|  |
| --- |
|  |
| **TFRS9 Model development document** |
| Credit Card Porfolio |
|  |
|  |
|  |

|  |
| --- |
|  |

Document History

This section documented the revision history and version control of this document. It shall record every major and minor revision of the model development regarding the development of Probability of Default (PD), Exposure at Default (EAD), and Loss Given Default (LGD) models modules which are used for the purpose of calculation of ECL of the TFRS9 accounting book.

Revision History

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Version** | **Phase** | **Revision Date** | **Summary of Changes** | **Page** | **File name** | **Changed by** |
| 1.0 | 3.2 | 14-Jun-19 | Initial Version |  | KBank\_TFRS9\_Model\_Development\_Document\_CC\_V1.0\_190614 | Pattadol B. |

Reviewer

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Version** | **Signature** | **Review Date** |
| Thepdanai Danswasvong | 1.0 |  |  |
| Panrit Tosukhowong | 1.0 |  |  |
| Nareerut Poolpun | 1.0 |  |  |

Approvals

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Version** | **Signature** | **Approve Date** |
| Duangporn Kit-o-pas | 1.0 |  |  |
| Sanphet Sukhapesna | 1.0 |  |  |

# Introduction

Probability of default (PD), Exposure at default (EAD), and Loss Given Default (LGD) are three components which are used to calculate credit risk capital and provision. In this document, we are focusing on Credit Card (CC) which has a portfolio size about 58,339,127,785 Baht.

In the current regulatory setting, the provisioning of expected credit loss is calculated from inherent risk parameters; the probability of default (PD), exposure at default (EAD) and loss given default (LGD). Each parameter is estimated by average through the economic cycle from each specific range of time which can be considered as a downturn situation. However, these risk parameters do not reflect future risk exposure. Hence, the TFRS9 which is a new accounting standard is introduced. The impairment under TFRS9 setting also covers forward-looking components which help to improve financial stability and improve bank credit risk.

This document outlines the development process of all model related to the TFRS9 calculation. For each risk component, this document shall clearly state the scope (model usage), methodology considered, model development approach, final model and the initial validation results.

# TFRS9 Expected Credit Loss

On July 24, 2014, the International Accounting Standards Board (IASB) issued the final version of the ‘International Financial Reporting Standard (IFRS) 9 – Financial Instruments’. As a primary component of the new accounting standard, the IASB introduced a forward looking impairment model. The IASB thereby reacted to delayed recognition of credit losses identified as a weakness of existing accounting standards during the course of the global financial crisis (of 2007/08). In particular, the biggest critique of incurred loss approach under IAS 39 was the recognition of credit losses only upon evidence of a trigger event. In this regard, IASB’s approach of forward looking credit loss estimation was evident from the below extract.

*“The new standard requires an entity to recognise expected credit losses at all times and to update the amount of expected credit losses recognised at each reporting date to reflect changes in the credit risk of financial instruments. This model is forward-looking and it eliminates the threshold for the recognition of expected credit losses, so that it is no longer necessary for a trigger event to have occurred before credit losses are recognised. Consequently, more timely information is required to be provided about expected credit losses.” [“Project summary – IFRS 9 Financial Instruments”, IFRS Foundation, 07/2014, p.14][[1]](#footnote-1)*

Following the publication of IFRS 9 Financial Instruments in July 2014, the Basel Committee on Banking Supervision issued their ‘Guidance on Credit Risk and Accounting for Expected Credit Losses’ (GCRAECL) in December 2015. This covers in particular the impairment (Expected Credit Losses) element and how it should be embedded in and supported by internal processes.

Thai Accounting Standards are substantially converged with IFRS Standards, though the financial instruments Standards that are part of IFRS Standards have not yet been adopted. Thai Accounting Standards include several national financial instruments standards that differ from IFRS Standards. Henceforth TFRS9 can be considered as an adaptation from IFRS9 and replaces the existing TAS101.

***Principle 5 states****–A bank should have policies and procedures in place to appropriately validate models used to assess and measure expected credit losses. This presentation will provide an overview of the scope of work and the proposed validation approach for KBank, based on further discussions we will provide a more detailed view of the approach based on the complexity and materiality of the underlying models.*

Changes due to ‘*IFRS 9 – Financial Instruments*’ can be grouped into three categories.

* ***Classification and measurement***: Classification determines how financial assets and liabilities are accounted for in financial statements and, in particular, how they are measured on an ongoing basis:
  + Assets: one classification approach
  + Liabilities: addressing the volatility in profit or loss caused by changes in the credit risk of financial liabilities that are measured at fair value
* ***Impairments***: Forward-looking impairment model based on expected losses:
  + The new model requires entities to recognise expected credit losses at all times (12-month or lifetime expected loss) which includes measurement of changes in expected credit losses
  + It is no longer necessary for a trigger event to have occurred before credit losses are recognised
  + The new model is also accompanied by improved disclosures about expected credit losses and credit risk
* ***Hedge accounting***: Clear alignment with risk management:
  + The rules allow components of non-financial items to be hedged (previously not allowed by IAS 39)
  + IFRS 9 eliminates the distinction between financial and non-financial items and looks at whether a risk component can be identified and measured and therefore reflected in management activities

The primary change from IAS 39 to IFRS 9 is the evolution from an incurred loss view to a forward looking expected loss view which needs to be accounted for in the impairment models.

This new accounting standard will be effective from 2020. The IFRS 9 standard provides a new set of regulations that the new loss provisioning process will need to satisfy.



Figure 1, From IAS39 to IFRS 9

In particular, the new impairment rules require that the lifetime credit risk of an account be assessed at each model run to determine if there has been a significant increase in credit risk since origination. For accounts where the credit risk has significantly increased (including defaults) the lifetime expected credit losses must be used. If the credit risk has not significantly increased, then only credit losses resulting from expected defaults in the next 12 months must be used. The approach is outlined in terms of a stage classification accounting for significant increase in credit risk as a pivotal element of IFRS 9.

**Table 1: Stages under IFRS 9**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Stage 1** | **Stage 2** | **Stage 3** |
| **Stage description** | Includes accounts for whom no significant increase in credit risk since initial recognition has been observed | Includes accounts whose credit risk has significantly increased since initial recognition but no objective evidence of impairment has been observed, with a rebuttal presumption that this occurs when the account reaches 30 days past due | Includes accounts where the objective evidence of impairment has been observed |
| 12-month expected credit losses, i.e. credit losses due to default events within subsequent 12 months, are recognised in balance sheet | Lifetime expected credit losses (LTECL) i.e. credit losses due to default events spanning the (expected) lifetime of the facility, are recognised in balance sheet | Lifetime expected credit losses (LTECL) are recognised in balance sheet |
| **Expected loss** | IFRS 9 guidelines require to assess the 1 year expected credit losses without prescribing the tangible estimation procedure    , , , represent marginal PD, EAD and LGD at time *t* | IFRS 9 guidelines require the lifetime expected credit losses without prescribing the tangible estimation procedure   Where, , , , represent marginal PD, EAD and LGD at time *t* and *T* represents the remaining lifetime of the account | Expected loss of a defaulted client given the loss rate, i.e. the shortfall in net present value of expected cash flows versus the carrying amount of the loan |

Table 1 Stages under IFRS 9

The exemplary expected loss (EL) assessment in Table 1 listed above is based on an estimate of 1 year parameters probability of default (PD), loss given default (LGD) and exposure at default (EAD) for stage 1 and multi-year PD, LGD and EAD assessments for stage 2 including a discount factor to the reporting date.

# Definition

| **Term** | **Acronym** | **Description** |
| --- | --- | --- |
| Days Past Due | DPD | The number of days that an account is currently in arrears |
| Delinquent |  | An asset is described as delinquent if it is associated with any amount of arrears |
| Expected Loss/Expected Credit Loss | EL/ECL | Interchangeable terms. EL = PD\*EAD\*LGD |
| Exposure at Default | EAD | Exposure at Default (EAD) is defined as the expected amount drawn by borrowers at the time of default. |
| Probability of Default | PD | Probability of default (PD) is the risk that the borrower will be unable or unwilling to repay its debt in full or on time. The risk of default is derived by analyzing the obligor’s capacity to repay the debt in accordance with contractual terms. PD is generally associated with financial characteristics such as inadequate cash flow to service debt, declining revenues or operating margins, high leverage or declining liquidity |
| Default customer |  | Default customer is customer who failed to make on-time repayment (>= minimum payment rate) of their loans for more than ninety consecutive days or three months. |
| Credit Conversion Factor | CCF | Credit Conversion Factor (CCF) is defined as percentage rate of current undrawn credit line that will likely be utilized by borrowers at the time of default |
| Lifetime Expected Credit Loss | LTECL / LEL | The Expected Credit Loss over the behavioural lifetime of an asset |
| Beta distribution |  | The beta distribution is a family of continuous [probability distributions](https://en.wikipedia.org/wiki/Probability_distribution) defined on the interval [0, 1] [parametrized](https://en.wikipedia.org/wiki/Statistical_parameter) by two positive [shape parameters](https://en.wikipedia.org/wiki/Shape_parameter), denoted by *α* and *β*, |

Table 2, Term definition and description

# Probability of Default

Probability of default (PD) is the risk that the borrower will be unable or unwilling to repay its debt in full or on time. The risk of default is derived by analyzing the obligor’s capacity to repay the debt in accordance with contractual terms. PD is generally associated with financial characteristics such as inadequate cash flow to service debt, declining revenues or operating margins, high leverage or declining liquidity.

With the new IFRS9 loan loss provision, loans are classified in three stages: stage 1 – initial recognition (yet to be impaired), stage 2 – significant increase in credit risk, and stage 3 – objective indicators of impairment. For loans in stage 1, banks need to estimate 1-year expected credit losses. On the other hand for loans in stage 2 and 3, banks need to provide provision and thus estimate expected credit losses for the whole lifetime of the loans. Both the 1-year and lifetime expected credit losses estimation shall reflect the banks’ forward looking macro-economic view.

KBank’s PD estimation for IFRS9 is modelled according to the following principles: (i) the PD estimation for IFRS9 should be point-in-time (PIT) and reflect current market conditions, (ii) the PD estimates should use structural and behavior information, and (iii) estimation of PD should include the forward looking aspect of the macroeconomic outlook specific to particular sector. For loans in stage 1 and 2, the modelled probability of default will be over 12 months and lifetime respectively. For stage 3, the probability default will be at 100%.

KBank leverages existing behavioral scorecard to construct a new credit rating system (supermaster rating) and then use it to create the probability of default term-structure of each obligor up to its maturity based on a continuous time homogeneous or non-homogeneous Markov transition matrix and then incorporate systematic risk into PD term-structure model via Vasicek model for calculating joint loss distribution of bank exposures (Vasicek, 2002).

This section outlines step by step Probability of Default (PD) model development methodology for the champion model and states its compliance with the current IFRS9 regulatory requirement. It also documented all the technical difficulties that the model developer experienced undergoing these model development steps.

## Scope

This probability of default model should be used to create the PD term structure for all valid stage-1 and stage-2 credit card instruments. For stage-3 instruments, they are automatically assigned to be at 100%.

## Modelling Methodology & Design

KBank leverages existing behavioral scorecard to construct a new credit rating system (supermaster rating) and then use it to create the probability of default term-structure of each obligor up to its maturity based on a continuous time homogeneous or non-homogeneous Markov transition matrix and then incorporate systematic risk into PD term-structure model via Vasicek model for calculating joint loss distribution of bank exposures (Vasicek, 2002).

The existing behavioral scorecard is regarded by the bank as a feeder model to the TFRS9 probability of default model. Hence, this document does not extend to the model development and validation of the current behavioral scorecard.

For KBank retail portfolio, we explore a way to construct PD term structure by applying Markov state transition model. The Markov property represents the assumption that the evolution of credit migration is independent of the past credit migration history. In other words, the probability of migration from rating class to rating class does not depend on any information from the past, only the current rating matters. To find out the best credit migration for our corporate lending, we consider several alternative approaches: (i) discrete time Markov transition, (ii) homogeneous continuous time Markov transition, and (iii) nonhomogeneous continuous time Markov transition.

### Methodology Review

#### **TTC PD: Discrete Time Markov Transition (DTMT)**

We first consider a discrete time Markov chain. Let be a discrete-time transition matrix

where denotes the probability of change in rating at the beginning of a year to rating class at year’s end. In total, we have ratings (states). The final rating represents a default state.

A useful property of the time homogeneous Markov transition matrix is the fact that the *n*-year transition matrix is simply given by the *n*th power of the one-year transition matrix , denoted by i.e.

[Matrix multiplication times]

The cumulative nth year default probabilities for a starting rating class is the last column of ,

The model assumes that probability of state (rating) transition only depends on the current information (current rating) i.e. rating transition exhibit Markov property. In reality, the rating migration needs not to be stationary or time-homogeneous and the maturity of loans need not be a yearly increment. We explore further to identify a more suitable approach to model the probability of default term structure.

#### **TTC PD: Homogeneous Continuous Time Markov Transition**

A continuous time, time-homogeneous Markov chain is predicated in terms of a symmetric generator matrix (for possible transitory states)

where for all , and for

The transition matrix for time interval is given by

where denotes the matrix exponential

However, there is evidence that the assumption of time-homogeneity is very likely not precisely true for real credit ratings migration. We therefore consider nonhomogeneous time Markov chain and would like to find that best represent the dynamics of PD migration. That is, we no longer assume that the transition rates  are constant over time. This will be done using parametric approach.

#### **TTC PD: Non Homogeneous Continuous Time Markov Transition**

We replace the homogeneous generator for the time interval by the time-dependent-generator:

where X denote matrix multiplication and is the diagonal matrix in with elements of as follow

The function is defined as follow

where ,are non-negative constant and time .

Finally, we decide to apply nonhomogeneous continuous time Markov transition to forecast the dynamics of PD migration for our lending portfolio, because it better reflects seasoning effects and improves the fit, especially on the lowest and highest ratings.

#### **PIT PD: Vasicek (2002) and Credit Adjustment Factor**

Previously, we derive TTC PD term structure or accumulative PD for, say, n years. The forward PD is then calculated and defined as the probability of the loan has defaulted by the end of year n, minus the probability of the loan had already defaulted before the beginning of year n. That is, the forward PD is the difference between the n-year and the n-1 year accumulative PD.

Our approach to convert forward PD to PIT PD is through the application of the Vasicek (2002) equation using a Cycle Adjustment Factor (CAF). The PIT PD for year t is given by:

Where is the cumulative standard normal distribution function

is the cumulative inverse standard normal distribution function

is the asset-segment correlation

or Cycle Adjustment Factor is the normalized risk factor that represents the point in the economic cycle

is the forward PD for an individual obligor/rating

is the PIT PD for an individual obligor/rating

We next find the fitted CAF model by seeking relationship between credit index (Z) and macroeconomic factors. That is, we regress the normalized historical credit index against the historical macroeconomic factors. The macroeconomic factors are assumed to be exogenous to the model and are listed in Table 3 together with their expected signs.

|  |  |  |
| --- | --- | --- |
| No. | Macroeconomic Factors | Expected Sign |
| 1 | Private Consumption | - |
| 2 | Investment | - |
| 3 | House Price Index | - |
| 4 | Unemployment Rate | + |
| 5 | Diesel Price | - |
| 6 | Household Debt | + |

Table 3, List of macroeconomic factors and their expected signs

*Grouping criteria and asset correlation*

Credit Adjustment Factors (CAF) in different homogeneous asset classes are naturally different based upon their economic drivers. Since commercial portfolio is based on systematic factors in industry, we develop CAF by Occupation: Salary Earner and Entrepreneur.

To develop the econometric model, a direct regression of the observed actual default against the macroeconomic factors is performed. The estimated regression equation is

Where measure the “credit cycle” at time t and is the OLS estimates which minimize the sum of squared error. In addition, the macroeconomic factors that are selected into the final fitted CAF model should fulfill the following conditions:

1) Its coefficients should be significant (p-value < 0.05)

2) Its coefficients should be consistent with economic theory as well as improve the overall model’s performance (e.g. adjusted R2).

Finally, given the forecasted macroeconomic factors, we can calculate CAF for the forecasted period. In conjunction with forward PD estimates, we then calculate forward-looking PIT PD using the above-mentioned Vasicek equation where is the best fit between actual and predicted model.

## Model Development Approach

For PD term structure model, KBank believe that the Markov state transition model family is the appropriate methodology for the construction of PD term structure. The inputs of the model and the development exercise align with the bank’s current data structure. The output of model also aligns with current business model and TFRS9 risk requirements in term of generating the default term structure.

KBank explore the possibility of instigating the non-homogeneous continuous time Markov state transition model (NHCTMT) as a champion model for Credit card PD term structure. But depending on the results of the curve adjustment optimization, KBank may opt to revert to the homogeneous model instead. Additionally, KBank also attempt to use the survival analysis methodology to create the default term structure as a challenger model.

Figure 2, Steps of PD term structure model development

This section briefly outlines the steps that have been taken in the development of the champion model (NHCTMT) for the TFRS9 probability of default term structure.

The model development process starts with data gathering and sample preparation. We used a long history of default performance, B-score and TDR-score in the derivation of our supermaster rating and 1-year through-the-cycle probability of default. This will be elucidate in detail in the next section of this document.

After the supermaster rating and scale has been well developed, we re-apply the rating criteria to our historical observations and observed the 1-year rating migrations over different observation periods. We constructed our through-the-cycle 1-year state transition probability based on the average migration rates across years of observation. The next step involves human intervention in order to smooth out and ensure that the TTC rating migration matrix comply with the necessary conditions of the NHCTMT model (will be discussed later).

After the TTC migration matrix has been finalized, we transform it to the generator matrix. We use this generator matrix together with an optimization algorithm to find appropriate curve adjustment parameters and . The objective of the optimization algorithm is to find the curve adjustment parameters such that the PD curves are the closest to and still cover the historical default rates across all ratings.

