.4198 | 0.0076 | 0.4049 | 0.4347 | 3050.92 | <.0001 | 1 | 0.588616509 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.6227 | 0.0063 | 0.6104 | 0.6349 | 9895.47 | <.0001 | 1 | 0.588616509 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.7561 | 0.0061 | 0.7441 | 0.768 | 15291.1 | <.0001 | 1 | 0.588616509 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var007 | 1 | 0.3122 | 0.0092 | 0.2942 | 0.3302 | 1160.67 | <.0001 | 2 | 0.598612054 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var004 | 1 | 0.407 | 0.0076 | 0.3921 | 0.4219 | 2860.06 | <.0001 | 2 | 0.598612054 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.5788 | 0.0063 | 0.5664 | 0.5912 | 8350.45 | <.0001 | 2 | 0.598612054 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.7249 | 0.0061 | 0.7129 | 0.7369 | 13934.5 | <.0001 | 2 | 0.598612054 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var007 | 1 | 0.3171 | 0.0092 | 0.2991 | 0.3351 | 1191.98 | <.0001 | 3 | 0.606284093 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var004 | 1 | 0.3989 | 0.0076 | 0.384 | 0.4138 | 2737.31 | <.0001 | 3 | 0.606284093 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var013 | 1 | 0.6463 | 0.0121 | 0.6227 | 0.67 | 2876.62 | <.0001 | 3 | 0.606284093 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var002 | 1 | 0.567 | 0.0064 | 0.5545 | 0.5795 | 7921.09 | <.0001 | 3 | 0.606284093 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.7277 | 0.0062 | 0.7156 | 0.7397 | 13992.1 | <.0001 | 3 | 0.606284093 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var007 | 1 | 0.1444 | 0.01 | 0.1249 | 0.1639 | 210.28 | <.0001 | 4 | 0.615017084 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var003 | 1 | 0.3096 | 0.0077 | 0.2946 | 0.3247 | 1626.69 | <.0001 | 4 | 0.615017084 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var004 | 1 | 0.3921 | 0.0076 | 0.3772 | 0.407 | 2658.69 | <.0001 | 4 | 0.615017084 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var013 | 1 | 0.6237 | 0.0121 | 0.6 | 0.6474 | 2663.93 | <.0001 | 4 | 0.615017084 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var002 | 1 | 0.5089 | 0.0065 | 0.4962 | 0.5215 | 6198.8 | <.0001 | 4 | 0.615017084 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.688 | 0.0062 | 0.6759 | 0.7 | 12456.3 | <.0001 | 4 | 0.615017084 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var007 | 1 | 0.1429 | 0.01 | 0.1233 | 0.1624 | 205.3 | <.0001 | 5 | 0.616413275 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var015 | 1 | 0.2346 | 0.0147 | 0.2059 | 0.2634 | 255.96 | <.0001 | 5 | 0.616413275 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var003 | 1 | 0.3101 | 0.0077 | 0.295 | 0.3252 | 1622.5 | <.0001 | 5 | 0.616413275 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var004 | 1 | 0.3903 | 0.0076 | 0.3753 | 0.4052 | 2625.22 | <.0001 | 5 | 0.616413275 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var013 | 1 | 0.637 | 0.0121 | 0.6132 | 0.6607 | 2756.18 | <.0001 | 5 | 0.616413275 