# Model Data

[Documentation should include discussion of historical data set construction, including data sources, adjustments to the data set, and documentation validating the use of any external data*.*

The purpose of this chapter is to provide an understanding of the data used for modeling, while also providing evidence that data weaknesses and limitations have been thoroughly considered and can be justified. If data was not used as part of the Modeling Approach, please write N/A and explain why data was not used or questions are not applicable.]

The purpose of this chapter is to provide an understanding of the technical aspects of the data used for modeling, while also providing evidence that data weaknesses and limitations have been thoroughly researched and mitigated within the scope of the model development process.

The model’s data sources and controls are discussed followed by an explanation as to how data is accessed, processed and centrally stored for all of the model development phases. Included is a description of the data reconciliation, preparation, and cleaning processes as well as any data sampling and transformations performed. Further, evidence of all business and technical data quality checks performed is included in this chapter. This chapter also includes a list of all relevant policy/operational events that impact the model results.

The key data enhancements attempted as part of the 2019 CCAR process includes the following-

1. Consistent product identification logic that align with the business definition and rendition
2. For 2019 CCAR model, one of the changes in the data source, specifically of the REL\_TRAN.LOSS table has been the inclusion of CPB loan data. In coordination with Risk Data Mart and CPB Risk team a restatement of the loss table for all CPB historical data was implemented in Jan 2018. This restatement addresses the data challenges that were reported on the previous CCAR models, where CPB was manually included in the loss table provided by the CPB Risk team. This restatement automated the process, added data quality check, reconciliation between other data source and now available in a single loss table. See official email from Data Mart attached.
3. All source input files have been copied to a frozen dataset to prevent potential model errors from version updates
4. Extended the model development data to include recent time to holistically capture the portfolio composition changes
5. Incorporated balance differences between CMI and CPB
6. Introduced new interaction effects and splines to improve the model’s performance in specific periods or segments
7. Imputed the missing values with a median, instead of omitting them and thereby reducing the possibilities of omitted variable bias
8. Capped the extremely large value of the continuous variable by 99th percentile plus 2 times the standard deviation
9. In PD model, for 2nd liens, instead of assuming a constant first lien balance when calculating its mark-to-market CLTV, assumed a constant junior ratio to impute the first lien balance and CLTV
10. Leveraged a new refreshed property valuation file, in line with what the business recommends and uses, for estimating the distressed values for all properties with/without prior distressed values
11. Included all zero losses in the development data pool for the Severity model

## Model Input

This section will provide additional information to the “Automated Document” Section A.3.: “Inputs, Outputs & Assumptions” (refer to the “Note on Model Risk Management – Document Automation”, in the beginning of this document)

The data development phase, review phase and sampling phase are designed to determine whether building a risk model is feasible and to set high-level parameters such as loan exclusions, target definition, development sample window and performance sample window parameters. These high-level parameters decrease the number of anomalies and improve the accuracy of the model, and are naturally supported by a sound rationale. Significant time and engagement with stakeholders is the key to building a successful risk model throughout the data development, review, and sampling phases.

For 2019 Method A RM Model, model developers sourced data to build the model using both internal and external data sources. The data is compiled for use in model development and multiple quality checks are conducted to ensure continued accuracy of the data. The following sections describe the data sources and controls in further detail.

### Data Sources and Controls

[Provide details regarding how the data was accessed, processed and stored.]

[What are the performance, segmentation and account data sources for the model?](#_Toc421107375)

[Document sources for all internal and external data used in model development. Describe each data source and whether the data is extracted in a manual or automatic manner. Mention whether the data has been used in the past for model development purposes.]

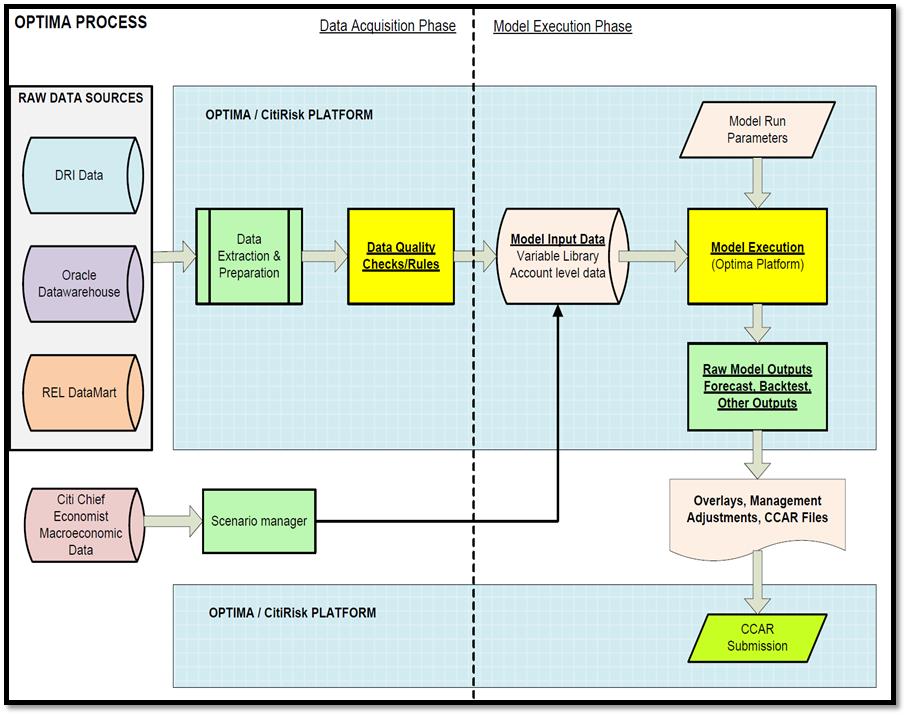
Model developers sourced data to build the model using both internal and external data sources. The data is compiled for use in model development and multiple quality checks are conducted to ensure accuracy of the data throughout the process. The following sections describe the data sources and quality control in further detail.

Significant time and engagement with stakeholders is key to building a successful risk model throughout the data development, review, and sampling phases. The process is designed to determine whether building a risk model is feasible and to set high-level parameters such as loan exclusions, target definition, development sample window, and performance sample window. These parameters decrease the number of anomalies while improving the accuracy of the model.

The 2019 Method A RM model development data and portfolio level data are all sourced (at the time of development) from robust internal data sources consisting of: SAS Grid/Optima/Model Execution Package and from an established external sources for macro-economic information (i.e. HPI, Unemployment, Interest Rate), Core Logic, Moody’s, and Global Consumer Risk Management (GCRM).

As result of the new global initiatives within the Global Consumer group, the 2019 NA Mortgage Method A RM models leveraged this new centralized data source called MEP Optima (see figure 4.1.1.1). This Global Database platform within the SAS Grid environment which hosts multiple raw data sources as well as macroeconomic variables. Within the grid, a Model Execution Platform (MEP) contains all the data required to execute the models which by design have undergone several data quality check and data proofing process to make sure to comply all the data requirements for all CCAR models and data quality principles.

Figure 4.1.1.1 MEP Optima Process



The 2019 Method A RM model’s objectives are to produce core line of business risk models with a product level focus on residential mortgage portfolio. This segmentation also aligns with the business strategy, and loss forecasting regulatory requirements. In addition, the data scope will be limited to BankMortgage Inc.(CMI), Bank Private Bank(CPB) and US Retail Bank (USRB) mortgage portfolios, including conventional mortgage loan products with non-zero balances, risk owned portfolio products.

MRM Question - Sponsor has mentioned creation of unique OOT sample is in response to limitation#19546. However, limitation #19546 states about inclusion of pre-2008 data in LGD model. Therefore, sponsor is requested to provide justification for not including pre-2008 data in LGD model development sample.

Answer - Although there are some reconciled issues caused by the misclassified loans mainly observed before May 2008, CAMU had set the observations date and performance window for development data from Feb 2006 to Dec 2017.  The decision was made mainly to ensure there are sufficient observations to meet the gating principles outlined by Model Risk Management Policy, and to cover both stress and non-stress periods within the U.S. housing price lifecycle.  However, abnormal actual losses were observed in snapshot Jan-2008 after removing these loans from the model back testing.   To avoid any misinterpretation of the model performance, shifting the reporting snapshot date from Jan-2008 to April-2008 is considered appropriate.  Please refer to Section 6.3 for the rationale of excluding 2008Q1 data from back testing.

**Table 4.1.1.1 Data Source, Location, and Frequency of Updates**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Location and Source** | **Owner** | **Frequency**  **Updated** |
| Loan performance and borrower data | SAS Grid/Model Execution Package | REL Risk Data Team | Monthly |
| Bureau data | SAS Grid/Model Execution Package | REL Risk Data Team | Monthly |
| Loan Status/Resolution data | SAS Grid/Model Execution Package | O&T Operations | Monthly |
| **Historical Macroeconomic data**  HPI  Unemployment  Interest Rate LIBOR  Income  VIX  S&P500 index  All Other Rates | Core Logic  (Moody’s) [www.economy.com](http://www.economy.com)  Bloomberg  Haver  Bank Velocity  SAS Grid/Model Execution Package | External Vendor/REL Risk Data Mart/ GCRM/Risk Portfolio Loss Forecasting | Monthly |
| Future Macroeconomic data | Forecast provided by Bank’s GCRM to a controlled access internal Share Point Site available for download ([https://globalconsumer.collaborationtools.consumer.Bankgroup.net/sites/CCAR\_RR/default.aspx](https://globalconsumer.collaborationtools.consumer.citigroup.net/sites/CCAR_RR/default.aspx)) | GCRM/Risk Portfolio Loss Forecasting | Semi-annually |

[What are the macroeconomic data sources for the model including stress scenarios?](#_Toc421107375)

[Document sources for all internal and external data used in model development. Verify data source consistency across models developed in the same country or business – macroeconomic data for the same variable (e.g. GDP) should be identical.

Describe each data source and whether the data is extracted in a manual or automated manner. Only Bank-verified data sources are acceptable. Mention whether the data has been used in the past for model development purposes.

If there were alternative sources of macroeconomic data, provide rationale for selecting the specific source. ]

***Example:*** *Macro-economic scenarios are sourced from Federal Reserve CCAR guidance, in conjunction with Bank-verified data.*

There are two main categories for the macroeconomic data used in the model development and sensitivity analysis namely: Historical and Future Outlook macroeconomic data. GCRM provides both historical and future macroeconomic variables for US State level GDP, S&P 500 index, Volatility Index (VIX), Income and Interest rates (Swap 1 year, 5 year and 10 year, 30 year Prime rate etc.), which are sourced originally, from Bloomberg, Bank Velocity and Haver. For historical Home Price Index (HPI), Risk Data Mart team managed and own the processing of the data originally sourced from Core Logic. Historical Unemployment rate, Risk Portfolio Loss Forecasting Team processes the data from Moody’s economy.com platform before loading to SAS Grid/Model Execution Package in Optima.

All future macroeconomic data are provided by GCRM and then process and managed by Risk Portfolio Loss Forecasting team leverage in BAU Loss Forecasting exercise and analysis. Both historical and future macroeconomic data undergo systematic control and data quality process before loading to SAS Grid/Model Execution Package in Optima for model development usage.

What process checks and controls are in place to ensure data integrity and that the data is properly sourced and reliable?

* What process checks and controls are in place to ensure data integrity and that the data is properly sourced and reliable?

[Describe the high-level data preparation and transmission process, with a focus on maker-checker responsibilities, accountability of key personnel during the process, and data versioning. Describe how the model sponsor ensured that developers received current data (including the sign-off process, and when this occurs), and that this current data was used in model development. Mention sign-offs and other process checks confirming data adequacy for the modeling purpose within each process step.

What data was considered to be potentially relevant but was not used? What is the sponsor’s view of the potential impact of excluding this data? [Please provide the details and the impact analysis]

Describe whether the process differs for model implementation data (compared to model development data). Describe the process both for performance data, and macroeconomic data and stress scenarios.

This section also includes controls for data processing scripts. If a standard version/revision control system is used (such as SVN, CVS, Git, ClearCase or Team Foundation Server), describe the way in which versioning control is used. Describe controls in place to:

* + - Track changes made to data extraction and processing scripts.
    - Track which version of the scripts was used to prepare a given version of the data.
    - Prevent inadvertent changes in data extraction scripts.

Describe if a code review was performed (script maker/programmer - script checker split).]

Historical datasets stored in SAS Grid Model Execution Package (MEP) in Optima and Risk REL Data Mart, undergo an automated and controlled extraction, transformation and loading (ETL) process, which include detail data validation and data quality procedure for all variables and attributes. In addition to data quality checks by the Bank’s MEP Optima team, Risk Management has quarterly Management Control Assessments (MCAs) setup to review risk associated with inadequate monitoring, improper data transformation, and improper changes to data, business policies, or business logic.

Additional layer of maker-checker control process is setup between GCRM, Risk Portfolio Loss Forecasting team and CAMU in handling Macroeconomic data. Please refer to attachment ‘4.1.1 GCRM Evidence\_Macro\_Inputs’ for the GCRM evidence around the usage of all macro-economic variables in the models. CAMU Data Quality team performs additional validation check before using them for Model development. The maker-checker process comprised of checking data integrity, independent validation of the SAS code logic and proper signoff on any data treatment and adjustment made on the original data source.

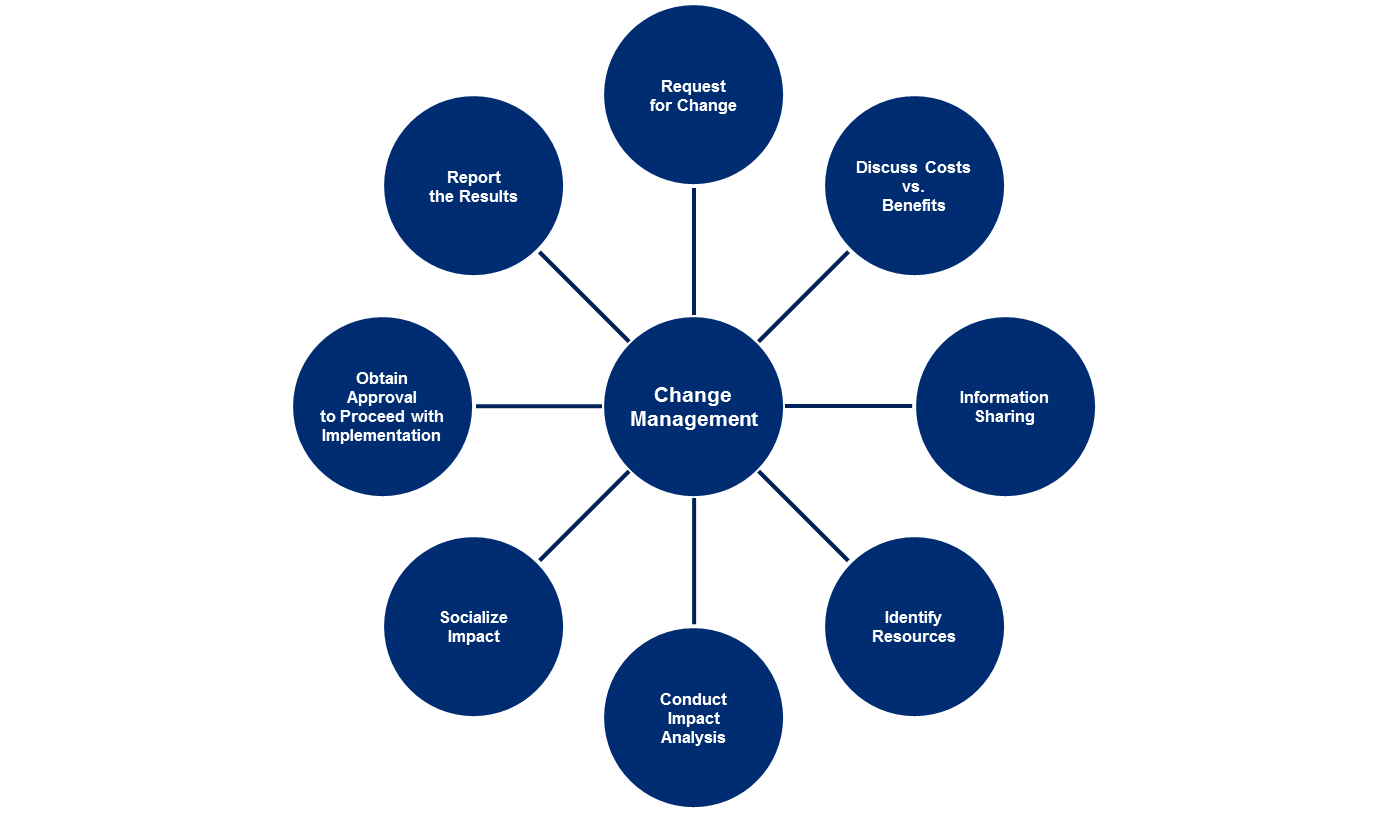
Furthermore, CAMU Data Quality team performs detail reconciliation from the original data source and collaborates with Risk Data Team to identify gaps and document resolution. CAMU is member of the data steward’s committee and work with other risk business partners to provide proper escalation and resolution on the all data issues. Moreover, CAMU production and implementation team prepare the data and provides clean and standardized dataset to modelers. Finally, modelers further evaluate the data by analyzing attributes, characteristics performance measures and key risk metrics to identify inherent potential data issue before using in model development.

**Table 4.1.1.2 MEP Optima Extract, Transform, and Load (ETL) Processing**

| **MEP Optima Extract, Transform, and Load (ETL) Processing** | |
| --- | --- |
| **Controls** | **Monitoring/MIS** |
| **1.** Autosys enables the establishment of automated schedules and user defined event rules for the execution of required programs. Rules and prerequisites can be loaded into Autosys that must be satisfied prior to the execution of a program. Autosys provides real-time alerts and user notifications through-out the execution of the job schedule. Every job has an assigned owner, primary monitor, secondary monitor, business admin, technical admin, and subject matter expert. The MEP Optima Team performs/monitor these controls utilizing Autosys Scheduler, Job Metadata SAS datasets and SAS Logs/Reports on RELSAS20 Unix server providing information about each job's scheduling details, assignments, and expected outcomes | i. Direct Negative Control Information ii. Activity logs stored in SAS datasets on SAS Grid and Unix server iii. Attributes, characteristics, or conditions satisfied: job is executed successfully with a passing grade. |
| **2.** A file transfer reconciliation is performed from data source to the data copied. This reconciliation deals with mainly record count. An e-check is in place to notify Credit Risk when a trend occurs in record count that is above a certain percentage limitation. | i. Direct Negative Control Information   1. ii. Activity logs stored in SAS datasets on SAS Grid and Unix server   iii. Attributes, characteristics, or conditions satisfied: job is executed successfully with a passing grade. |
| **3.** Decisions reviewed during the Risk Data Stewards meeting are to be acknowledged and approved by Charter Members of the Data Stewards by returning an email with their approval when meeting minutes are distributed. Email approvals are obtained from the appropriate end users management team prior to releasing monthly MEP Optima globally. | i. Timely approval of MEP Optima changes/updates by appropriate approval authorities. ii. To make sure changes to the MEP Optima are being communicated via Data Stewards meetings.  iii. Monthly release of Model Execution Platform (MEP) is being approved by appropriate resources.  iv. Share Point Data Steward meeting minutes are up-to-date and available. |

Further, there is significant data governance for the overall management of the availability, usability, integrity, and security of the MEP Optima platform. The below process flow clearly outlines the flow for data governance and data change management within the repository:

**Figure 4.1.1.2 Data Governance Flow for MEP Optima**



* Is the macroeconomic data source and data itself approved for use in Bank?

[Describe if the macroeconomic data is acceptable for use in Bank models (e.g. that a sufficient forecast period is available for the data, that the source is considered reliable, etc.)

Yes. All macroeconomic sources mentioned above are approved for use in Bank CCAR Mortgage Risk Models, both historical and future macroeconomic data sources either are directly from GCRM or processed by Risk Portfolio Loss Forecasting team with corresponding approval and signoff confirmation from GCRM. Furthermore, macroeconomic stress scenarios are provided uses the Federal Reserve CCAR guidance in conjunction with Bank verified data.

### Data Reconciliation

[The following template can be used to document reconciliation and as a guide through the reconciliation process. If no data was used in the model development, please answer “N/A – no data used” for all questions in this section.]

Is the data source regularly reconciled to another independent data source for all key model variables?

[Respond “Yes” or “No”. If “Yes”, describe the independent (finance, controlling etc.) system, the reconciliation process and sign-off process. Include how often the reconciliation is performed. Attach evidence either in the appendix of this document or reconciliation templates. If “No”, provide an explanation for the lack of reconciliation, along with compensating controls and other measures made to ensure data reliability.

All key model variables should be reconciled. For example, if the model calculates a PD rate based on the ratio of active and default accounts, then the number of defaults and the number of active accounts should be reconciled. Reconciling only GCL and NCL is insufficient.]

Yes. Bank reconciles data at minimum on a quarterly basis as evidenced within the reconciliation template.

Bank’s Risk Analysis and Reporting Team reconcile its Monthly CCAR Fed Submission (14M report) to the Federal Reserves Consolidated Financial Statements for Holding Companies Report (FR Y-9C). There are some differences between the two reports which are due to varying accounting treatment of loans as reflected in the 14M and FRY-9C reporting, which are discussed in more detail below.

In addition, Optima group performs reconciliation and variable profiling for all root source variables uploaded into SAS Grid MEP in Optima. Detail data quality check are performed to make sure all the data elements are within the proper expected domain values and monitor any outliers or unexpected data issue encountered.

If the data source is not regularly reconciled, how was the data used for model development reconciled?

[Describe the independent (finance, controlling etc.) data system. It should be clear that the system is independent. It should be a system used by another department, preferably used to prepare financial statements.

Describe the reconciliation process. Attach evidence either in the appendix of this document or the reconciliation templates. Describe any differences in the process between development and implementation reconciliation.