## Data Management

### Overview of Input Data Set

The first step of our model development is the derivation of supermaster scale. We gathered Credit card portfolio data and B-score of each instrument from 2012-01 to 2016-05.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Field\_Name | Existing/New | Description |
| 1. | POS\_DT | Existing | Data as of observation point |
| 2. | LPM\_NO | Existing | Customer ID |
| 3. | ASSET\_CLASS | Existing | Asset class of customer is to identify Good/Bad customer. If the asset class contains “NPL” then it is defined as bad if not then it is defined as good. |
| 4. | B\_score | Existing | Customer Level B-Score (minimum from account level) |
| 5. | Supermaster\_rating | New | Supermaster rating derived from 2.1 |

Table 4, Input data

From the data we observe portfolio actual default and construct a new super master rating. Subsequent to the completion of the supermaster rating, we observe a rating migration of the same observation and performance period. This process is outlined in the figure below.

Figure 3, Observation and outcome of default event and rating transition

We also utilize occupation data to explore the possibility of segmentation of the migration matrices.

### Development Sample Design

The observation point is every quarter starting from January 2012 to May 2016. The performance period for observation of actual default or rating migration is 12 month from the observation point i.e. January 2012 to May 2017.

### Data Cleansing and Exception Handling

In the construction of our development sample, at every observation point we exclude records using the following rules

|  |  |  |  |
| --- | --- | --- | --- |
| # | Issue | Detail | Handling |
| 1 | NPL record | The record/customer is NPL at the observation | Remove from sample |
| 2 | Open less than 3 months | The record/customer is new and there is no behavioral score | Remove from sample |
| 3 | No performance | The record/customer with no performance i.e. closure, write-off etc. | Remove from sample |

Table 5, Data cleansing and exception handling rules

### Final Development Sample

This section show step by step derivation of our final model development sample. Starting from the entire data set and working the way through record exclusion and association of final development sample with the bad rate.

Tables below show the number of customers in each of the observation point before and after exclusion.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Asset Class/Month | Jan-12 | April-12 | July-12 | Oct-12 | Jan-13 | April-13 | July-13 | Oct-13 |
| Total | 1,329,503 | 1,416,762 | 1,516,803 | 1,616,041 | 1,703,711 | 1,745,023 | 1,835,809 | 1,898,265 |
| Exclusion | | | | | | | | |
| NPL | 15,289 | 16,060 | 19,548 | 24,914 | 28,022 | 30,628 | 37,415 | 42,629 |
| Final Model Development Sample | | | | | | | | |
| Total | 1,314,214 | 1,400,702 | 1,497,255 | 1,591,127 | 1,675,689 | 1,714,395 | 1,798,394 | 1,855,636 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Asset Class/Month | Jan-14 | April-14 | July-14 | Oct-14 | Jan-15 | April-15 | July-15 | Oct-15 | Jan-16 | April-16 |
| Total | 1,959,013 | 2,059,847 | 2,122,108 | 2,168,803 | 2,210,056 | 2,255,059 | 2,304,387 | 2,347,721 | 2,342,648 | 2,343,296 |
| Exclusion | | | | | | | | | | |
| NPL | 52,109 | 74,621 | 84,705 | 93,522 | 88,110 | 98,575 | 113,996 | 124,857 | 116,855 | 125,631 |
| Final Model Development Sample | | | | | | | | | | |
| Total | 1,906,904 | 1,985,226 | 2,037,403 | 2,075,281 | 2,121,946 | 2,156,484 | 2,190,391 | 2,222,864 | 2,225,793 | 2,217,665 |

Table 6, Observations and development samples

The performances of the development sample (i.e. number of bad customer within the defined performance period) are as per the table below. Please note that there are two default performance definitions for Credit card portfolio. The first definition is 90+ DPD as reflected by BOT class. The second definition is TDR loans behavioral default. For detail on the definition, please refer to the definition section of this document.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Asset Class | % Bad | | | | | | | |
| Q1 Jan-12 | Q2  Apr-12 | Q3 July-12 | Q4  Oct-12 | Q1 Jan-13 | Q2  Apr-13 | Q3 July-13 | Q4  Oct-13 |
| All | 2.93% | 3.26% | 3.55% | 3.80% | 4.18% | 4.52% | 4.59% | 4.62% |
| PL Normal |  |  |  |  |  |  |  |  |
| SMA/SMQ |  |  |  |  |  |  |  |  |
| TDR |  |  |  |  |  |  |  |  |
| Watchlist/Reschedule |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Asset Class | % Bad | | | | | | | | | |
| Q1 Jan-14 | Q2  Apr-14 | Q3 July-14 | Q4  Oct-14 | Q1 Jan-15 | Q2  Apr-15 | Q3 July-15 | Q4  Oct-15 | Q1  Jan-16 | Q2  Apr-16 |
| All | 4.47% | 5.42% | 5.58% | 5.67% | 6.12% | 6.08% | 5.81% | 6.03% | 6.42% | 5.60% |
| PL Normal |  |  |  |  |  |  |  |  |  |  |
| SMA/SMQ |  |  |  |  |  |  |  |  |  |  |
| TDR |  |  |  |  |  |  |  |  |  |  |
| Watchlist/Reschedule |  |  |  |  |  |  |  |  |  |  |

Table 7, Default performance of the development sample

## Model Development

### Supermaster Scale and Rating

The first step in our PD term structure model development is the construction of supermaster rating and supermaster scale PD. For current loan and TDR customer, we use scoring model to create ratings and thus assign a suitable rating for each customer. For other asset classes, we directly observe the long run historical default rate within each asset class and assign them to an appropriate rating in the supermaster scale.



Figure 4, Overview of Data Input

The outcome of this exercise is the supermaster rating and scale as shown below. The PD in each rating are calibrated to the long run default performance up to June 2016.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Supermaster Rating** | **Asset\_class** | **B-Score Range** | | | | **PD Master Scale**  **30/1/2018** | **201612**  **Actual DR** |
| **HUUtd** | **HUDelq** | **LUUtd** | **LUDelq** |
| 1 | PL Normal |  |  | >1430 |  | 0.31% | 0.28% |
| 2 | PL Normal (Inactive) |  |  |  |  | 0.51% | 0.41% |
| 3 | PL Normal |  |  | 1357-1430 |  | 0.97% | 0.79% |
| 4 | PL Normal |  |  | 1287-1345 |  | 1.98% | 1.22% |
| PL Normal (New) |  |  |  |  |
| 5 | PL Normal | >750 |  | 1176-1286 |  | 3.94% | 3.23% |
| 6 | Reschedule |  |  |  |  | 8.70% | 7.27% |
| PL Normal | 674-750 |  | 1114-1175 | >981 |
| 7 | PL Normal | 592-673 |  | 938-1113 |  | 14.18% | 11.36% |
| TDR2 |  |  |  |  |
| Watch list |  |  |  |  |
| 8 | TDR4 |  |  |  |  | 21.05% | 18.18% |
| TDR |  |  |  |  |
| PL Normal | 400-591 | >482 | 793-937 | 813-980 |
| 9 | TDR1 |  |  |  |  | 34.67% | 29.88% |
| TDR3 |  |  |  |  |
| TDR6 |  |  |  |  |
| PL Normal | <399 | 321-482 | <792 | 651-812 |
| 10 | PL Normal |  | <320 |  | <650 | 50.69% | 46.00% |
| TDR5 |  |  |  |  |
| 11 | SMQ |  |  |  |  | 68.95% | 66.21% |
| SMA |  |  |  |  |

Table 8, Supermaster Scale

### Through-the-cycle rating transition matrix

After the supermaster rating and scale are finalized, we re-apply the rating criteria to our sample and observe the migration of supermaster rating. By taking a simple average across each observation point, we obtain the following obtain normalized long-run average (observed) transition matrix or through-the-cycle (TTC) transition matrix.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | Default |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |  |  |  |  |  |  |
| Default |  |  |  |  |  |  |  |  |  |  |  |  |

Table 9, Observed transition matrix

It is necessary that our transition matrices must possess the property of diagonal dominance. The probability of rating transition from the initial state should be a bell curve i.e. the diagonal transition probability (identical initial and final state) should be the highest transition rate and then subsequently decreasing according to the distance from the starting state (rating). We smooth out our transition rates to obtain such property using concepts from Moody’s.

The key concept of transition smoothing is to maintain the average movement from the observed matrix while lowering a variance in each rating. In addition, the left most or the right most rating should be the highest average movement since they can be only upgraded or downgraded. On the other hand, the middle rating should be the lowest average movement since there is a same probability to move upward or downward, having said that, the adjustment process involves of expert judgement and is a manual process.

Average movement for rating can be calculated through

Standard deviation for rating is then expressed as follow

It is then possible to compare the average movement and standard deviation of transition of the observed and the adjusted transition matrix‘s average movement. Note that the goal of the smoothing process is to find a transition matrix with good property while minimizing the changes to the average movement and standard deviation from the original matrix. The outcome of this stage are the adjusted and normalized transition matrices (final) as follow

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | D |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |  |  |  |  |  |  |
| D |  |  |  |  |  |  |  |  |  |  |  |  |

Table 10, Smoothed transition matrix

The new matrices have the following average movement and standard deviation which are comparable to the original observed matrices.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Rating | Average Movement | | | Standard Deviation | | |
| Observed | Smooth | Smooth  (Normalized) | Observed | Smooth | Smooth  Normalized |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |

Table 11, Pre and post adjustment: average movement and S.D.

### Generator matrix Q

This part will describe how to derive the generator matrix as continuous time transition matrix from the matrix M (the discrete matrix) within the acceptable range of error. Please see the methodology review section on the property of the generator matrix.

We use a library in R called “ctmcd” and “expm” in order to transform matrix M to matrix Q. More specifically, we choose “gmDA” function available in “ctmcd” library to find generator matrix Q and “expm” is the function to calculate Taylor series of matrix exponential. In addition, we need to find the generator matrix Q such that it minimizes the error between the discrete matrix and the continuous time matrix (after exponential or natural logarithm)

Where is our acceptable level of error

With the default “gmDA” function, we obtain the following value of the objective function: 0.00285 which is intolerable. In order to reduce this error, we use a scaling matrix (a diagonal matrix) to adjust the generators slightly.

The new objective function becomes

The optimization yields the following value of and the desired generator matrices for salary earner and entrepreneur respectively

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | D |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |  |  |  |  |  |  |
| D |  |  |  |  |  |  |  |  |  |  |  |  |

Table 12, Adjustment Marix A

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | D |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |  |  |  |  |  |  |
| D |  |  |  |  |  |  |  |  |  |  |  |  |

Table 13, Final Generator matrix

### Non-homogeneous curve adjustment

In this section, we will explain how to obtain an optimal choice of alpha and beta to best fit the model. Suppose that is output of taking an exponential of the non-homogenous transition matrix as follow

The output is highly dependent on the curve adjustment function as reviewed in the methodology review section. Mathematically speaking, one can construct the optimization equation to find the scalar of alpha and beta such that the modelled probability term structure is the closest to the actual cumulative default rate at any time i.e.

where is the rating in consideration

is the historically cumulative default rate of each rating at time

is the PD for each rating from the generator matrix Q at time .

The inequality constraint is to ascertain that probability of default from the model will always be higher than the actual DR at any point on the term structure. We adopt R programming with function “solnp” available within the “RSolnp” library as our optimization tool.

The  is obtained from observing an average 5 year historically marginal default rate of performing account in the sample between 201201-201206 and then sum up each year to obtain cumulative. The default definition is the same as in the previous exercise i.e. 90+ DPD and behavioral default (from 2016 onwards). For example, for observe point 201201, the marginal default rate for rating 1 can be illustrated as per the following figure



Figure 5, Observable default rate term structure

The cumulative default rate for each rating for salary earner and entrepreneur that will be use in the calibration of our PD term structure is as follow

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Supermaster Rating | Year | | | | | |
| 0 | 1 | 2 | 3 | 4 | 5 |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |

Table 14, Default term structure

By solving optimization above, the result of alpha and beta associated to each rating is given as the following. However, a big gap on the first year especially on the lower rating between PD model and actual default leads to a substantial amount of sum square error across all ratings of approximately 0.60731. We obtained the following curve adjustment parameter

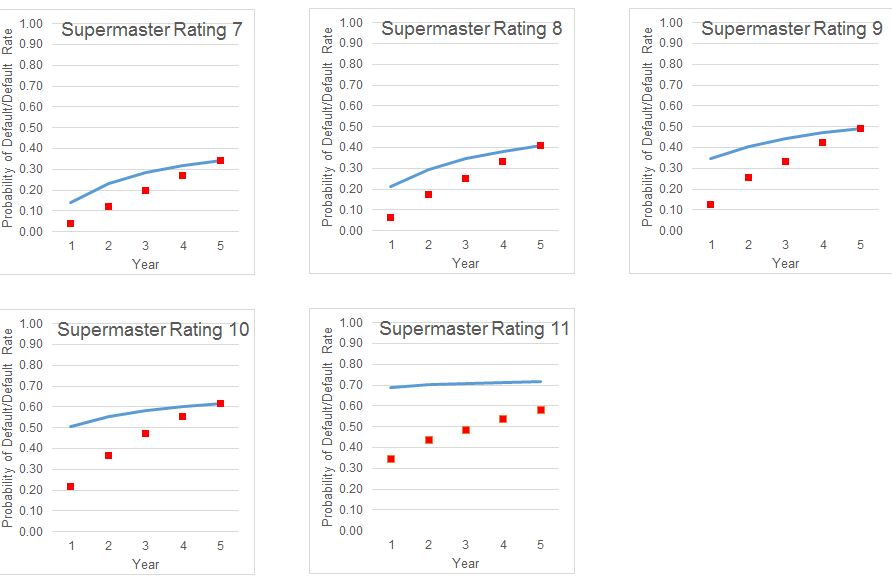
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Curve Adjustment Parameter |  | |  | | Supermaster Rating | | | | | | | | | |
| 1 | 2 | | 3 | | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | D |
|  |  |  | |  | |  |  |  |  |  |  |  |  |  |
|  |  |  | |  | |  |  |  |  |  |  |  |  |  |

Table 15, Curve adjustment parameter

Subsequently, one can obtain the cumulative TTC PD term structure for 5 years from the model follow

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Supermaster Rating | Actual Cumulative Default Rate | | | | | | Model TTC Cumulative Default Rate | | | | | |
| Year | | | | | | Year | | | | | |
| 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |  |  |  |  |  |  |

Table 16, Cumulative default rate: actual vs. model



### Conversion from TTC to PIT PD

In this section, we will explain how to convert TTC PD obtained from the previous steps to PIT PD via the Vasicek equation. There are 2 components namely the and that would need to be identified. We suspect that the default rate of customer in each occupation segment i.e. salary earner and entrepreneur of the customer may have different relationship with macro variables. We segmented our samples into two group based on occupation. The definition of occupation is grouped as follow: entrepreneur and freelance are flagged as “Entrepreneur”, otherwise they are “Salary Earner”.

The data preparation is to identify which macro variable is suited for each sub-segmentation and it can be generated by seeking the relationship between actual default and macro-economic index via R programming. Specifically, the observation period is quarterly from Jan-2012 to April-2016 and the performance period is 1 year sliding window from Jan-2013 to April-2017. The good and bad definition is exactly the same as curve fitting in Non-homogenous section.