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var002 | 1 | 0.5003 | 0.0065 | 0.4876 | 0.513 | 5937.28 | <.0001 | 5 | 0.616413275 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6811 | 0.0062 | 0.669 | 0.6932 | 12141.4 | <.0001 | 5 | 0.616413275 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var007 | 1 | 0.1157 | 0.01 | 0.0962 | 0.1352 | 135.12 | <.0001 | 6 | 0.623830018 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var015 | 1 | 0.2563 | 0.0147 | 0.2275 | 0.2851 | 304.29 | <.0001 | 6 | 0.623830018 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var003 | 1 | 0.2924 | 0.0077 | 0.2773 | 0.3074 | 1447.71 | <.0001 | 6 | 0.623830018 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var009 | 1 | 0.5106 | 0.0115 | 0.4881 | 0.5332 | 1964.8 | <.0001 | 6 | 0.623830018 | WOE\_BC\_NB\_INQ\_LAST12 |
| var013 | 1 | 0.5932 | 0.0121 | 0.5695 | 0.617 | 2394.07 | <.0001 | 6 | 0.623830018 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var004 | 1 | 0.3919 | 0.0076 | 0.377 | 0.4067 | 2671.05 | <.0001 | 6 | 0.623830018 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.4772 | 0.0065 | 0.4645 | 0.4899 | 5407.72 | <.0001 | 6 | 0.623830018 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6699 | 0.0061 | 0.6578 | 0.6819 | 11870.1 | <.0001 | 6 | 0.623830018 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var007 | 1 | 0.1111 | 0.01 | 0.0916 | 0.1306 | 124.46 | <.0001 | 7 | 0.623989945 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var012 | 1 | 0.1938 | 0.0127 | 0.1689 | 0.2187 | 232.62 | <.0001 | 7 | 0.623989945 | WOE\_BC\_TOT\_CLLC\_BAL |
| var015 | 1 | 0.2295 | 0.0148 | 0.2004 | 0.2585 | 240.37 | <.0001 | 7 | 0.623989945 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var003 | 1 | 0.2804 | 0.0077 | 0.2653 | 0.2955 | 1319.37 | <.0001 | 7 | 0.623989945 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var009 | 1 | 0.4938 | 0.0116 | 0.4712 | 0.5165 | 1820.45 | <.0001 | 7 | 0.623989945 | WOE\_BC\_NB\_INQ\_LAST12 |
| var013 | 1 | 0.6029 | 0.0121 | 0.5791 | 0.6267 | 2464.68 | <.0001 | 7 | 0.623989945 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var004 | 1 | 0.3924 | 0.0076 | 0.3775 | 0.4073 | 2671.91 | <.0001 | 7 | 0.623989945 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.4696 | 0.0065 | 0.4568 | 0.4823 | 5216.2 | <.0001 | 7 | 0.623989945 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6639 | 0.0062 | 0.6519 | 0.676 | 11625.8 | <.0001 | 7 | 0.623989945 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var007 | 1 | 0.112 | 0.0099 | 0.0926 | 0.1315 | 127.56 | <.0001 | 8 | 0.625374362 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var012 | 1 | 0.1836 | 0.0127 | 0.1588 | 0.2084 | 210.57 | <.0001 | 8 | 0.625374362 | WOE\_BC\_TOT\_CLLC\_BAL |
| var015 | 1 | 0.2227 | 0.0148 | 0.1937 | 0.2516 | 227.31 | <.0001 | 8 | 0.625374362 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var011 | 1 | 0.4451 | 0.014 | 0.4177 | 0.4726 | 1007.79 | <.0001 | 8 | 0.625374362 | WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC |
| var003 | 1 | 0.2822 | 0.0077 | 0.2671 | 0.2973 | 1341.54 | <.0001 | 8 | 0.625374362 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var013 | 1 | 0.4809 | 0.0127 | 0.456 | 0.5057 | 1438.71 | <.0001 | 8 | 0.625374362 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var009 | 1 | 0.4763 | 0.0115 | 0.4537 | 0.4989 | 1704.25 | <.0001 | 8 | 0.625374362 | WOE\_BC\_NB\_INQ\_LAST12 |
| var004 | 1 | 0.3932 | 0.0076 | 0.3784 | 0.4081 | 2703.31 | <.0001 | 8 | 0.625374362 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.4561 | 0.0065 | 0.4434 | 0.4689 | 4937.6 | <.0001 | 8 | 0.625374362 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6561 | 0.0061 | 0.6441 | 0.6681 | 11441.3 | <.0001 | 8 | 0.625374362 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var008 | 1 | 0.0647 | 0.0105 | 0.044 | 0.0853 | 37.7 | <.0001 | 9 | 0.625041867 | WOE\_BC\_NB\_IN\_TRD\_30P\_EVER |
| var007 | 1 | 0.1075 | 0.01 | 0.088 | 0.127 | 116.56 | <.0001 | 9 | 0.625041867 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var012 | 1 | 0.1763 | 0.0127 | 0.1514 | 0.2012 | 192.13 | <.0001 | 9 | 0.625041867 | WOE\_BC\_TOT\_CLLC\_BAL |
| var015 | 1 | 0.2128 | 0.0149 | 0.1837 | 0.242 | 205.18 | <.0001 | 9 | 0.625041867 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var011 | 1 | 0.4447 | 0.014 | 0.4172 | 0.4722 | 1005.77 | <.0001 | 9 | 0.625041867 | WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC |
| var003 | 1 | 0.273 | 0.0079 | 0.2576 | 0.2884 | 1207.99 | <.0001 | 9 | 0.625041867 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var013 | 1 | 0.4839 | 0.0127 | 0.459 | 0.5088 | 1453.51 | <.0001 | 9 | 0.625041867 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var009 | 1 | 0.4715 | 0.0116 | 0.4488 | 0.4942 | 1661.28 | <.0001 | 9 | 0.625041867 | WOE\_BC\_NB\_INQ\_LAST12 |
| var004 | 1 | 0.3937 | 0.0076 | 0.3789 | 0.4086 | 2705.04 | <.0001 | 9 | 0.625041867 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.453 | 0.0065 | 0.4402 | 0.4658 | 4840.59 | <.0001 | 9 | 0.625041867 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6532 | 0.0062 | 0.6412 | 0.6653 | 11272.6 | <.0001 | 9 | 0.625041867 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var008 | 1 | 0.0725 | 0.0105 | 0.0518 | 0.0932 | 47.26 | <.0001 | 10 | 0.626631102 | WOE\_BC\_NB\_IN\_TRD\_30P\_EVER |
| var015 | 1 | 0.1283 | 0.0157 | 0.0976 | 0.159 | 66.97 | <.0001 | 10 | 0.626631102 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var007 | 1 | 0.1116 | 0.01 | 0.0921 | 0.1312 | 125.45 | <.0001 | 10 | 0.626631102 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var012 | 1 | 0.1747 | 0.0127 | 0.1497 | 0.1996 | 188.35 | <.0001 | 10 | 0.626631102 | WOE\_BC\_TOT\_CLLC\_BAL |
| var006 | 1 | 0.1826 | 0.0109 | 0.1613 | 0.2039 | 281.9 | <.0001 | 10 | 0.626631102 | WOE\_BC\_TOT\_AMT\_AVAIL |
| var011 | 1 | 0.4407 | 0.0141 | 0.4131 | 0.4682 | 980.74 | <.0001 | 10 | 0.626631102 | WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC |
| var003 | 1 | 0.2713 | 0.0079 | 0.2558 | 0.2867 | 1187.04 | <.0001 | 10 | 0.626631102 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var013 | 1 | 0.4818 | 0.0127 | 0.4569 | 0.5068 | 1437.19 | <.0001 | 10 | 0.626631102 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var009 | 1 | 0.4795 | 0.0116 | 0.4567 | 0.5022 | 1705.69 | <.0001 | 10 | 0.626631102 | WOE\_BC\_NB\_INQ\_LAST12 |
| var004 | 1 | 0.3941 | 0.0076 | 0.3793 | 0.409 | 2705.72 | <.0001 | 10 | 0.626631102 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.4171 | 0.0068 | 0.4037 | 0.4304 | 3728.82 | <.0001 | 10 | 0.626631102 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6512 | 0.0062 | 0.6392 | 0.6633 | 11184.9 | <.0001 | 10 | 0.626631102 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var008 | 1 | 0.0763 | 0.0106 | 0.0556 | 0.097 | 52.22 | <.0001 | 11 | 0.627424519 | WOE\_BC\_NB\_IN\_TRD\_30P\_EVER |
| var015 | 1 | 0.1259 | 0.0157 | 0.0951 | 0.1567 | 64.05 | <.0001 | 11 | 0.627424519 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var007 | 1 | 0.1017 | 0.01 | 0.0821 | 0.1213 | 103.72 | <.0001 | 11 | 0.627424519 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var006 | 1 | 0.141 | 0.0111 | 0.1192 | 0.1629 | 160.24 | <.0001 | 11 | 0.627424519 | WOE\_BC\_TOT\_AMT\_AVAIL |
| var012 | 1 | 0.1733 | 0.0127 | 0.1484 | 0.1983 | 185.05 | <.0001 | 11 | 0.627424519 | WOE\_BC\_TOT\_CLLC\_BAL |
| var010 | 1 | 0.2345 | 0.0132 | 0.2087 | 0.2603 | 316.97 | <.0001 | 11 | 0.627424519 | WOE\_BC\_AVG\_MOS\_BC\_TRD |
| var011 | 1 | 0.3896 | 0.0143 | 0.3615 | 0.4178 | 738.01 | <.0001 | 11 | 0.627424519 | WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC |
| var013 | 1 | 0.4645 | 0.0128 | 0.4395 | 0.4895 | 1326.56 | <.0001 | 11 | 0.627424519 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var003 | 1 | 0.29 | 0.0079 | 0.2744 | 0.3056 | 1333.76 | <.0001 | 11 | 0.627424519 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var009 | 1 | 0.4492 | 0.0117 | 0.4262 | 0.4722 | 1465.66 | <.0001 | 11 | 0.627424519 | WOE\_BC\_NB\_INQ\_LAST12 |
| var004 | 1 | 0.3927 | 0.0076 | 0.3779 | 0.4076 | 2685.19 | <.0001 | 11 | 0.627424519 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.4101 | 0.0068 | 0.3967 | 0.4235 | 3589.38 | <.0001 | 11 | 0.627424519 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6508 | 0.0062 | 0.6387 | 0.6629 | 11157.6 | <.0001 | 11 | 0.627424519 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var005 | 1 | -0.062 | 0.0101 | -0.0818 | -0.0422 | 37.53 | <.0001 | 12 | 0.627269907 | WOE\_BC\_PCT\_TRD\_NEVER\_DELQ |
| var008 | 1 | 0.09 | 0.0108 | 0.0689 | 0.1112 | 69.5 | <.0001 | 12 | 0.627269907 | WOE\_BC\_NB\_IN\_TRD\_30P\_EVER |
| var015 | 1 | 0.1315 | 0.0157 | 0.1007 | 0.1624 | 69.75 | <.0001 | 12 | 0.627269907 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var007 | 1 | 0.0977 | 0.01 | 0.0781 | 0.1173 | 95.46 | <.0001 | 12 | 0.627269907 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var006 | 1 | 0.1492 | 0.0112 | 0.1272 | 0.1712 | 176.84 | <.0001 | 12 | 0.627269907 | WOE\_BC\_TOT\_AMT\_AVAIL |