If the model is prepared by an external model developer, or if there is a complex data handover, describe how the model developer verified that the data used for model development is complete and matches data extracted by country. The suggested approach is to obtain data reconciliation from the model sponsor and verify that the reconciliation matches data used for actual model development. A complex data handover takes place if the model is developed out of country, e.g., by a shared service center or Bank model development hub.

All key model variables must be reconciled. For example, if the model calculates a PD rate based on the ratio of active and default accounts, then the number of defaults and the number of active accounts must be reconciled. Reconciling only GCL and NCL is insufficient.]

Not applicable. See reconciliation evidence as provided in the required template.

Were there any reconciliation differences?

[Describe any reconciliation differences encountered during the process, high-level causes of the differences, and suggested remediation. It must be clear whether the data can or cannot be used for model development, and whether it is sufficiently accurate. Quantify the impact of differences.]

The first difference is that the 14M report leverages UPB (unpaid balances) while the FR Y-9C Financial Report leverages ENR (ending net receivables). UPB calculations do not allow for movement of a loan from one portfolio to another. Although the mortgage portfolios do not have a large number of internal portfolio transfers on a regular basis, there is a segment of HELOC loans (which at any given month are converting to fully amortized loans) which can cause some insignificant discrepancies quarter-over-quarter.

Secondly, CAMU would like to point out these distinct areas of insignificant differences: 1) FAS91 (Accounting for Nonrefundable Fees and Costs Associated with Originating or Acquiring Loans and Initial Direct Costs of Leases), 2) LHS Clearing loans (loans in process pipeline) and 3 Offline FAS91 (HELOC Mods)

Therefore, in the enclosed reconciliation template, the spreadsheet attempts to reconcile these differences within a tolerance of less than 1%.

In addition to the overall ENR balance reconciliation between Risk system and Financial General Ledger, individual key attributes were also tested for consistency and accuracy by going through the development sample attributes and compared against the original data source from the SAS Grid/MEP Optima over a snapshot period consistent in the development time frame; overall results are consistent with the official data source, with very minor volume of difference due to timing or re-statement of the original data source system. No serious pattern or data quality anomalies were observed.

The reconciliation process includes frequency matching using a sampling methodology from the development sample used in the modeling process with specific focus on the data quality and consistency of the data used in the model development and original data source.

Remaining variances are mainly due to a timing difference between the time the modeler used the dataset and the time the validation check was performed. The DataMart team is continuously improving and updating the source data file based on the latest and most accurate information. This may result in historical data adjustments or re-statements. (See attached files named ‘4.1.2 CCAR Data Reconciliation…’ for detail results).

### Data Characteristics, Fields and Definitions

[Provide a description of data preparation, including the steps followed in assembling the data, and the number of records (observations) per account. Consider including stylized or actual example records in order to make the structure of the data set transparent to the reviewer. Describe each input explicitly, including dependent/independent variables (for regressions), data sources, frequency of data pull, contact person, format (e.g. Boolean, character, integer, floating-point number, fixed-point number and length, size of the variable and position of the decimal point), proper unit (e.g. thousands, millions, percentage), data type (external, internal), derived variables construction methodology, and output. For monetary variables, indicate whether USD or another currency is used and label all graphs and tables accordingly.]

#### Please provide descriptive statistics for the underlying data including histograms/frequency tables, and trends over time.

[Please provide mean, standard deviation, median, extreme values, mode, percentage of missing records, and other statistical descriptive measures (as appropriate), for each input.

Comment on whether the values are in line with expectations given the composition of the portfolio. Analyze outliers.]

##### PD Model

The initial monthly data for Method A RM PD model is from Jan 2005 through Dec 2017 yielded 77,969,449 records with 4,785,882 unique accounts before data filtering in Table 4.1.3.1. The data is constructed as a standard panel dataset and the performance of the monthly data is followed month over month until the onset of one of these statuses: involuntary payoff (IVP), or voluntary payoff (PO), or survived at the end of Dec 2017.

After data is extracted from the risk database, steps are taken to remove and/or clean erroneous data. Any lines with a data quality issue are excluded from the data water fall. In addition to the data clean up, CAMU determined to exclude all observations for loans received treatment since there are models designed particularly for this segment. CAMU also collaborated with the risk team to ensure the correct population is included in the development data. The detail of all data quality issues leading to exclusions are noted in the data waterfall in 4.1.3.1.

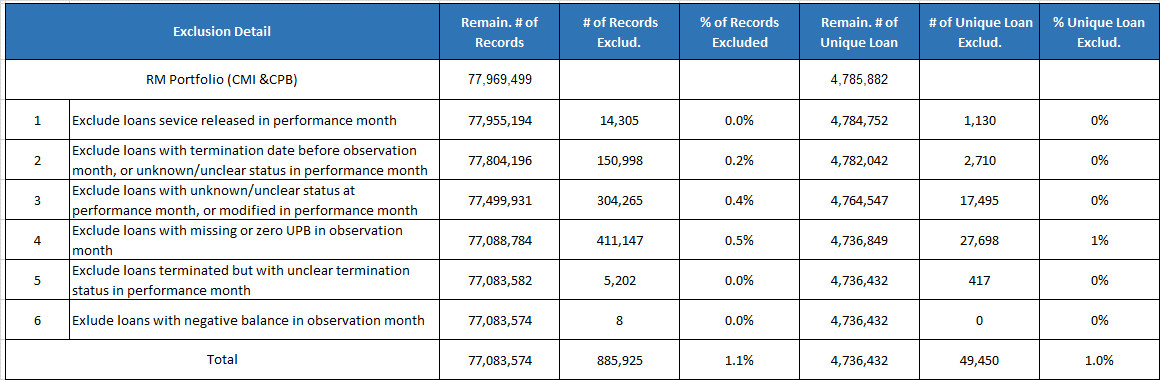
The exclusions were generally very limited and designed in alignment with business knowledge/intuition. The criteria can be categorized into three exclusion reasons.

Exclusion Reason 1: Criterion # 1- These are excluded in order to define the model development population that best aligns with CCAR objectives. Particularly the asset sale Sold and Released (SR) or Sold and Serviced (SS) records in the performance months are excluded, for the reason that SR do not have reliable performance, while SS is not deemed as portfolio risk hence considered outside CCAR scope. Indeed, the drop of 14,305 unique loans at this step is primarily caused by excluding sold and released – they only have two performance records, first in the observation month is portfolio loan, and second in the performance month when the asset becomes Sold and Service Released.

Exclusion Reason 2: Criterion #s 2, 4,5 and 6 -These are excluded to improve data quality and remove the records with missing performance or negative balances or those outside the scope of model development period.

Exclusion Reason 3: Criterion # 3 - When a loan is successfully modified, its delinquency level will in general be set to current. This is not due to borrower’s self-curing behavior and should not be modeled as a curing event. The Risk Loss Mitigation Policy instead of borrower behavior primarily determined the volume of modifications. The modified pools of loans are modeled within the scope of the Modified loan model suite and hence are excluded from the RM development data.

**Table 4.1.3.1 Data Water Fall for Method A RM** **Model**



From Jan 2005 to Dec 2017, there are a total of 4,785,882 unique accounts with 77,969,499observations pulled from the monthly Residential Mortgage portfolio of which 75,514,848 observations are CMI and 2,454,651 are from CPB first Lien. After the data waterfall, the data contains 4,736,432 unique accounts with a total of 77,083,574 observations.

In the above panel data, depending on the start buckets, five major scenarios can typically describe a borrower’s behavior that include (1) unchanged delinquency status, (2) cure or partial cure status: from a higher delinquency status to buk01 or to a relative lower delinquency status, (3) status goes worse: from a lower delinquency status to a higher delinquency status, (4) involuntary payoff, or (5) voluntary payoff.

The 7 starting buckets are defined as the delinquency state at the observation month:

1. Buk 01: 0 to 29 days past due

2. Buk 2: 30 to 59 days past due

3. Buk 3: 60 to 89 days past due

4. Buk 4: 90 to 119 days past due

5. Buk 5: 120 to 149 days past due

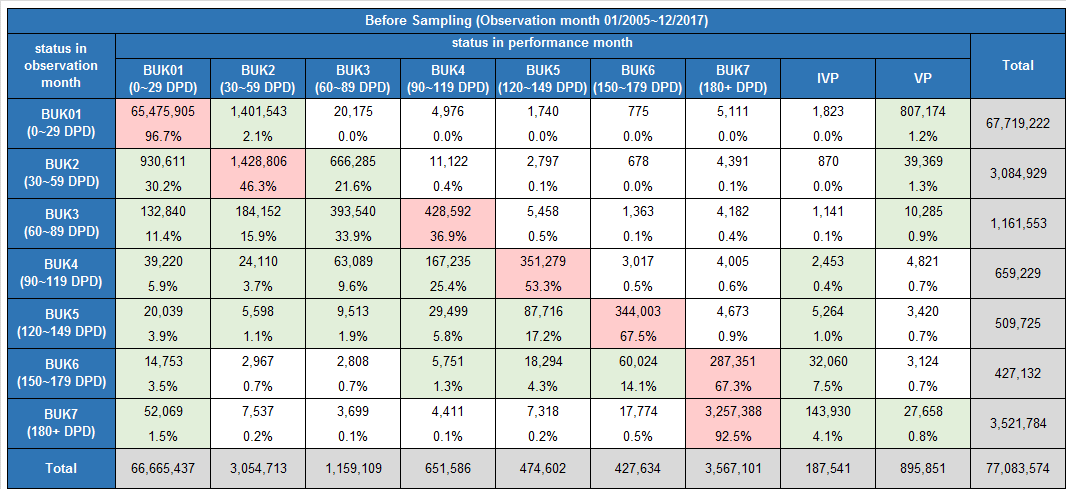
6. Buk 6: 150 to 179 days past due

7. Buk 7: at least 180 days past due

The model development sample is then segmented in Table 4.1.3.2 by delinquency status that can be represented by a 7X9 transition matrix in which loans transitioned from one of the seven delinquency states in the observation month (t) to the nine possible transition states including the terminal events of charge off (IVP) and payoff (VP) in the performance month (t+1).

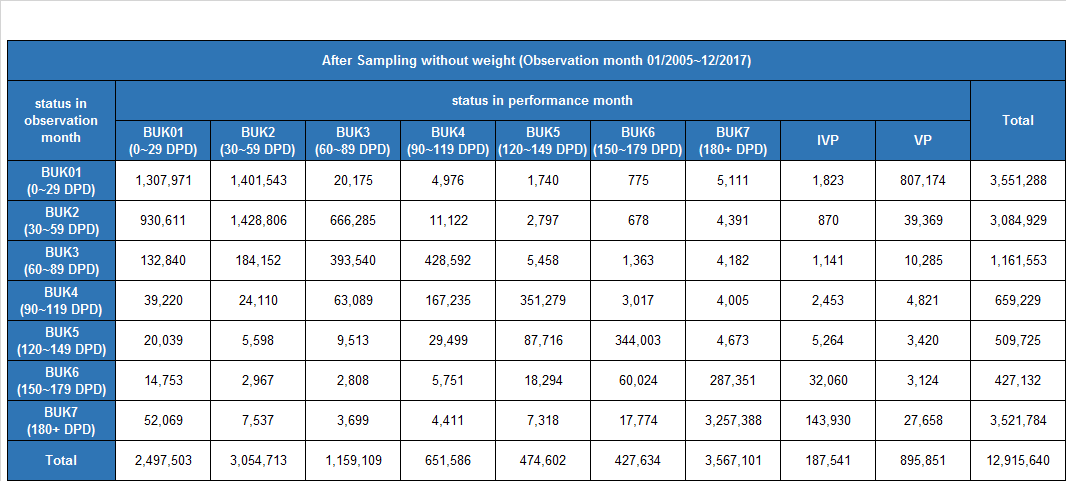
In Table 4.1.3.2 below, all cells highlighted in green are the transitions that will be estimated as a model with key risk drivers (modeled transition), and the cells in pink are being used as reference category when develop the model. In addition, the cells in white are the non-modeled transitions to be developed as a lookup table with constant numbers for stress and non-stress period respectively.

**Table 4.1.3.2 Transition Roll Rate Matrix of RM panel data before Sampling on Current Bucket**

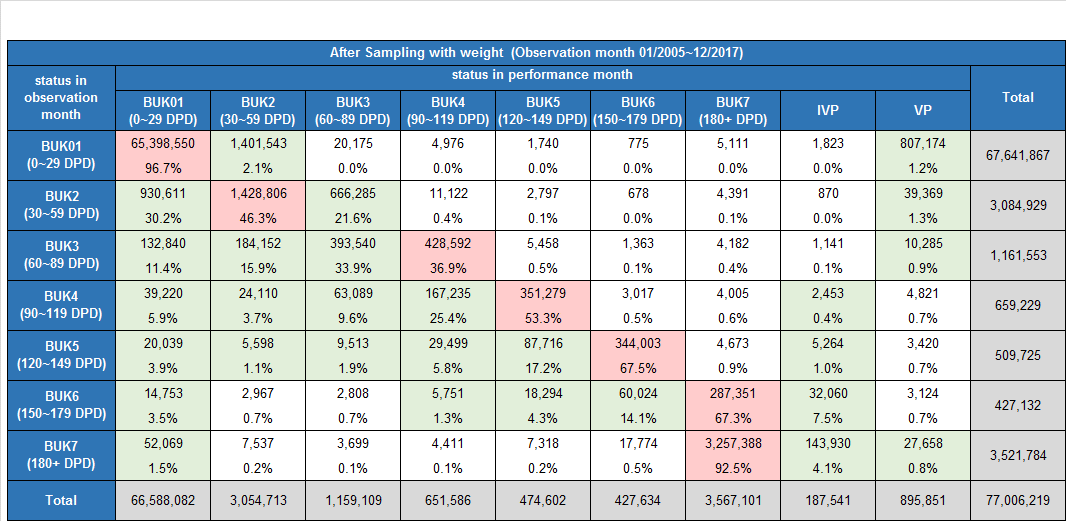


CAMU conducted random sampling on the population that stays in BUK01 in both the observation month and performance month with a 2% sampling rate. As shown in table 4.1.3.3, number of records in the cell of “BUK01-BUK01” decreased from 65,475,905 to 1,307,971. After sample down, the total number of records drops to 12,915,640 of which 12,794,442 are CMI and 121,198 from CPB first lien. Table 4.1.3.4 gives the records and the transition rate on the sampled data with weight, which shows that the randomly sampling significantly reduces number of records to be included in modeling while maintaining the same transition rate.

**Table 4.1.3.3 Number of Records in RM data after 2% Sampling on Current Bucket**



**Table 4.1.3.4 Transition Roll Rate Matrix of RM panel data after Sampling on Current Bucket with** **weight**



At the model development stage, it is further determined to exclude the records collected on observation month before Feb 2006 because of the relatively low data quality of the credit bureau attributes such as FICO during that period. It’s also decided that the model to be developed only on loans within the range of CMI or CPB official loan list. The official loan list is the list that Risk Loss Forecasting team uses to report their number for CCAR. Official loan list is reconciled with PEARL system and also reconciled in Optima as part of the data quality check. This is the main reason the CCAR model is developed on the official loan list to make sure it is based on reconciled data which is reported by regulators, thus showing correct and consistent population for model development.

Hence 1,727,861 more records were removed from the sampled data with 11,187,779 observations left for model development. By applying weight on the BUK01->BUK01 sample, the total number of observation is 63,657,028.

All records that have gone through the above waterfall and sampling process are them divided into model development (DEV), in-time hold out (INT) validation and out-of-time (OOT) validation sample. First, data collected between Feb 2006 and Dec 2011, and Apr 2014 and Dec 2017 are used as the in time sample, and those between Jan 2012 and Mar 2014 are kept as the out-of-time validation sample.

In this version of model development, CAMU holds out an OOT sample solely to comply with MRM’s OOT test requirement although CAMU believes that holding out OOT data will potentially result in loss of performance and portfolio change information. Due to this concern, CAMU carefully chose Jan-2012 to March-2014 period as the out-of-time validation sample and retained data between Feb 2006 - Dec 2011 and Apr 2014 -Dec 2017 as model development data due to the following considerations –

1. First and foremost, as per limitation # 19546, it was recommended to include the most recent time period as part of the model development sample. To comply with MRM’s requirements, CAMU did not utilize the hold out most recent data as OOT. CAMU also agreed that holding out recent performance data would disable us from capturing the most recent macro-economic trend, such as interest rate increase, portfolio and underwriting policy change. The inclusion of the recent period data helped capture the go-forward state of the business in terms of origination profile and portfolio composition mix.

2. Excluding 2006-2008 data is not an option because it will result in the loss of a portion of the stress period performance. This leaves CAMU no choice but to carve out a middle period as the OOT hold out sample to follow MRM’s OOT requirement.

3. To ensure parameter stability and OOT hold out didn’t cause significant information loss, CAMU conducted further parameter stability analysis and revised/dropped instable parameters. The parameter stability analysis was two-fold including:

a. CAMU estimated the same specification on the 20% INT sample to ensure consistent parameter sign, significance and reasonable magnitude

b. CAMU further estimated the same specification on the entire data including the OOT period to ensure consistent parameter sign, significance and reasonable magnitude

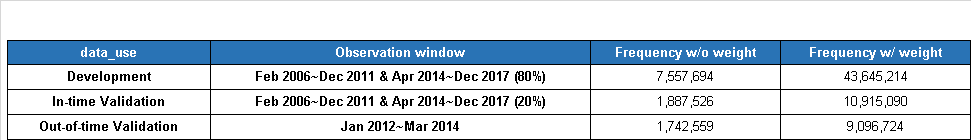
Figure 4.1.3.1a demonstrates how the default rate in the development data captures the change of swap interest rates during the stress and non-stress periods. When swap rates dropped significantly in 2009 and remained low between 2009 and 2011, higher default rate was observed. Likewise, when the interest rates increased slightly after the US economy stress period, the default rate continued to drop over time.

Figure 4.1.3.1a Residential Mortgage Development Default Rate and Swap Trends

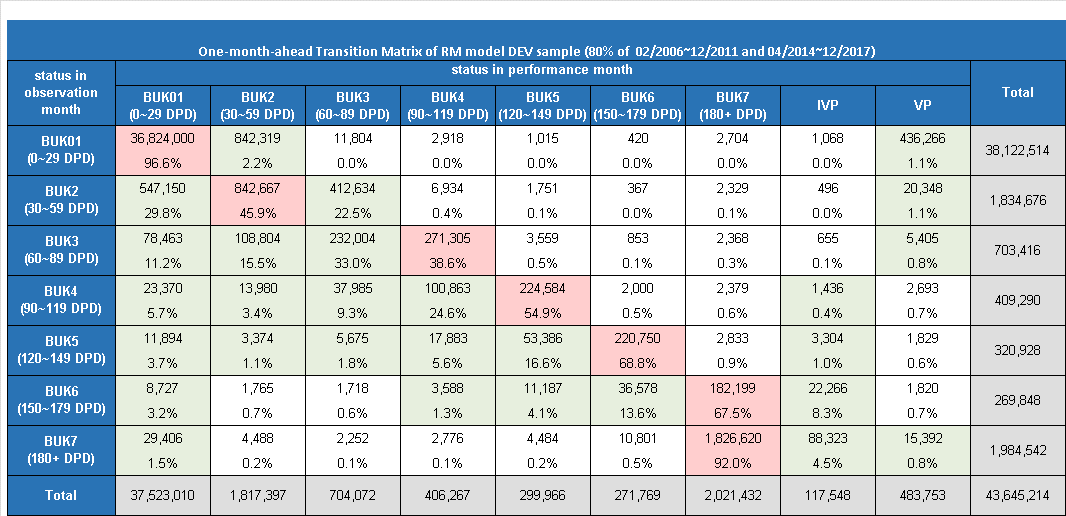
Second, within the in-time sample, 80% of unique accounts with 43,645,214 observations were used as development samples vs 20% of unique accounts with 10,915,090 observations as validation. Table 4.1.3.5 givens the total number of records in the final model DEV, INT and OOT sample on the sampled data before and after applying weight.

It is worth noting though, that as CAMU expected, holding out the OOT data could potentially result in loss of performance information, especially given that the Jan2012-March2014 period represented the initial phase of economic recovery with high volume of various types of loan modifications and refinancing due to extremely low interest rate and foreclosure settlement (2012) etc. However, given the model’s strong performance (backtest) for the OOT period, it can be concluded that the impact from such exclusion is expected to be small as the model development data still comprised of sufficient performance data covering both economic boom and bust. The model also included a full interest rate cycle in its development data set which included periods of rising and falling rates. Please refer to Chapter 2 for additional details on the interest rate cycle. A few transitions are heavily impacted by portfolio management policy over time, such as IVP transitions, are always statistically challenging to model in the first place. For additional details on the OOT backtest, please refer to Section 6.3 of the MDD. Please note the OOT sample only excluded model performance data specific to that period, loans that originated during the OOT period and were active post the conclusion of the OOT period are included as part of the development data sample.

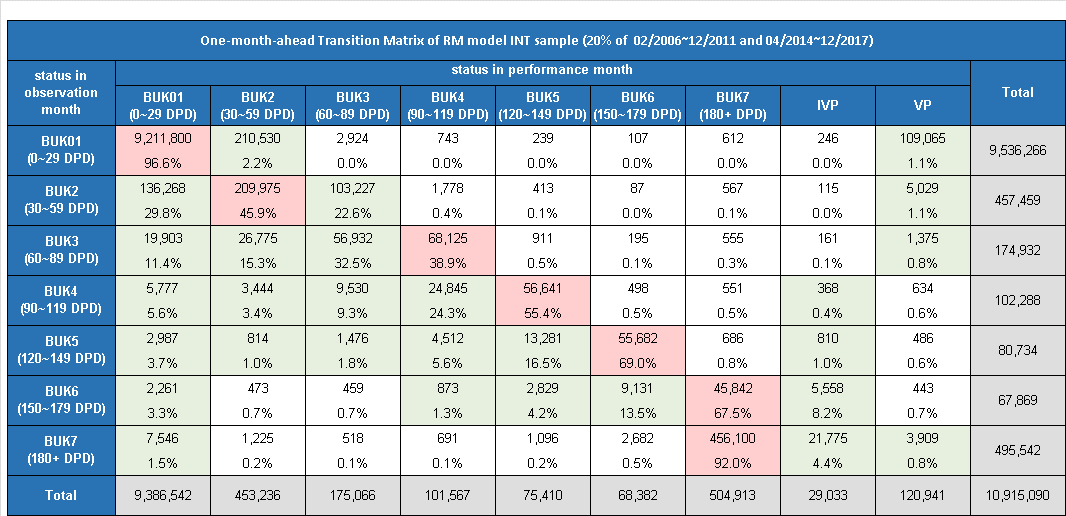
**Table 4.1.3.5 Sample size in Development/In-time Validation/Out-of-Time Validation Sample**

Table 4.1.3.6.1-3 provides the final one-month-ahead transition matrix of the model development, in-time-hold-out and out-of-time validation samples that are used estimate the CCAR 2019 never-modified RM PD model.