The historical macro variable index is based on the Bank’s internal organization research or “K-Research”. Since PIT PD is a forward-looking concept then the point of macro-economic index will be the next 12 months from observation point. The greater detail is shown as the following table:

|  |  |  |
| --- | --- | --- |
| No. | Name | Description |
| 1. | Date | Data monthly as of observation point from 200901 to 201603 |
| 2. | DR\_Salary\_Earner | 1 year actual default rate of Salary Earner from Date |
| 3. | DR\_Self\_Employed | 1 year actual default rate of Entrepreneur from Date |
| 4. | GDP\_2002 | Historical Thai GDP data as of (Date + 12 months) |
| 5. | PRIVATE\_CONSUM | Private Consumption index data as of (Date + 12 months) |
| 6. | INVESTMENT | Investment index data as of (Date + 12 months) |
| 7. | GOV\_SPEND | Government spending index data as of (Date + 12 months) |
| 8. | EXPORT | Export index data as of (Date + 12 months) |
| 9. | IMPORT | Import index data as of (Date + 12 months) |
| 10. | UNEMPLOY\_RATE | Unemployment rate index data as of (Date + 12 months) |
| 11. | HEADLINE\_CPI | Headline CPI index data as of (Date + 12 months) |
| 12. | CORE\_CPI | Core CPI index data as of (Date + 12 months) |
| 13. | DIESEL\_PRICE | Diesel Price index data as of (Date + 12 months) |
| 14. | MLR | MLR index data as of (Date + 12 months) |
| 15. | HOUSE\_PRICE | Houseprice index data as of (Date + 12 months) |
| 16. | HOUSEHOLD\_DEBT | Household debt index data as of (Date + 12 months) |
| 17. | CCI | CCI index data as of (Date + 12 months) |

Table 17, Input data for macro-overlay model

We also randomly spit sample into testing and training i.e. 80 to 20 respectively. After that, we fit a linear model to the time series data to establish the relationship between macro variables and the default rates over time. The selection of independent variables is done via a backward selection process. However, the final set of macro-economic variables is selected into the final model are based on economic intuition, predictive significance of the independent variable and a tolerable level of model accuracy (Adjusted Rsquare).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product | Occupation | *β0* | *βUnemp.Rate* | *βHHDebt* | *Adjusted* |
| Credit card | Salary Earner |  |  |  | 40-80% |
| Entrepreneur |  |  |  | 40-80% |

Table 18, Final coefficient of macro-overlay model

After obtaining CAF, the next part is to calculate that is the best fit to the model. In other word, we put the actual default as PD PIT and PD TTC as forward-looking PD with calculated CAF from above then one can solve optimization problem to obtain . However, due to data availability with more prudential correlation, we use a benchmark from Moody’s for salary earner and empirical observation for entrepreneur result of 4 and 10 percent respectively. This is because we believe that entrepreneur is more sensitive to systematic risk than salary earner.

|  |  |  |
| --- | --- | --- |
| Product | Occupation | Correlation |
| Credit card | Salary Earner | 4% |
| Entrepreneur | 10% |

Table 19, Correlation factor

## Pre-Validation

### TTC Model

For the 1-year TTC PD in the master scale we have validate it using a binomial test

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Supermaster Rating | Asset\_class | PD Master Scale  30/1/2018 | Critical PD  30/1/2018 | Actual DR 201612 | Binomial Test |
| 1 | PL Normal | 0.31% | 0.32% | 0.28% | Accept |
| 2 | PL Normal | 0.51% | 0.53% | 0.41% | Accept |
| 3 | PL Normal | 0.97% | 1.01% | 0.79% | Accept |
| 4 | PL Normal | 1.98% | 2.05% | 1.22% | Accept |
| PL Normal | Accept |
| 5 | PL Normal | 3.94% | 4.01% | 3.23% | Accept |
| 6 | Reschedule | 8.70% | 8.92% | 7.27% | Accept |
| PL Normal | Accept |
| 7 | PL Normal | 14.18% | 14.54% | 11.36% | Accept |
| TDR | Accept |
| Watch list | Accept |
| 8 | TDR | 21.05% | 21.67% | 18.18% | Accept |
| TDR | Accept |
| PL Normal | Accept |
| PL Normal | Accept |
| 9 | TDR | 34.67% | 35.77% | 29.88% | Accept |
| TDR | Accept |
| TDR | Accept |
| PL Normal | Accept |
| PL Normal | Accept |
| 10 | PL Normal | 50.69% | 51.36% | 46.00% | Accept |
| PL Normal | Accept |
| TDR | Accept |
| 11 | SMQ | 68.95% | 69.47% | 66.21% | Accept |
| SMA | Accept |

Table 20, Binomial test of Supermaster Scale

For the TTC PD term structure, the generator matrix and the curve adjustment parameters are calibrated to the PD term structure from Jan-2013 to Jan-2017. At the time of development, we have yet to identify an appropriate measure to validate the performance of the PD term structure. The optimization was performed to ensure that the modelled term structure closely matches that of the observations. Based on the sum square error of the actual vs. model prediction, the sum of squared error across all rating is attributed to approximately 0.60731.

|  |  |  |
| --- | --- | --- |
| Curve Adjustment | SSE | MAPE |
| All | <0.8 | <80% |

**4.5.2 PIT Model**

The performance of the (linear model) is validated via adjusted . Both the salary earner and entrepreneur has an acceptable range of adjusted as follow

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product | Occupation | *β0* | *βUnemp.Rate* | *βHHDebt* | *Adjusted* |
| Credit card | Salary Earner |  |  |  | 40-80% |
| Entrepreneur |  |  |  | 40-80% |

Table 21, Adjusted for macro-overlay models

# Exposure at Default (EAD)

Exposure at Default (EAD) is defined as the expected outstanding balance at the time of default. In practice, the EAD of products with explicit credit line, such as credit card, tends to be the same as its balance including the potential increase in the outstanding balance from a reference date to the time of default. KBank estimates the EAD for credit card or other revolving credit products by using indirect approach that focus on evaluating Credit Conversion Factor (CCF)

For an active revolving instrument (non-NPL), the EAD at reporting date can be described as

Where, is current principal amount at time

is current limit amount at time

is Credit Conversion Factor (CCF)

is Effective Interest Rate (EIR)

is a variable acting as an on/off switch for interest calculation

is the number of month for accrued interest to be calculated

is realized accrued interest at time

CCF model predicts expected utilization if a customer defaults. Thus, validation is performed only on the defaulted population i.e. the NPLs. We use the following definition of actual CCF for the purpose of our model development and model validation.

For an instrument , the actual Credit Conversion Factor (CCF) is defined as percentage rate of current undrawn credit line that will likely be utilized by borrowers at the time of default i.e.

Where is the time of observation

is the time of default

is the prediction window (usually 12)

## Scope

The CCF models used to calculate EAD are available to all KBank’s credit card customers whose cards are valid and statuses are not in non-performing stage (non-NPL only).These models are developed under two segments of Kbank’s credit card customer, namely High utilization(HU) and Low utilization & inactive (LU). Different CCF values will be assigned to the KBank’s customers in different segments.

## Methodology Review

### Constant Through-the-Cycle CCF

The constant Through-the-Cycle CCF value for each segment will be the average of all historical CCF values in that segment.

### Scorecard for Through-the-Cycle CCF

Scorecard for credit conversion factor represents different characteristics of Kbank customer’s credit card demographics and spending behaviors (current age, age of account, unused credit line and so on). Kbank implements this scorecard model to convert customer’s credit card characteristic factors into CCF scores and then assign CCF rating corresponding to the CCF scores.

An example of KBank’s credit scorecard is shown in the table below

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Characteristic | Attribute | Score |
|  | Constant |  |  |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |

Table **22**, TTC CCF Scorecard

### Point-in-Time (PIT) CCF

To comply with IFRS9, Kbank develops a Point-in-time CCF model that takes into account cyclical effects or macroeconomic factors. The appropriate macroeconomic factors will be selected using variable analysis and the logistic regression will be performed to find the specific value of coefficient (beta) for each macroeconomics factor.

To properly handle our model, these assumptions are made,

1. Kbank assumes our historical CCF can reflect future CCF.
2. Kbank assumes our credit card spending behavior can solely be assigned into two groups, HU and LU& inactive.

## Model Development Approach

This section will provides a step-by-step guide of how the EAD is calculated

The input of actual CCF data will be prepared and performed an exception handling under development preparation and data cleansing then this data will be mapped with customer behavior data in order to classify the segments of the customers (HU or LU & inactive). The data after mapping will be grouped as a customer level and pass through a modeling section where CCF model was developed using this sample input data. Our CCF model will be validated with CCF validation sample before implementation.

## Data Management

### Overview of Input Data

CCF model development is performed based on a sample of defaulted instrument. In addition to specification of default definition and default time period, the input for actual CCF calculation includes outstanding balance at default, past outstanding balance and past limit. In summary, the data requirement for the calculation of actual CCF calculation is shown as follows:

Figure 6, Data required for actual CCF calculation

### Sample Design

In model development, there are two specific time periods must be considered; observation and performance period. The observation period consists of a specific period of time prior to the point of default that used to characterize a customer’s past behavior. The performance period consists of a specific time to default. Data used to develop model is listed in Table 23.

|  |  |  |
| --- | --- | --- |
| Period | Timeframe | Data |
| Observation Period (tr) | January 2007- June 2015 | * Information of customers at 12 or 6 months prior to the time they default. * Macroeconomic factors |
| Performance Period (td) | January 2008- December 2015 | * Total outstanding balances of customers who default for the first time * Actual CCF calculated by using formula in TTC-EAD\_Model Development for Credit Card Document |

Table 23, Development data sets

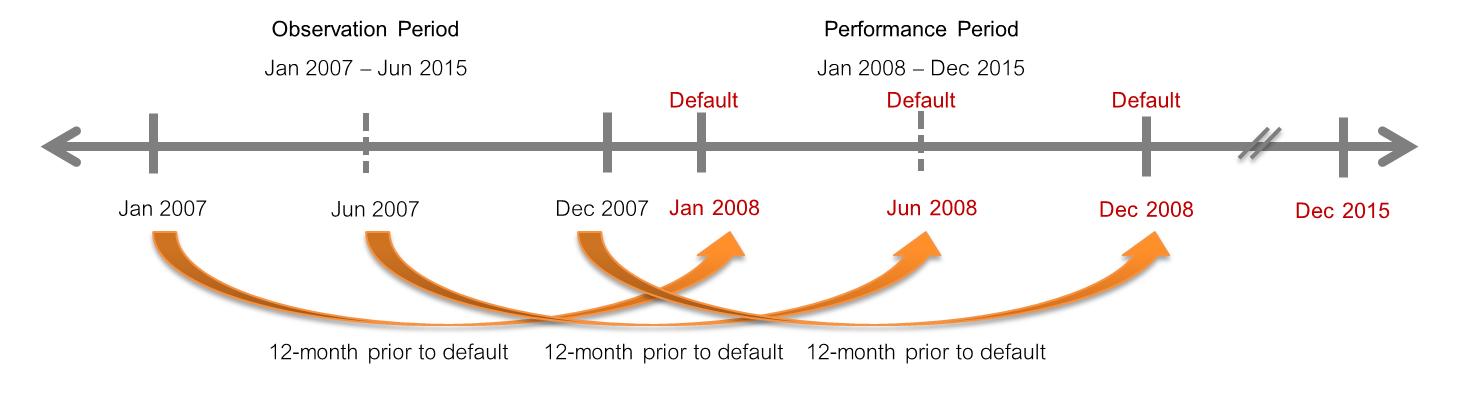


Figure 7, Development data sets periods

The data used in PIT-CCF model development is time series data which is the data collected on the same observational unit at multiple time periods. The time periods are ninety-six months from January 2007 to December 2014 which is the observation period. Therefore, the data set for developing model consists of ninety-six observations.

### Data Cleansing and Exception Handling

Kbank’s development data will enter into exclusion rule and data handling process to filter out undesired customers and reassign value to the specific characteristic factors in this data.

#### Exclusion Rule

Some customers may have some conditions that might not be appropriate to be used in CCF model development, such as missing credit line, months on book less than 6 months or etc. Thus, need to be excluded from the model development

The customers with all conditions listed in the table below will be excluded from the model development.

|  |  |
| --- | --- |
| No. | Exclusion Rules |
| 1 | Corporate cards  Card type code not in (840, 940) |
| 2 | Total outstanding balances before write-off at the time of default <= 2,000 THB |
| 3 | Missing credit line or credit line <= 0 at observation period(tr) |
| 4 | Missing open date |
| 5 | Months on book of customer < 6 months |
| 6 | MIN\_AGE\_OF\_ACC < 3 months |
| 7 | Flagged with bad BLOCK\_CODE  W, A, Z, B, K, M, O, I, Y, L, E, F, S, U, X, J, D |
| 8 | Cannot identify segment  HU, LU, Inactive |

Table 24, Exclusion Rules for the model development

#### Data Handling

Because our CCF score card characteristic factors may contain null values or values that need to be converted before being used, so these values must be reassigned before developing our model.

An example of how to handle data

|  |  |  |  |
| --- | --- | --- | --- |
| if | BLOCK\_CODE = ‘W’ | then | BLOCK\_CODE = 1 |
| else if | BLOCK\_CODE = ‘A’ | then | BLOCK\_CODE = 2 |
| else if | BLOCK\_CODE = ‘Z’ | then | BLOCK\_CODE = 3 |
| else if | BLOCK\_CODE = ‘B’ | then | BLOCK\_CODE = 4 |

Please refer to an [Appendix A](#_A._Full_reassignment) for full reassignment of CCF score card characteristic factors.

### Final Development Sample

#### **Through-the-Cycle (TTC) CCF**

The default customer data set is separated into two segments; HU and LU & Inactive and then randomly separated into two groups; training sample (80%) and testing sample (20%) as shown in **Error! Reference source not found.**.

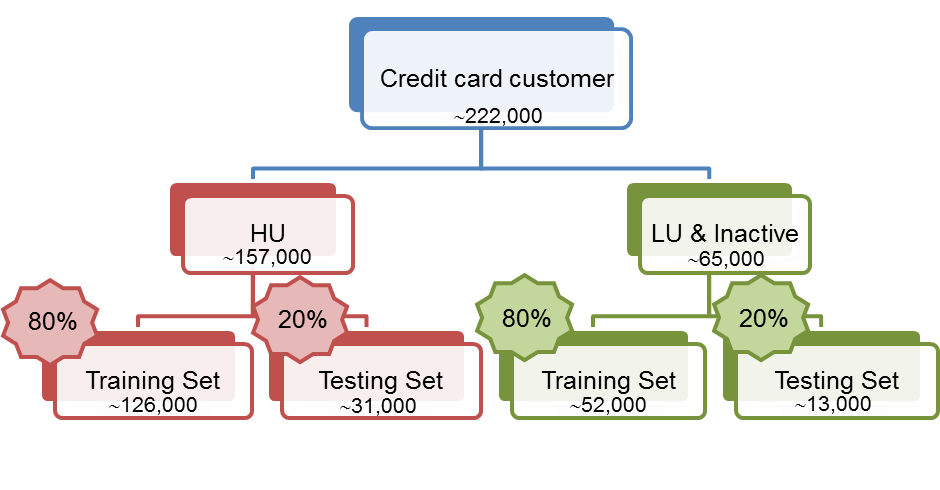


Figure 8, Sample size

#### **Point-in-Time (PIT) CCF**

Kbank develops PIT-CCF models based on customer credit usage rate and TTC-CCF rating. There are 5 sub-models for CCF PIT. Each model is for each TTC-CCF rating.

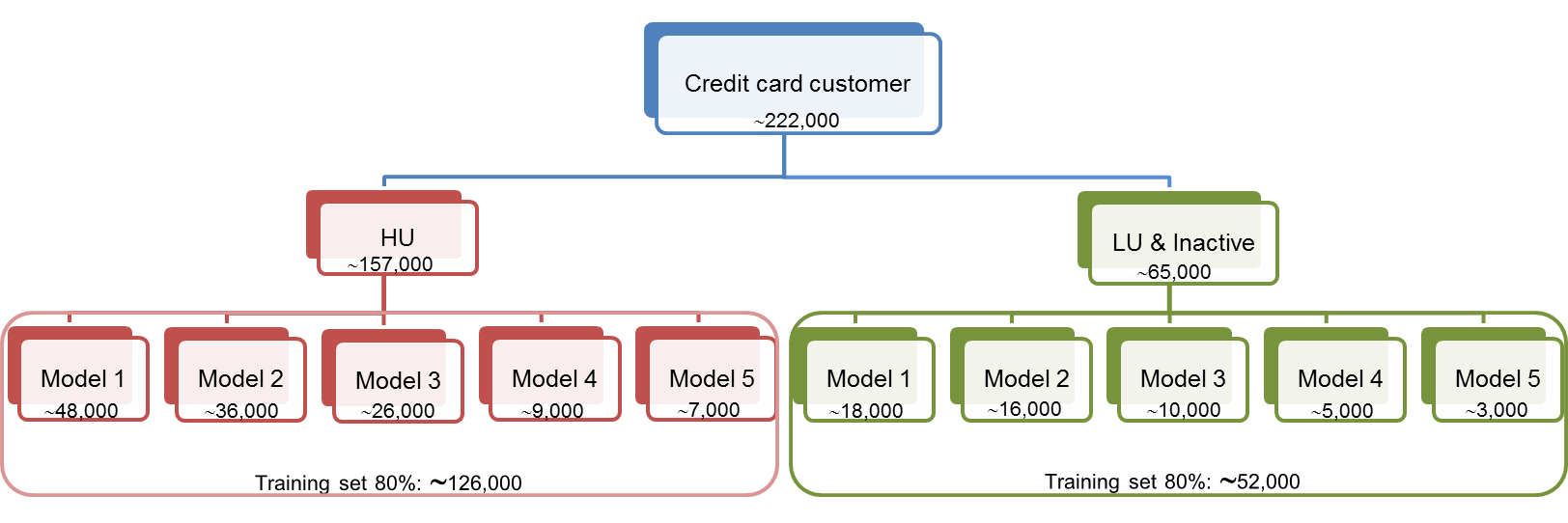


Figure 9, PIT Sample size

## Model Development

Credit Conversion Factor (CCF) is defined as percentage rate of current undrawn credit line that will likely be utilized by borrowers at the time of default.

The actual CCF is calculated as follow,

Where, td = time when the customer default.

tr = observation time (usually 6 or 12 months prior to the default time).

This actual CCF will be used in Through-the-cycle CCF and Point-in-Time CCF.

### Through-the-Cycle CCF Model

This section describes the statistical modeling approach that is used in the estimation of CCF in order to obtain EAD against the candidate drivers. Kbank has developed the CCF models based on the two different philosophies of credit risk rating. The first one, Through-the-Cycle (TTC), is independent from cyclical changes in the creditworthiness of a customer. The second one, Point-in-Time (PIT), takes into account cyclical effects or macroeconomic factors.

The CCF modeling has posed substantial challenged with regards to predicting the exposure at default time because the distribution of CCF does not conform to normal distribution. As shown in Figure below, the CCF distribution tends to be highly bimodal with two peaks at zero and one (no change in balance and balance has gone up to the credit line respectively), and a relatively flat distribution between those peaks.