| var012 | 1 | 0.1796 | 0.0128 | 0.1545 | 0.2046 | 197.37 | <.0001 | 12 | 0.627269907 | WOE\_BC\_TOT\_CLLC\_BAL |
| var010 | 1 | 0.2473 | 0.0133 | 0.2211 | 0.2734 | 343.82 | <.0001 | 12 | 0.627269907 | WOE\_BC\_AVG\_MOS\_BC\_TRD |
| var011 | 1 | 0.391 | 0.0143 | 0.3629 | 0.4191 | 743.2 | <.0001 | 12 | 0.627269907 | WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC |
| var003 | 1 | 0.3254 | 0.0098 | 0.3062 | 0.3447 | 1099.04 | <.0001 | 12 | 0.627269907 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var013 | 1 | 0.4592 | 0.0128 | 0.4342 | 0.4843 | 1291.93 | <.0001 | 12 | 0.627269907 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var009 | 1 | 0.4451 | 0.0117 | 0.4221 | 0.4682 | 1436.58 | <.0001 | 12 | 0.627269907 | WOE\_BC\_NB\_INQ\_LAST12 |
| var004 | 1 | 0.3937 | 0.0076 | 0.3788 | 0.4085 | 2697.73 | <.0001 | 12 | 0.627269907 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.415 | 0.0069 | 0.4015 | 0.4286 | 3625.53 | <.0001 | 12 | 0.627269907 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6523 | 0.0062 | 0.6402 | 0.6644 | 11191.6 | <.0001 | 12 | 0.627269907 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |
| var005 | 1 | -0.0697 | 0.0101 | -0.0896 | -0.0498 | 47.23 | <.0001 | 13 | 0.627431227 | WOE\_BC\_PCT\_TRD\_NEVER\_DELQ |
| var015 | 1 | 0.1207 | 0.0158 | 0.0898 | 0.1516 | 58.69 | <.0001 | 13 | 0.627431227 | WOE\_BC\_MOS\_SNC\_RCNT\_RE\_TRD\_ACT |
| var008 | 1 | 0.0864 | 0.0108 | 0.0652 | 0.1075 | 63.98 | <.0001 | 13 | 0.627431227 | WOE\_BC\_NB\_IN\_TRD\_30P\_EVER |
| var007 | 1 | 0.0991 | 0.01 | 0.0796 | 0.1187 | 98.63 | <.0001 | 13 | 0.627431227 | WOE\_BC\_TOT\_AMT\_PAST\_DUE |
| var014 | 1 | 0.1503 | 0.0149 | 0.1211 | 0.1796 | 101.5 | <.0001 | 13 | 0.627431227 | WOE\_PASF\_NB\_PRDT\_MC\_ACTIF |
| var006 | 1 | 0.1464 | 0.0112 | 0.1244 | 0.1684 | 170.46 | <.0001 | 13 | 0.627431227 | WOE\_BC\_TOT\_AMT\_AVAIL |
| var012 | 1 | 0.1752 | 0.0128 | 0.1502 | 0.2002 | 188.32 | <.0001 | 13 | 0.627431227 | WOE\_BC\_TOT\_CLLC\_BAL |
| var010 | 1 | 0.2446 | 0.0133 | 0.2185 | 0.2708 | 337.14 | <.0001 | 13 | 0.627431227 | WOE\_BC\_AVG\_MOS\_BC\_TRD |
| var011 | 1 | 0.3586 | 0.0147 | 0.3299 | 0.3873 | 598.67 | <.0001 | 13 | 0.627431227 | WOE\_TMPS\_NB\_MOIS\_OUVERT\_CLNT\_BLC |
| var003 | 1 | 0.3301 | 0.0098 | 0.3108 | 0.3493 | 1130.12 | <.0001 | 13 | 0.627431227 | WOE\_BC\_MOS\_SNC\_RCNT\_DELQ |
| var013 | 1 | 0.4554 | 0.0128 | 0.4303 | 0.4804 | 1270.58 | <.0001 | 13 | 0.627431227 | WOE\_NB\_MOIS\_ECHEANCE\_AMORT |
| var009 | 1 | 0.4431 | 0.0117 | 0.4201 | 0.4661 | 1425.41 | <.0001 | 13 | 0.627431227 | WOE\_BC\_NB\_INQ\_LAST12 |
| var004 | 1 | 0.3963 | 0.0076 | 0.3815 | 0.4112 | 2735.44 | <.0001 | 13 | 0.627431227 | WOE\_PASF\_MNT\_TOT\_DELQ\_ACTUEL |
| var002 | 1 | 0.4106 | 0.0069 | 0.3971 | 0.4241 | 3540.39 | <.0001 | 13 | 0.627431227 | WOE\_BC\_NET\_FRACTN\_BC\_TRD\_BURDN |
| var001 | 1 | 0.6476 | 0.0062 | 0.6355 | 0.6597 | 11011.2 | <.0001 | 13 | 0.627431227 | WOE\_NB\_JR\_DELQ\_MAX\_6MOIS |