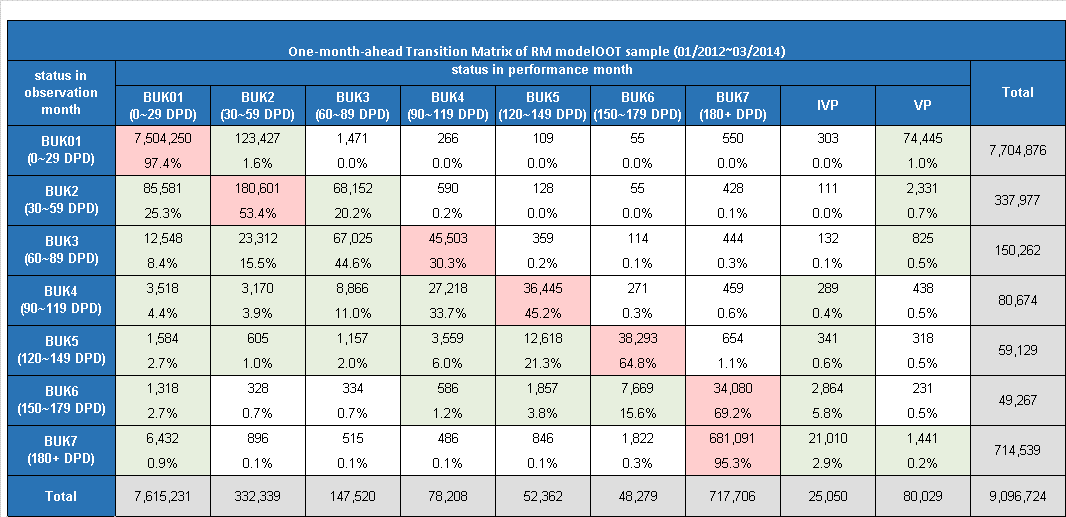
**Table 4.1.3.6.1 Transition matrix of RM Development Sample**



**Table 4.1.3.6.2 Transition matrix of RM In-Time Validation Sample**



**Table 4.1.3.6.3 Transition matrix of RM Out-of-time Validation Sample**



##### LGD Model

All samples in the severity dataset experienced involuntary payoff (IVP) resulting in terminating by default between Feb2005 -June2017. An IVP event is identified primarily by one of the following statuses: Short Sale (SS), Deed in Lieu of Foreclosure (DL), Foreclosure – Third Party Sale (3F), REO (RO), Charge Off (CO), or REO Sold (RS) in the Inactive Detail file located in CCR Risk MEP Optima; or having foreclosure/ short sale activities in foreclosure related tables in DRI. To ensure the correct population used in the severity model, CAMU collaborated with the risk team and only selected loans as the development samples if they satisfied the data water fall filtering in Table 4.1.3.7. As showed in below table, there were 582,301 unique accounts that defaulted from February 2005 to June 2017. After the data waterfall, there were 304,926 accounts remained for development, validation, and out-of-time checking.

**Table 4.1.3.7.: Data Water Fall for Severity Model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Exclusion Detail | Remain. # Unique Line | #Unique Records Exclude. | % Unique Records Exclude. |
|  | Loans with IVP terminal event as of Feb 2018 | **582,301** | - | - |
| 1 | Excluded Loans with inactive date not in the preliminary development period: Feb2005~Jun2017 | **575,413** | 6,888 | 1.2% |
| 2 | Excluded non-risk loans | **426,932** | 148,481 | 25.5% |
| 3 | Excluded repurchase loans | **421,710** | 5,222 | 0.9% |
| 4 | Excluded government incentive loans | **416,843** | 4,867 | 0.8% |
| 5 | Excluded loans with modification after inactive | **415,319** | 1,524 | 0.3% |
| 6 | Excluded non-VA government loans | **355,052** | 60,267 | 10.3% |
| 7 | Excluded loans without loss | **346,215** | 8,837 | 1.5% |
| 8 | Excluded loans with missing original balance or property amount | **343,998** | 2,217 | 0.4% |
| 9 | Excluded loans corresponding to one-timer loss adjustment1 | **337,994** | 6,004 | 1.0% |
| 10 | Further excluded loans before Jan 2008 to improve data quality | **304,926** | 33,068 | 5.7% |
|  | Total | 304,926 | 277,375 | 47.6% |

For additional information on one-timer events and treatments, please refer to attachment “3.2\_Memorandum\_One-Timer Exclusions - 111417.docx”

The NCL Severity model data segmented by three different loan types: (1) First Lien, (2) Second Lien, and (3) VA Loan. Except VA loans which are VA-insured, all models were designed for portfolios with loans not insured by government agency. Presented below in Tables 4.1.3.8 -4.1.3.11 illustrates the defaulted loan distribution by Modified vs Non-Modified loans, by product types (Residential vs HELOC vs FRHEL)

and by model development and validation samples. Please note within the scope of the Method A RM Model suite, we consider the non-modified residential mortgages only.

**Table 4.1.3.8: Summary of Defaulted Loan Distribution by MOD vs Non-MOD**



**Table 4.1.3.9: Summary of Defaulted Loan Distribution by Product Type**



**Table 4.1.3.10: Summary of Defaulted Loan Distribution(Non-CPB) by Development & Validation Samples**



**Table 4.1.3.11: Summary of Defaulted Loan Distribution(CPB) by Development & Validation Samples**



After the additional filtering by modification status and product type, the final sample size for the First Lien segment was approximated around 64,817 of which 68% (44,296) was reserved for model development. Tables 4.1.3.10 and 4.1.3.11 illustrates the distribution of defaults with the in time samples. The next set of Figures and Tables illustrate the loss trend and loss type by lien position and VA for the model development sample. For additional details, please refer to attachment ‘4.1.3 –Severity Trends’.xlsx’.

**Table 4.1.3.12: Distribution of Defaults by Loss Types (First Lien) for In Time Sample**



**Table 4.1.3.13: Distribution of Defaults by Loss Types (Second Lien) for In Time Sample**

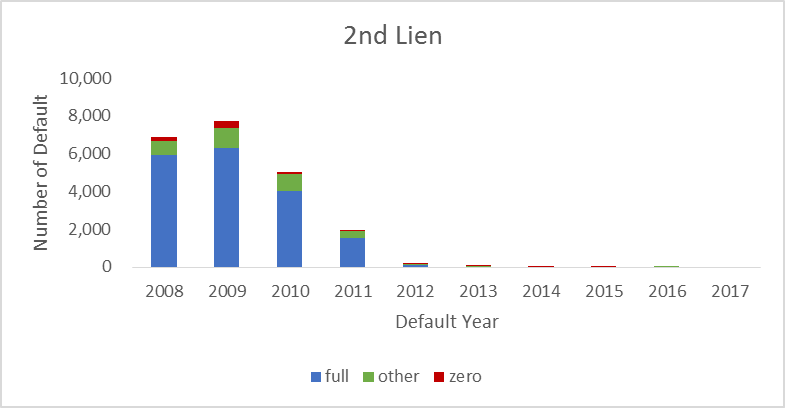


**Table 4.1.3.13: Distribution of Defaults by Loss Types (VA) for In Time Sample**



**Figure 4.1.3.1: Count of First Lien Defaults by Loss Types between Jan-2008 and June -2017.**

**Figure 4.1.3.2: Count of Second Lien Defaults by Loss Types between Jan-2008 and June -2017.**



**Figure 4.1.3.3: Count of VA Defaults by Loss Types between Jan-2008 and June -2017.**

The modeling team provides a data dictionary and the detail of summary statistics for all input variables used in the Method A RM Model suite. Overall, the predictor variables can be divided into three categories: macroeconomic variables, loan-level variables, and bureau variables. A macroeconomic variable, also known as a ‘state’ variable, changes over time. Account-level loan attributes are either constant over time or time-invariant. Certain time-varying account-level variables, such as current CLTV, are derived by combining a static account-level variable with a dynamic collateral variable. Both decision FICO and refresh FICO are used in the Method A RM model, the former is time-invariant while the latter changes over time. All these variables have in-depth description listed in the data dictionary and the statistical summary tables listed as below. Please refer to the Gating Principles within Chapter 3 for additional details on the listed variables.

To warrant high data quality, CAMU performs extreme value analysis to compliment business insight. When missing or outliers are identified, appropriate treatments are provided case by case. For instance, to smooth outliers, CAMU imputes data with caps and floors to keep the data fell within a reasonable range. After imputation, CAMU will inspect if there is any exclusion bias by examining the extreme values and the deviation from the mean. If no severe deviation found, it indicates the impact is not significant and the data is reliable. For scores that fall out of the range, they are replaced by the median value and a dummy variable is created to identify the imputation. Detail about the logic, reason of clean up, and the value before and after imputation are described in section 4.1.6. Please refer to attachment ‘4.1.3 - RM\_Method\_A\_Data\_Dictionary.xlsx’ for the RM PD model data dictionary