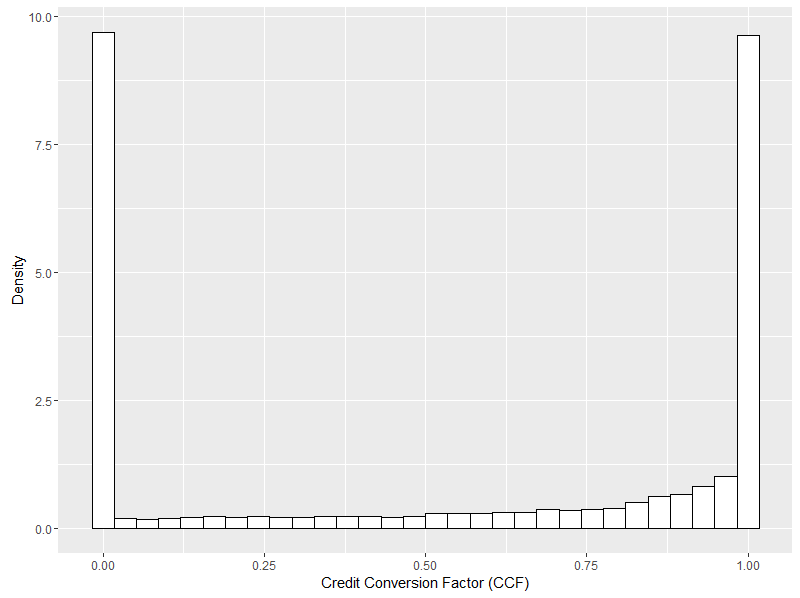


Figure 10, Distribution of the credit conversion factor

Using traditional linear regression models may be less suitable for the CCF because of the non-normal distribution and bounded response variable, leading to inaccurate predicted value; CCF is less than zero or greater than one. Therefore, Kbank uses binary logit model in an attempt to resolve this issue by grouping the observations for the CCF into two categories. For the binary response variable, the actual CCF is split into two levels as follows:

Good: Actual CCF >= Average actual CCF

Bad: Actual CCF < Average actual CCF

Kbank performs logistic regression with divergence objective function in FICO Model Builder. The divergence technique is the objective function to maximize the divergence of the model, meaning that it concentrates on how well the model can discriminate good customers from bad customers and maximize that. This can be said that bad customers would be clustered in the low range score and vice versa as shown in Figure below.

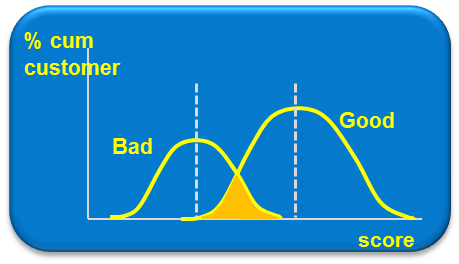


Figure 11, Divergence technique

Furthermore, FICO Model Builder uses the Marginal Contribution (MC) to measure how much effect does each variable has on the model and manage the relationship of variables in the model (multicollinearity) automatically. This is shown as a bar of greenish and reddish color in the GUI, where the color indicates the difference between the MC of training and testing sample set and the length indicates how large is the MC of each variable is for the model.

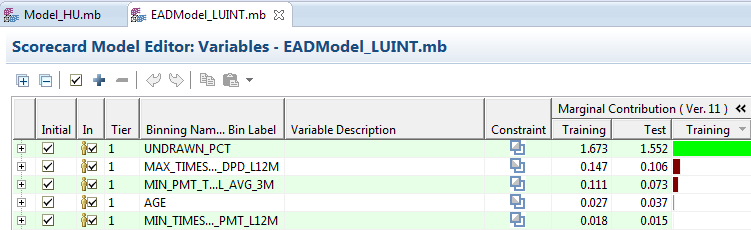


Figure 12, Marginal Contriubiton display in FICO Model Builder

Characteristic Analysis

Scorecard consists of a group of characteristics statistically determined to be predictive in separating good and bad customers. To analyze strength of a variable in the ability to separate good customers from bad customers, the following criteria are considered:

Weight of Evidence (WOE)

WOE is used to measure predictive power for each attribute.

WOE = ln(Distribution Good / Distribution Bad)

The higher WOE, the lower risk the customer becomes bad and vice versa.

Information Value (IV)

IV is used to measure predictive power for each characteristic.

IV = Sum [(Distribution Good – Distribution Bad) \* WOE]

Rules of Thumb for IV:

< 0.02 : Unpredicted

0.02-0.1 : Weak predictive power

> 0.1-0.3 : Medium predictive power

> 0.3 : Strong predictive power

Business sense

There is reasonable trend of good/bad odds by characteristics or attributes. To identify the trend of WOE value of attributes in each characteristic, the following things should be considered in binning the characteristic.

* Each attribute should have a population of at least 5%
* There should not be an attribute which has good customers of zero or bad customers of zero
* For any characteristics, WOE of each attribute should differ significantly.

Kbank performs characteristic analysis by using FICO Model Builder and represents the results in Table 25 and **Error! Reference source not found.** for HU model and LU & Inactive model respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Characteristic | IV | IV Meaning |
| 1 | DRAWN\_PCT |  | Medium |
| 2 | UNDRAWN\_PCT |  | Weak |
| 3 | UNDRAWN\_AMT |  | Weak |
| 4 | MIN\_SCORE |  | Weak |
| 5 | REF\_PRINC |  | Weak |
| 6 | MIN\_RISKGRADE |  | Weak |
| 7 | MAX\_TIMES\_X\_PLUS\_DPD\_L8M |  | Weak |
| 8 | MAX\_TIMES\_X\_PLUS\_DPD\_L3M |  | Weak |
| 9 | MIN\_PMT\_TO\_PREV\_BAL\_AVG\_3M |  | Weak |
| 10 | MIN\_PMT\_TO\_PREV\_BAL\_AVG\_6M |  | Weak |
| 11 | MIN\_MONTHS\_SINCE\_LAST\_DELIN |  | Weak |
| 12 | MAX\_CASH\_ADV\_TO\_SPEND\_AVG\_6M |  | Unpredicted |
| 13 | MAX\_DAYS\_SINCE\_LAST\_PURCHASE |  | Unpredicted |
| 14 | MIN\_TIMES\_FULL\_PMT\_L6M |  | Unpredicted |
| 15 | MIN\_TIMES\_PART\_PMT\_L6M |  | Unpredicted |
| 16 | MIN\_AGE\_OF\_ACC |  | Unpredicted |
| 17 | CUR\_AGE |  | Unpredicted |
| 18 | MAX\_CASH\_ADV\_TO\_LIMIT\_AVG\_3M |  | Unpredicted |
| 19 | MAX\_TIMES\_GREATER\_30\_DPD\_L8M |  | Unpredicted |
| 20 | MAX\_TIMES\_GREATER\_30\_DPD\_L3M |  | Unpredicted |
| 21 | MAX\_BAL\_TO\_LIMIT\_AVG\_3M |  | Unpredicted |
| 22 | MAX\_DAYS\_SINCE\_LAST\_PMT |  | Unpredicted |
| 23 | MAX\_TIMES\_OVER\_LIMIT\_L3M |  | Unpredicted |
| 24 | MAX\_TIMES\_OVER\_LIMIT\_L6M |  | Unpredicted |
| 25 | REF\_LINE |  | Unpredicted |
| 26 | MAX\_TIMES\_60\_TO\_89\_DPD\_L6M |  | Unpredicted |
| 27 | GENDER |  | Unpredicted |
| 28 | MAX\_TIMES\_GREATER\_60\_DPD\_L6M |  | Unpredicted |

Table 25, Charactertistic analysis results for HU model

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Characteristic | IV | IV Meaning |
| 1 | MAX\_TIMES\_X\_PLUS\_DPD\_L8M |  | Weak |
| 2 | MAX\_TIMES\_X\_PLUS\_DPD\_L3M |  | Weak |
| 3 | MIN\_SCORE |  | Weak |
| 4 | MIN\_RISKGRADE |  | Weak |
| 5 | UNDRAWN\_AMT |  | Weak |
| 6 | MIN\_MONTHS\_SINCE\_LAST\_DELIN |  | Weak |
| 7 | DRAWN\_PCT |  | Weak |
| 8 | UNDRAWN\_PCT |  | Weak |
| 9 | REF\_PRINC |  | Weak |
| 10 | MIN\_PMT\_TO\_PREV\_BAL\_AVG\_6M |  | Weak |
| 11 | MAX\_CASH\_ADV\_TO\_SPEND\_AVG\_6M |  | Weak |
| 12 | MIN\_PMT\_TO\_PREV\_BAL\_AVG\_3M |  | Weak |
| 13 | MAX\_DAYS\_SINCE\_LAST\_PMT |  | Unpredicted |
| 14 | MAX\_BAL\_TO\_LIMIT\_AVG\_3M |  | Unpredicted |
| 15 | REF\_LINE |  | Unpredicted |
| 16 | MIN\_TIMES\_FULL\_PMT\_L6M |  | Unpredicted |
| 17 | MAX\_DAYS\_SINCE\_LAST\_PURCHASE |  | Unpredicted |
| 18 | CUR\_AGE |  | Unpredicted |
| 19 | MAX\_CASH\_ADV\_TO\_LIMIT\_AVG\_3M |  | Unpredicted |
| 20 | MIN\_AGE\_OF\_ACC |  | Unpredicted |
| 21 | MIN\_TIMES\_PART\_PMT\_L6M |  | Unpredicted |
| 22 | MAX\_TIMES\_OVER\_LIMIT\_L6M |  | Unpredicted |
| 23 | MAX\_TIMES\_OVER\_LIMIT\_L3M |  | Unpredicted |
| 24 | GENDER |  | Unpredicted |

Table 26, Charactertistic analysis results for LU & Inactive model

#### **Segmentation Analysis (if applicable)**

The objective of the analysis is to determine the optimal number of scorecards to develop, along with the best segmentations. The sub-populations must contain adequate number of observations for the development, and must distinguish predictive patterns and population distributions. By applying business purpose and expert judgment, data is classified into two segments due to credit usage rate as follows:

1. High Utilization (HU) Customers with average utilization rate > 35% in previous 3 months.
2. Low Utilization (LU) and Inactive Customers with average utilization rate <= 35% in previous 3 months and customers without spending in previous 8 months respectively.

#### **Final constant Through-the-Cycle CCF**

Kbank average all CCF values for all development samples in each segment and obtain the final constant Through-the-Cycle CCF as follows,

|  |  |
| --- | --- |
| Segment | CCF avg |
| HU | 0.35 |
| LU & Inactive | 0.55 |

Table 27, constant TTC CCF by segment

Kbank use constant Through-the-Cycle CCF model as a champion model.

#### **Final Through-the-Cycle CCF Scorecard**

After performing characteristic and segmentation analysis, the final Through-the-Cycle CCF Scorecard are carried out per each segment as follow,

***CCF scorecard for HU customer***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Characteristic | Attribute | Score | %Proportion |
|  | Constant |  |  |  |
| 1 | UNDRAWN\_PCT (%) |  |  |  |
| 2 | MAX\_TIMES\_X\_PLUS\_DPD\_L8M |  |  |  |
| 3 | MIN\_PMT\_TO\_PREV\_BAL\_AVG\_3M (%) |  |  |  |
| 4 | CUR\_AGE (years) |  |  |  |
| 5 | MIN\_AGE\_OF\_ACC (months) |  |  |  |

Table 28, Score of grouped variables in HU scorecard

After calculating CCF score for all credit card customers, Kbank assigns different CCF rating for each customer by using score range criteria as shown in table

|  |  |  |  |
| --- | --- | --- | --- |
| CCF Rating | CCF Score Interval | CCFAVG | %Proportion |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |

Table 29, CCF score and CCF rating mapping for HU model

***CCF scorecard for LU & Inactive customer***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Characteristic | Attribute | Score | %Proportion |
|  | Constant |  |  |  |
| 1 | UNDRAWN\_PCT (%) |  |  |  |
| 2 | MAX\_TIMES\_X\_PLUS\_DPD\_L8M |  |  |  |
| 3 | MAX\_CASH\_ADV\_TO\_SPEND\_AVG\_6M (%) |  |  |  |
| 4 | CUR\_AGE (years) |  |  |  |
| 5 | MIN\_PMT\_TO\_PREV\_BAL\_AVG\_3M (%) |  |  |  |

Table 30, Score of grouped variables in LU & Inactive scorecard

|  |  |  |  |
| --- | --- | --- | --- |
| CCF Rating | CCF Score Interval | CCF\_AVG | %Proportion |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |

Table 31, CCF score and CCF rating mapping for LU & Inactive model

### Point in time CCF model

#### **Beta distribution fitting**

In each sub-segment, Kbank observes that the actual CCF distribution tends to be highly bimodal with two peaks at zero and one. Therefore, Kbank decides to fit all actual CCF in each sub-segment with the beta distribution.

To fit the beta distribution in each sub-segment, Kbank need to solve for parameter alpha (, beta () from the following equation,

Where are parameters in beta distribution

i = ith sub-segment

The result is as follows,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sub-segment | Utilization | CCF Rating | alpha | beta |
| 1 | HU | 1 |  |  |
| 2 | HU | 2 |  |  |
| 3 | HU | 3 |  |  |
| 4 | HU | 4 |  |  |
| 5 | HU | 5 |  |  |
| 6 | LU | 1 |  |  |
| 7 | LU | 2 |  |  |
| 8 | LU | 3 |  |  |
| 9 | LU | 4 |  |  |
| 10 | LU | 5 |  |  |

Table 32, Alpha, Beta under beta distribution for each sub-segment

**Note,** the alpha, beta from Table 32 will be used in the transformation of beta distribution to normal distribution in charactertistic analysis section.

**Characteristic Analysis**

For any CCF value in each sub-segment, the response variables are then obtained by calculating average actual CCF by month over the observation period as shown in the equation (1)

(1)

Where, m =each month in observation period.  
n = Number of customers in observation period m for each sub-segment.

i = ith sub-segment

Then, transform beta distribution of actual CCF values into normal distribution.

Using

Where,

= inverse standard normal distribution function.

= the beta distribution function with parameter alpha (a) and beta (b) from Table 32

t = reporting time

After our actual CCF are converted into normal distribution variable, we then search for the proper macroeconomic factors as follows,

For independent variables, data is collected on candidate macroeconomic factors as shown in Table 2. They are monthly historical macroeconomic factors from January 2001 to December 2014.

|  |  |
| --- | --- |
| No. | Variables |
| 1 | Change in Thai Gross Domestic Product (GDP)-YoY |
| 2 | Private consumption |
| 3 | Investment |
| 4 | Government spending |
| 5 | Export |
| 6 | Import |
| 7 | Unemployment rate |
| 8 | Headline Consumer Price Index (CPI) |
| 9 | Core Consumer Price Index (CPI) |
| 10 | Diesel price |
| 11 | MLR |
| 12 | House price |
| 13 | Household debt |
| 14 | Consumer Confidence Index (CCI) |

*Table 33*

, Candidate independent variables considered for the model

To select appropriate independent variables; macroeconomic factors, included in the regression models, the following criteria are considered:

1. Those macroeconomic factors affect consumption of retail credit card customers.

2. Those macroeconomic factors describe linear relationship between them and CCF.

After perform variable analysis, the candidate macroeconomic factors selected to be in regression model are listed in Table 34.Table 34, Selected macroeconomic factors

|  |  |  |
| --- | --- | --- |
| No. | Macroeconomic factors | Expected sign\* |
| 1 | GDP | - |
| 2 | Unemployment rate | + |
| 3 | Household debt | + |
| 4 | CCI | - |

Table 34, Selected macroeconomic factors

Note1: +/- signs describe direction of linear relationship between those factors and CCF.

#### **Final Point-in-Time CCF**

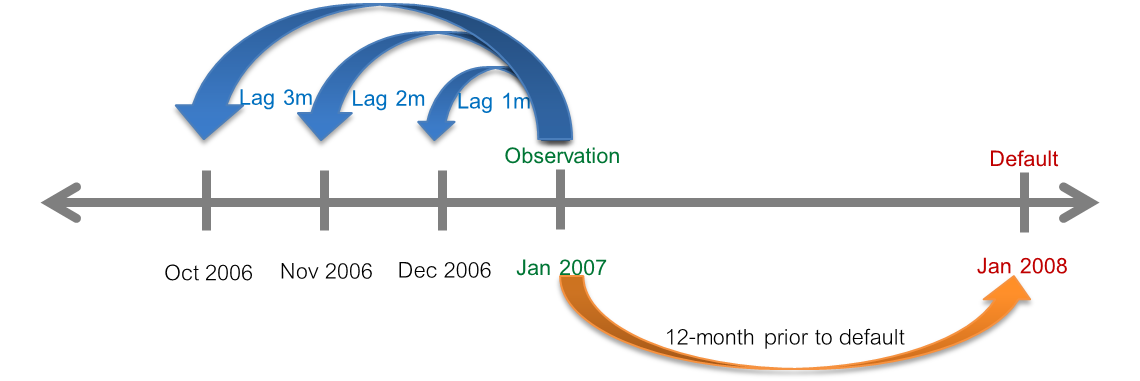


Figure 13, Lagging macroeconomics factors

The general point-in-time CCF equation is defined as,

Where, t = current observation time

= number of months that we observe back from our current observation time for macroeconomic factors i.

= coefficients of macroeconomic factors i.

= macroeconomics factors i at time t-

= inverse beta transformation with parameter alpha (a) and beta (b) from Table 32.

= standard normal distribution with mean=0 and variance =1

After performing characteristic and segmentation analysis, Kbank finally choose the lagging macroeconomic factors and coefficients for each customer segment as follows,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | TTC-CCF rating | GDP | Unemployment rate | Household debt | CCI |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |

Table 35, PIT-CCF HU model: Lagging macroeconomic factors

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | TTC-CCF rating |  |  |  |  |  |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |

Table 36, PIT-CCF HU model: Coefficient

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | TTC-CCF rating | GDP | Unemployment rate | Household debt | CCI |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |

Table 37, PIT-CCF LU & Inactive model: Lagging macroeconomic factors

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | TTC-CCF rating |  |  |  |  |  |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |

Table 38, PIT-CCF LU & Inactive model: Coeffiecinet

**Note.** The code steps and code for model finalization will be shown in [Appendix B.](#_B.__Finalize)

## Pre-Validation

### TTC Model

#### **Validation Methodology**

Kbank compares R-squared (R2), Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) of training sample and testing sample to ensure performance and robustness of the model.

R2: A statistical measure of how close the data is to fitted regression line. The larger R2, the better model it is.

MSE: A statistical measure of difference between actual value and predicted value.

MAPE: A statistical measure for forecasting error (percentage error of predicted value compared to actual value).

Moreover, Kbank analyzes trend over the time of estimated PIT-CCF whether it is consistent with actual CCF value or not.