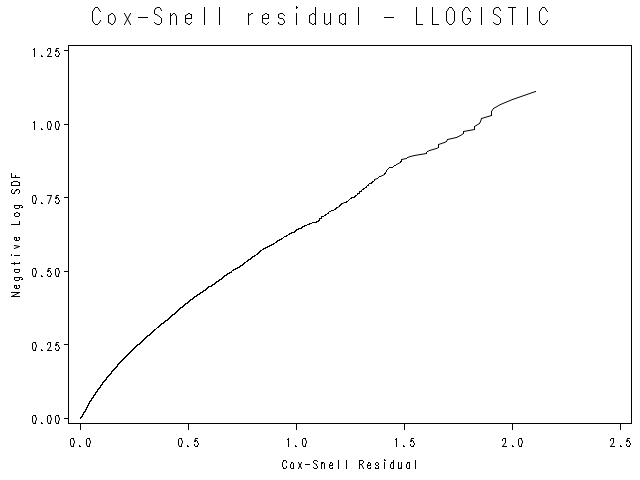
**Residual Analysis**

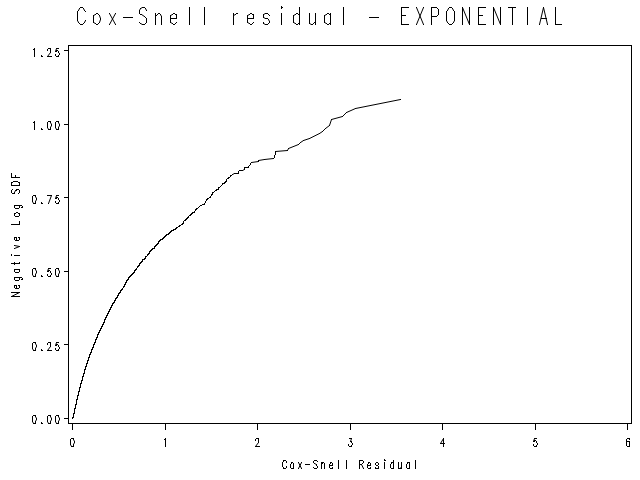




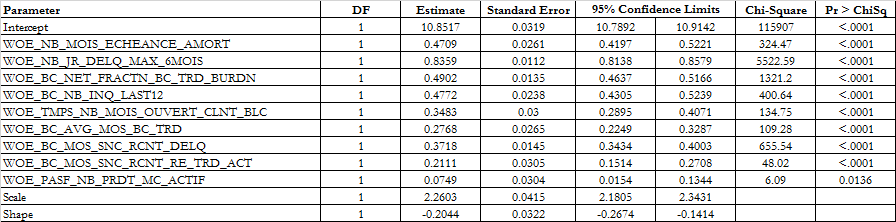




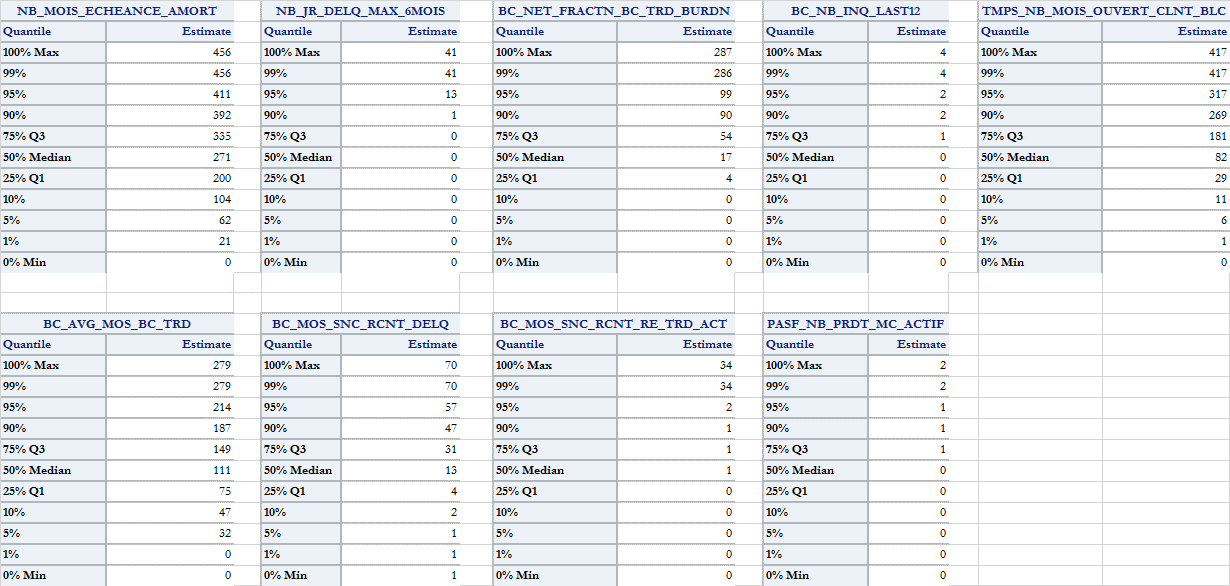




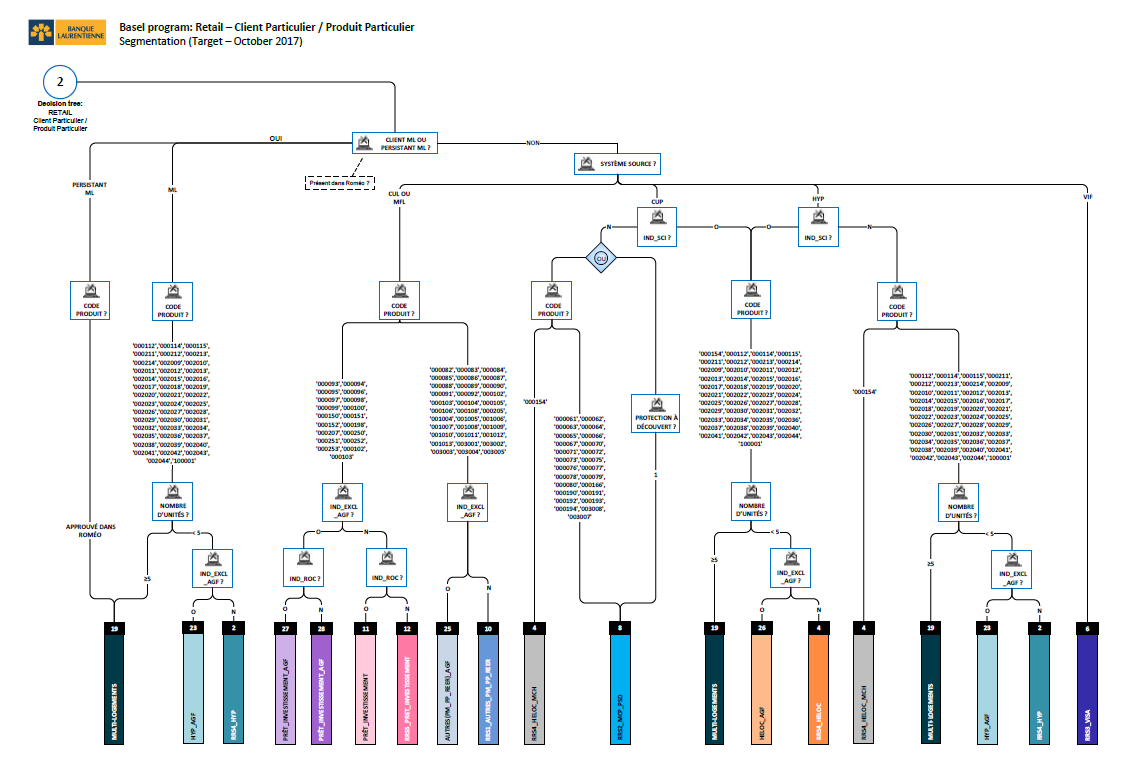
**Preliminary model - population Test (Validation Test)**



### Appendix 12: Distribution of explanatory variables



### Appendix 13: Decision tree (segmentation)



### Appendix 14: Performance testing approach chart

See section 8.4 of the documentation.

### Appendix 15: Discussion with the business lines

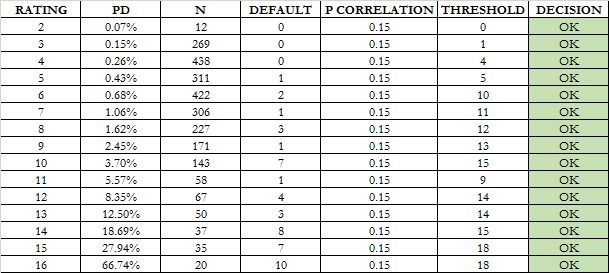
There is no discussion with the business lines in the PD mortgage development.