##### Table 4.1.3.13: Data Dictionary of Method A RM PD Model with Model Drivers

| **Output variable name** | **Description and definition** | **Related fields (if any)** | **Source field (or formula if calculated field)** |
| --- | --- | --- | --- |
| ARM\_5Yr | arm for 5 yr. | fixed\_rate\_ind, arm\_type\_year | B\_S\_ARM\_IND=FIXED\_RATE\_IND='N'; ARM\_5Yr=B\_S\_ARM\_IND=1 and ARM\_TYPE\_YEAR in ('5YR'); |
| ARM\_GT5Yr | arm > 5yr | fixed\_rate\_ind, arm\_type\_year | ARM\_GT5Yr=B\_S\_ARM\_IND=1 and (ARM\_LE3Yr^=1) and (ARM\_5Yr^=1); |
| B\_M\_coborrower\_IND | indicator of coborrower | ssn\_id | if in\_sec=1 or ssn\_id="1" then B\_M\_coborrower='1'; else B\_M\_coborrower='0'; |
| B\_M\_EVER\_MBA\_60P\_IND | indicator of ever mba past 60 days | ever\_mba\_60p\_ind |  |
| B\_M\_IN\_TRIAL\_IND | indicator of trial loan | Trial\_init\_mon, Trial\_comp\_mon | B\_M\_IN\_TRIAL\_IND= (.<TRIAL\_INIT\_MON\_ <= file\_dt <= TRIAL\_COMP\_MON\_   or (.<TRIAL\_INIT\_MON\_ <= file\_dt and TRIAL\_COMP\_MON\_=.)); |
| B\_S\_ARM\_IND | indicator of arm loans | fixed\_rate\_ind | renamed arm\_ind = B\_S\_ARM\_IND |
| B\_S\_BROK\_IND | indicator of channel via broker | channel\_cd | B\_S\_BROK\_IND=C\_S\_CHANNEL\_CD='BROK'; |
| B\_S\_CORR\_IND | indicator of channel via correspondence | channel\_cd | B\_S\_CORR\_IND=C\_S\_CHANNEL\_CD='CORR'; |
| B\_S\_DocFull\_IND | indicator of income in full at application | income\_doc\_cd | if INCOME\_DOC\_CD in ("CSTD","FULL") then B\_S\_DocFull='1'; else B\_S\_DocFull='0'; |
| B\_S\_DocLow\_IND | indicator of low income | income\_doc\_cd |  |
| B\_S\_DocMISS\_IND | indicator of missing income document at application | income\_doc\_cd | if INCOME\_DOC\_CD in (" ",".") then B\_S\_DocMISS='1'; else B\_S\_DocMISS='0'; |
| B\_S\_ge7YR\_ARM\_IND | indicator of arm year >= 7Yr | arm\_type\_year, orig\_term | B\_S\_ge7YR\_ARM\_IND = 0; if Fixed\_Rate\_IND='Y' then do;  if N\_S\_ORIG\_TERM<=180 then B\_S\_le15YR\_FRM\_IND = 1;else B\_S\_gt15YR\_FRM\_IND = 1; end; else do;  if ARM\_TYPE\_YEAR in ('1YR','2YR','3YR','0MO','1MO','2MO','3MO','6MO','4YR') then B\_S\_lt5YR\_ARM\_IND=1;else   if ARM\_TYPE\_YEAR = '5YR' then B\_S\_5YR\_ARM\_IND = 1;else   if ARM\_TYPE\_YEAR in ('7YR','8YR','OTHER','10YR','6YR','') then B\_S\_ge7YR\_ARM\_IND = 1; end; |
| B\_S\_gt15YR\_FRM\_IND | indicator or fixed rate term >= 15 Yr | arm\_type\_year, orig\_term | B\_S\_ge7YR\_ARM\_IND = 0; if Fixed\_Rate\_IND='Y' then do;  if N\_S\_ORIG\_TERM<=180 then B\_S\_le15YR\_FRM\_IND = 1;else B\_S\_gt15YR\_FRM\_IND = 1; end; else do;  if ARM\_TYPE\_YEAR in ('1YR','2YR','3YR','0MO','1MO','2MO','3MO','6MO','4YR') then B\_S\_lt5YR\_ARM\_IND=1;else   if ARM\_TYPE\_YEAR = '5YR' then B\_S\_5YR\_ARM\_IND = 1;else   if ARM\_TYPE\_YEAR in ('7YR','8YR','OTHER','10YR','6YR','') then B\_S\_ge7YR\_ARM\_IND = 1; end; |
| B\_S\_INT\_ONLY\_IND | indicator of trial loan | Trial\_init\_mon, Trial\_comp\_mon | B\_M\_IN\_TRIAL\_IND= (.<TRIAL\_INIT\_MON\_ <= file\_dt <= TRIAL\_COMP\_MON\_   or (.<TRIAL\_INIT\_MON\_ <= file\_dt and TRIAL\_COMP\_MON\_=.)); |
| B\_S\_le15YR\_FRM\_IND | Indicator of fixed rate term <= 15 yr | arm\_type\_year, orig\_term | if Fixed\_Rate\_IND='Y' then do;  if N\_S\_ORIG\_TERM<=180 then B\_S\_le15YR\_FRM\_IND = 1;else B\_S\_gt15YR\_FRM\_IND = 1; end; |
| B\_S\_lienfirst\_IND | indicator of first lien | lien\_cd | Lienfirst\_ind=(B\_S\_lien\_CD='0'); |
| B\_S\_lien2nd\_IND | indicator of 2nd lien | lien\_cd | Lien2nd\_ind=(B\_S\_lien\_CD='1'); |
| B\_S\_LoanPurpMiss\_IND | indicator of missing loan purchase info | loan\_purp\_cd | if LOAN\_PURP\_CD in ("."," ") then B\_S\_LoanPurpMiss='1'; else B\_S\_LoanPurpMiss='0'; |
| B\_S\_LoanPurpPurch\_IND | indicator of purchase loans | loan\_purp\_cd | if LOAN\_PURP\_CD in ("PURCH") then B\_S\_LoanPurpPurch='1'; else B\_S\_LoanPurpPurch='0'; |
| B\_S\_LoanPurpWork\_IND | indicator of work out | loan\_purp\_cd | if LOAN\_PURP\_CD in ("WORKO") then B\_S\_LoanPurpWork='1'; else B\_S\_LoanPurpWork='0' |
| B\_S\_lowIncDoc\_IND | indicator of low income | apprved\_type\_cd | if approval\_type\_cd in(3,10,12,13,14,18,24,25,53,54,57,60,61,64,70,73,75,76,84,87,90,91,92,93,101,102,  103,104,105, 106,107,117,118,119,120,121,122,123,126,128,129,130,131,133,137,138,200,204,205,206)  or approval\_type\_cd in(85,86,88,89,124,201,202) then B\_S\_lowIncDoc='1'; else B\_S\_lowIncDoc='0';/\*11/19/2016 Shuangxi changed\*/ |
| B\_S\_missS\_M\_FicoRefresh | Indicator of both decision fico and refresh fico are missing | Dcsn\_fico, beacon50\_score | B\_S\_missS\_M\_FicoRefresh = 0; S\_M\_FicoRefresh2 =S\_M\_FicoRefresh; if miss\_S\_S\_FicoDcsn = 1 and miss\_S\_M\_FicoRefresh = 1 then B\_S\_missS\_M\_FicoRefresh = 1; else if miss\_S\_S\_FicoDcsn = 0 and miss\_S\_M\_FicoRefresh = 1 then do;  if N\_M\_MOB <= 3 then S\_M\_FicoRefresh2 = S\_S\_FicoDcsn;  else do;  S\_M\_FicoRefresh2 = S\_S\_FicoDcsn;  B\_S\_missS\_M\_FicoRefresh = 1;  end; end; |
| B\_S\_OCC\_IV\_IND | indicator of occupancy = investment | occ\_type\_cd | B\_S\_OCC\_IV\_IND=C\_S\_OCC\_TYPE\_CD='IV'; |
| B\_S\_OCC\_OO\_IND | indicator of occupancy = owner occupied | occ\_type\_cd | B\_S\_OCC\_OO\_IND=C\_S\_OCC\_TYPE\_CD='OO'; |
| B\_S\_OCC\_SH\_IND | indicator of occupancy = second home | occ\_type\_cd | B\_S\_OCC\_SH\_IND=C\_S\_OCC\_TYPE\_CD='SH'; |
| B\_S\_pmi\_IND | indicator of private mortage insurance coverage | loan\_type\_cd | if loan\_type\_cd=8 then B\_S\_pmi='1'; else B\_S\_pmi='0'; |
| B\_S\_PropCondoCP\_IND | indicator of property = condo | prop\_type\_cd | if PROP\_TYPE\_CD in ("COOP","COND","MAFG") then B\_S\_PropCondoCP='1'; else B\_S\_PropCondoCP='0'; |
| B\_S\_PropMulti\_IND | indicator of property = multiple | prop\_type\_cd | if PROP\_TYPE\_CD in ("2FAM","3FAM","4FAM","5FAM","MULT") then B\_S\_PropMulti='1'; else B\_S\_PropMulti='0'; |
| B\_S\_PropSFR\_IND | indicator of single family | prop\_type\_cd | if PROP\_TYPE\_CD in ("SFR") then B\_S\_PropSFR='1'; else B\_S\_PropSFR='0'; |
| B\_S\_WLST\_IND | indicator of channel via WLST | channel\_cd | B\_S\_WLST\_IND=C\_S\_CHANNEL\_CD='WLST'; |
| B\_S\_WLST\_IND\_2nd |  |  | B\_S\_WLST\_IND\_2nd=B\_S\_WLST\_IND\*B\_S\_lien2nd\_IND; |
| Balloon | indicator of balloon loan | balloon\_ind | Balloon=(balloon\_ind='Y'); |
| Balloon\_Mature | Indicator of balloon loan after loan is mature | balloon\_ind, N\_M\_MOB, N\_S\_Orig\_term | Balloon\_Mature=Balloon\*(N\_M\_MOB\_cap>=N\_S\_ORIG\_TERM); |
| community\_loan\_ind | Indicator or community loans |  |  |
| cpb\_ind | indicator CPB loan |  |  |
| curr\_neg | HPI in negative | HPI |  |
| D\_M\_PRIN\_BAL\_10k\_SP5LSP15\_CMI | principle balance for CPB loans capped at 150K and floored at 50K | Prin\_bal | D\_M\_PRIN\_BAL\_in10k\_SP5LSP15 = min(max(D\_M\_PRIN\_BAL\_in10k,5),15); D\_M\_PRIN\_BAL\_10k\_SP5LSP15\_CMI = D\_M\_PRIN\_BAL\_in10k\_SP5LSP15\*(cpb\_ind=0); |
| D\_M\_PRIN\_BAL\_in10K | principle balance in 10K | Prin\_bal | D\_M\_PRIN\_BAL\_in10K=D\_M\_PRIN\_BAL/1e4; |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_first | principle balance in 10K for first lien loans (non-CPB) | Prin\_bal | D\_M\_PRIN\_BAL\_in10K\_CMI\_first =D\_M\_PRIN\_BAL\_in10K\_CMI\*(B\_S\_Lien2nd\_IND=0); |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_2nd | principle balance in 10K for 2nd lien loans (non-CPB) | Prin\_bal | D\_M\_PRIN\_BAL\_in10K\_CMI\_2nd =D\_M\_PRIN\_BAL\_in10K\_CMI\*(B\_S\_Lien2nd\_IND=1); |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_CONV | principle balance in 10K for non GOV (non-CPB) | Prin\_bal | D\_M\_PRIN\_BAL\_in10K\_CMI=D\_M\_PRIN\_BAL\_in10K\*(CPB\_IND=0); D\_M\_PRIN\_BAL\_in10K\_CMI\_CONV=D\_M\_PRIN\_BAL\_in10K\_CMI\*(GOV=0); |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_HPIDecr | principle balance in 10K for non CPB loans with HPI in past 12mos <= 1 (non-CPB) | Prin\_bal | D\_M\_PRIN\_BAL\_in10K\_CMI\_HPIDecr =D\_M\_PRIN\_BAL\_in10K\_CMI\*(HPI\_tm12\_ratio<=1); |
| D\_M\_PRIN\_BAL\_in10k\_CPB | principle balance in 10k for CPB loans | Prin\_bal | D\_M\_PRIN\_BAL\_in10K\_CPB=D\_M\_PRIN\_BAL\_in10K\*CPB\_IND; |
| D\_M\_PRIN\_BAL\_in10k\_LSP10 | spline principle balance in 10k <= 100K | Prin\_bal | D\_M\_PRIN\_BAL\_in10K\_LSP10=min(D\_M\_PRIN\_BAL\_in10K,10) |
| D\_M\_PRIN\_BAL\_in10k\_LSP20\_CMICONV | spline principle balance in 10k <= 200K (non-CPB and non-GOV) | Prin\_bal | D\_M\_PRIN\_BAL\_in10k\_LSP20 = min(D\_M\_PRIN\_BAL\_in10k,20); D\_M\_PRIN\_BAL\_in10k\_LSP20\_CMI = D\_M\_PRIN\_BAL\_in10k\_LSP20\*(cpb\_ind=0); D\_M\_PRIN\_BAL\_in10k\_LSP20\_CMICONV = D\_M\_PRIN\_BAL\_in10k\_LSP20\_CMI\*(GOV=0); |
| D\_M\_PRIN\_BAL\_in10k\_SP10\_GOV | spline principle balance in 10K >=100K (GOV) | Prin\_bal, portfolio\_type\_cd | D\_M\_PRIN\_BAL\_in10K\_SP10=max(D\_M\_PRIN\_BAL\_in10K,10); D\_M\_PRIN\_BAL\_in10K\_SP10\_GOV = D\_M\_PRIN\_BAL\_in10K\_SP10\*(GOV=0); |
| D\_M\_PRIN\_BAL\_in10k\_SP20\_CMICONV | spline principle balance in >=200K (non-CPB and non-GOV) | Prin\_bal, portfolio\_type\_cd | D\_M\_PRIN\_BAL\_in10k\_SP20\_CMICONV = D\_M\_PRIN\_BAL\_in10k\_SP20\_CMI\*\*(GOV=0); |
| D\_M\_PRIN\_BAL\_LE2K | principle balance <= 2K | Prin\_bal | D\_M\_PRIN\_BAL\_LE2K=D\_M\_PRIN\_BAL\_in10K<=0.2; |
| D\_M\_PRIN\_BAL\_LE5K | principle balance <= 5K | Prin\_bal | D\_M\_PRIN\_BAL\_LE5K=D\_M\_PRIN\_BAL\_in10K<=0.5; |
| D\_M\_PRIN\_BAL\_SP100K\_CMI\_first | principle balance >=100K (non-CPB and non-GOV) | Prin\_bal | D\_M\_PRIN\_BAL\_SP100K\_CMI = D\_M\_PRIN\_BAL\_SP100K\*(cpb\_ind=0); D\_M\_PRIN\_BAL\_SP100K\_CMI\_first =D\_M\_PRIN\_BAL\_SP100K\_CMI\*(B\_S\_Lien2nd\_IND=0); |
| D\_M\_PRIN\_BAL\_SP100K\_CMI\_HPIDecr | spline principle balance in 10K > 100K for non CPB loans with HPI in past 12mos <= 1 (non-CPB) | Prin\_bal , lien\_cd | D\_M\_PRIN\_BAL\_SP100K\_CMI\_HPIDecr =D\_M\_PRIN\_BAL\_SP100K\_CMI\*(HPI\_tm12\_ratio<=1); |
| DEC | Indicator of month Dec | file\_dt | MON=MONTH(FILE\_DT); DEC=MON=12; |
| FCL\_Moratorium\_Dec09 | foreclosure moratorium is enacted in Dec'09 | file\_dt | FCL\_Moratorium\_JF09 =('01jan2009'd<=file\_dt<='01feb2009'd); |
| FCL\_Moratorium\_JF09 | foreclosure moratorium is enacted between Jan'09 and Feb'09 | file\_dt | FCL\_Moratorium\_Dec09 = file\_dt = mdy(12,1,2009); |
| fcl\_time | foreclosre time |  |  |
| fcl\_time\_CONV | foreclosre time interacts with non-GOV indicator | fcl\_sale\_dt, fcl\_begin\_dt, portfolio\_type\_cd | fcl\_time\_CONV = fcl\_time\*(GOV=0); |
| FEB | indicator of month Februrary | file\_dt | MON=MONTH(FILE\_DT); FEB=MON=2; |
| FRM\_LE15Yr | fixed rate term <=15 Yr | fixed\_rate\_ind | FRM\_LE15Yr=B\_S\_ARM\_IND=0 and N\_S\_ORIG\_TERM<=180; |
| GOV | indicator of GOV loans | portfolio\_type\_cd | GOV=C\_M\_PORTFOLIO\_TYPE\_CD='GOV'; |
| HPI\_APP | HPI index (app) associates with time of application | HPI | HPI\_APP=HPI\_M0\_v4/HPI\_Orig\_v4; |
| HPI\_APP\_LE1 | HPI index (app) <=1 | HPI | HPI\_APP\_LE1=HPI\_APP<=1; |
| HPI\_APP\_LSP1\_CMI | HPI index (app) <=1 for non-CPB loans | HPI | HPI\_APP\_LSP1\_CMI =HPI\_APP\_LSP1\*(CPB\_IND=0); |
| HPI\_tm12\_ratio | HPI index (tm12) associates with past 12 mos | HPI |  |
| HPI\_tm12\_ratio\_CMI | HPI index (tm12) associates with past 12 mos for non-CPB | HPI | HPI\_tm12\_ratio\_CMI= HPI\_tm12\_ratio\*(CPB\_IND=0); |
| HPI\_tm12\_ratio\_LE1 | indicator of HPI index (tm12) <=1 | HPI | HPI\_TM12\_RATIO\_LE1=HPI\_TM12\_RATIO<=1; |
| HPI\_tm12\_ratio\_LSP1 | HPI index (tm12) <=1 | HPI | HPI\_tm12\_ratio\_LSP1=min(HPI\_tm12\_ratio,1); |
| HPI\_tm12\_ratio\_LSP1\_CMI\_CONV | HPI index (tm12) <=1 interacts with non-CPB and non-GOV loans indicators | HPI, portfolio\_type\_cd | HPI\_tm12\_ratio\_LSP1\_CMI= HPI\_tm12\_ratio\_LSP1\*(CPB\_IND=0); HPI\_tm12\_ratio\_LSP1\_CMI\_CONV = HPI\_tm12\_ratio\_LSP1\_CMI\*(GOV=0); |
| HPI\_tm12\_ratio\_SP1 | HPI index (tm12) >=1 | HPI | HPI\_tm12\_ratio\_SP1=max(HPI\_tm12\_ratio,1); |
| HPI\_tm12\_ratio\_SP1\_2nd | HPI index (tm12) >=1 for 2nd lien | HPI, lien\_cd | HPI\_tm12\_ratio\_SP1\_2nd = HPI\_tm12\_ratio\_SP1\*(B\_S\_Lien2nd\_IND=1); |
| HPI\_tm12\_ratio\_SP1\_MOBGT120 | interaction of HPI index (tm12) with MOB > 120 | HPI | HPI\_tm12\_ratio\_SP1\_MOBGT120=HPI\_tm12\_ratio\_SP1\*N\_M\_MOB\_GT120; |
| HPI\_tm12\_ratio\_WLST\_CMI | HPI index (tm12) assocates with non-CPB loans via Wall Street channel | HPI, channel\_cd | B\_S\_WLST\_IND=C\_S\_CHANNEL\_CD='WLST'; HPI\_tm12\_ratio\_WLST=HPI\_tm12\_ratio\*B\_S\_WLST\_IND; HPI\_tm12\_ratio\_WLST\_CMI = HPI\_tm12\_ratio\_WLST\*(CPB\_IND=0); |
| income\_12m\_ratio | Ratio of Income in past 12 mos | income | if income\_12m^=. then income\_12m\_ratio=income/income\_12m; |
| income\_12m\_ratio\_LSP1 | Ratio of income in past 12 mos with spine <= 1 | income | income\_12m\_ratio\_LSP1=min(income\_12m\_ratio,1 ); |
| INX\_SP500\_12m\_ratio | ratio of SP500 in past 12 mos | SP500 | if INX\_SP500\_12m^=. then INX\_SP500\_12m\_ratio=INX\_SP500/INX\_SP500\_12m; |
| INX\_SP500\_12m\_ratio\_LSP1 | ratio of SP500 in past 12 mos with spine <= 1 | SP500 | INX\_SP500\_12m\_ratio\_LSP1 = min(INX\_SP500\_12m\_ratio,1); |
| INX\_SP500\_12m\_ratio\_SP1 |  | SP500 | INX\_SP500\_12m\_ratio\_SP1=max(INX\_SP500\_12m\_ratio,1); |
| INX\_SP500\_3m\_ratio | ratio of SP500 in past 3 mos | SP500 | INX\_SP500\_3m=lag3(INX\_SP500); if INX\_SP500\_3m^=. then INX\_SP500\_3m\_ratio=INX\_SP500/INX\_SP500\_3m; |
| INX\_SP500\_3m\_ratio\_CONV | ratio of SP500 in past 3 mosfor non-GOV loans | SP500, portfolio\_type\_cd | INX\_SP500\_3m\_ratio\_CONV= INX\_SP500\_3m\_ratio\*(GOV=0); |
| INX\_SP500\_tm\_CONV | SP500 of the month for non-GOV loans | SP500, portfolio\_type\_cd | INX\_SP500\_tm =INX\_SP500; INX\_SP500\_tm\_CONV=INX\_SP500\_tm\*(GOV=0); |
| JAN | indicator of month January | file\_dt | MON=MONTH(FILE\_DT); JAN=MON=1; |
| judicial | judicial states | prop\_state\_cd | judicial= (C\_S\_PROP\_STATE\_CD2 in ('CT', 'DE', 'FL', 'IL', 'IN', 'IA', 'KS', 'KY', 'LA', 'MA', 'MD', 'ME', 'ND',   'NJ', 'NM', 'NY', 'OH', 'OK', 'PA', 'PR', 'SC', 'SD', 'UT', 'VT', 'VI', 'WI')); |
| judicial\_CONV | Indicator of judicial state associates with non-GOV loans | prop\_state\_cd, portfolio\_type\_cd | judicial\_CONV = judicial\*(GOV=0); |
| JUL | indicator of month July | file\_dt | MON=MONTH(FILE\_DT); JUL=MON=7; |
| JUN | indicator of month June | file\_dt | JUN=MON=6; |
| Katrina | indicator of area covered by storm Katrina | prop\_zip\_cd | Katrina=(Katrina\_area and mdy(8,1,2005)<=file\_dt<=mdy(7,1,2007)); |
| MAR | Indicator of month March | file\_dt | MAR=MON=3; |
| MAY | indicator of month May | file\_dt | MAY=MON=5; |
| miss\_N\_M\_Age\_Oldest\_Mtg | indicator of missing related info | cmi16 | miss\_N\_M\_Age\_Oldest\_Mtg = missing(N\_M\_Age\_Oldest\_Mtg); |
| miss\_N\_M\_age\_oldest\_trd | indicator of missing related info | var\_k | miss\_N\_M\_age\_oldest\_trd=missing(N\_M\_age\_oldest\_trd); |
| miss\_N\_M\_n\_open\_mort | indicator of missing related info | var\_a39 | miss\_N\_M\_n\_open\_mort=missing(N\_M\_n\_open\_mort); |
| MTM\_CLTV | ratio of MTM to CLTV | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_LTV=D\_M\_PRIN\_BAL/(D\_S\_ORIG\_PROP\_AMT\*HPI\_M0/HPI\_ORIG)\*100; MTM\_CLTV=MTM\_LTV; if B\_S\_lienfirst^='1' then do;  if D\_S\_OrigBalance>0 and D\_S\_FST\_MTG\_BAL>0 then do;  Junior\_Ratio\_Orig=min(max(0,D\_S\_OrigBalance/(D\_S\_FST\_MTG\_BAL+1e-6)),10);   MTM\_CLTV=MTM\_LTV\*(1+1/Junior\_Ratio\_Orig);  end;else   MTM\_CLTV=MTM\_LTV\*(P\_S\_OrigCLTV/P\_S\_OrigLTV); end; |
| MTM\_CLTV\_GT100\_GOV | Market to market CLTV with spline > 100 (GOV) | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV, portfolio\_type\_cd | MTM\_CLTV\_GT100\_GOV=(MTM\_CLTV>100)\*(GOV=1); |
| MTM\_CLTV\_LE10 | indicator of arket to market CLTV <=10 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_LE10=MTM\_CLTV<=10; |
| MTM\_CLTV\_LSP100 | Market to market CLTV with spline <=100 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_LSP100=min(MTM\_CLTV,100); |
| MTM\_CLTV\_LSP60 | Market to market CLTV with spline <=60 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_LSP60=min(MTM\_CLTV,60); |
| MTM\_CLTV\_MOBGT120 | interaction of market to market CLTV with MOB >120 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_MOBGT120=MTM\_CLTV\*N\_M\_MOB\_GT120; |
| MTM\_CLTV\_SP40\_first\_CONV | Market to market CLTV with spline >=40 for non-CPB first lien loans | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV, portfolio\_type\_cd | MTM\_CLTV\_SP40=max(MTM\_CLTV,40); MTM\_CLTV\_SP40\_first = MTM\_CLTV\_SP40 \* B\_S\_lienfirst\_IND; MTM\_CLTV\_SP40\_first\_CONV= MTM\_CLTV\_SP40\_first\*(GOV=0); |
| MTM\_CLTV\_SP40LSP120 | Market to market CLTV with spline capped at 120 and floored at 40 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_SP40LSP120=min(max(MTM\_CLTV,40),120); |
| MTM\_CLTV\_SP40LSP120\_GOV | Market to market CLTV with spline capped at 120 and floored at 40 (GOV) | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV, portfolio\_type\_cd | MTM\_CLTV\_SP40LSP120\_GOV=MTM\_CLTV\_SP40LSP120\*GOV |
| MTM\_CLTV\_SP60 | Market to market CLTV with spline >=60 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_SP60=max(MTM\_CLTV,60); |
| MTM\_CLTV\_SP60\_HPIDcr | Market to market CLTV >=60 with HPI in past 12mos <= 1 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_SP60\_HPIDcr = MTM\_CLTV\_SP60\*(HPI\_tm12\_ratio<=1); |
| MTM\_CLTV\_SP60LSP80\_CPB | Market to market CLTV capped at 80 and floored at 60 for CPB loans | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_SP60LSP80\_CPB = MTM\_CLTV\_SP60LSP80\*cpb\_ind; |
| MTM\_CLTV\_SP80 | Market to market CLTV with spline >=80 | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV | MTM\_CLTV\_SP80=max(MTM\_CLTV,80); |
| MTM\_CLTV\_SP80\_first\_CONV | Market to market CLTV with spline >=80 for non-GOV first lien loan | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV, portfolio\_type\_cd | MTM\_CLTV\_SP80\_first = MTM\_CLTV\_SP80 \* B\_S\_lienfirst\_IND MTM\_CLTV\_SP80\_first\_CONV= MTM\_CLTV\_SP80\_first\*(GOV=0); |
| MTM\_CLTV\_SP80\_2nd\_CONV | Market to market CLTV with spline >=80 for non-GOV 2nd lien loan | prin\_bal, orig\_prop\_amt, HPI,fst\_mtg\_bal, origBalance, origCLTV, origLTV, portfolio\_type\_cd | MTM\_CLTV\_SP80\_2nd\_CONV= MTM\_CLTV\_SP80\_2nd\*(GOV=0); |
| N\_M\_Age\_Oldest\_Mtg | age of oldest mortgage (bureau) | cmi16 | N\_M\_Age\_Oldest\_Mtg = input(CMI16,20.); |
| N\_M\_age\_oldest\_trd | age of oldest trade (bureau) | var\_k | age\_oldest\_trd=input(var\_k,20.); |
| N\_M\_Curr\_lt\_Orig\_Rate\_LSP48 | current note rate > origination rate with spline <= 48 | curr\_note\_rate, | N\_M\_Curr\_lt\_Orig\_Rate\_LSP48 = min(N\_M\_Curr\_lt\_Orig\_Rate,48); |
| N\_M\_logHpiChange\_Orig | change of HPI at origination with log transformation | HPI | if (hpi\_m0\*hpi\_orig > 0) then N\_M\_logHpiChange\_Orig=log(hpi\_m0/hpi\_orig);  else N\_M\_logHpiChange\_Orig = 1; |
| N\_M\_logHpiChange\_Orig\_LSP0 | change of HPI associates with the origination with log transformation. Spline sets to <= 0 | HPI | N\_M\_logHpiChange\_Orig\_LSP0=min(N\_M\_logHpiChange\_Orig,0); |
| N\_M\_logHpiChange\_Orig\_SP0 | change of HPI associates with the origination with log transformation. Spline sets to >= 0 | HPI | N\_M\_logHpiChange\_Orig\_SP0=max(N\_M\_logHpiChange\_Orig,0); |
| N\_M\_MOB\_cap\_WLST\_CONV | MOB with cap associates with non-GOV loans from WLST channel | loan\_orig\_dt, file\_dt, portfolio\_type\_cd | N\_M\_MOB\_cap=min(N\_M\_MOB,180); N\_M\_MOB\_cap\_WLST=N\_M\_MOB\_cap\*B\_S\_WLST\_IND; N\_M\_MOB\_cap\_WLST\_CONV= N\_M\_MOB\_cap\_WLST\*(GOV=0); |
| N\_M\_MOB\_LSP36 | MOB <= 36 | loan\_orig\_dt, file\_dt | N\_M\_MOB\_LSP36=min(N\_M\_MOB\_cap,36); |
| N\_M\_MOB\_LSP36\_GOV | MOB <= 36 (GOV) | loan\_orig\_dt, file\_dt, portfolio\_type\_cd | N\_M\_MOB\_LSP36\_GOV=N\_M\_MOB\_LSP36\*GOV; |
| N\_M\_MOB\_LSP72 | MOB <= 72 | loan\_orig\_dt, file\_dt | N\_M\_MOB\_LSP72=min(N\_M\_MOB\_cap,72); |
| N\_M\_MOB\_SP36 | MOB >= 36 | loan\_orig\_dt, file\_dt | N\_M\_MOB\_SP36=max(N\_M\_MOB\_cap,36); |
| N\_M\_MOB\_SP36LSP180 | MOB capped at 180 and floored at 36. | loan\_orig\_dt, file\_dt | N\_M\_MOB\_SP36LSP180=min(max(N\_M\_MOB,36),180); |
| N\_M\_n\_open\_mort | number of open mortgage (bureau) | var\_a39 | n\_open\_mort=input(var\_a39,20.); |
| never\_neg | indicator for HPI before negative | hpi | never\_neg=((.<file\_dt<firstdate\_neg) or firstdate\_neg=.); |
| OCT | indicator of month October |  | OCT=MON=10; |
| P\_AND\_I\_ratio\_ARM | ratio of principal interest amount with the amount in past 6 mos for ARM loans | principal\_interest\_amt, curr\_note\_rate, prin\_bal, fixed\_rate\_ind | P\_AND\_I\_ratio\_ARM=P\_AND\_I\_ratio\*B\_S\_ARM\_IND; |
| P\_AND\_I\_ratio\_IO | ratio of principal interest amount with the amount in past 6 mos for loans with IO | principal\_interest\_amt, curr\_note\_rate, prin\_bal, int\_only\_ind | P\_AND\_I\_ratio\_IO=P\_AND\_I\_ratio\*B\_S\_INT\_ONLY\_IND; |
| P\_M\_CURR\_NOTE\_RATE\_ARM | Interaction of current note rate and ARM indicator | curr\_note\_rate, fixed\_rate\_ind | P\_M\_CURR\_NOTE\_RATE\_ARM=P\_M\_CURR\_NOTE\_RATE\*B\_S\_ARM\_IND; |
| P\_M\_CURR\_NOTE\_RATE\_ARM\_CONV | Interaction of current note rate and ARM indicator for non-GOV loans | curr\_note\_rate, fixed\_rate\_ind, portfolio\_type\_cd | P\_M\_CURR\_NOTE\_RATE\_ARM=P\_M\_CURR\_NOTE\_RATE\*B\_S\_ARM\_IND; P\_M\_CURR\_NOTE\_RATE\_ARM\_CONV=P\_M\_CURR\_NOTE\_RATE\_ARM\*(GOV=0); |
| P\_M\_PresIntSpread\_LSP3pt5\_CMI | Interest spread <=3.5 for non-CPB | curr\_note\_rate, swap | P\_M\_PresIntSpread\_LSP3pt5 = min(P\_M\_PresIntSpread,3.5); P\_M\_PresIntSpread\_LSP3pt5\_CMI = P\_M\_PresIntSpread\_LSP3pt5\*(cpb\_ind =0); |
| P\_M\_PresIntSpread\_LSP3pt5\_FRM | Interaction of Interest spread with fixed rate indicator | curr\_note\_rate, swap, FIXED\_RATE\_IND | P\_M\_PresIntSpread\_LSP3pt5\_FRM = P\_M\_PresIntSpread\_LSP3pt5\*B\_S\_FIXED\_RATE\_IND; |
| P\_M\_PresIntSpread\_LSP5\_CPB | Interest spread <=5 (CPB) |  | P\_M\_PresIntSpread\_LSP5 = min(P\_M\_PresIntSpread,5); P\_M\_PresIntSpread\_LSP5\_CPB = P\_M\_PresIntSpread\_LSP5\*cpb\_ind; |
| P\_M\_PresIntSpread\_LSP5\_PRA | interest spread spline interacts with pre-arm reset indicator |  | P\_M\_PresIntSpread\_LSP5\_PRA = P\_M\_PresIntSpread\_LSP5\*(B\_S\_PreResetArm\_IND); |
| P\_M\_PresIntSpread\_LSP5\_Pre2010 | interest spread spline interacts with prior 2010 indicator |  | P\_M\_PresIntSpread\_LSP5\_Pre2010 = P\_M\_PresIntSpread\_LSP5\*(Post2010\_Orig=0); |
| P\_M\_State\_Unemp\_Rate | unemployment rate in state level | unemployment |  |
| P\_M\_State\_Unemp\_Rate\_2nd | Unemployment in State level interacts with 2nd lien indicator | unemployment, lien\_cd | P\_M\_State\_Unemp\_Rate\_2nd=P\_M\_State\_Unemp\_Rate\*(B\_S\_Lien2nd\_IND=1); |
| P\_M\_State\_Unemp\_Rate\_currneg | Unemployment in State level interacts with current non-negative indicator | unemployment, HPI | P\_M\_State\_Unemp\_Rate\_currneg=P\_M\_State\_Unemp\_Rate\*curr\_neg. |
| P\_M\_State\_Unemp\_Rate\_lag12\_GOV | unemployment in State level with 12 mos lag interacts with GOV indicator | unemployment, portfolio\_type\_cd | P\_M\_State\_Unemp\_Rate\_lag12=P\_M\_State\_Unemp\_Rate/R\_M\_State\_UnempB12M; P\_M\_State\_Unemp\_Rate\_lag12\_GOV= P\_M\_State\_Unemp\_Rate\_lag12\*GOV; |
| P\_S\_BE\_DEBT\_RATIO | ratio of income paying debt | be\_debt\_ratio | rename BE\_DEBT\_RATIO = P\_S\_BE\_DEBT\_RATIO; |
| P\_S\_OrigIntSpread | interest spread at origination | orig\_note\_rate, swap | P\_S\_OrigIntSpread=orig\_note\_rate - market\_O; |
| P\_S\_OrigIntSpread\_FRM | interest spread interacts with fixed rate indicator | orig\_note\_rate, swap, fixed\_rate\_ind | P\_S\_OrigIntSpread\_FRM=P\_S\_OrigIntSpread\*(1-B\_S\_ARM\_IND);; |
| POST\_ARM\_12M\_IND | indicator of 6 mos before ARM | arm\_type\_year | if B\_S\_ARM\_IND = 1 and POST\_ARM\_IND = 1 and -12<=ARM\_MonthLeft<=0 then POST\_ARM\_12M\_IND = 1; |
| POST\_ARM\_6M\_IND | indicator of 6 mos before ARM | arm\_type\_year | IF ARM\_TYPE\_YEAR = '1YR' THEN RESET\_period=12; ELSE  IF ARM\_TYPE\_YEAR = '1MO' THEN RESET\_period=1; ELSE  IF ARM\_TYPE\_YEAR = '10YR' THEN RESET\_period=120; ELSE  IF ARM\_TYPE\_YEAR = '2YR' THEN RESET\_period=24; ELSE  IF ARM\_TYPE\_YEAR = '2MO' THEN RESET\_period=2; ELSE  IF ARM\_TYPE\_YEAR = '3YR' THEN RESET\_period=36; ELSE  IF ARM\_TYPE\_YEAR = '3MO' THEN RESET\_period=3; ELSE  IF ARM\_TYPE\_YEAR = '4YR' THEN RESET\_period=48; ELSE  IF ARM\_TYPE\_YEAR = '5YR' THEN RESET\_period=60; ELSE  IF ARM\_TYPE\_YEAR = '6YR' THEN RESET\_period=72; ELSE  IF ARM\_TYPE\_YEAR = '6MO' THEN RESET\_period= 6; ELSE  IF ARM\_TYPE\_YEAR = '7YR' THEN RESET\_period=84; ELSE  IF ARM\_TYPE\_YEAR = '8YR' THEN RESET\_period=96; ELSE  IF ARM\_TYPE\_YEAR ='OTHER' THEN RESET\_period= 1; ELSE   RESET\_period=999;   if RESET\_period=999 then ARM\_MonthLeft=999;  else ARM\_MonthLeft=RESET\_period-N\_M\_MOB;  POST\_ARM\_IND=.<ARM\_MonthLeft<=0;  POST\_ARM\_6M\_IND=-6<ARM\_MonthLeft<=0; |
| POST\_ARM\_IND | Post arm indicator | arm\_type\_year | POST\_ARM\_IND=.<ARM\_MonthLeft<=0; |
| POST\_ARM\_IND\_WLST | interaction of post arm indicator with WLST indicator | fixed\_rate\_ind, channel\_cd | POST\_ARM\_IND\_WLST=POST\_ARM\_IND\*B\_S\_WLST\_IND; |
| POST\_IO\_6M\_IND | indicator of post IO in 6 mos | int\_only\_year | IF B\_S\_INT\_ONLY\_IND = 1 AND -6<=IO\_MonthLeft<=0 then POST\_IO\_6M\_IND = 1; |
| POST\_IO\_IND | indicator of post IO period | int\_only\_year | POST\_IO\_IND=.<IO\_MonthLeft<=0; |
| post\_neg\_CONV | HPI change to negative | HPI, portfolio\_type\_cd | post\_neg=0; if file\_dt>=firstdate\_back\_pos>. then do;  post\_neg=1; end; post\_neg\_CONV=post\_neg\*(GOV=0); |
| Post2010\_Orig | Origination date after year 2010 | loan\_orig\_dt | Post2010\_Orig=N\_S\_LOAN\_Orig\_Dt>=mdy(1,1,2010); |
| Post2010\_Orig\_CONV | Origination date after year 2010 interacts with non-GOV indicator | loan\_orig\_dt, portfolio\_type\_cd | Post2010\_Orig\_CONV= Post2010\_Orig\*(GOV=0); |
| PPP\_IND | PPP indicator | PPP\_ind |  |
| PRE\_ARM\_12M\_IND | indicator of pre arm in past 12 mos | fixed\_rate\_ind | PRE\_ARM\_12M\_IND = 0; if B\_S\_ARM\_IND = 1 and POST\_ARM\_IND = 0 and 0<=ARM\_MonthLeft<=12 then PRE\_ARM\_12M\_IND = 1; |
| PRE\_ARM\_6M\_IND | indicator of pre arm in past 6 mos | fixed\_rate\_ind | PRE\_ARM\_6M\_IND = 0; if B\_S\_ARM\_IND = 1 and POST\_ARM\_IND = 0 and 0<=ARM\_MonthLeft<=6 then PRE\_ARM\_6M\_IND = 1; |
| PRE\_IO\_12M\_IND | indicator of pre io in past 12 mos | INT\_ONLY\_IND | PRE\_IO\_12M\_IND = 0; IF B\_S\_INT\_ONLY\_IND = 1 AND 0<=IO\_MonthLeft<=12 then PRE\_IO\_12M\_IND = 1; |
| PRE\_IO\_6M\_IND | indicator of pre io in past 6 mos | INT\_ONLY\_IND | PRE\_IO\_6M\_IND = 0; IF B\_S\_INT\_ONLY\_IND = 1 AND 0<=IO\_MonthLeft<=6 then PRE\_IO\_6M\_IND = 1; |
| Pre\_Maturity\_12M\_IND | indicator of pre maturity in past 12 mos | orig\_term | Pre\_Maturity\_12M\_IND = 0<= N\_S\_ORIG\_TERM-N\_M\_MOB<=12; |
| Pre\_Maturity\_3M\_IND | indicator of pre maturity in past 3 mos | orig\_term | Pre\_Maturity\_3M\_IND = 0<= N\_S\_ORIG\_TERM-N\_M\_MOB<=3; |
| Pre\_Maturity\_6M\_IND | indicator of pre maturity in past 6 mos | orig\_term | Pre\_Maturity\_6M\_IND = 0<= N\_S\_ORIG\_TERM-N\_M\_MOB<=6; |
| R\_M\_fcl\_time\_B12M\_LE1\_GOV | interaction of foreclosure past 12 mos indicator with GOV indicator | unemployment, portfolio\_type\_cd | R\_M\_fcl\_time\_B12M\_LE1 = R\_M\_fcl\_time\_B12M<=1; R\_M\_fcl\_time\_B12M\_LE1\_GOV = R\_M\_fcl\_time\_B12M\_LE1\*GOV; |
| R\_M\_State\_UnempB12M | unemployment in State level in past 12 mos | Unemployment | if (unemployment\_t>0 and unemployment\_tm12>0) then R\_M\_State\_UnempB12M=unemployment\_t/unemployment\_tm12;  else R\_M\_State\_UnempB12M = 1; |
| R\_M\_State\_UnempB12M\_BROK\_CONV | convicator of r m\_state\_unempb12m co past BROK mos | unemployment, channel\_cd, portfolio\_type\_cd | R\_M\_State\_UnempB12M\_BROK =R\_M\_State\_UnempB12M\*B\_S\_BROK\_IND; R\_M\_State\_UnempB12M\_BROK\_CONV=R\_M\_State\_UnempB12M\_BROK\*(GOV=0); |
| R\_M\_State\_UnempB12M\_CMI | cmiicator of r m\_state cm past UnempB12 mos | unemployment | R\_M\_State\_UnempB12M\_CMI =R\_M\_State\_UnempB12M\*(CPB\_IND=0); |
| R\_M\_State\_UnempB12M\_CORR | corricator of r m\_state co past UnempB12 mos | unemployment, channel\_cd | R\_M\_State\_UnempB12M\_CORR=R\_M\_State\_UnempB12M\*B\_S\_CORR\_IND; |
| R\_M\_State\_UnempB12M\_IO\_CONV | convicator of r m\_state\_unempb12m co past IO mos | unemployment, portfolio\_type\_cd, int\_only\_year, int\_only\_ind | R\_M\_State\_UnempB12M\_IO =R\_M\_State\_UnempB12M\*B\_S\_INT\_ONLY\_IND; R\_M\_State\_UnempB12M\_IO\_CONV=R\_M\_State\_UnempB12M\_IO\*(GOV=0); |
| R\_M\_State\_UnempB12M\_SP1 | unemployment in State level in past 12 mos | unemployment | R\_M\_State\_UnempB12M\_SP1 = max(R\_M\_State\_UnempB12M,1); |
| R\_M\_State\_UnempB12M\_SP1\_MOBGT120 | mobgt120icator of r m\_state\_unempb12m mo past SP1\_MOBGT120 mos | unemployment, loan\_orig\_dt | R\_M\_State\_UnempB12M\_SP1\_MOBGT120=R\_M\_State\_UnempB12M\_SP1\*N\_M\_MOB\_GT120; |
| R\_M\_US\_GDPB1M | GDP in US level | GDP |  |
| S\_M\_FicoRefresh | refresh fico | beacon50\_score | S\_M\_Fico\_P=input(beacon50\_score,20.);/\*11/21/2016 Shuangxi changed beacon50\_score+0\*/  S\_M\_Fico\_S=input(beacon50\_score\_s,20.);/\*11/21/2016 Shuangxi changed\*/  if in\_prim eq 1 and S\_M\_Fico\_P<=0 then B\_M\_FicoMissing1='1'; else B\_M\_FicoMissing1='0'; /\*11/21/2016 Shuangxi changed\*/   if in\_prim ne 1 then B\_M\_FicoMissing2='1'; else B\_M\_FicoMissing2='0'; /\*11/21/2016 Shuangxi changed\*/ |
| S\_M\_FicoRefresh\_LSP640 | refresh fico floored at 640 | beacon50\_score | S\_M\_FicoRefresh\_LSP640=min(S\_M\_FicoRefresh,640); |
| S\_M\_FicoRefresh\_SP640 | refresh fico capped at 640 | beacon50\_score | S\_M\_FicoRefresh\_SP640=max(S\_M\_FicoRefresh,640); |
| S\_M\_FicoRefresh\_SP720 | refresh fico capped at 720 | beacon50\_score | S\_M\_FicoRefresh\_SP720=max(S\_M\_FicoRefresh,720); |
| S\_S\_FicoDcsn | decision fico | Dcsn\_fico | S\_S\_FicoDcsn=0;  B\_S\_FicoDcsnMiss='0';/\*11/19/2016 Shuangxi changed 0 to '0' \*/  IF DCSN\_FICO>=300 and DCSN\_FICO<=850 then S\_S\_FicoDcsn=DCSN\_FICO; |
| Sandy | indicator of property in area hit by the storm Sandy | prop\_zip\_cd |  |
| state\_AZ\_CONV | Indicator of AZ state | prop\_state\_cd, portfolio\_type\_cd | state\_AZ = C\_S\_PROP\_STATE\_CD2 = 'AZ'; state\_AZ\_CONV = state\_AZ\*(GOV=0); |
| state\_FL\_CONV | indicator of FL state | prop\_state\_cd, portfolio\_type\_cd | state\_FL = C\_S\_PROP\_STATE\_CD2 = 'FL'; state\_FL\_CONV = state\_FL\*(GOV=0); |
| state\_NJ\_CONV | indicator of NJ state | prop\_state\_cd, portfolio\_type\_cd | state\_NJ = C\_S\_PROP\_STATE\_CD2 = 'NJ'; state\_NJ\_CONV = state\_NJ\*(GOV=0); |
| state\_NY\_CONV | indicator of NJ state | prop\_state\_cd, portfolio\_type\_cd | state\_NY = C\_S\_PROP\_STATE\_CD2 = 'NY'; state\_NY\_CONV = state\_NY\*(GOV=0); |