#### **Validation Result**

The result in Table 39 indicates that there is no significant difference between two samples.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Predictive Power Statistic | Training (80%) | Testing (20%) |
| HU | R2 | 70-100% | 70-100% |
| MSE | 10-40% | 10-40% |
| LU & Inactive | R2 | 70-100% | 70-100% |
| MSE | 10-40% | 10-40% |

Table 39, Scorecard validation results

### Model Precision Test: Actual vs. Forecasted

**Model validation result: HU model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TTC-CCF rating 1** | | **TTC-CCF rating 2** | | |
|  | |  | | |
| MAPETTC-CCF = 6.14 | MAPEPIT-CCF = 5.54 | MAPETTC-CCF = 7.73 | | MAPEPIT-CCF = 5.23 |
|  |  |  | |  |
|  |  |  | |  |
| **TTC-CCF rating 3** | | **TTC-CCF rating 4** | | |
|  | |  | | |
| MAPETTC-CCF = 5.36 | MAPEPIT-CCF = 4.80 | MAPETTC-CCF = 8.64 | | MAPEPIT-CCF = 6.16 |
|  | | | | |
| **TTC-CCF rating 5** | | | | |
|  | | | | |
| MAPETTC-CCF = 8.28 | | | MAPEPIT-CCF = 5.34 | |

**Model validation result: LU & Inactive model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TTC-CCF rating 1** | | **TTC-CCF rating 2** | | |
|  | |  | | |
| MAPETTC-CCF = 8.38 | MAPEPIT-CCF = 8.39 | MAPETTC-CCF = 8.17 | | MAPEPIT-CCF = 7.43 |
|  |  |  | |  |
| **TTC-CCF rating 3** | | **TTC-CCF rating 4** | | |
|  | |  | | |
| MAPETTC-CCF = 6.94 | MAPEPIT-CCF = 6.21 | MAPETTC-CCF = 7.78 | | MAPEPIT-CCF = 6.51 |
|  | | | | |
| **TTC-CCF rating 5** | | | | |
|  | | | | |
| MAPETTC-CCF = 10.30 | | | MAPEPIT-CCF = 9.68 | |

### PIT Model

There is no model pre-validation for Point-in-Time CCF model.

# Loss Given Default

Loss given default, LGD, can be defined as the share of a defaulted exposure that will never be recovered by the lenders. The loss given default shall be assessed in an economic sense rather than a mere accounting perspective. That said the discount effect associated with the recovery cash flow and cost associated with collecting recoveries shall be considered.

LGD amount is the total EAD amount subtracted by the expected recovery amount and plus by the collection cost. Recoveries must include both late payments and the sale price of collateral when liquidated. The discount rate used should be the cost of equity, and recoveries should be discounted back to the date of the first missed payment leading to default. Then, to get the % LGD, the LGD amount is divided by the total EAD amount as shown in the equation below,

Where,

= LGD of customer *i* at default time *t*

= recovered cash flows and predicted cash flows of customer *i* at time *j*

= present value function using appropriate discount rate

= direct and indirect costs associated with the collection process

= outstanding of customer *i* at the time of default

Recovery cash flow is the amount of loan or obligation that will possibly be repaid to creditors in the event of a default then deducted by the direct and indirect costs associated with the collection process.

For the formula of %LGD calculation as above will be applied to both customers whose cash flow process has already finished and those with unfinished cash flow process. With unfinished cash flow process customers, expected recovery cash flow was therefore required to calculate and aggregate with the actual cash flow (see more details in section 6.5).

Then post-default state of customer will be defined which can be separated into 4 pathways as following:

|  |  |
| --- | --- |
| Type of Default | Description |
| 1. Turn Good | Customers who, after defaulted for 3 months, the BOT class for the next 6 months is monitored to be 1 and the current BOT class is 1. |
| 2. Restructure | Customers who need restructured process, consisting of four subgroups |
| 2.1 Success | Customers who, after restructured for 3 months and is already NPL, the BOT class for the next 12 months is monitored to be 1 and the current BOT class is 1. |
| 2.2 Not Success | Customers who, after restructured for 3 months and is already NPL, the BOT class for the next 12 months is monitored to be more than or equal 2. |
| 2.3 Finished | Customers who restructured and complete the recovery process |
| 2.4 Litigation | Customers who restructured but is already litigated |
| 2.5 Others | Customers who restructured and is NPL for less than 15 months |
| 3. Litigation | Customers who need litigation process |
| 4. Unclassified | Customers who are not in the first five groups |

Table 40, Pathways and definitions

## Scope

Through-the-cycle LGD will be applied identically in all asset classes in credit card portfolio.

## Methodology Review

#### **Constant Through-the-Cycle LGD**

The constant Through-the-Cycle LGD is calculated as the average of historical actual LGD of all customers in credit card portfolio.

#### **Scorecard for Through-the-Cycle LGD**

Scorecard for loss given default represents different characteristics of Kbank customer’s credit card demographics and spending behaviors (age of customer, occupation, EAD percentage and so on).

Kbank implements this scorecard model to convert customer’s credit card characteristic factors into LGD scores and then assign LGD values to these scores.

An example of KBank’s LGD scorecard is shown in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Characteristics |  | Attributes | Score | %Population |
|  | Constant |  |  |  |  |
| 1 | EAD percentage (%) |  |  |  |  |
| 2 | Occupation |  |  |  |  |
| 3 | Age of customer |  |  |  |  |

Table 41, Example of KBank’s characteristic factors

#### **Point-in-Time (PIT) LGD**

To comply with IFRS9, Kbank develops a Point-in-time LGD model that takes into account cyclical effects or macroeconomic factors. The appropriate macroeconomic factors will be selected using variable analysis and the logistic regression will be performed to find the specific value of coefficient (beta) for each macroeconomics factor.

To properly handle our model, these assumptions are made,

1. Kbank assumes our historical LGD can reflect future LGD

2. Kbank use constant Through-the-Cycle LGD model as a champion model.

## Model Development Approach

Development sample was prepared through the process of data cleansing in the first step. The group sample needed to match with relevant cash flow table source which was later on aggregated into a customer level. The following data was then used to calculate LGD prior to any validation and reporting.

## Data Management

### Overview of Input Data Set

The input for actual LGD calculation includes outstanding balance at default, state of default, final flag and actual cumulative recovery amount. The sum of the present value (PV) of these cash flows amounted to the accumulated recovery amount which is then used to calculate the value of LGD. The state of default indicate a posterior state of action after the customer has defaulted. The final flag indicates whether there is an on-going collection attempt and thus whether the state indicated are final. In summary, the data requirement for the calculation of actual LGD is shown as follows:

Figure 14, Data required for actual LGD calculation

### Sample Design

To calculate the LGD value of each customer, the recovery cash flow is collected from the first day that the customer becomes default until the most recent data (Dec 2016). In addition, to estimate robust models, the observation period should cover the whole business cycle which is usually considered between 3 to 5 years.

|  |  |  |
| --- | --- | --- |
| Period | Timeframe | Data |
| Observation Period | Jan 2002 – Aug 2016 | Default customers information |
| Recovery Period | Jan 2002 – Dec 2016 | Recovery cash flow of default customers |

Table 42, Observation and recovery period for recovery curve

To develop a LGD scorecard, the most recent demographic characteristics and the behavior characteristics at one month prior to default of the customers is analyzed together with the LGD value calculated from the recovery cash flow. However, we do not have the factors as old as the default customers’ information so the early ones are excluded. Note that the behavior characteristics of default customers are captured at one month prior to default because once the customer becomes default; no more behavior factors are collected.

|  |  |  |
| --- | --- | --- |
| Period | Timeframe | Data |
| Observation Period | Feb 2007 – Aug 2016 | Default customers information |
| Recovery Period | Feb 2007 – Dec 2016 | Recovery cash flow of default customers |

Table 43, Observation and recovery period for score development

### Data Cleansing and Exception Handling

#### **Exclusion Rules**

The accounts that meet the exclusion rules will not be included in the data sample for the scorecard development. The exclusion status is based on the customer status at the observation point.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Exclusion Rules | # Customers | Proportion |
|  | **Total Population** | **319,708** | **100.0%** |
| 1 | Not default customers | 47,620 | 14.9% |
| 2 | Not CC only customers | 2,632 | 0.8% |
| 3 | Default outstanding <= 2,000 | 5,657 | 1.8% |
| 4 | Outlier cash flow | 11,772 | 3.7% |
| 5 | Partial recovery but has 100% recovery | 6,049 | 1.9% |
|  | **Population for recovery curve and forecasting** | **245,978** | **76.9%** |
| 6 | Incomplete factors for score development | 64,017 | 20.0% |
| *6.1* | *- All factors are missing value* | *63,141* | *19.7%* |
| *6.2* | *- Credit line is missing value* | *876* | *0.3%* |
|  | **Population for LGD score development** | **181,961** | **56.9%** |

Table 44, Eligible population after exclusions

Additional notes for each exclusion rules:

1. Only default customers are selected in order to observe the actual recovery of each customer.
2. KBank selected customers who only hold credit card product and no other products because the cash flow in the RI database is stored at a customer level and it cannot be identified whether the input cash flow is to pay the debt for which product.
3. Customers with default outstanding less than 2,000 baht are excluded because it is likely that the input cash flow will cover only the interest of their debt and not the actual debt.
4. Outlier cash flow refers to customers whose recovery rate shoots up abnormally fast without being written off.
5. Customers who are flagged to still have a partial recovery but already have a recovery rate of 100% just from the discounted existing cash flow is strange so they are excluded.

#### **Exception Handling**

##### For recovery curve and forecasting

Firstly, each customer is flagged with a default pathway according to the definition in Table 40. Then, the customers who meet the condition 1 through 5 in Table 44 are excluded. In addition, it is noted that the collected cash flow does not include the amount of write-off that occurs and because the write-off process (selling NPL accounts) is done in a batch manner, the average recovery rate of each write-off batch is calculated. Though, KBank does not have the write-off amount for each customer, there is still a file containing all the customers who are written off and can be used to flag the SELL\_FLAG variable in

Table 45.

|  |  |  |
| --- | --- | --- |
| No | Variables | Description |
| 1 | TYPE | Default pathway |
| 2 | FST\_OUTS | Default outstanding |
| 3 | FINAL\_FLAG | Flag for complete (1) or incomplete (0) recovery  **Rules:**  if customer is in write-off file then **1**  else **FINAL\_FLAG** (keeps original value) |
| 4 | FST\_NPL\_DATE | First default date |
| 5 | BEF\_FINAL\_DATE | Date of complete recovery stored in BE format (if the customer is still in the collection process, this will be 0)  **Rules:**  if customer is in write-off file then **WRITEOFF\_DATE**  else if BEF\_FINAL\_DATE = 0 then **0**  else **BEF\_FINAL\_DATE – 54300** (convert to AD) |
| 6 | SELL\_FLAG | Flag for write-off (1) or not write-off (0)  **Rules:**  if customer is in write-off file then **1**  else **0** |
| 7 | CASH\_FLOW\_yyyymm | Cash flow amount for each month |

Table 45, Example variables for calculation of recovery rate

Note. For all variables will be shown in an [Appendix C](#_C._Full_data).

In addition, the cash flow at each month for each customer is cleaned in the following manners:

1. Negative cash flow is changed to zero.
2. If the cash flow month is not in between the FST\_NPL\_DATE and BEF\_FINAL\_DATE, they are changed to zero. There is a possibility that the customer gets better along the way and the bank decided to approve for additional loan where in this case the cash flow would be referred to the customer’s usual payment and the recovery rate might be over-estimated.
3. If the customer is in write-off file, the average recovery rate of the write-off batch that the customer belongs to is used to multiply with the remaining outstanding and put in as cash flow on the customer’s write-off date.

##### For LGD score development

Table 45 shows a total of 43 independent variables for LGD score model, consisting of 1 default utilization (No.1), 25 behavioral variables (No.2-26) and 17 demographic variables (No.26-43). They are generated on a monthly basis to support the data mapping where for each default customer, the default utilization and behavior variables at 1 month prior to the default month is collected together with the most recent value of the demographic variables. In addition, as the original data of the behavior variables are collected at account level, they are aggregated into customer level on risk-average concept as stated in

Table 45. It is also noted that some variables need special value handling as listed in Table 46 and this should be performed before data aggregation.

|  |  |  |
| --- | --- | --- |
| No | Variables | Description |
| 3 | AGE\_OF\_ACC  (VAR\_3) | Months on book for each account  **Special Handling:**  if VAR\_3 is missing then **1000**  **Data Aggregation:**  Min(VAR\_3) |
| 4 | DAYS\_SINCE\_LAST\_PURCHASE  (VAR\_4) | Number of days since last purchase  **Special Handling:**  if VAR\_4 is missing then -**1000**  else if VAR\_4 = **999** then -**999**  Data Aggregation:  Max(VAR\_4) |

Table 46, Example variables for LGD score development

After all the data variables are assembled, missing values of each variable is investigated in order to filter out the one that has too many missing values. Therefore, the demographic variables from 35 through 43 are excluded since they contain more than 40 percent of missing values as shown in

Table 47.

|  |  |  |
| --- | --- | --- |
| Factors | Number of Missing | Percent of Missing |
| RESIDENTIAL\_STATUS | 117,798 | 47.89% |
| EMPLOYMENT\_PERIOD | 117,173 | 47.64% |
| EMPLOYMENT\_PHONE\_FLAG | 245,978 | 100.00% |
| TOTAL\_EMPLOYMENT\_PERIOD | 117,173 | 47.64% |
| VERIFIED\_INCOME | 117,376 | 47.72% |
| MAILING\_ADDRESS | 117,801 | 47.89% |
| PAYMENT\_METHOD | 162,772 | 66.17% |
| FICO\_SCORE | 164,904 | 67.04% |
| DEBT\_BURDEN | 233,915 | 95.10% |

Table 47, Number and percent of missing records

Furthermore, in order to perform divergence technique, a good/bad indicator is necessary as the response variable of the model. Therefore, the development dataset is converted from a continuous target of LGD value to a binary target. The customers with LGD value apart from 0 or 1 are split into 2 cases, with the same explanatory variables – one with a response variable of good (1) and a weight of 1-LGD, the other with a response of bad (0) and a weight of LGD as shown in Figure 8. For customers with LGD value of 1, a response variable is assigned to 0 and customers with LGD value of 0, a response variable is assigned to 1.

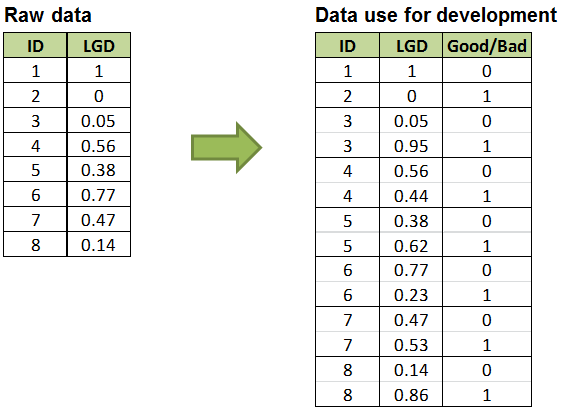
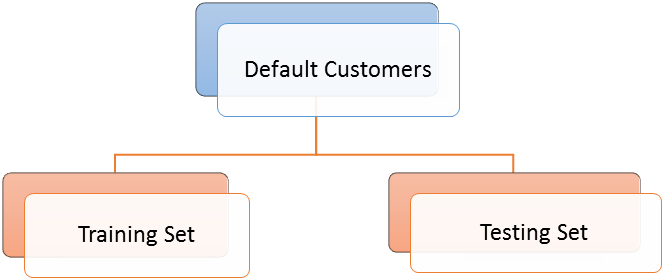


Figure 15, Conversion of continuous target to binary target

### Final Development Sample

For calculating the recovery curve, the sample data is not segmented. However, for LGD score development, the default customer data is separated into 2 sets; training set (80%) and testing set (20%) as shown below



181,961

145,568

36,393

80%

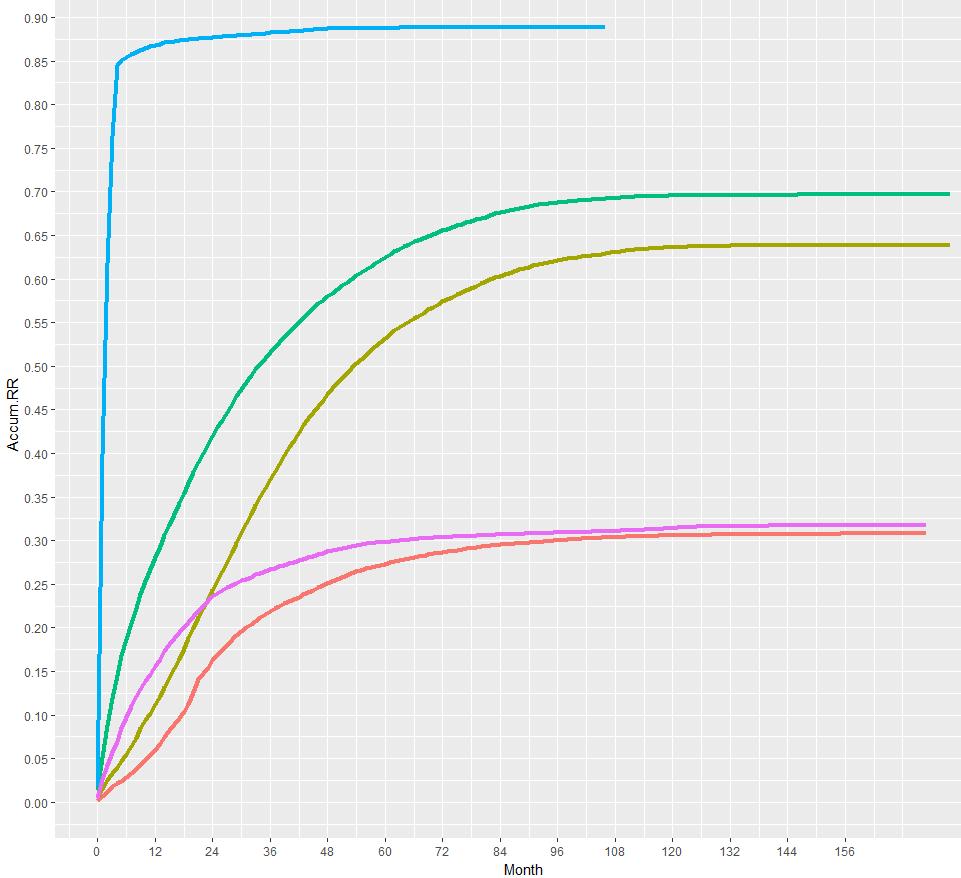
20%

## Model Development

After customers default, Kbank will collect monthly cash flow from the customers, once KBank has the monthly cash flow for each default customer, the monthly present value of each cash flow is calculated with a discount rate of 20% per year for each customer. Then, the customers are separated into two groups; closed customers and non-closed customers.