### Appendix 16: List of products in the modeling database

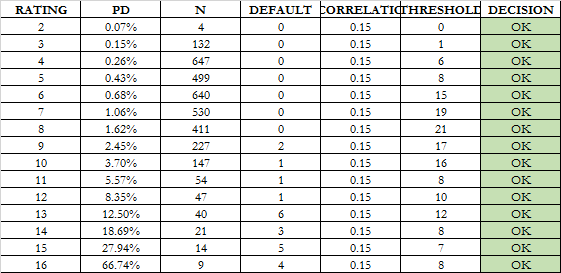
|  |  |  |
| --- | --- | --- |
| **ID\_SYS\_PROD** | **CD\_PRODUIT** | **NOM\_PROD\_AN** |
| HYP | 000112 | RESIDENTIAL MORTGAGE CAD (1 TO 4 APP.) |
| HYP | 000114 | RESIDENTIAL MORTGAGE |
| HYP | 000211 | PERSONALIZED MORTGAGE LOAN |
| HYP | 000212 | PERSONALIZED MORTGAGE LOAN |
| HYP | 000213 | PERSONALIZED MORTGAGE LOAN |
| HYP | 000214 | PERSONALIZED MORTGAGE LOAN |

### Appendix 17: Backtesting result for AGF

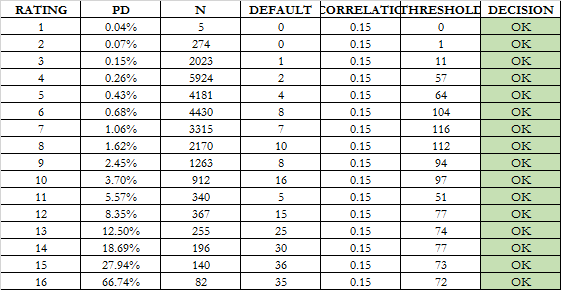
**Binomial test with correlation**



### Appendix 18: Backtesting result for ALT\_A



### Appendix 19: Backtesting result for B2B

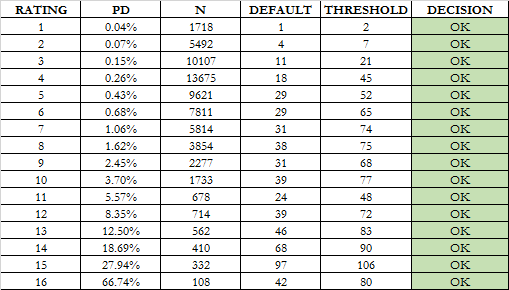


### Appendix 20: Backtesting result without conservatism

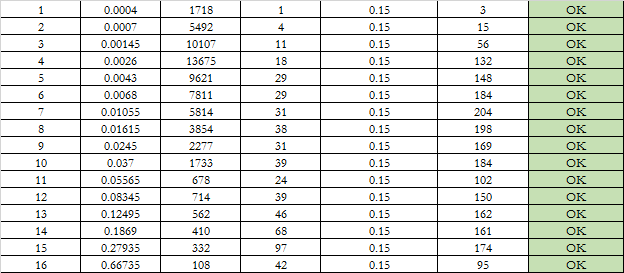
**AUROC**



**Binomial test without correlation**



**Vasicek**



**Blochinger-Leippold**



1. **Federal Reserve Bank**, *Risk-Based Capital Standards: Advanced Capital Adequacy Framework*. 2006 [↑](#footnote-ref-1)
2. Working Paper 14 “*Studies on the Validation of Internal Rating Systems*” is published by BCBS in May 2005. [↑](#footnote-ref-2)
3. Basel Committee on Banking Supervision, 2001 [↑](#footnote-ref-3)
4. **Sources:**

   ***Statistique Canada.*Table   379-0031** -  Produit intérieur brut (PIB) aux prix de base, selon le Système de classification des industries de l'Amérique du Nord (SCIAN), mensuel (dollars x 1 000 000).

   ***Statistique Canada.*Table   282-0087** -  Enquête sur la population active (EPA), estimations selon le sexe et le groupe d'âge, désaisonnalisées, mensuel. [↑](#footnote-ref-4)
5. This method diminishes the risk of having too much correlation between the covariates. [↑](#footnote-ref-5)
6. Some transformations were tested when a certain variable was close to be statistically significant. [↑](#footnote-ref-6)
7. An additional criterion was used that pertains to certain variables only. In the phase of building characteristics, multiple “versions” of certain characteristics were built. In such a case, we would not allow both variables to enter the model because they are intrinsically the same variable, but with different construction-methodologies. [↑](#footnote-ref-7)
8. For more detail, please see Allison, Paul D. 1995. *Survival Analysis Using SAS: A Practical Guide*. Cary, NC: SAS Institute Inc. [↑](#footnote-ref-8)