##### Table 4.1.3.14: Descriptive Data Summary of PD Model Drivers (Continuous Variable)

| **Variable** | **N** | **N Miss** | **Minimum** | **Mean** | **Maximum** | **STD** |
| --- | --- | --- | --- | --- | --- | --- |
| ARM\_5Yr | 7,557,694 | 0 | 0 | 0.07 | 1 | 0.26 |
| ARM\_GT5Yr | 7,557,694 | 0 | 0 | 0.03 | 1 | 0.17 |
| D\_M\_PRIN\_BAL\_10k\_SP5LSP15\_CMI | 7,557,694 | 0 | 0 | 10.01 | 15 | 4.23 |
| D\_M\_PRIN\_BAL\_in10K | 7,557,694 | 0 | 1E-06 | 15.15 | 813.65 | 20.52 |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_first | 7,557,694 | 0 | 0 | 13.89 | 144.57 | 15.76 |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_2nd | 7,557,694 | 0 | 0 | 0.33 | 99.48 | 1.65 |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_CONV | 7,557,694 | 0 | 0 | 11.51 | 144.57 | 16.27 |
| D\_M\_PRIN\_BAL\_in10K\_CMI\_HPIDecr | 7,557,694 | 0 | 0 | 9.55 | 144.57 | 14.20 |
| D\_M\_PRIN\_BAL\_in10K\_CPB | 7,557,694 | 0 | 0 | 0.93 | 813.65 | 14.35 |
| D\_M\_PRIN\_BAL\_in10K\_LSP10 | 7,557,694 | 0 | 1E-06 | 7.65 | 10 | 3.05 |
| D\_M\_PRIN\_BAL\_in10k\_LSP20\_CMICONV | 7,557,694 | 0 | 0 | 8.22 | 20 | 7.42 |
| D\_M\_PRIN\_BAL\_in10K\_SP10\_GOV | 7,557,694 | 0 | 0 | 14.44 | 813.65 | 20.55 |
| D\_M\_PRIN\_BAL\_in10k\_SP20\_CMICONV | 7,557,694 | 0 | 0 | 18.66 | 144.57 | 15.29 |
| D\_M\_PRIN\_BAL\_SP100K\_CMI\_first | 7,557,694 | 0 | 0 | 15.73 | 144.57 | 14.81 |
| D\_M\_PRIN\_BAL\_SP100K\_CMI\_HPIDecr | 7,557,694 | 0 | 0 | 10.89 | 144.57 | 13.88 |
| fcl\_time | 7,557,694 | 0 | 1 | 7.79 | 55 | 7.59 |
| fcl\_time\_CONV | 7,557,694 | 0 | 0 | 5.76 | 55 | 6.85 |
| HPI\_APP | 7,557,694 | 0 | 0.36 | 1.11 | 21.13 | 0.49 |
| HPI\_APP\_LSP1\_CMI | 7,557,694 | 0 | 0 | 0.91 | 1 | 0.16 |
| HPI\_tm12\_ratio | 7,557,694 | 0 | 0.63 | 0.97 | 1.43 | 0.08 |
| HPI\_tm12\_ratio\_CMI | 7,557,694 | 0 | 0 | 0.96 | 1.43 | 0.12 |
| HPI\_tm12\_ratio\_LSP1 | 7,557,694 | 0 | 0.63 | 0.95 | 1 | 0.06 |
| HPI\_tm12\_ratio\_LSP1\_CMI\_CONV | 7,557,694 | 0 | 0 | 0.73 | 1 | 0.40 |
| HPI\_tm12\_ratio\_SP1 | 7,557,694 | 0 | 1 | 1.02 | 1.43 | 0.03 |
| HPI\_tm12\_ratio\_SP1\_2nd | 7,557,694 | 0 | 0 | 0.08 | 1.43 | 0.28 |
| HPI\_tm12\_ratio\_SP1\_MOBGT120 | 7,557,694 | 0 | 0 | 0.16 | 1.39 | 0.37 |
| HPI\_tm12\_ratio\_WLST\_CMI | 7,557,694 | 0 | 0 | 0.22 | 1.43 | 0.40 |
| income\_12m\_ratio | 7,557,694 | 0 | 0.89 | 1.01 | 1.14 | 0.03 |
| income\_12m\_ratio\_LSP1 | 7,557,694 | 0 | 0.89 | 0.99 | 1 | 0.02 |
| INX\_SP500\_12m\_ratio | 7,557,694 | 0 | 0.55 | 1.04 | 1.50 | 0.21 |
| INX\_SP500\_12m\_ratio\_LSP1 | 7,557,694 | 0 | 0.55 | 0.93 | 1 | 0.13 |
| INX\_SP500\_12m\_ratio\_SP1 | 7,557,694 | 0 | 1 | 1.10 | 1.50 | 0.11 |
| INX\_SP500\_3m\_ratio | 7,557,694 | 0 | 0.70 | 1.01 | 1.25 | 0.10 |
| INX\_SP500\_3m\_ratio\_CONV | 7,557,694 | 0 | 0 | 0.78 | 1.25 | 0.43 |
| INX\_SP500\_tm\_CONV | 7,557,694 | 0 | 0 | 1013.25 | 2673.61 | 619.89 |
| MTM\_CLTV | 7,557,694 | 0 | 0 | 76.72 | 300 | 36.89 |
| MTM\_CLTV\_GT100\_GOV | 7,557,694 | 0 | 0 | 0.07 | 1 | 0.25 |
| MTM\_CLTV\_LE10 | 7,557,694 | 0 | 0 | 0.06 | 1 | 0.23 |
| MTM\_CLTV\_LSP100 | 7,557,694 | 0 | 0 | 71.44 | 100 | 28.93 |
| MTM\_CLTV\_LSP60 | 7,557,694 | 0 | 0 | 52.08 | 60 | 16.19 |
| MTM\_CLTV\_MOBGT120 | 7,557,694 | 0 | 0 | 6.36 | 300 | 18.57 |
| MTM\_CLTV\_SP40\_first\_CONV | 7,557,694 | 0 | 0 | 54.26 | 300 | 43.76 |
| MTM\_CLTV\_SP40LSP120 | 7,557,694 | 0 | 40 | 77.94 | 120 | 26.28 |
| MTM\_CLTV\_SP40LSP120\_GOV | 7,557,694 | 0 | 0 | 19.05 | 120 | 37.36 |
| MTM\_CLTV\_SP60 | 7,557,694 | 0 | 60 | 84.63 | 300 | 26.63 |
| MTM\_CLTV\_SP60\_HPIDcr | 7,557,694 | 0 | 0 | 57.94 | 300 | 49.18 |
| MTM\_CLTV\_SP60LSP80\_CPB | 7,557,694 | 0 | 0 | 0.62 | 80 | 6.37 |
| MTM\_CLTV\_SP80 | 7,557,694 | 0 | 80 | 92.46 | 300 | 20.81 |
| MTM\_CLTV\_SP80\_first\_CONV | 7,557,694 | 0 | 0 | 63.44 | 300 | 44.80 |
| MTM\_CLTV\_SP80\_2nd\_CONV | 7,557,694 | 0 | 0 | 7.85 | 300 | 27.77 |
| N\_M\_Age\_Oldest\_Mtg | 7,557,694 | 0 | 1 | 116.82 | 482 | 52.85 |
| N\_M\_age\_oldest\_trd | 7,557,694 | 0 | 0 | 202.39 | 752 | 65.95 |
| N\_M\_Curr\_lt\_Orig\_Rate\_LSP48 | 7,557,694 | 0 | 0 | 3.04 | 48 | 10.76 |
| N\_M\_logHpiChange\_Orig | 7,557,694 | 0 | -1.01 | 0.04 | 3.05 | 0.33 |
| N\_M\_logHpiChange\_Orig\_LSP0 | 7,557,694 | 0 | -1.01 | -0.10 | 0 | 0.17 |
| N\_M\_logHpiChange\_Orig\_SP0 | 7,557,694 | 0 | 0 | 0.14 | 3.05 | 0.24 |
| N\_M\_MOB\_cap\_WLST\_CONV | 7,557,694 | 0 | 0 | 10.72 | 180 | 25.78 |
| N\_M\_MOB\_LSP36 | 7,557,694 | 0 | 1 | 31.51 | 36 | 8.59 |
| N\_M\_MOB\_LSP36\_GOV | 7,557,694 | 0 | 0 | 7.53 | 36 | 14.33 |
| N\_M\_MOB\_LSP72 | 7,557,694 | 0 | 1 | 50.30 | 72 | 22.35 |
| N\_M\_MOB\_SP36 | 7,557,694 | 0 | 36 | 72.29 | 180 | 42.33 |
| N\_M\_MOB\_SP36LSP180 | 7,557,694 | 0 | 36 | 72.29 | 180 | 42.33 |
| N\_M\_n\_open\_mort | 7,557,694 | 0 | 0 | 1.14 | 99 | 0.87 |
| P\_AND\_I\_ratio\_ARM | 7,557,694 | 0 | 0 | 0.23 | 2.93 | 0.42 |
| P\_AND\_I\_ratio\_IO | 7,557,694 | 0 | 0 | 0.09 | 2.93 | 0.28 |
| P\_M\_CURR\_NOTE\_RATE\_ARM | 7,557,694 | 0 | -0.13 | 1.51 | 18.75 | 3.01 |
| P\_M\_CURR\_NOTE\_RATE\_ARM\_CONV | 7,557,694 | 0 | -0.13 | 1.42 | 18.75 | 2.97 |
| P\_M\_PresIntSpread\_LSP3pt5\_CMI | 7,557,694 | 0 | -5.73 | 2.94 | 3.5 | 1.00 |
| P\_M\_PresIntSpread\_LSP3pt5\_FRM | 7,557,694 | 0 | -5.73 | 2.33 | 3.5 | 1.47 |
| P\_M\_PresIntSpread\_LSP5\_CPB | 7,557,694 | 0 | -2.56 | 0.02 | 5 | 0.28 |
| P\_M\_PresIntSpread\_LSP5\_PRA | 7,557,694 | 0 | -5.52 | 0.46 | 5 | 1.35 |
| P\_M\_PresIntSpread\_LSP5\_Pre2010 | 7,557,694 | 0 | -5.73 | 3.65 | 5 | 1.54 |
| P\_M\_State\_Unemp\_Rate | 7,557,694 | 0 | 2.10 | 7.26 | 14.6 | 2.61 |
| P\_M\_State\_Unemp\_Rate\_2nd | 7,557,694 | 0 | 0 | 0.55 | 14.6 | 2.00 |
| P\_M\_State\_Unemp\_Rate\_currneg | 7,557,694 | 0 | 0 | 3.91 | 14.6 | 4.56 |
| P\_M\_State\_Unemp\_Rate\_lag12\_GOV | 7,557,694 | 0 | 0 | 1.61 | 14.60 | 3.23 |
| P\_S\_BE\_DEBT\_RATIO | 7,557,694 | 0 | 0 | 28.49 | 100 | 20.27 |
| P\_S\_OrigIntSpread | 7,557,694 | 0 | -9.41 | 3.03 | 12.09 | 2.10 |
| P\_S\_OrigIntSpread\_FRM | 7,557,694 | 0 | -9.41 | 2.47 | 12.09 | 2.30 |
| R\_M\_State\_UnempB12M | 7,557,694 | 0 | 0.38 | 1.13 | 2.53 | 0.29 |
| R\_M\_State\_UnempB12M\_BROK\_CONV | 7,557,694 | 0 | 0 | 0.10 | 2.53 | 0.34 |
| R\_M\_State\_UnempB12M\_CMI | 7,557,694 | 0 | 0 | 1.12 | 2.53 | 0.31 |
| R\_M\_State\_UnempB12M\_CORR | 7,557,694 | 0 | 0 | 0.44 | 2.53 | 0.57 |
| R\_M\_State\_UnempB12M\_IO\_CONV | 7,557,694 | 0 | 0 | 0.10 | 2.53 | 0.35 |
| R\_M\_State\_UnempB12M\_SP1 | 7,557,694 | 0 | 1 | 1.18 | 2.53 | 0.26 |
| R\_M\_State\_UnempB12M\_SP1\_MOBGT120 | 7,557,694 | 0 | 0 | 0.17 | 2.53 | 0.41 |
| R\_M\_US\_GDPB1M | 7,557,694 | 0 | -0.08 | 0.03 | 0.08 | 0.03 |
| S\_M\_FicoRefresh | 7,557,694 | 0 | 300 | 575.74 | 850 | 90.70 |
| S\_M\_FicoRefresh\_LSP640 | 7,557,694 | 0 | 300 | 559.76 | 640 | 64.44 |
| S\_M\_FicoRefresh\_SP640 | 7,557,694 | 0 | 640 | 655.99 | 850 | 38.83 |
| S\_M\_FicoRefresh\_SP720 | 7,557,694 | 0 | 720 | 724.35 | 850 | 15.98 |
| S\_S\_FicoDcsn | 7,557,694 | 0 | 300 | 632.87 | 850 | 72.59 |