Closed customers are customers who have a complete recovery data or have fully recovered their debt. Non-closed customers are those who have partial recovery data and are still in the collection process.

For closed customer, a recovery curve is plotted in accordance to the number of months after default for each default pathway as shown in Figure 16



Turn Good

Litigation and Restructure

Restructure

Litigation

Unclassified

Figure 16, Recovery curve of each default pathway

In addition, for closed customers, two pieces of information is observed which are stated below together with the value of them in Table 48.

1. Flat period: The number of months after default that customers in each default pathway will reach a stable period, a period where there would likely be no more recovery afterward.
2. The accumulated recovery rate at that flat period.

|  |  |  |
| --- | --- | --- |
| Default Pathway | Flat Period | Recovery Rate @ Flat Period |
| Litigation | 148 |  |
| Litigation and Restructure | 130 |  |
| Restructure | 128 |  |
| Turn Good | 60 |  |
| Unclassified | 146 |  |

Table 48,Flat period and its recovery rate

Lastly, the recovery rate at flat period of each default pathway is used to forecast the final recovery amount of each non-closed customer using the equation below:

Where Potential recovery = RR at Flat period – RR at Last cash flow period. To calculate the final recovery rate, the final recovery amount is simply divided by the initial default outstanding. The example for forecasting the final recovery amount is shown in Figure 17, Example for forecasting final recovery.

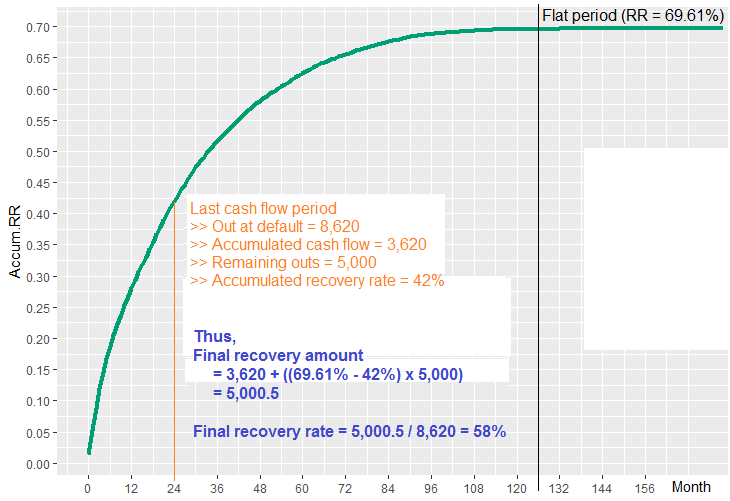


Figure 17, Example for forecasting final recovery

It is noted that the final recovery amount of a non-closed customer whose default period is already more than the flat period is not forecasted. Moreover, the final recovery rate is capped at 100%. Once the forecasting process is done, the data for both closed and non-closed customers are combined for further data preparation in the next section. The LGD value is also calculated by subtracted the recovery rate from 1 (LGD = 1-RR).

### TTC Model

#### **Characteristic Analysis**

To analyze the strength of a variable in the ability to separate good customers from bad customers, the following criteria are considered:

1. Weight of Evidence (WOE)

This measures the predictive power of each attribute. The higher the WOE, the lower risk the customer goes bad and vice versa.

WOE = ln(Distribution Good / Distribution Bad)

1. Information Value (IV)

This measures the predictive power of each characteristic with the rule of thumb as shown in Table 49.

IV = Sum [(Distribution Goodi – Distribution Badi) \* WOEi]

|  |  |
| --- | --- |
| Information Value | Predictive Power |
| < 0.02 | Useless for prediction |
| 0.02 – 0.1 | Weak predictor |
| > 0.1 – 0.3 | Medium predictor |
| > 0.3 – 0.5 | Strong predictor |

Table 49, Rule of thumb for information value

1. Business Definition

To identify the trend of WOE value of attributes in each characteristic, the following things should be considered in binning the characteristic.

* Each attribute should have a population of at least 5%.
* There should not be an attribute which has good customers of zero or bad customers of zero.
* For any characteristics, WOE of each attribute should differ significantly.

The result of characteristic analysis for LGD scorecard model is shown in

Table 50.

#### **Multivariate Regression Modeling for LGD Scorecard**

KBank uses FICO Model Builder as a tool for developing LGD scorecard.

#### **Final Scorecard**

| **No.** | **Characteristics** | **Attributes** | **Score** | **%Population** |
| --- | --- | --- | --- | --- |
|  | Constant |  |  |  |
| 1 | EAD percentage (%) |  |  |  |
| 2 | Occupation |  |  |  |
| 3 | Age of customer |  |  |  |
| 4 | Education |  |  |  |
| 5 | Average payment to previous |  |  |  |
|  | balance in previous 6 months (%) |  |  |  |
| 6 | Age of account (months) |  |  |  |

Table 50, LGD scorecard

#### **LGD Master Scale**

After getting the final scorecard, KBank divided the score range into 5 grades for the convenience in using these LGD grade for strategic decision and portfolio management.

|  |  |  |
| --- | --- | --- |
| LGD Grade | LGD Score Range | %LGD |
| 1 | < 475 |  |
| 2 | 475 – 484 |  |
| 3 | 485 – 489 |  |
| 4 | 490 – 499 |  |
| 5 | >= 500 |  |

Table 51,LGD score and LGD grade mapping

**Final Constant LGD**

#### The constant LGD will be the average of all actual LGD across all developments sample which resulted in a constant value of 0.6813.

Kbank then uses a constant LGD model as a champion model.

### PIT Model

After the LGD master scale is developed, it will be used to develop an econometric model to estimate the PIT LGD. The LGD grade is adjusted against the uncertainty in the future, taking into consideration the factors that may have an effect on the payment ability and loan loss recovery. Macroeconomic factors are assumed to be exogenous to the model and the ones that are possibly related to credit card are listed in Table 52 together with their expected signs.

|  |  |  |
| --- | --- | --- |
| No. | Macroeconomic Factors | Expected Sign |
| 1 | GDP | - |
| 2 | Unemployment Rate | + |
| 3 | Household Debt | + |
| 4 | CCI | - |

Table 52, List of macroeconomic factors and their expected signs

To analyze strength of independent variables in explanatory power or ability to predict response variable, the following criteria should be considered:

* 1. Regression coefficients and p-values

The p-value for each term of independent variable tests the null hypothesis that the coefficient is equal to zero (no effect on response variable). A low p-value (< 0.05) indicates rejecting the null hypothesis. In other words, an independent variable with low p-value is likely to be a meaningful addition to the model because change in the independent variable’s value is associated with change in response variable’s value.

* 1. Adjusted R-squared

The adjusted R-squared is used to compare the explanatory power of regression models that contain different numbers of independent variable. It is a modified version of R-squared which increases only if the new independent variable improves the performance of model more than would be expected by chance.

* 1. Appropriate lagging-time of macroeconomic factor

Kbank develops several models with different lagging macroeconomic factors to improve performance of the models. **Error! Reference source not found.** illustrates the way to find lagging macroeconomic factors.

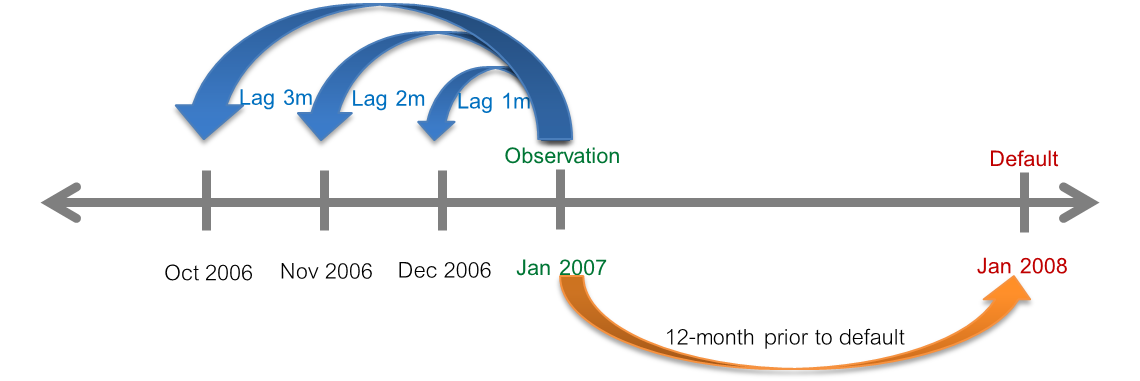
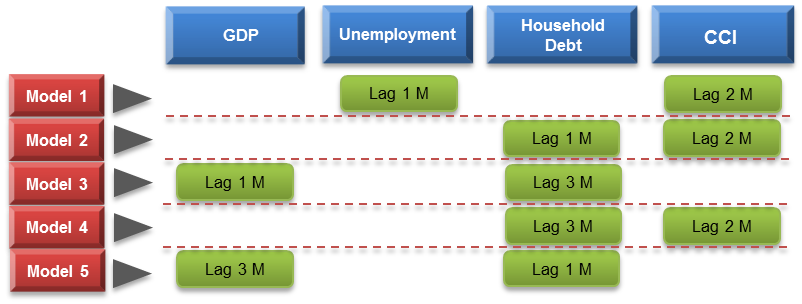
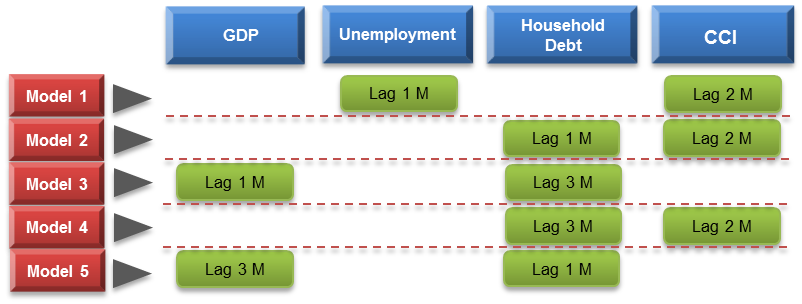


Figure 18, Lagging macroeconomic factors



*LGDPITt = β0 + β1Macro1t +… + βnMacront + εi*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. | TTC LGD Range | β0 | β1 | β2 | β3 | β4 |
| 1 | < 475 |  | NA |  | NA |  |
| 2 | 475 – 484 |  | NA | NA |  |  |
| 3 | 485 – 489 |  |  | NA |  | NA |
| 4 | 490 – 499 |  | NA | NA |  |  |
| 5 | >= 500 |  |  | NA |  | NA |

PIT-LGD is derived from doing a regression on TTC-LGD together with macroeconomic factors that affect consumption of retail credit card customers

## Pre-Validation

### TTC Model

KBank uses the Gini coefficient of the training and testing data set to measure the strength or efficiency of the scorecard. It measures how effectively the scorecard separates the goods from the bads. Table 53 shows the Gini coefficient of both the training and testing data set of LGD scorecard.

|  |  |
| --- | --- |
| Sample | Gini |
| Training set (80%) | 10-30% |
| Testing set (20%) | 10-30% |

Table 53, Gini coefficient of LGD scorecard

In addition, since KBank uses FICO Model Builder for developing the scorecard, ROC curve is also generated for the model as shown in Figure 19.

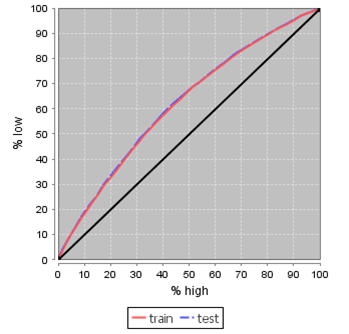


Figure 19, ROC curve of LGD scorecard

### PIT Model

There is no pre-validation for PIT model.

# Criteria for a Significant Increase in Credit Risk (SICR)

With the new IFRS9 loan loss provision, loans are classified in three stages: stage 1 – initial recognition (yet to be impaired), stage 2 – significant increase in credit risk, and stage 3 – objective indicators of impairment. For loans in stage 1, banks need to estimate 1-year expected credit losses. On the other hand for loans in stage 2 and 3, banks need to provide provision and thus estimate expected credit losses for the whole lifetime of the loans. Both the 1-year and lifetime expected credit losses estimation shall reflect the banks’ forward looking macro-economic view.

The criteria for significant increase in credit risk (stage transfer criteria) will be used for the classification of loans between stage 1 and stage 2. KBank employed both the qualitative and the quantitative criteria for stage classification. An example of qualitative stage transfer criteria stage is the 30+ days past due (DPD), fraud, black list from the revenue department and so on. Thus, all instruments with DPD>30 days are automatically classified as stage 2. This model development document solely focuses on the formulation of the quantitative stage transfer criteria.

This section of the document outlines the methodology review and the development including any expert opinions and judgements of our champion significant increase in credit risk criteria. KBank explore four methodologies to develop the quantitative criteria the significant increase in credit risk namely the rating downgrade criteria, the remaining lifetime PD criteria, the forward run test and the high credit risk region. In each of the methodology, KBank leverages existing behavioral scorecard, supermaster rating transition matrix and PD term structure to induce statistical inference and the formulation of the criteria itself.

## Scope

Both the qualitative and quantitative stage transfer criteria are applied to all non-NPL instruments. All non-performing loan instruments are automatically assigned to stage 3.

## Modelling Methodology Review

### Rating Downgrade

The rating downgrade is a quantitative stage transfer criteria that contemplate on the rating downgrade compared to the rating assigned at the loan origination. If the number of rating downgraded of an instrument is greater than a given threshold then the instrument is deemed to be a significant increase in credit risk and assigned to stage 2. The number of rating downgrade can be identified illustratively in the figure below.

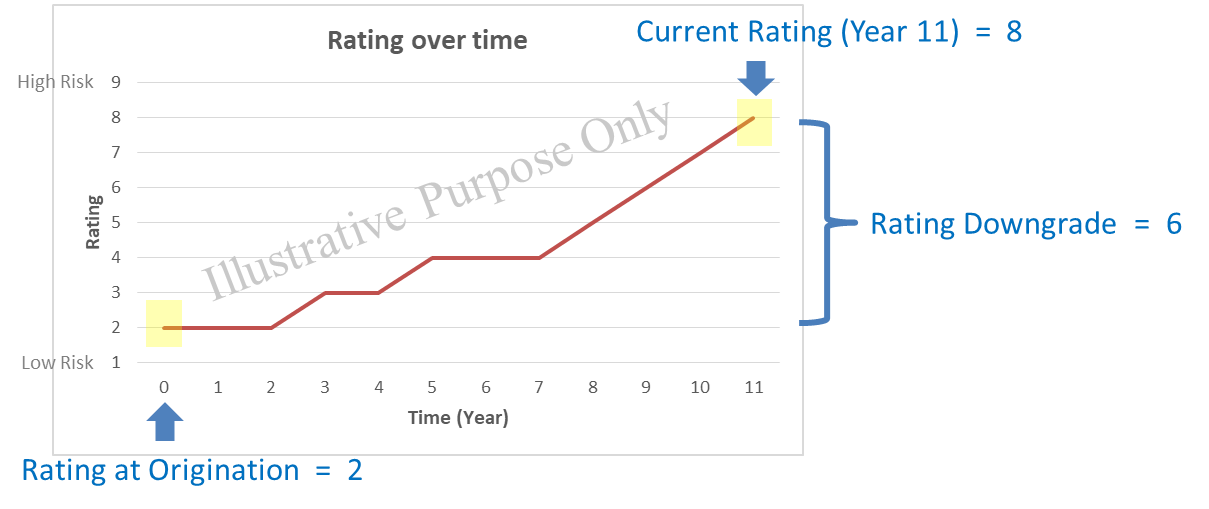


Figure 20, Rating observation: origination vs. reporting

For retail portfolio, we use an empirical rating migration from bootstrapped samples to construct the threshold for notch change (between 75-85 percentiles depending on the portfolio).

### Remaining Lifetime Probability of Default

The remaining Lifetime PD stage transfer criteria recognize if the remaining lifetime probability of default at the reporting date has deviated statistically from the lifetime probability of default estimated at origination. This measure is highly sensitive to our choice of PD term structure model. By default this compares the lifetime PD generated from the champion model of each portfolio.