##### Table 4.1.3.15: Descriptive Data Summary of PD Model Drivers (Indicator Variable)

| Variable | N | Count (=1) | Count % |
| --- | --- | --- | --- |
| B\_M\_coborrower\_IND | 7,557,694 | 3,583,150 | 47.41% |
| B\_M\_IN\_TRIAL\_IND | 7,557,694 | 218,481 | 2.89% |
| B\_S\_ARM\_IND | 7,557,694 | 1,708,771 | 22.61% |
| B\_S\_BROK\_IND | 7,557,694 | 820,767 | 10.86% |
| B\_S\_CORR\_IND | 7,557,694 | 2,989,494 | 39.56% |
| B\_S\_DocFull\_IND | 7,557,694 | 3,648,964 | 48.28% |
| B\_S\_DocLow\_IND | 7,557,694 | 722,387 | 9.56% |
| B\_S\_DocMISS\_IND | 7,557,694 | 2,107,628 | 27.89% |
| B\_S\_ge7YR\_ARM\_IND | 7,557,694 | 232,821 | 3.08% |
| B\_S\_gt15YR\_FRM\_IND | 7,557,694 | 4,884,936 | 64.64% |
| B\_S\_INT\_ONLY\_IND | 7,557,694 | 646,719 | 8.56% |
| B\_S\_le15YR\_FRM\_IND | 7,557,694 | 963,987 | 12.76% |
| B\_S\_lienfirst\_IND | 7,557,694 | 6,948,155 | 91.93% |
| B\_S\_lien2nd\_IND | 7,557,694 | 609,539 | 8.07% |
| B\_S\_LoanPurpMiss\_IND | 7,557,694 | 29,336 | 0.39% |
| B\_S\_LoanPurpPurch\_IND | 7,557,694 | 2,737,019 | 36.22% |
| B\_S\_LoanPurpWork\_IND | 7,557,694 | 20,646 | 0.27% |
| B\_S\_lowIncDoc\_IND | 7,557,694 | 581,680 | 7.70% |
| B\_S\_missS\_M\_FicoRefresh | 7,557,694 | 254,499 | 3.37% |
| B\_S\_OCC\_IV\_IND | 7,557,694 | 174,048 | 2.30% |
| B\_S\_OCC\_OO\_IND | 7,557,694 | 7,293,549 | 96.50% |
| B\_S\_OCC\_SH\_IND | 7,557,694 | 90,097 | 1.19% |
| B\_S\_pmi\_IND | 7,557,694 | 220,638 | 2.92% |
| B\_S\_PropCondoCP\_IND | 7,557,694 | 668,513 | 8.85% |
| B\_S\_PropMulti\_IND | 7,557,694 | 143,987 | 1.91% |
| B\_S\_PropSFR\_IND | 7,557,694 | 6,581,345 | 87.08% |
| B\_S\_WLST\_IND | 7,557,694 | 1,748,937 | 23.14% |
| B\_S\_WLST\_IND\_2nd | 7,557,694 | 249,365 | 3.30% |
| Balloon | 7,557,694 | 268,864 | 3.56% |
| Balloon\_Mature | 7,557,694 | 6,801 | 0.09% |
| community\_loan\_ind | 7,557,694 | 210,243 | 2.78% |
| cpb\_ind | 7,557,694 | 71,201 | 0.94% |
| curr\_neg | 7,557,694 | 3,467,751 | 45.88% |
| D\_M\_PRIN\_BAL\_LE2K | 7,557,694 | 104,228 | 1.38% |
| D\_M\_PRIN\_BAL\_LE5K | 7,557,694 | 170,220 | 2.25% |
| DEC | 7,557,694 | 642,979 | 8.51% |
| FCL\_Moratorium\_Dec09 | 7,557,694 | 129,920 | 1.72% |
| FCL\_Moratorium\_JF09 | 7,557,694 | 209,217 | 2.77% |
| FEB | 7,557,694 | 617,761 | 8.17% |
| FRM\_LE15Yr | 7,557,694 | 963,987 | 12.76% |
| GOV | 7,557,694 | 1,675,436 | 22.17% |
| HPI\_APP\_LE1 | 7,557,694 | 3,594,934 | 47.57% |
| HPI\_TM12\_RATIO\_LE1 | 7,557,694 | 4,831,683 | 63.93% |
| JAN | 7,557,694 | 562,314 | 7.44% |
| judicial | 7,557,694 | 3,738,748 | 49.47% |
| judicial\_CONV | 7,557,694 | 2,886,385 | 38.19% |
| JUL | 7,557,694 | 636,734 | 8.42% |
| JUN | 7,557,694 | 639,514 | 8.46% |
| MAR | 7,557,694 | 605,059 | 8.01% |
| MAY | 7,557,694 | 635,319 | 8.41% |
| miss\_N\_M\_Age\_Oldest\_Mtg | 7,557,694 | 2,076,411 | 27.47% |
| miss\_N\_M\_age\_oldest\_trd | 7,557,694 | 3,553,795 | 47.02% |
| miss\_N\_M\_n\_open\_mort | 7,557,694 | 3,514,185 | 46.50% |
| MTM\_CLTV\_LE10 | 7,557,694 | 423,305 | 5.60% |
| never\_neg | 7,557,694 | 3,823,082 | 50.59% |
| OCT | 7,557,694 | 639,846 | 8.47% |
| POST\_ARM\_12M\_IND | 7,557,694 | 155,895 | 2.06% |
| POST\_ARM\_6M\_IND | 7,557,694 | 79,753 | 1.06% |
| POST\_ARM\_IND | 7,557,694 | 672,837 | 8.90% |
| POST\_ARM\_IND\_WLST | 7,557,694 | 196,861 | 2.60% |
| POST\_IO\_6M\_IND | 7,557,694 | 13,133 | 0.17% |
| POST\_IO\_IND | 7,557,694 | 68,869 | 0.91% |
| post\_neg\_CONV | 7,557,694 | 147,996 | 1.96% |
| Post2010\_Orig | 7,557,694 | 146,710 | 1.94% |
| Post2010\_Orig\_CONV | 7,557,694 | 113,931 | 1.51% |
| PPP\_IND | 7,557,694 | 926,851 | 12.26% |
| PRE\_ARM\_12M\_IND | 7,557,694 | 165,138 | 2.19% |
| PRE\_ARM\_6M\_IND | 7,557,694 | 81,995 | 1.08% |
| PRE\_IO\_12M\_IND | 7,557,694 | 31,690 | 0.42% |
| PRE\_IO\_6M\_IND | 7,557,694 | 16,657 | 0.22% |
| Pre\_Maturity\_12M\_IND | 7,557,694 | 79,808 | 1.06% |
| Pre\_Maturity\_3M\_IND | 7,557,694 | 31,316 | 0.41% |
| Pre\_Maturity\_6M\_IND | 7,557,694 | 47,575 | 0.63% |
| R\_M\_fcl\_time\_B12M\_LE1\_GOV | 7,557,694 | 1,011,091 | 13.38% |
| state\_AZ\_CONV | 7,557,694 | 115,187 | 1.52% |
| state\_FL\_CONV | 7,557,694 | 479,720 | 6.35% |
| state\_NJ\_CONV | 7,557,694 | 135,589 | 1.79% |
| state\_NY\_CONV | 7,557,694 | 334,913 | 4.43% |

Please refer to attachment ‘4.1.3 Severity Method A Data Dictionary.xlsx’ for Severity model data dictionary.

##### Table 4.1.3.16: Descriptive Data Summary of LGD Continuous Model Drivers

| **Variable** | **N** | **N Miss** | **Minimum** | **Mean** | **Maximum** | **Std Dev** |
| --- | --- | --- | --- | --- | --- | --- |
| **first Lien** |  |  |  |  |  |  |
| cLTV\_MTM\_unemp\_down | 65055 | 0 | 0 | 39.04 | 300 | 51.06 |
| log\_curr\_bal | 65055 | 0 | 0 | 11.48 | 16.02 | 1.41 |
| log\_curr\_bal\_SP\_gt300k | 65055 | 0 | 12.61 | 12.68 | 16.02 | 0.21 |
| log\_curr\_bal\_SP\_le300k | 65055 | 0 | 0 | 11.41 | 12.61 | 1.33 |
| log\_HPI\_t\_orig\_ratio | 65055 | 0 | -1.00 | -0.16 | 2.46 | 0.32 |
| log\_HPI\_tm12\_ratio\_SP\_le0 | 65055 | 0 | -0.46 | -0.07 | 0 | 0.09 |
| LTV\_mtm\_SP\_gt90 | 65055 | 0 | 90 | 106.83 | 300 | 23.96 |
| prin\_reduction | 65055 | 0 | 0 | 0 | 0 | 0 |
| UnempRate | 65055 | 0 | 0 | 8.69 | 14.60 | 2.34 |
| UnempRate\_SP\_gt9 | 65055 | 0 | 9 | 9.83 | 14.60 | 1.21 |
| **2nd Lien** |  |  |  |  |  |  |
| cLTV\_MTM\_SP\_gt80\_le120 | 21902 | 0 | 80 | 108.06 | 120 | 15.06 |
| junior\_ratio\_SP\_lep20 | 21902 | 0 | 0.01 | 0.19 | 0.2 | 0.03 |
| log\_curr\_bal\_SP\_gt10k\_le150k | 21902 | 0 | 9.21 | 10.53 | 11.92 | 0.70 |
| log\_curr\_bal\_SP\_gt150k | 21902 | 0 | 11.92 | 11.92 | 12.94 | 0.05 |
| log\_fst\_mtg\_bal | 21902 | 0 | 5.96 | 11.62 | 14.46 | 1.00 |
| log\_HPI\_tm12\_ratio | 21902 | 0 | -0.46 | -0.12 | 0.21 | 0.11 |
| prin\_reduction | 21902 | 0 | 0 | 0 | 0 | 0 |
| R\_M\_UnempB12M\_SP\_gt1\_le1p2 | 21902 | 0 | 1 | 1.14 | 1.2 | 0.08 |
| UnempRate | 21902 | 0 | 2.70 | 8.71 | 14.6 | 2.29 |

##### Table 4.1.3.17: Descriptive Data Summary of LGD Binary Model Drivers

| Variable | N | Count | Count % |
| --- | --- | --- | --- |
| first Lien |  |  |  |
| BROK\_IND | 65055 | 8380 | 12.9% |
| FHL\_ind | 65055 | 0 | 0.0% |
| HLC\_ind | 65055 | 0 | 0.0% |
| judicial | 65055 | 27173 | 41.8% |
| pmi | 65055 | 1982 | 3.0% |
| 2nd Lien |  |  |  |
| CBNA\_ind | 21902 | 0 | 0.0% |
| deficiency | 21902 | 15981 | 73.0% |
| judicial | 21902 | 9757 | 44.5% |
| lpi | 21902 | 73 | 0.3% |
| REL\_ind | 21902 | 21902 | 100.0% |
| wasUSRB | 21902 | 0 | 0.0% |

#### If a vendor model is used, describe any model-specific data requirements.

[Describe potential requirements and limitations, including dependencies on legacy formats (e.g., a limit of 65,535 lines per file). Describe how the data was managed to match the requirements or overcome the limitations.]

Not Applicable. The Method A RM model suite is not a vendor model.

1. MRM Question - As per the model limitations, there was no suggestion/comment about changing the selection of OOT sample in the previous limitations from standard practice. Further, based on confirmation/attestation provided by model sponsor, updated method A is supposed to remediate limitation of both Method A and Method B models, except limitation about alternative modeling approach. One of the limitation (limitation id 16856) in FRHEL method B is following “*No consideration of recent 2 years of data either for model development or model stability analysis*”. In this limitation also, there was no suggestion/comment about changing the selection of OOT sample from standard practice.

Hence, follow-up questions from MRM are as follows:

* Sponsor needs to further justify creation of this non-standard OOT sample.
* Based on the above limitation (#19546), sponsor is requested to provide justification for not including pre-2008 data in LGD model development sample.

Answer –

**Justification for the selection of the OOT Period** - In this version of model development, CAMU holds out an OOT sample solely to comply with MRM’s OOT test requirement although CAMU believes that holding out OOT data will potentially result in loss of performance and portfolio change information. Due to this concern, CAMU carefully choose Jan-2012 to March-2014 period as out-of-time validation sample and retaining data between Feb 2006 - Dec 2011 and Apr 2014 -Dec 2017 as model development data due to the following considerations –

1. First and foremost, as per the limitation on the prior model suite around the non-inclusion of recent period data, it was recommended to include the most recent time period as part of the model development sample. To comply with MRM’s requirements, CAMU did not choose to hold out most recent data as OOT. CAMU also agrees with MRM that holding out recent performance data will disable us from capturing the most recent macro-economic trend, such as interest rate increase, portfolio and underwriting policy change. The inclusion of the recent period data helped in capturing the go-forward state of the business in terms of origination profile and portfolio composition mix.

2. Excluding 2006-2008 data is not an option either because it will result in loss of part of the stress period performance. This leaves CAMU at no choice but to carve out a middle period as OOT hold out sample to follow MRM’s OOT requirement.

3. To ensure the parameter stability and OOT hold out does not cause significant information loss, CAMU further conducted parameter stability analysis and revise/dropped instable parameters. The parameter stability analysis was two-folds including

a. CAMU estimates the same specification on the 20% OOS sample to ensure consistent parameter sign, significance and reasonable magnitude

b. CAMU further estimates the same specification on the entire data including the OOT period to ensure consistent parameter sign, significance and reasonable magnitude

It is worth noting though, as CAMU expected, holding out OOT data could potentially result in loss of performance information, especially given that the Jan2012-March2014 period represented the initial phase of economic recovery with high volume of various types of loan modifications and refinancing due to extremely low interest rate and foreclosure settlement(2012) etc. However, given the model’s strong performance (backtest) for the OOT period, it can be concluded that the impact from such exclusion is expected to be small as the model development data still comprised of sufficient performance data covering both economic boom and bust. A few transitions that are heavily impacted by portfolio management policy over time, such as IVP transitions, are always statistically challenging to model in the first place. For additional details on the OOT backtest, please refer to Section 6.3 of the MDD. Also please be aware that the OOT period exclusion only pertains to the performances observed during this OOT period. Loans that originated during this period and stayed active behind the OOT period, are included as part of model development sample.

**Justification for not including pre-2008 data in LGD model development sample**- As per Table 4.1.3.4 pertaining to the LGD Exclusion Waterfall; exclusion # 10 excludes all loans prior to 2008. This was done as the historical loss performance during the earlier development period of the DataMart (pre-2008) exhibited high discrepancy with finance system mainly attributed to lack of a reconciliation process which was establish and enhanced early in 2009 between Risk DataMart and Finance. Based on a deep dive analysis that was conducted in collaboration with the REL DataMart team, it was revealed several discrepancies within the Risk Loss data prior to 2008, attributable to either manual entry errors or top line adjustments to actual loss numbers. Since the reconciliation error between Risk DataMart’s Loss file and PEARL (master data file) for the pre-2008 loss data exceeded well over the 1.5% threshold set by REL DataMart team, it was considered prudent to not use the pre-2008 data as part of the Severity model development. For additional details, please look at the attachment – ‘1.1 Model development data sources and Mortgage Transformation considerations’ for pertinent details.

2. MRM Question - In addition to the above points, Sponsors are required to provide rationale/supporting evidence on the below details:

* Rationale for not considering the long sample as the development sample in case recent period data is needed to capture the recent trend

Answer – Please note that the model development data includes the recent most period as part of its development data pool. Hence this question is not valid.

* Analysis and evidence how the development data completes an economic cycle which is a GP1 requirement. Please note that Jan 2012- Mar 2014 period was initial phase of economic recovery with high volume of various types of loan modifications and refinancing due to extremely low interest rate, foreclosure settlement in 2012 etc. Dropping this period may not capture the impact of these important events. Also, there was a big change in portfolio composition between Jan 2012– Mar 2014.

Answer – Bank’s Portfolio composition continues to evolve with business strategy leading to changes over time. The current model is built based on loan characteristics thereby allowing it to capture responsiveness to loan characteristics.

In addition, changes in the macroeconomic environment also lead to different borrower behaviors over time. One such instance is observed during the selected OOT period for the model, where higher rates of voluntary prepayment (VP) were observed over many months during a phase of economic recovery. As noted above, the model is developed using loan characteristics as well as observed borrower behavior and macroeconomic drivers over a long time span, and as discussed in section 6.3.2.5 (see also attachment 6.3.2.5 Model OOT Backtest.xlsx), the model is able to predict this rising VP trend within the threshold of error (per MRM Testing Guidance).

* Please provide details on how development data captures increasing interest rate period.

Answer – Please see the swap rate trends over time. Swap rates are the rate attributes used in MOD model to factor in the interest rate movements in the economy which includes the increasing rate environment starting 2015Q4.

* Details on how the panel data was created.

Answer - The panel data creation process (which includes the raw data and the subsequent loan exclusions) have been discussed in Section 4.1.3. Please refer to this section for additional details.

* Provide details of the OOT sample to specify whether only new originations or data from seasoned loans were included.

Answer – As stated many times before, the OOT period exclusion only pertains to the performances observed during this OOT period. Loans that originated during this period and stayed active behind the OOT period, are included as part of model development sample.

3. MRM Question - For the PD non-modeled transitions, the model uses the data from Apr2009-Dec2011 for stress roll rate & Apr2014-Dec2017 for non-stress roll rate. Please provide the rationale for this selection

Answer - The Stress period values are volume weighted mean of transition rates based on data from 200904 (Please note that the MOD Program was initiated Q2-2009 and data on MODs became available starting 2009Q4). US HPI reaches its peak in the second quarter of 2006 and started dropping over time. It hit the bottom in 2009 and remained low until the first quarter of 2012. Instead of using the exact 27month stress period defined by MRM, the stress value selection covered this entire period of HPI decrease to avoid over-fitting the model performance due to data volatility or at odd with MRM’s “blind back test” requirement.

### Data Sampling Methodology and Results

[If the model is based on a sample, describe the data sampling methodology/scheme (e.g. stratified random sample based on defined strata, etc.), sampling results, sample validation techniques (e.g. out-of-sample/ hold-out sample, out-of-time, and cross-sample validation). Provide evidence that the chosen sample is representative of the total population. In addition to describing the sampling methodology used, justify its appropriateness by comparison of the total population vs sampled population (for e.g. population distribution on key risk drivers, descriptive statistics etc.)].

Modelers commonly use sampling strategy to reduce extremely large sample sizes and improve computational efficiency in the modeling procedures. For the 2019 Method A RM PD model development, modelers used sampling strategy to create a sample for the following reference transition- BUK01->BUK01. On the other hand, observations collected on all other transitions were fully sampled in to the model development process.

All rationale behind the selection of the appropriate transition bucket(s) for the sampling process, the selection of relevant weights and the comparison of key risk drivers before and after sampling have been thoroughly discussed as evidenced below.

The main reasons why BUK01->BUK01 was selected for sampling down because 1) this specific bucket has extremely large sample size (Table 4.1.4.1 shows 65,475,905 loans in BUK01->BUK01); hence sampling down this transition had helped with improving the model efficiency significantly; 2) BUK01->BUK01 is the most common and benign transition used as reference cell in respective logistic regressions. Sampling down this specific transition was expected to have minimal impact to the information enrichment and completeness of potential risk drivers in subsequent modeling.

Based on the rationales stated above, the BUK01-> BUK01 was randomly sampled down by 2%. The sampling rate was so chosen to ensure that the final sample had an adequate number of observations to correctly draw all inferences. Based on existing literature review and research conducted on the prevailing industry trends, the preferred approach suggests that sampling should not adversely affect the power of the statistical test.

As discussed in section 4.1.3, the sampling process was conducted on the data collected on observation month between Jan 2005 and Dec 2017 that has gone through the initial data waterfall process. As shown in table 3,4,5 in section 4.1.3, before the 2% sampling down process, BUK01->BUK01 transition has 65,475,905 records and the total number of observations including all transitions was 77,083, 574. After sampling, the numbers reduced to 1,307,971 in the BUK01->BUK01 and 12,915,640 in the entire datasets.

To justify the appropriateness of the sampling, modelers have examined the population distributions of key risk drivers before vs. after sampling, where risk drivers include marked to market CLTV, Refresh FICO, current principal balance, HPI 12m change, and Unemployment 12m change. This analysis is shown below in Table 4.1.4.1 for BUK01->BUK01 transition, which clearly demonstrates that for the BUK01->BUK01 population, the distributions of all the key risk drivers are very comparable before vs. after sampling. This analysis has provided solid evidence that the sampling have no significant impact to the key risk drivers as model inputs.

Table 4.1.4.1: Comparison of Major Risk Drivers before and after Sampling



Sampling strategy was not applicable for the 2019 Method A RM Mortgage LGD model.

### Data cleaning and preparation

What were the data cleaning steps?

[Describe outlier treatment, observation removal, missing data treatment.

Also describe remediation of any data quality issues.]

In the 2019 Method A RM PD and LGD modeling process, prudent and thorough data checks have been conducted to confirm overall sound data quality, as already described in section 4.1.1 to 4.1.3. Based on the observations of scrutinized data reconciliation / quality check reports and descriptive statistics analysis, data cleaning treatments were performed to mitigate the impact of missing values to ensure continuity in the data. Treatments include the following considerations.

1. For the continuous variables except for FICO, missing values were imputed by the median value of the data that has gone through the initial waterfall process and a corresponding missing control variable was created. In particular, considering the significant differences between the CMI and CPB portfolio in terms of property value, loan size and borrower credit profile, missing value on CMI and CPB loans was imputed by their respective portfolio median.

A relatively high missing rate was observed on Refresh FICO (S\_M\_FICORefresh), particularly among the loan just booked into the portfolio. So for the newly originated loans (month on book<=3), if the refresh FICO was missing, CAMU used decision FICO to feed in the value. Otherwise the portfolio median was used for the missing value imputation and in this case a missing control variable would be created.

The use of a missing control variable means the choice of imputation default value had no impact on the model. Any combination of the missing control variable and default value had the same result. Missing values are always imputed instead of deleting the records with missing values from the development data, with the benefits that it can retain the maximum available data coverage, and more importantly, it keeps the potential unique prediction power of the missing values in the modeling. Second, the missing values are imputed by the median, because the median is deemed as the “neutral” value of the valid range, which is not affected by large outliers. Third, using missing indicator as control variable allows modelers to accurately model the marginal effect of the missing values for the variable independently from the effect of its valid values, and differentiate the true mean values records from the imputed records, which also serve as reassurance to eliminate potential estimation bias due to missing imputation by the mean.

1. Limited capping methods were used to treat outliers. Cap was chosen as the 99% percentile + 2\*STD and was calculated for CMI and CPB respectively.
2. No observation was blindly removed due to missing value alone. Instead, the population exclusion waterfall was fully analyzed and any exclusions were supported with sound rationales (please refer to next section of data exclusion).
3. Within the Severity model, for loans (primarily first lien) with Private Mortgage Insurance (PMI) coverage, if the PMI coverage ratio is missing, it is substituted with a 25% imputed rate. On the entire LGD development data of 304,926 accounts, 1970 carry PMI coverage. PMI coverage percentage ranges from 6% to 50%, Table 4.1.5.1. below illustrates the PMI coverage percentage (PMI\_pct) univariate distribution on LGD development data, which shows that 25% is the median value and most common of PMI coverage percentage.

**Table 4.1.5.1: PMI Coverage Percentage Univariate Distribution**



5. Within the Severity model, for all first mortgage loans with missing balances, these are also adjusted as UPB/0.25. For the first mortgage loans with missing balances, it is imputed with a 25% rate for the LGD development sample. This is based on the empirical analysis of the LGD development data. As shown, the 25% rate is the median value.

**Table 4.1.5.2: First Mortgage Balance Univariate Distribution**



1. For all junior liens within the LGD development pool, the junior ratio is imputed at 20%,presented below is the empirical evidence .

**Table 4.1.5.3: Junior Ratio Univariate Distribution**



1. For the RM PD model – second lien loans, CAMU assumed a constant junior ratio to derive the balance of the first lien to determine the most updated mark-to-market CLTV of the loan. For most of the second liens, a special challenge to determine the CLTV is the dynamic value of the first lien balance associated with the property is not available in the source data. The only available information is the first lien balance when this second lien was originated. To address the issue, CAMU assumes that the ratio between the first lien and second lien stays constant throughout the life of the second lien.

For example, suppose a second lien was originated with 20K UPB, and at its origination, there was also an 80K UPB first lien associated with the property, then the junior ratio at origination is 20K/80K=25%. Then at any observation month, if the UPB of this second lien drops to 10K, then CAMU assume its first lien balance also proportionally reduces to 10/25%=40K. Then the total of 50K will be used to calculate the combined CLTV on this second lien.

Refer to section 4.1.4, Data Water Fall for RM never-modified Loans for the data cleaning or exclusion. Each model driver that required imputation or treatment has been listed in Table 4.1.5.4. To keep the variables in a reasonable range, CAMU imputed variables on a case-by-case basis. There was no sign of bias by examining the deviation of mean values after imputation from the values before imputation.

**Table 4.1.5.4. Method A RM** **PD Variable Treatment**

|  | **Variable** | **Treatment** | **% missing before treatment** | **Median** | **Maximum before treatment** | **Maximum after treatment (99th pct+2 StdDev)** | **Mean before treatment** | **Mean after treatment** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CMI Only** | D\_M\_PRIN\_BAL | cap | 0% | 111,436 | 13,025,000 | 1,445,666 | 181,673 | 179,915 |
| P\_M\_PresIntSpread | cap | 0% | 3.13 | 16.34 | 14.88 | 3.45 | 3.45 |
| P\_S\_OrigIntSpread | cap | 0% | 2.39 | 71.81 | 12.09 | 2.67 | 2.67 |
| S\_M\_FicoRefresh | missing value | 12% | 660 | 850 | 850 | 669 | 671 |
| P\_M\_CURR\_NOTE\_RATE | missing value, cap | 0% | 6.63 | 21.00 | 18.75 | 6.89 | 6.89 |
| P\_S\_ORIG\_NOTE\_RATE | missing value, cap | 1% | 6.84 | 76.88 | 19.16 | 7.22 | 7.22 |
| P\_S\_BE\_DEBT\_RATIO | missing value, cap | 9% | 33.65 | 995.00 | 100.00 | 30.85 | 30.95 |
| D\_S\_ORIG\_PROP\_AMT | missing value, cap | 1% | 175,000 | 32,599,119 | 2,861,921 | 308,379 | 301,717 |
|  | | | | | | | | |
| **CPB Only** | D\_M\_PRIN\_BAL | cap | 0% | 693,020 | 53,400,000 | 8,136,479 | 998,808 | 977,318 |
| P\_M\_CURR\_NOTE\_RATE | cap | 0% | 4.50 | 10.88 | 9.55 | 4.49 | 4.49 |
| P\_S\_ORIG\_NOTE\_RATE | cap | 0% | 4.88 | 17.00 | 10.50 | 4.81 | 4.81 |
| P\_M\_PresIntSpread\* | cap | 0% | 1.94 | 8.31 | 8.31 | 2.03 | 2.03 |
| P\_S\_OrigIntSpread\* | cap | 0% | 1.83 | 6.97 | 6.97 | 1.86 | 1.86 |
| S\_M\_FicoRefresh | missing value | 14% | 747 | 850 | 850 | 744 | 775 |
| P\_S\_BE\_DEBT\_RATIO | missing value, cap | 2% | 29.58 | 100.00 | 100.00 | 29.64 | 29.64 |
| D\_S\_ORIG\_PROP\_AMT | missing value, cap | 0% | 1,320,000 | 153,840,000 | 18,974,945 | 2,053,902 | 2,011,603 |
|  | | | | | | | | |
| **All** | N\_M\_age\_oldest\_trd | missing value, cap | 54% | 215 | 983 | 752 | 237 | 225 |
| N\_M\_n\_open\_mort | missing value | 54% | 1 | 79 | 79 | 1.5 | 1.24 |
| N\_M\_Age\_Oldest\_Mtg | missing value, cap | 36% | 124 | 613 | 482 | 127 | 126 |

Refer to Table 4.1.3.7.: Data Water Fall for Severity Model for the data exclusions pertaining to the Severity model. For all other data cleaning and preparation, any model driver that required imputation or treatment has been listed in Table 4.1.5.5. To keep the variables in a reasonable range, CAMU imputed variables case by case. For instance, principal balance (prin\_bal) was capped at 99% percentile + 2\*STD to remove the outlier values. There was no sign of bias as can be seen by examining the deviation of mean values after imputation from the values before imputation.

**Table 4.1.5.5. LGD Model Variable Treatment**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Reason for Cleaning | Data Cleaning Logic | %Missing | Mean before Imputation | Mean after Imputation |
| prin\_bal | out-of-range | cap at 99%+2STD | 0% | $143,744 | $142,287 |
| fts\_mtg\_bal | out-of-range, missing | cap at 99%+2STD, impute with medium value | 26% | $201,446 | $160,051 |

Were any data proxies used?

[A data proxy is an alternate variable used to calculate missing values. Also include macroeconomic data proxies.]

***Example:*** *90 day delinquent loans may be used as a proxy for default.*

No data proxies were used.

Was any data excluded for data quality or business reasons? What was the rationale for the exclusion?

[Describe data exclusions due to quality concerns or due to business reasons (e.g., loan modifications, discontinued products). Provide the rationale and explain why the exclusion is reasonable. Provide a waterfall chart based on exclusions and filtering (e.g., with number of observations.)

Also describe if any macroeconomic data was excluded for data quality reasons.]

There was no macroeconomic data exclusions made for quality reasons. For both PD and LGD model exclusions, please refer to Section 4.1.3. The data exclusions stated above in table 4.1.3.1 and table 4.1.3.4 for data waterfall exclusion details for PD and LGD model respectively, were mainly due to out of scope population, not part of the official loan list, different accounting treatment for Government loans and known data quality of the loss table in the early period of its development prior to 2008. The overall exclusion is relatively small and does not represent the current state of the data in the portfolio. In particular, for 180+DPD to VP model, government loans prior to Jan 2008 were excluded from model development because abnormally high 180+DPD to VP rate, which is 4.5%, were observed in the data during that period.

What data discrepancies were observed and what actions were taken to mitigate them? What was the impact of data discrepancies, cleaning, data exclusions and remediation?

[Describe each significant discrepancy identified during data quality checks and how it was treated. Include missing data (e.g., missing FICO scores).

Describe how the treatment affected the portfolio. Comment on changes in GCL/ENR, number of accounts, etc. and potential impact on the model. Quantify the impact.

It must be clear whether the data can still be used for model development, and that the remediation was conservative and prudent. Also include the impact of data exclusions.]

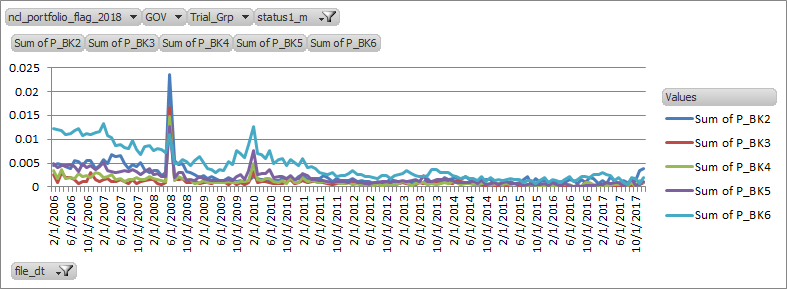
##### PD Development data

During the course of 2019 Method A RM model development, as part of the rigorous data quality checks and descriptive statistical analyses, few spikes were observed from the one-month-ahead roll rate time curve. Such spikes indicate the natural intervention of business actions taken at that time. Therefore the spikes were documented and actively communicated with business users to seek policy explanations.

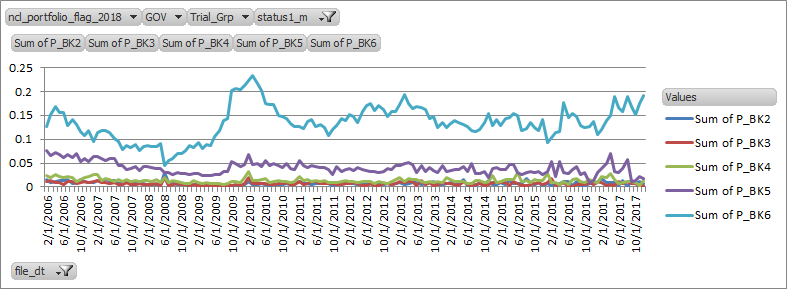
There was no action taken in modeling practice for these spikes for two reasons. First, these spikes occurred infrequently in the performance history data. So they are expected to have minimal impact on model forecasting. Second, these spikes may represent valid data points at various periods. Therefore, it is preferable to keep them intact in the modeling data.

1. Spike in roll rate from BUK7 to BUK 2,3,4,5,6 in June’08– During the RM PD model development process, ,two spikes were noted in the partial cure roll rates from BUK7 to BUKs2-6. These spikes could be attributable to the debt repayment plans that were put in place for the bankruptcies claimed under Chapters 13/11, loan modifications and subsequent payment adjustments made. These spikes affected the Method A -Residential Mortgage first lien delinquencies.
2. Spike in roll rate from BUK6, 7 & BUKs 2,3,4,5,6 in Feb’10 –Spikes were noted in the partial cure roll rates from Buk7,6 to Buks2-6. These spikes, as before were attributable to the payment adjustments made on these loans

**Table 4.1.5.1: BUK7 -> Partial Cure BUKs 2-6: Two spikes in 200806 and 201002**

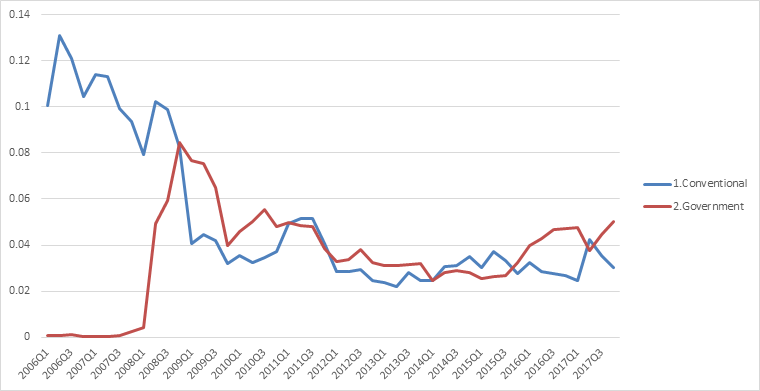


**Table 4.1.5.2: BUK6 -> Partial Cure BkT2-6: one spike in 201002**



3) Increase in GOV IVP Rate since 2016 - A Government loans is typically characterized having 97 LTV at origination with corresponding high DTI (Debt-to-income) ratio. Operationally the business has always faced challenges in government loan liquidation as the process to liquidate government loans is not the same as conventional loans. Government loans are insured, and in order to file the insurance claim and obtain the insurance proceeds the property must be conveyed to HUD. As part of this process, Bank must meet HUD’s guidelines for conveyance. These guidelines require eviction to be complete, as well as for the property to be in “conveyance condition.” This is a higher standard than what is required for conventional loans, where there are no similar requirements around eviction and property condition. For conventional loans elements such as occupancy and property condition impact REO price but are not preconditions for a loan to be sold through REO. These problems in conveying government loans are caused by two factors: 1) changing and inconsistent interpretations of HUD requirements, impacting all servicers, and 2) idiosyncratic challenges Bank has faced in the management of HUD conveyance where the operational group has been transitioned between sites and has undergone multiple leadership changes. Policy wise, when a government loan becomes IVP and goes into subsequent foreclosure sale; and if BankMortgage wins the bid during the foreclosure induced sale; this loan would tend to remain active on the Bank’s system (data warehouse) until conveyed to HUD (Dept. of Housing & Urban Development). But the loan even though showing ‘active’ status is not a real loan as the property was already foreclosed and the collateral is owned by BankMortgage. This observable data trend with the uptick in the IVP rates for Government loans, contradicts with the prevailing macro-economic trends in recent time periods which are marked with significant and consistent home price appreciation and unprecedentedly low unemployment rates. On a strategic note, as mentioned earlier, there were a transfer of middle office functionalities, who managed the foreclosure of the government loans, which lead to instability in the management process and the subsequent issuance of MRA’s from OCC around the management of debts previously contracted (DPC). Although CMI had increased performance since then, especially in regards to the increased management and regulatory scrutiny in this area, particularly related to debts previously contracted (DPC) exposure, this also has played a role in the uptick of the IVP rates.

**Table 4.1.5.3: Trend in GOV IVP Rate**



##### LGD Development data

For LGD Model development data, the relevant population exclusions are discussed above in the waterfall logic. In particular, the business maintains a one-timer loan level list that lists all loans with abnormal losses. For additional information on one-timer events and treatments, please refer to attachment “3.2\_Memorandum\_One-Timer Exclusions - 111417.docx”. Since these losses do not conform to business rationale or economic intuition, all such loans are excluded from the development data. The one timer list also includes loss spike(s)/dip(s) that occur due to strategic business intervention decision.

### Data transformation

[Describe any statistical systematic transformations applied to the performance, segmentation, account or macroeconomic data, e.g., expansion/mapping of national (or international regional) macro drivers to state-level (or country level) macro drivers, lags, differencing related to stationarity or seasonality.]

What is the rationale for performance, segmentation and account data transformation?

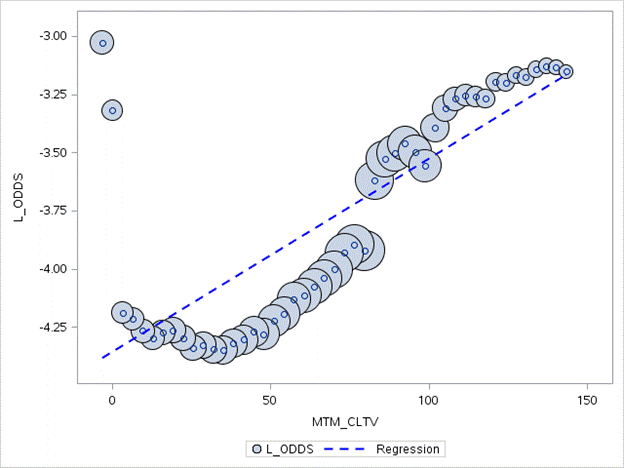
The 2019 Method A RM PD and LGD variable transformations include all the new variable transformations that were conducted this modeling cycle based on economic expectations and comprehensive statistical analyses during the variable selection process. These variable transformations were performed to improve model performance particularly on granular segments, improve the model’s sensitivity to certain key risk drivers, capture the non-linearity of some specific trends, address previous feedbacks from model reviewers / regulators and business users and lastly improve the model’s overall performance. The detailed economic rationale and related analyses for these variable transformations are further discussed in section 5.1.4 of variable selection.

The variable transformations that were attempted this modeling cycle on this specific model suite included the following –

* Adding Caps and/or Floors – Caps and Floors sets upper and lower bounds to the values that the numerical attributes can take. Caps and Floors are primarily done to align the model’s performance with business intuition. A good example to illustrate here would be usage of the floor for the marked-to-market LTV in the Severity model. For added rationale on the use of the floor at MTM LTV = 100, please refer to Section 3.1. of the MDD.
* Introducing Interaction Effects- - Interaction variables were created in the model suite to account for the differential credit risk effects between the stress / non-stress cycles. Another reason of interaction variables is to model the differentiated sensitivities to risk drivers by key sub-groups without separate segmentation. For example, within the RM PD model BUK01->BUK2 transition, the state level unemployment rate was interacted with the CMI portfolio, whereas the unemployment 12 month ago was interacted with the CPB portfolio indicator, as it’s observed that historically, the CPB portfolio demonstrated a delayed response to the change in macro-economic environment along with a different magnitude of sensitivity.