In short, the significant increase in credit risk under this criteria is SICR is reflected by statistical deviation in **[1] the remaining Lifetime PD estimated at origination** compared to **[2] Lifetime PD at reporting date**. The statistical deviation in this criteria is founded on a confidence band based on the average deviations in the marginal probabilities over time (for each supermaster rating). The standard deviation for our confidence band is illustratively measured as per the following figure

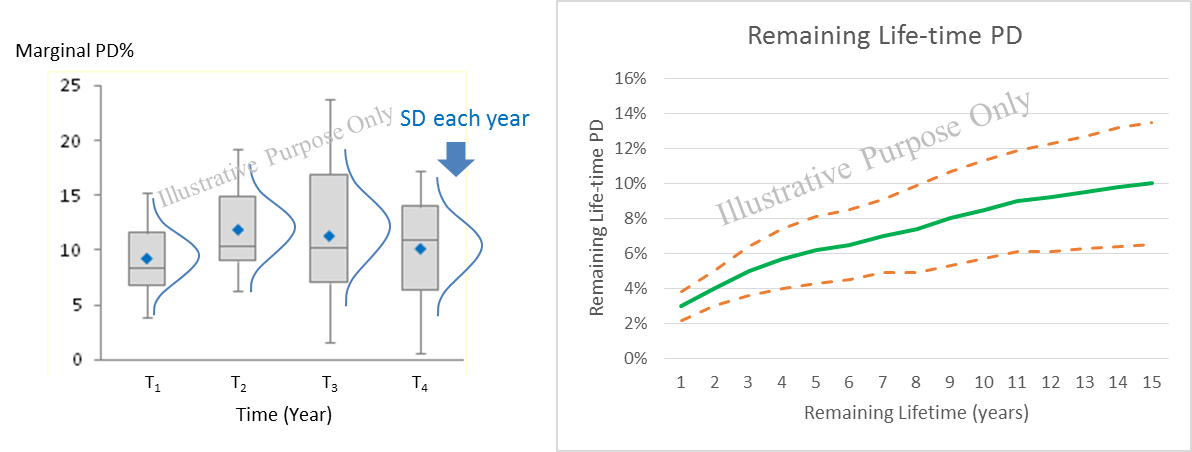


Figure 21, Standard deviation and confidence band for Lifetime PD

After the average standard deviation of marginal probabilities has been defined . We generate the confidence band for each of the remaining lifetime as follow

Where is the average standard deviation of the mariginal PD for rating

is the cumulative PD at estimated origination (origination rating ) upto maturity date

is z-value of a confindence level (by default 80%, two-tailed)

### Forward Run Test

The forward run test criteria check if the Forward PD at reporting date has significantly increased from the Forward PD at the similar point of time estimated at origination. If we were to plot forward PDs of two time series [1] Forward PD from origination and [2] Forward PD at reporting date, we can compare the two time series using run test to see if they are statistically different as illustrated in the figure below.

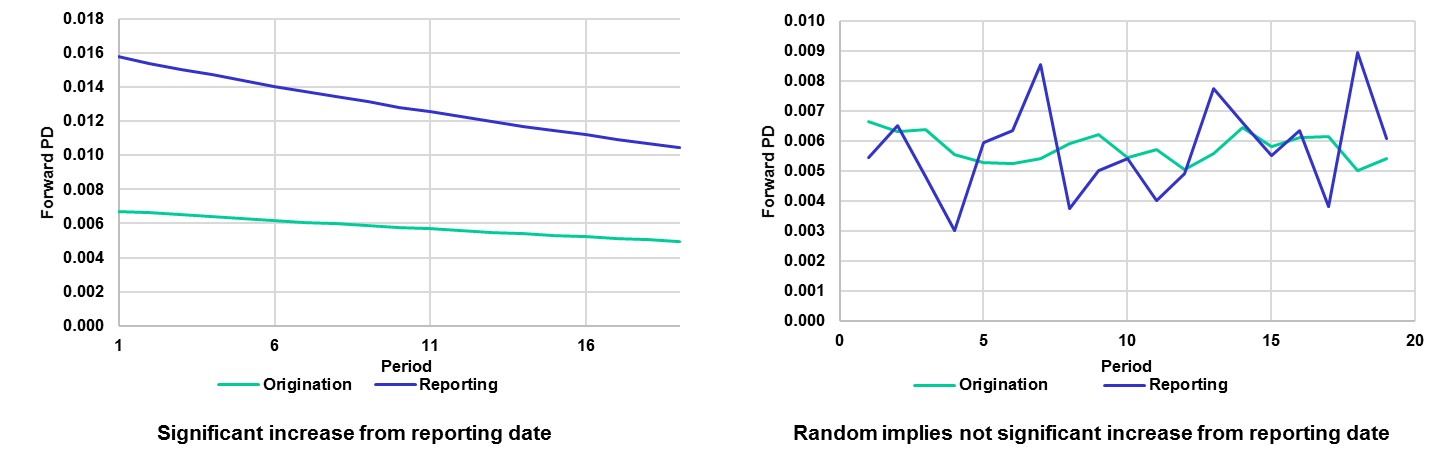


Figure 22, forward PD run test

The hypothesis of the forward PD run test is defined as follow

*H0: Forward PDs at origination date and at reporting date are not different*

Test statistic

Where is the observed number of runs

are the number of origination and reporting date values

### High Credit Risk Region

The high credit risk region is our final significant increase in credit risk criteria which may be use in conjunction with other criteria. The high credit risk region is defined as a supermaster rating threshold. If a credit rating of an instrument is greater than the prespecified value then it is automatically assigned to stage-2.

The conception of the high credit risk region would start with an observation of the average rating of a portfolio. We’ll then use an associated rating downgrade threshold as the high credit risk region for the portfolio. The setting is further validated by testing the equivalent of survival curves i.e. between the low credit risk and high credit risk region.

The hypothesis of the equivalency of survival curve is defined as follow

*H0: The customers with credit rating higher than the HCRR threshold have equivalent survival curve to those customers with ratings better than the threshold.*

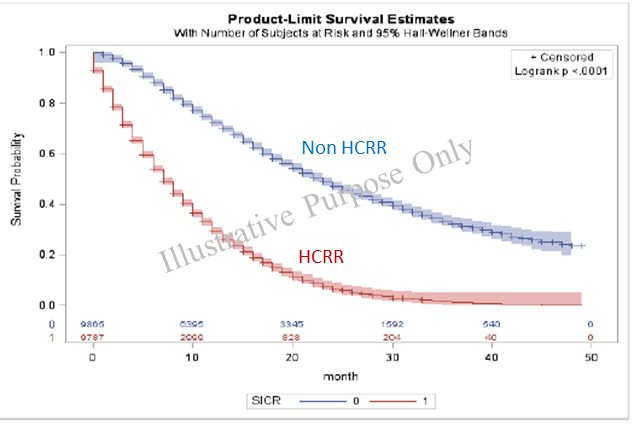


Figure 23, Equivalency of survival curves between HCRR and non-HCRR

## Model Development Approach

For significant increase in credit risk criteria, KBank believe that the rating downgrade and the high credit risk region are the appropriate methodologies to be used as a champion criteria. The main reason for this is the fact that the methodology is lower in complexity in terms of both the data requirement and system implementation. The input of the model and the development exercise aligns with the bank’s current data structure. The output of model also aligns with current business model and TFRS9 risk requirements in term of classification of instruments whose credit risk have increase significantly. Other methodologies are explored and implemented as challenger SICR criteria.

This section briefly outline the steps that have been taken in the development of the champion SICR criteria i.e. rating downgrade and high credit risk region. The development SICR model is done successively to the development of PD model and will be using an identical sampling data. Figure below outline the steps taken in the development of rating downgrade and high credit risk region.

Figure 24, Steps of development: stage transfer criteria

.

The development process starts with consolidation of portfolio data across time into a single data set (these are the same data set which are used in the PD model development). Then development samples are created by bootstrap sampling from the large data set. We then observe the 75 to 85 percentile of migration for each of the initial rating in each of the sample. The observed mode of the 75-85 migration percentiles (based on expert opinion) across multiple bootstrapped samples are then selected as a downgrade criteria.

For the high credit risk region, we observe the average supermaster rating of an entire portfolio in the data set then we use the downgrade criteria for that rating as a high credit risk region.

## Data Management

For data management, please refer to the Data Management section for the modelling of the probability of default term structure.

## Model Development

### Rating Downgrade

The development of rating downgrade criteria starts with the bootstrapped sampling to observe the rating (supermaster) migrations. For each initial rating, we then observe the distribution of the 12-month rating change as per the following figure.

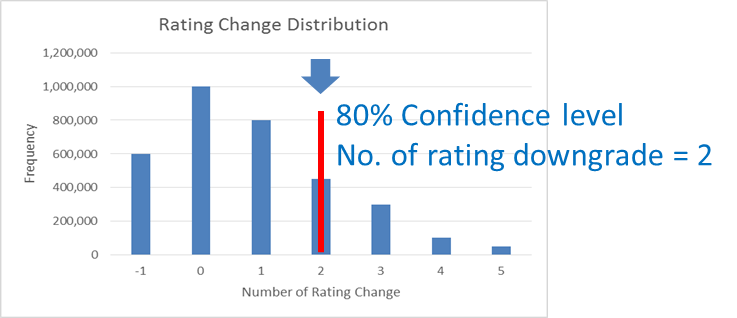


Figure 25

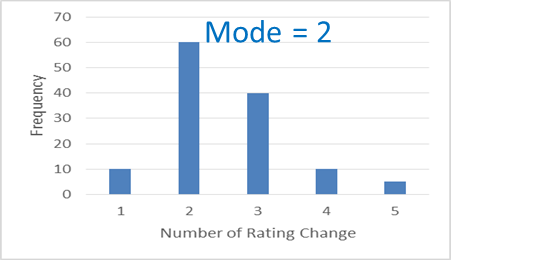


Figure 26, Rating change distribution and mode of bootstrapped sample

The steps are repeated across multiple (100) bootstrapped samples to observe (100) of 80th percentile of the 12-month rating change. The mode of these 80th percentile numbers is then chosen as downgrade criteria. The result of this exercise is the number of downgrade rating change for each of the initial (origination) rating. The result is then overridden with few expert judgement to maintian the monotonicity of the downgrade for each of the initial rating. The results are as follow

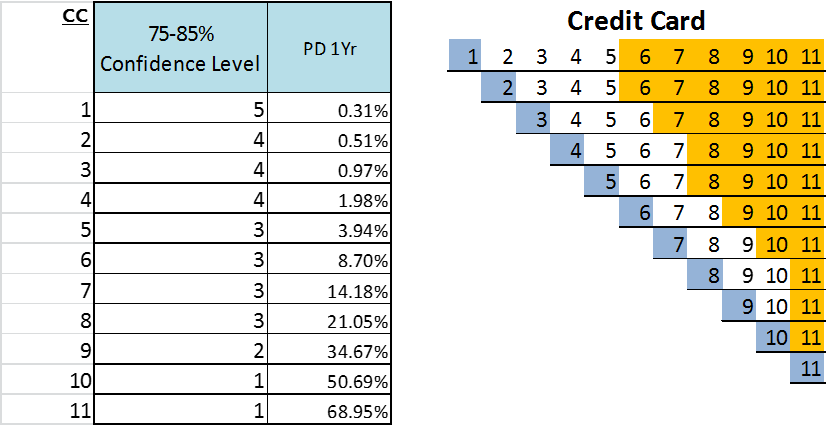


Figure 27, Final rating downgrade criteria

### High Credit Risk Region

As mentioned previously, the high credit risk region is essentially the rating downgrade (change) of the average portfolio rating which in this case is rating 4. The high credit risk region is thus rating 8 or greater. The distribution of customer that falls under and in the high credit risk region are shown in the table below.

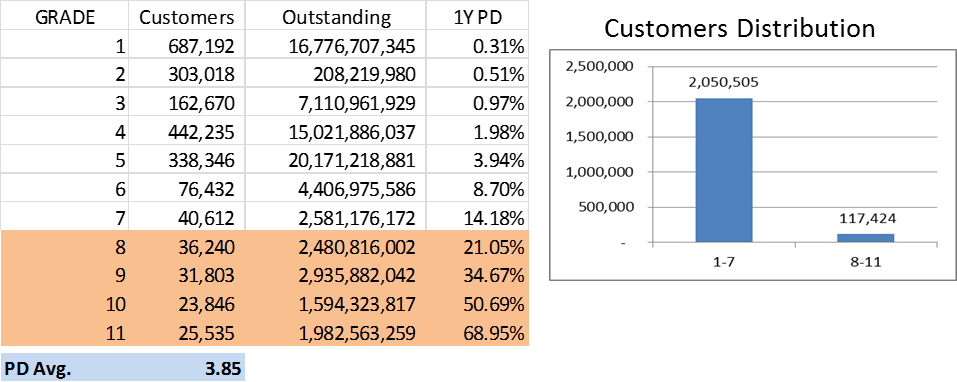


Figure 28, Final high credit risk region criteria

We validate this setting by considering the survival curves of each segment i.e. low credit risk region vs. high credit risk region. The validation results show that the survival curves are vastly different and the high credit risk region setting is appropriate. Results of validation are shown below.

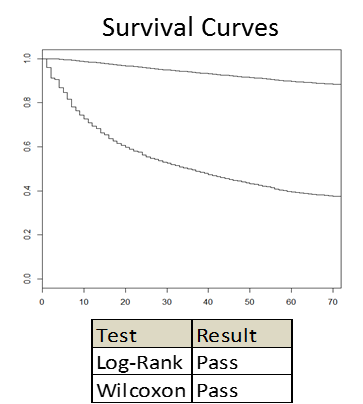


Figure 29, Validation of HCRR, equivalency of survival curves

# Appendix

### A. Full reassignment of CCF score card characteristic factors

BLOCK\_CODE

|  |  |  |  |
| --- | --- | --- | --- |
| if | BLOCK\_CODE = ‘W’ | then | BLOCK\_CODE = 1 |
| else if | BLOCK\_CODE = ‘A’ | then | BLOCK\_CODE = 2 |
| else if | BLOCK\_CODE = ‘Z’ | then | BLOCK\_CODE = 3 |
| else if | BLOCK\_CODE = ‘B’ | then | BLOCK\_CODE = 4 |
| else if | BLOCK\_CODE = ‘K’ | then | BLOCK\_CODE = 5 |
| else if | BLOCK\_CODE = ‘M’ | then | BLOCK\_CODE = 6 |
| else if | BLOCK\_CODE = ‘O’ | then | BLOCK\_CODE = 7 |
| else if | BLOCK\_CODE = ‘I’ | then | BLOCK\_CODE = 8 |
| else if | BLOCK\_CODE = ‘Y’ | then | BLOCK\_CODE = 9 |
| else if | BLOCK\_CODE = ‘R’ | then | BLOCK\_CODE = 10 |
| else if | BLOCK\_CODE = ‘T’ | then | BLOCK\_CODE = 11 |
| else if | BLOCK\_CODE = ‘H’ | then | BLOCK\_CODE = 12 |
| else if | BLOCK\_CODE = ‘L’ | then | BLOCK\_CODE = 13 |
| else if | BLOCK\_CODE = ‘E’ | then | BLOCK\_CODE = 14 |
| else if | BLOCK\_CODE = ‘F’ | then | BLOCK\_CODE = 15 |
| else if | BLOCK\_CODE = ‘S’ | then | BLOCK\_CODE = 16 |
| else if | BLOCK\_CODE = ‘G’ | then | BLOCK\_CODE = 17 |
| else if | BLOCK\_CODE = ‘C’ | then | BLOCK\_CODE = 18 |
| else if | BLOCK\_CODE = ‘P’ | then | BLOCK\_CODE = 19 |
| else if | BLOCK\_CODE = ‘Q’ | then | BLOCK\_CODE = 20 |
| else if | BLOCK\_CODE = ‘U’ | then | BLOCK\_CODE = 21 |
| else if | BLOCK\_CODE = ‘X’ | then | BLOCK\_CODE = 22 |
| else if | BLOCK\_CODE = ‘J’ | then | BLOCK\_CODE = 23 |
| else if | BLOCK\_CODE = ‘V’ | then | BLOCK\_CODE = 24 |
| else if | BLOCK\_CODE = ‘N’ | then | BLOCK\_CODE = 25 |
| else if | BLOCK\_CODE = ‘D’ | then | BLOCK\_CODE = 26 |
| else if | (BLOCK\_CODE = ‘-1’) or  (BLOCK\_CODE is missing) | then | BLOCK\_CODE = 27 |
| else |  |  | BLOCK\_CODE = 99 |

AGE\_OF\_ACC

|  |  |  |  |
| --- | --- | --- | --- |
| if | AGE\_OF\_ACC is missing | then | AGE\_OF\_ACC = 999 |

DAYS\_SINCE\_LAST\_PURCHASE

|  |  |
| --- | --- |
| if | (DAYS\_SINCE\_LAST\_PURCHASE is missing) or (DAYS\_SINCE\_LAST\_PURCHASE = 999) |
|  | then DAYS\_SINCE\_LAST\_PURCHASE = -999 |

BAL\_TO\_LIMIT\_AVG\_3M

|  |  |
| --- | --- |
| if | (BAL\_TO\_LIMIT\_AVG\_3M is missing) or (BAL\_TO\_LIMIT\_AVG\_3M = 999) |
|  | then BAL\_TO\_LIMIT\_AVG\_3M = -999 |
| else if | BAL\_TO\_LIMIT\_AVG\_3M < 0 |
|  | then BAL\_TO\_LIMIT\_AVG\_3M = 0 |
| else if | BAL\_TO\_LIMIT\_AVG\_3M > 100 |
|  | then BAL\_TO\_LIMIT\_AVG\_3M = 100 |

CASH\_ADV\_TO\_LIMIT\_AVG\_3M

|  |  |
| --- | --- |
| if | (CASH\_ADV\_TO\_LIMIT\_AVG\_3M is missing) or  (CASH\_ADV\_TO\_LIMIT\_AVG\_3M = 999) |
|  | then CASH\_ADV\_TO\_LIMIT\_AVG\_3M = -999 |
| else if | CASH\_ADV\_TO\_LIMIT\_AVG\_3M < 0 |
|  | then CASH\_ADV\_TO\_LIMIT\_AVG\_3M = 0 |
| else if | CASH\_ADV\_TO\_LIMIT\_AVG\_3M > 100 |
|  | then CASH\_ADV\_TO\_LIMIT\_AVG\_3M = 100 |