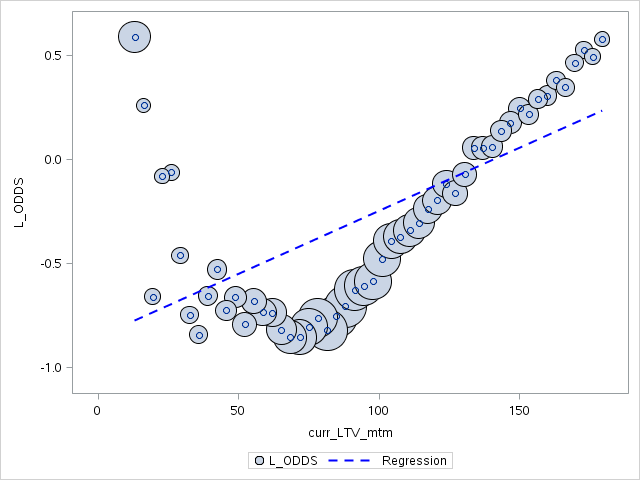
Spline Transformations- Spline transformed variables are transformations that are used to capture non-linear relationships of the root variables, the choices of spline knots are based on business / economic intuitions and validated by bivariate analysis as visualized in bubble plots. For example within the Severity stage two model a spline is created for the marked-to-market CLTV at 90 which aligns with the business policy requirements that changes for CLTV 90 and over(Figure 4.1.6.2). For the RM PD model, as shown in the bubble plot below (Figure 4.1.6.1), BUK01->BUK2 rate increases with CLTV when CLTV is greater than 40, but the slope is less steep when CLTV is above 80. In this case, mark-to-market CLTV splines were created at 40 and 80 to capture the differentiated sensitivities of the delinquency rate changes to a unit change of CLTV within different ranges of CLTV.

**Figure 4.1.6.1 – Spline Transformation Justification for the RM PD Model – Mark-to-Market CLTV**

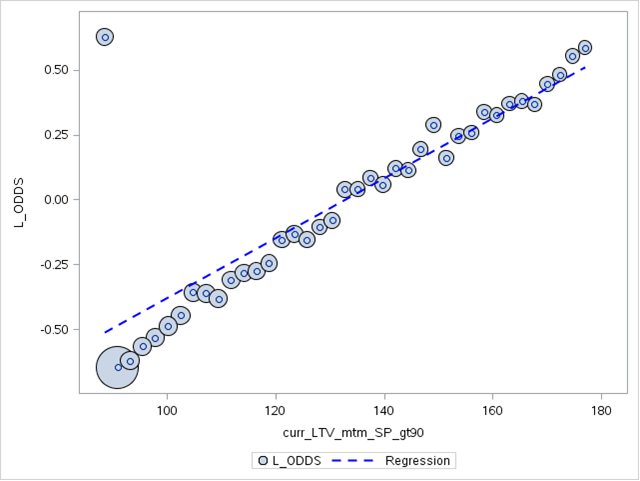


**Figure 4.1.6.2 – Spline Transformation Justification for the LGD Model – Mark-to-Market CLTV**

**Before Introduction of Splines**



**After Introduction of Splines (Monotonicity is maintained)**

****

Presented below are the variable transformations attempted on the 2019 Method A RM PD and Severity models respectively.

##### Table 4.1.6.3: 2019 Residential Mortgage PD Variable Transformation





##### Table 4.1.6.4: 2019 LGD Variable Transformation





#### Does the approach to macroeconomic variable selection or transformation significantly differ from approaches in similar countries in the region?

The U.S. real estate market is both mature and idiosyncratic. The approaches taken in nearby countries were not considered.

Macroeconomic variables such as HPI, unemployment rate, interest rate, income, GDP are typical inputs to loss forecasting models. Usage of such variables are often recommended within the realm of credit risk models to correctly gauge the effect of the economy (global and local) on the credit worthiness of the portfolios. For instance, yearly change of HPI Levels is commonly used in models instead of levels and such transformation is not significantly different from approaches practiced in similar countries in the region.

#### Are data transformations based on business intuition?

Yes it is based on business intuition. If the splines have different trends, we only keep the one that aligns with business intuition. A good example to illustrate here would be the monotonic ranking of borrower marked-to-market LTV with a floor set at 100 with respect to the LGD model (log odds ratio). For additional details, please see Section 3.1.

#### Do data transformations satisfy statistical requirements?

[If appropriate, describe statistical requirements for data transformation.]

The statistical criteria for selecting transformed variables are the same as other variables 1) correct sign 2) significance 3) improved model performance and most importantly 4) align with common business intuition. If any of the transformed variables is insignificant (p>0.05), has small chi-square or has high correlation (VIF> 50%) with other key predictors, it will be dropped from the model.

#### Are data transformations applied consistently across similar segments, particularly in the case of macroeconomic variables?

***Example****: The same types of data transformations were applied across all segments.*

Yes the variables input pool for both models (PD + LGD) are defined consistently across all segments.

### Data sufficiency

What data quality checks have been done?

[Provide evidence of consistency and integrity checks and describe how the data was tested. The tests should include final development data, as well as account-level data.

Data should be analyzed for sudden changes, missing and outlier values, and inconsistent fields within and across records.

Data quality checks should address the following areas:

1. **Business data quality checks –** quick overview of data based on charts to visually detect any outliers or anomalies
2. **Technical data quality checks** – detailed data analysis including:
   * Primary keys and duplicates check
   * Statistical checks -- univariate distribution (mean, median, minimum, maximum, number of observations, and where applicable frequencies) for segmentation, dependent, final independent variables
   * Missing or zero values check, date / time conformity, value format conformity
   * Account level data checks (checks for default values, duplicates, record mapping -- joins, completeness of fields, data and business logic integrity within and across records and frequency checks)

Data quality checks are specified in the following templates:

Data quality checks must include interpretation and analysis of results. Reconciliation is not a data quality check.]

CAMU works closely with the MEP Optima team to ensure data integrity. CAMU examine the distribution of data for both continuous and binary variables, identify outliers and anomalies, and provide treatment to improve data quality. There are two types of data quality checks performed namely: Business Data Quality Check and Technical Data Quality Check.

The Business Data Quality Check provides an overview of the portfolio including key attributes over the performance period consistent in the development time frame of the model. In addition, key portfolio segmentation was considered to identify the robustness and data quality over the same time period. Upon examining the portfolio trend, these key attributes and segments are observed with potential anomalies and portfolio breaks, providing sufficient justification and explanation of potential inconsistency of the data. Data Quality manager and Business owner or Model Sponsor, attest and review the overall trend to make sure that all observations are consistent with the observed portfolio trend.

The other data quality check performed is the Technical Data Quality, this process consist of examining the development data and the original source data, including original data attributes used in the model development, primary keys, statistical checking and account level check with the data source subject matter experts (SME). Variables used in the development sample are verified and went through a vetting process, checking for inconsistency and data issues. This process includes working closely with the MEP Optima team, to verify frequency check, data consistency, completeness, logic and conformity. Once the account level information has been tested and DM team records the frequency counts and matching for all the variables and explains the results and further attest for the quality of the data for further use in model development. Upon testing, DM team provides feedback to the modelers and data quality managers to address issues or confirm reconciliation results.

For further detail results, please see attachment files named ‘4.1.7 CCAR Business Data Quality…’ for Business Data Quality and ‘4.1.7 CCAR 2018 Technical Data Quality - Optima VarLib - Severity\_CMI\_pt1\_pt2’ for Technical Data Quality results.

What statistical tests were run to verify whether the available data quality is sufficient to develop the model? What were the results of those tests?

[Where applicable, refer to Gating Principles.]

Various Statistical Tests are conduct by the modelers to identify outliers and completeness of the data. Frequency testing, Mean and Standard Deviation analysis are performed as well as testing outlined in the Gating Principles (see Chapter 3.2.1).

Were there any significant events that have affected data quality in the past?

[Describe any known events that might affect data quality. These should not be economic or common portfolio changes but rather technical events such as data lost during data migration.]

Not applicable. The data used for model development are sourced from official and approved data sources with appropriate formal control and data quality check in place. CAMU data quality and Risk DataMart team performed detail business and technical data quality check as part of the maker-checker control process and data variance remained within established control threshold. Other observed data issues are discussed and documented in more detail in the Data Assumption and Data Limitation section in 4.2 and the corresponding mitigation applied.

Is the data coverage sufficient?

[The modeler should engage businesses to understand whether data coverage is sufficient. In cases where the data sample is limited, the modeler must determine whether available external data can augment Bank’s internal data.

Aggregated industry data is one option for augmenting Bank’s internal data. In order to use industry data, the modeler must first characterize the data from a business and statistical perspective, in order to validate that it is appropriate for the intended use. If the length of Bank’s internal data source is too short in duration and if the modeler is unable to source external data that is comparable to Bank’s internal data in both business function and statistical properties, then the modeler should proceed to document data shortcomings, including sign-offs from the business that additional relevant data is not available. If a model is developed for areas with insufficient data, please justify and explain.]

Yes. Overall development data performance window has been extended to a cover both pre-crisis and post crisis performance which capture sufficiently full business cycle performance during the stress and recovery period.

For NA Mortgage Method A Residential Mortgage model, observation and performance window used in the development extend from Feb 2006 to Dec 2017 and Jan 2009 to Jun 2017, for the Severity model covering both pre-crisis environment and recovery period post 2008. This rich data coverage captured all the policy changes, business strategy and customer behavior within the full business cycle of the mortgage portfolio.

### Vendor-model specific inputs

[If a vendor model is used, please provide the tuning and dialing parameters.

This should include parameters and settings of vendor software needed to replicate loading data into the vendor model such as data filters (within the vendor model), exclusions of certain data ranges and so on.]

The severity model is not a vendor model. There are no vendor-model specific inputs.

### Code and Data Guidance

[Model Sponsors must adhere to the coding practices and standards as mentioned in the Code and Data section of the Model Testing Guidance. The information regarding the flow of execution of codes and datasets must be provided in the attached template. All the materials must be shared as per the data sharing process. All the embedded templates should be mandatorily provided. Please refer to code and data guidance section in the Model Testing Guidance for more details]

* Provide details on what software was used for model development.

The SAS programming language was used in model development.

* Provide the location and access to all the codes and datasets. Also, attach the code run book as required by the code and data guidance.

Relevant code and datasets were uploaded to sub-directories within the following MRM server location /ccr/ccar\_mvg/1\_businessdata/US\_Secured/ccar2019/NA\_Mtg\_Method\_A/. Please refer to attachments named ‘4.1.9….xlsx’ for detailed descriptions of uploaded contents.



## Data Assumptions and Data Limitations

* What are the key performance, segmentation and account data assumptions?

[Include any implied data assumptions or expectations regarding data use.]

The implicit understanding for the model development data is that the loan-level, macro-economic and bureau information (if applicable) used for segmentation and performance measurement (where applicable) is accurate and reflects true account statuses and other characteristics that affect credit performances across time.

It is important to note here that there has been a significant improvement in data-quality control and checks, since the transition to Optima environment and as such the data has been deemed to be robust enough to support CCAR requirements.

As such, CAMU had compiled all the relevant data sources together in one comprehensive document, with description of the data type, type(s) of changes made to the data, frequency of such changes, & archival policy details. This compiled data had been validated by CAMU’s internal Data Quality team and subsequently transferred to a single data repository which had been time-stamped and frozen to minimize data pull errors and inconsistencies associated with periodic data refresh. All of these data quality management initiatives that CAMU conducted helped foster consistency and comparability across independent model suites and model usages and aided the built of a model suite that would be able to capture the portfolio trends and the associated risks.

**Table 4.2.1: Data Type Overview**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Data Type | Type(s) of changes over time | Frequency of changes | Archival frequency? | Examples |
| 1 | Static Loan Level Data | 1. Historical Restatements | 1. Rare restatements | Monthly | Loan Origination Channel: Correspondent, Retail, Wall Street etc. |
| 2 | Dynamic Loan Level Data | 1. **Incrementally added records** 2. Historical Restatements | 1. Monthly 2. Rare restatements | Monthly | Bureau Data, Delinquency status: bk1, bk2 etc. |
| 3 | Cumulative Loan Level Data | 1. **Event based** 2. Historical Restatements | 1. Event based 2. Rare restatements | Monthly | INACTIVE DETAIL: Charge Off, Pay Off etc. |
| 4 | Home Price Index | 1. **Incrementally added records** 2. Historical Restatements | 1. Monthly 2. Frequent restatements | Monthly | N/A |
| 5 | Unemployment Rate | 1. **Incrementally added records** 2. Historical Restatements | 1. Monthly 2. Rare restatements | Monthly | N/A |
| 6 | Interest Rate | 1. **Incrementally added records** 2. Historical Restatements | 1. Semi-annually/Quarterly 2. Rare restatements | Monthly | N/A |
| 7 | GDP | 1. **Incrementally added records** 2. Historical Restatements | 1. Semi-annually/Quarterly 2. Rare restatements | Monthly | N/A |
| 8 | Income | 1. **Incrementally added records** 2. Historical Restatements | 1. Semi-annually/Quarterly 2. Rare restatements | Monthly | N/A |
| 9 | SP 500 Index | 1. **Incrementally added records** 2. Historical Restatements | 1. Semi-annually/Quarterly 2. Rare restatements | Monthly | N/A |
| 10 | Volatility Index | **1. Incrementally**  **added records**  2. Historical  Restatements | 1. Semi-annually/Quarterly   2. Rare  restatements | Monthly | N/A |

Please note that all the files have been internally verified and frozen as of the last modified date to negate the possibilities of data update related errors/inconsistencies.

Other key assumptions which relate to modeling inputs are presented below.

1. The data used for the CCAR modeling exercise is considered sufficient to fulfil all CCAR and Non-CCAR reporting requirements.

2. The business’s current portfolio composition and volume are reflective of the go-forward states.

3. The model is dependent on future macro-economic values. Even if it were perfect, the accuracy of the model is only as good as that of macro-economic forecasts used to operationalize it.

* What are the key macroeconomic data assumptions?

[Include any implied data assumptions or expectations regarding data use.]

The key macro-economic data assumption is that they reflect the true underlying economic conditions. The Method A RM model suite leverages several macroeconomic attributes to gauge the effect of these economic indicators on the credit worthiness of the real estate portfolio. In the case of interest rates, which are market determined and recorded regularly, this constitutes a small modeling assumption. The importance of the unemployment rate assumption which is official government data based on surveys is comparatively much larger. Similar to unemployment level, state level income is based on the information gathered by the respective states from their Current Population Annual Survey which is based on voluntary participation of the state residents. The assumption effect is largest for HPI information which is proprietary and vendor supplied (CoreLogic - <https://www.corelogic.com/downloadable-docs/solutions/loan-performance-secondary-market-analytics-for-capital-markets/capital-markets-real-estate-analytics-suite.pdf>, - <https://www.corelogic.com/products/corelogic-hpi.aspx>).  In addition, for both Unemployment and HPI, there is an additional implicit assumption that the appropriate geographic level granularity is optimal.  Unemployment, GDP and Income are used at the state level, and HPI, is measured at the CBSA level, which consists of one or more counties (or equivalents) anchored by an urban center of at least 10,000 people along with all adjacent counties that are socioeconomically tied to the urban center by commuting. The S&P500 and VIX (tested but not used in the final model specification) are measured at the US level.

* What are the key limitations of the data?

[Include any data limitations.]

The implicit understanding for the model development data is that the loan-level details used for performance, risk drivers and segmentation (where applicable) are accurate and reflects true account status and characteristics across time.

It is important to note here, that with the migration of data to the SAS Grid/Optima/MEP environment, there has been a significant improvement in data-quality control and checks. The Optima environment leverages a centralized data source which undergoes robust data quality checks that adheres to a strict data requirement design which is mandatory for all CCAR models. This has been evidenced in detail in the data quality principles (see 4.1.1 Data Sources and Controls section for pertinent details).

CAMU acknowledges the importance of having a holistic view of the data used in the model development scope to build a robust model. The accuracy of data and transparency of related controls and quality checks around it drive the choice of models built, the performance metrics generated by the system as well as the usability of the end-to-end system by the model users. Therefore, a good ontology of the data is extremely critical for the quality of the work.

To determine if there exist true limitations around the model development data, CAMU diligently researched and executed the following checks to make sure that the data used for the 2019 CCAR process adhered to the highest quality. These checks are listed as below-

* Verified the authenticity and topography of all the listed data attributes.
* Assessed the scope of the data, especially over time, so that the model could avoid the seasonality bias.
* Checked for missing values, identify them, and assess their impact on the overall analysis
* Confirmed that the available pool of development data was large enough to build a robust model. For portfolios or segments that demonstrated sparse data, proactive quantitative decisions were made to use special analytical framework (coping mechanism) to model these loans.
* Made sure data type (numeric, character variables and so forth) is correct and set the upper and lower bounds of possible values.
* Paid extra attention to data integration given that the data comes from multiple sources (REL Datamart, GCRM Office for macroeconomic inputs, CoreLogic for HPI, etc.)

Apart from the quality checks listed above, CAMU did additional due diligence on its own or by collaborating with relevant business partners/teams to understand the reason behind missing values, inconsistencies in the data, presence of duplicates or outliers. Based on further discussions with the business partners, model end users and model reviewers, mitigating actions have been developed to address these data discrepancies to support the CCAR modeling requirements. As such steps were taken in conjunction with the model end users and data subject matter expects (SME) to ensure that the final data set was robust enough for model development. This was either achieved by normalizing/transforming the existing data to suit the purpose of model development and/or creating/leveraging a consistent set of logic to define populations segments across all models and its usages. All these steps were appropriately documented and approvals obtained before model is put into use.

Some of the prominent data weaknesses noticed this CCAR cycle are listed below, along with the corresponding mitigation plans that was adopted-

1. **Low quality of trial information for certain modification program:** The trial information such as month in trial, success/fail in trial, payment and term etc., are the proven to be valuable risk drivers especially for loans in the middle/high delinquency buckets. The predicting power for such attributes may be reduced somewhat because of the trial information is noisy or sometimes even completely miss-recorded.

**Mitigation:** Trial information data has improved overtime. Furthermore, few mods are expected in the future.

1. **Missing bureau information:** Significantly higher percentage of missing values for the no-FICO bureau attributes such as “number of open mortgage” (N\_M\_n\_open\_mort), “age of oldest trade line“ (N\_M\_age\_oldest\_trd) as late as April 2009 were observed

**Mitigation**: Missing data in the older vintage is not a current phenomenon, which can be addressed in estimation via missing value imputation. The model development data had more than adequate coverage in terms of sufficient parameters and other variables to capture the credit risk of the portfolio. Sufficient care has been taken to minimize the use of the bureau attributes by assessing their overall impact to model performance. Bureau attributes that are not adding much to the model’s performance have been dropped. For bureau attributes that are considered important and has missing values, these missing values have been imputed using a median value, thus reducing the potential of omission bias due to omitting them.

1. **Resets around revolving products (ARM/IO)**: Given the rising rate environment, there is an imminent chance of reset risks for specific home loan programs, such as adjustable-rate mortgages or interest-only mortgages that start with a low initial teaser rate and have the potential to reset to much higher rates once the fixed period ends. Such payment shocks are considered a risk factor within the model.

**Mitigation:** The model leverages several reset/payment shock attributes to assess the riskiness associated with reset. Specific timing and magnitude variables have been created which measures and quantifies the effect of resets on the loan’s performance pre and post ( within six months of payment shock) reset.

1. **Recognition of One Timer Loss Events:** One timer loss events are considered outlier observation(s) since they do not represent a normal business process. Initial data exploratory analysis on the loss data had revealed some unusual losses which resulted into significant loss amount or corresponding loss adjustment in a given time period. This was verified by the Risk Portfolio team, the Model Sponsor, who recognized similar events in their loss reporting structure. The following reasons have been narrowed down as the probable causation(s) for these one-time events.
   1. Adjustment due to principal write-downs,
   2. Modification Impact
   3. HELOC conversion,
   4. Credit from legal settlements,
   5. FFIEC Adjustments(Contra-accounting)
   6. Gain on sales

**Mitigation:** While relevant for the LGD model, one-timer loss events were imputed and removed from the development sample since they do not represent the normal business process. Initial analysis was performed and verified by model user in identifying the one timer list. For additional information on one-timer events and treatments, please refer to attachment “3.2\_Memorandum\_One-Timer Exclusions - 111417.docx”.

1. **Exclusion of data** - The exclusion of data from 2012/01-2014/03 in the PD estimation sample. The exclusion of data prior to 2008 in the LGD estimation sample. The exclusion of data prior to 2009 in the DV Logic Haircut assumption estimation.

**Mitigation** - MRM cited a limitation on the current model for not using most recent data to capture recent portfolio mix, performance behavior and macroeconomic trend with the exclusion of the data from 2012/01-2014/03 in the PD estimation sample, data prior to 2008 in the LGD estimation sample and the exclusion of data prior to 2009 in the Haircut assumption estimation. CAMU agrees with MRM that the recent data should be included in the model development sample. The quality of the data prior to 2008 is weak, 2008-2010 is CMI portfolio stress period and needed for a robust stress model. These aspects limit the use of two ends of development data for OOT exclusions. That OOT validation is a regulatory mandate. Therefore, the choice was always limited to pick up a phase in between. The period selected for OOT is excluded from development while variable stability analysis was performed on the full data and various back tests were performed to show overall satisfactory performance across time. Please refer to Section 4.1.3 for additional justification around the OOT data sample.

* Were any exogenous events/shocks evident in the historical data? How were they accounted for?

Exogenous events such as one-time accounting loss restatements have no effect on the development of the transition model which is built on delinquency and terminal events. These one-time losses are excluded from the LGD model development to reduce biases within the model’s performance. Please see LGD Development data waterfall in MDD Section 4.1.3. CAMU’s modeling practices align with Bank’s accounting policies as reflected in Global Consumer Credit Fraud Risk Policy (GCCFRP) Chapter 20.

* What factors, if any, mitigate the limitations of the data?

[Include a sub-section for each identified limitation of the data (e.g., short data period, potential inconsistencies, etc.). If none exists, provide a rationale for why one is not needed.

Please see mitigation plans attached as part of the data limitations discussion, as enumerated above.