CASH\_ADV\_TO\_SPEND\_AVG\_6M

|  |  |
| --- | --- |
| if | (CASH\_ADV\_TO\_SPEND\_AVG\_6M is missing) or  (CASH\_ADV\_TO\_SPEND\_AVG\_6M = 999) |
|  | then CASH\_ADV\_TO\_SPEND\_AVG\_6M = -999 |
| else if | CASH\_ADV\_TO\_SPEND\_AVG\_6M < 0 |
|  | then CASH\_ADV\_TO\_SPEND\_AVG\_6M = 0 |
| else if | CASH\_ADV\_TO\_SPEND\_AVG\_6M > 100 |
|  | then CASH\_ADV\_TO\_SPEND\_AVG\_6M = 100 |

TIMES\_OVER\_LIMIT\_L3M

|  |  |
| --- | --- |
| if | TIMES\_OVER\_LIMIT\_L3M is missing |
|  | then TIMES\_OVER\_LIMIT\_L3M = -999 |

TIMES\_OVER\_LIMIT\_L6M

|  |  |
| --- | --- |
| if | TIMES\_OVER\_LIMIT\_L6M is missing |
|  | then TIMES\_OVER\_LIMIT\_L6M = -999 |

DAYS\_SINCE\_LAST\_PMT

|  |  |
| --- | --- |
| if | (DAYS\_SINCE\_LAST\_PMT is missing) or (DAYS\_SINCE\_LAST\_PMT = 999) |
|  | then DAYS\_SINCE\_LAST\_PMT = -999 |

PMT\_TO\_PREV\_BAL\_AVG\_3M

|  |  |
| --- | --- |
| if | PMT\_TO\_PREV\_BAL\_AVG\_3M is missing |
|  | then PMT\_TO\_PREV\_BAL\_AVG\_3M = 999 |
| else if | PMT\_TO\_PREV\_BAL\_AVG\_3M < 0 |
|  | then PMT\_TO\_PREV\_BAL\_AVG\_3M = 100 |
| else if | PMT\_TO\_PREV\_BAL\_AVG\_3M > 100 |
|  | then PMT\_TO\_PREV\_BAL\_AVG\_3M = 100 |

PMT\_TO\_PREV\_BAL\_AVG\_6M

|  |  |
| --- | --- |
| if | PMT\_TO\_PREV\_BAL\_AVG\_6M is missing |
|  | then PMT\_TO\_PREV\_BAL\_AVG\_6M = 999 |
| else if | PMT\_TO\_PREV\_BAL\_AVG\_6M < 0 |
|  | then PMT\_TO\_PREV\_BAL\_AVG\_6M = 100 |
| else if | PMT\_TO\_PREV\_BAL\_AVG\_6M > 100 |
|  | then PMT\_TO\_PREV\_BAL\_AVG\_6M = 100 |

TIMES\_FULL\_PMT\_L6M

|  |  |
| --- | --- |
| if | TIMES\_FULL\_PMT\_L6M is missing |
|  | then TIMES\_FULL\_PMT\_L6M = 999 |

TIMES\_PART\_PMT\_L6M

|  |  |
| --- | --- |
| if | TIMES\_PART\_PMT\_L6M months is missing |
|  | then TIMES\_PART\_PMT\_L6M months = 999 |

MONTHS\_SINCE\_LAST\_DELIN

|  |  |
| --- | --- |
| if | MONTHS\_SINCE\_LAST\_DELIN is missing |
|  | then MONTHS\_SINCE\_LAST\_DELIN = 999 |

TIMES\_X\_PLUS\_DPD\_L3M

|  |  |
| --- | --- |
| if | TIMES\_X\_PLUS\_DPD\_L3M is missing |
|  | then TIMES\_X\_PLUS\_DPD\_L3M = -999 |

TIMES\_X\_PLUS\_DPD\_L8M

|  |  |
| --- | --- |
| if | TIMES\_X\_PLUS\_DPD\_L8M is missing |
|  | then TIMES\_X\_PLUS\_DPD\_L8M = -999 |

TIMES\_GREATER\_30\_DPD\_L3M

|  |  |
| --- | --- |
| if | TIMES\_GREATER\_30\_DPD\_L3M is missing |
|  | then TIMES\_GREATER\_30\_DPD\_L3M = -999 |

TIMES\_GREATER\_30\_DPD\_L8M

|  |  |
| --- | --- |
| if | TIMES\_GREATER\_30\_DPD\_L8M is missing |
|  | then TIMES\_GREATER\_30\_DPD\_L8M = -999 |

TIMES\_GREATER\_60\_DPD\_L6M

|  |  |
| --- | --- |
| if | TIMES\_GREATER\_60\_DPD\_L6M is missing |
|  | then TIMES\_GREATER\_60\_DPD\_L6M = -999 |

TIMES\_60\_TO\_89\_DPD\_L6M

|  |  |
| --- | --- |
| if | TIMES\_60\_TO\_89\_DPD\_L6M is missing |
|  | then TIMES\_60\_TO\_89\_DPD\_L6M = -999 |

INACTIVE\_ACCOUNT\_L8M

|  |  |
| --- | --- |
| if | INACTIVE\_ACCOUNT\_L8M is missing |
|  | then INACTIVE\_ACCOUNT\_L8M = 999 |

Group handled data in section 7.1 by LPM\_NO and performs logic as follows:

* min(BLOCK\_CODE)
* min(AGE\_OF\_ACC)
* max(DAYS\_SINCE\_LAST\_PURCHASE)
* max(BAL\_TO\_LIMIT\_AVG\_3M)
* max(CASH\_ADV\_TO\_LIMIT\_AVG\_3M)
* max(CASH\_ADV\_TO\_SPEND\_AVG\_6M)
* max(TIMES\_OVER\_LIMIT\_L3M)
* max(TIMES\_OVER\_LIMIT\_L6M)
* max(DAYS\_SINCE\_LAST\_PMT)
* min(PMT\_TO\_PREV\_BAL\_AVG\_3M)
* min(PMT\_TO\_PREV\_BAL\_AVG\_6M)
* min(TIMES\_FULL\_PMT\_L6M)
* min(TIMES\_PART\_PMT\_L6M months)
* min(MONTHS\_SINCE\_LAST\_DELIN)
* max(TIMES\_X\_PLUS\_DPD\_L3M)
* max(TIMES\_X\_PLUS\_DPD\_L8M)
* max(TIMES\_GREATER\_30\_DPD\_L3M)
* max(TIMES\_GREATER\_30\_DPD\_L8M)
* max(TIMES\_GREATER\_60\_DPD\_L6M)
* max(TIMES\_60\_TO\_89\_DPD\_L6M)
* min(INACTIVE\_ACCOUNT\_L8M)

### B. Finalize Model for point-in-time CCF

After Kbank selects meaningful independent variables that identify economic environment prior to default time. The model is then developed by using R 3.3 software.

HU model Run file “1-EADmodel\_HU.R”.

LU & Inactive model Run file “2-EADmodel\_LU.R”

Note2: R Programming files are in path [\\a011122j\2016\04\_EAD\20161125\_PiT](file:///\\a011122j\2016\04_EAD\20161125_PiT).

To analyze strength of independent variables in explanatory power or ability to predict response variable, the following criteria should be considered:

* 1. Regression coefficients and p-values

The p-value for each term of independent variable tests the null hypothesis that the coefficient is equal to zero (no effect on response variable). A low p-value (< 0.05) indicates rejecting the null hypothesis. In other words, an independent variable with low p-value is likely to be a meaningful addition to the model because change in the independent variable’s value is associated with change in response variable’s value.

* 1. Adjusted R-squared

The adjusted R-squared is used to compare the explanatory power of regression models that contain different numbers of independent variable. It is a modified version of R-squared which increases only if the new independent variable improves the performance of model more than would be expected by chance.

* 1. Appropriate lagging-time of macroeconomic factor

Kbank develops several models with different lagging macroeconomic factors to improve performance of the models. **Error! Reference source not found.** illustrates the way to find lagging macroeconomic factors.

### C. Full data variables for LGD scorecard development

**Table 7** Available data variables for scorecard development

|  |  |  |
| --- | --- | --- |
| No | Variables | Description |
| 1 | OUT\_PCT | The percentage of default outstanding by the credit line |
| 2 | BLOCK\_CODE  (VAR\_2) | Code indicating the status of the account (if the account is normal, the value should be -1 or missing value)  **Special Handling:**  if VAR\_2 = 'W' then **1**  else if VAR\_2 = 'A' then **2**  else if VAR\_2 = 'Z' then **3**  else if VAR\_2 = 'B' then **4**  else if VAR\_2 = 'K' then **5**  else if VAR\_2 = 'M' then **6**  else if VAR\_2 = 'O' then **7**  else if VAR\_2 = 'I' then **8**  else if VAR\_2 = 'Y' then **9**  else if VAR\_2 = 'R' then **10**  else if VAR\_2 = 'T' then **11**  else if VAR\_2 = 'H' then **12**  else if VAR\_2 = 'L' then **13**  else if VAR\_2 = 'E' then **14**  else if VAR\_2 = 'F' then **15**  else if VAR\_2 = 'S' then **16**  else if VAR\_2 = 'G' then **17**  else if VAR\_2 = 'C' then **18**  else if VAR\_2 = 'P' then **19**  else if VAR\_2 = 'Q' then **20**  else if VAR\_2 = 'U' then **21**  else if VAR\_2 = 'X' then **22**  else if VAR\_2 = 'J' then **23**  else if VAR\_2 = 'V' then **24**  else if VAR\_2 = 'N' then **25**  else if VAR\_2 = 'D' then **26**  else if VAR\_2 = '-1' or  VAR\_2 is missing then **27**  **Data Aggregation:**  Min(VAR\_2) |
| 3 | AGE\_OF\_ACC  (VAR\_3) | Months on book for each account  **Special Handling:**  if VAR\_3 is missing then **1000**  **Data Aggregation:**  Min(VAR\_3) |
| 4 | DAYS\_SINCE\_LAST\_PURCHASE  (VAR\_4) | Number of days since last purchase  **Special Handling:**  if VAR\_4 is missing then -**1000**  else if VAR\_4 = **999** then -**999**  Data Aggregation:  Max(VAR\_4) |
| 5 | BAL\_TO\_LIMIT\_AVG\_3M  (VAR\_5) | Average of account balance divide by credit limit in previous 3 months  Special Handling:  if VAR\_5 is missing then -**1000**  else if VAR\_5 = **999** then -**999**  else if VAR\_5 < **0** then **0**  else if VAR\_5 > **100** then **100**  Data Aggregation:  Max(VAR\_5) |
| 6 | CASH\_ADV\_TO\_LIMIT\_AVG\_3M  (VAR\_6) | Average of cash advance amount divide by credit limit in previous 3 months  Special Handling:  if VAR\_6 is missing then -**1000**  else if VAR\_6 = **999** then -**999**  else if VAR\_6 < **0** then **0**  else if VAR\_6 > **100** then **100**  Data Aggregation:  Max(VAR\_6) |
| 7 | CASH\_ADV\_TO\_SPEND\_AVG\_6M  (VAR\_7) | Average of cash advance amount divide by purchasing amount in previous 6 months  Special Handling:  if VAR\_7 is missing then -**1000**  else if VAR\_7 = **999** then -**999**  else if VAR\_7 < **0** then **0**  else if VAR\_7 > **100** then **100**  Data Aggregation:  Max(VAR\_7) |
| 8 | TIMES\_OVER\_LIMIT\_L3M  (VAR\_8) | Number of times that account balance exceeds credit limit in previous 3 months  Special Handling:  if VAR\_8 is missing then -**1000**  Data Aggregation:  Max(VAR\_8) |
| 9 | TIMES\_OVER\_LIMIT\_L6M  (VAR\_9) | Number of times that account balance exceeds credit limit in previous 6 months  Special Handling:  if VAR\_9 is missing then -**1000**  Data Aggregation:  Max(VAR\_9) |
| 10 | DAYS\_SINCE\_LAST\_PMT  (VAR\_10) | Number of days from latest payment  Special Handling:  if VAR\_10 is missing then -**1000**  else if VAR\_10 = **999** then -**999**  Data Aggregation:  Max(VAR\_10) |
| 11 | PMT\_TO\_PREV\_BAL\_AVG\_3M  (VAR\_11) | Average of payment amount divide by account balance in previous 3 months  Special Handling:  if VAR\_11 is missing then **1000**  else if VAR\_11 = **999** then **999**  else if VAR\_11 < **0** then **100**  else if VAR\_11 > **100** then **100**  Data Aggregation:  Min(VAR\_11) |
| 12 | PMT\_TO\_PREV\_BAL\_AVG\_6M  (VAR\_12) | Average of payment amount divided by account balance in previous 6 months  Special Handling:  if VAR\_12 is missing then **1000**  else if VAR\_12 = **999** then **999**  else if VAR\_12 < **0** then **100**  else if VAR\_12 > **100** then **100**  Data Aggregation:  Min(VAR\_12) |
| 13 | TIMES\_FULL\_PMT\_L6M  (VAR\_13) | Number of times that customer makes a full payment in previous 6 months  Special Handling:  if VAR\_13 is missing then **1000**  Data Aggregation:  Min(VAR\_13) |
| 14 | TIMES\_PART\_PMT\_L6M  (VAR\_14) | Number of times that customer makes a partial payment in previous 6 months  Special Handling:  if VAR\_14 is missing then **1000**  Data Aggregation:  Min(VAR\_14) |
| 15 | TIMES\_PMT\_MADE\_L6M  (VAR\_15) | Number of times that customer makes a payment (both full and partial) in previous 6 months  Special Handling:  if VAR\_15 is missing then **1000**  Data Aggregation:  Min(VAR\_15) |
| 16 | MONTHS\_SINCE\_LAST\_DELIN  (VAR\_16) | Number of months from latest delinquency  Special Handling:  if VAR\_16 is missing then **1000**  Data Aggregation:  Min(VAR\_16) |
| 17 | TIMES\_X\_DPD\_L3M  (VAR\_17) | Number of times that bill is past due in previous 3 months  Special Handling:  if VAR\_17 is missing then -**1000**  Data Aggregation:  Max(VAR\_17) |
| 18 | TIMES\_X\_DPD\_L8M  (VAR\_18) | Number of times that bill is past due in previous 8 months  Special Handling:  if VAR\_18 is missing then -**1000**  Data Aggregation:  Max(VAR\_18) |
| 19 | TIMES\_GREATER\_30\_DPD\_L3M  (VAR\_19) | Number of times that bill is 30-plus day past due in previous 3 months  Special Handling:  if VAR\_19 is missing then -**1000**  Data Aggregation:  Max(VAR\_19) |
| 20 | TIMES\_GREATER\_30\_DPD\_L8M  (VAR\_20) | Number of times that bill is 30-plus day past due in previous 8 months  Special Handling:  if VAR\_20 is missing then -**1000**  Data Aggregation:  Max(VAR\_20) |
| 21 | TIMES\_GREATER\_60\_DPD\_L6M  (VAR\_21) | Number of times that bill is 60-plus day past due in previous 6 months  Special Handling:  if VAR\_21 is missing then -**1000**  Data Aggregation:  Max(VAR\_21) |
| 22 | TIMES\_60\_TO\_89\_DPD\_L6M  (VAR\_22) | Number of times that bill is 60 to 89 day past due in previous 6 months  Special Handling:  if VAR\_22 is missing then -**1000**  Data Aggregation:  Max(VAR\_22) |
| 23 | INACTIVE\_L8M  (VAR\_23) | Flag stating whether the account is inactive in the previous 8 months (1) or not (0)  Special Handling:  if VAR\_23 is missing then **1000**  Data Aggregation:  Min(VAR\_23) |
| 24 | UTIL\_STATUS  (VAR\_24) | Utilization status of each account  Data Aggregation:  Min(VAR\_24) |
| 25 | SCORE  (VAR\_25) | Behavior score of each account  Data Aggregation:  Min(VAR\_25) |
| 26 | RISKGRADE  (VAR\_26) | Behavior risk grade of each account  Data Aggregation:  Min(VAR\_26) |
| 27 | AGE | Age of customer |
| 28 | GENDER | Gender of customer |
| 29 | EDUCATION | Highest education of customer |
| 30 | OCCUPATION | Occupation of customer |
| 31 | OCCUPATION\_CHAR | Occupation characteristics of customer |
| 32 | LIVING\_AREA | Living area of customer which is flagged from the current address zip code (CTC\_ADR\_ZIP\_CODE aka VAR\_31)  Rules:  if VAR\_31 is missing then **NULL**  else if VAR\_31 > **10999** then **Upcountry**  else if VAR\_31 in (**10130**, **10270**, **10290**, **10540**, **10560**) then **Upcountry**  else if VAR\_31 between [**10000,10999]** then **Bangkok**  else **NULL** |
| 33 | MARITAL\_STATUS | Marital status |
| 34 | SAVING\_DEPOSIT | Average wealth last 3 months |
| 35 | RESIDENTIAL\_STATUS | Current residential status |
| 36 | EMPLOYMENT\_PERIOD | Current employment period (month) |
| 37 | EMPLOYMENT\_PHONE\_FLAG | Current employment phone flag |
| 38 | TOTAL\_EMPLOYMENT\_PERIOD | Total employment period (month) |
| 39 | VERIFIED\_INCOME | Verified monthly income |
| 40 | MAILING\_ADDRESS | Mailing address |
| 41 | PAYMENT\_METHOD | Payment method of each account |
| 42 | FICO\_SCORE | FICO score |
| 43 | DEBT\_BURDEN | Debt burden |

1. http://www.IFRS.org/current-projects/iasb-projects/financial-instruments-a-replacement-of-ias-39-financial-instruments-recognitio/documents/IFRS-9-project-summary-july-2014.pdf [↑](#footnote-ref-1)