White Paper: Modeling Loss Given Default for CCAR, IFRS 9 and CECL for Retail Portfolios

Mohammad Amir Sarosh

Analytics Partner, Quantitative Risk Management

Genpact

Abstract

Loss Given Default or LGD is a key parameter in the expected loss framework for stress testing and allowance calculation for banks. The current regulatory paradigm both in the US and Europe expects banks to develop a suite of robust granular account level PD, LGD and EAD models for their retail portfolios for stress testing as well as allowances under IFRS 9 and CECL. These models are expected to capture both endogenous risk drivers like loan level characteristics and the exogenous factors of the prevailing macroeconomic conditions to generate quarterly forecasted parameters. With the said goal in mind this discusses a framework for account level time hazard PD, LGD and EAD models that serve the objective of fixed window stress testing applications as well life of the loan, loss calculation under CECL and IFRS 9. This paper discusses the model design and architecture of account level time hazard LGD model.

Introduction

LGD is a part of EL framework where EL = PD*LGD*EAD and LGD model is expected to provide a time varying forecast of loss severity to be applied to the respective quarterly PD and EAD forecasts for stress testing and allowance calculations. LGD is preferred to be a macro sensitive statistical model to comply with the enhanced requirement of stress testing and life time loss estimation.

The current state of the industry for retail portfolio particularly unsecured is to use simplistic long run average for LGD estimation. The average in some cases are calculated separately for benign and crisis period to addresses the requirement of macro sensitivity in stress testing and for IFRS 9 and CECL loss calculations. Some banks also use another simplistic approach of Frye-Jacob estimation to bring in the required macro sensitivity in loss severity forecasts in the absence of robust macro sensitive LGD model.

None of the above approaches explicitly captures or models the macroeconomic relationship at the account or segment level. This puts the onus of driving macro sensitivity in the expected loss estimation framework on PD and EAD models. Given that the next crisis is always different than the previous one backward looking approach can at best try to mimic history but falls short of producing a robust and a credible forecast.

Given the above need of large and mid-sized banks, the proposed approach of an account level time hazard macro sensitive LGD modeling framework is most suitable as it addresses the uses case of stress testing and allowance calculations under IFRS 9 and CECL.

The objective of this white paper is to provide a high-level understanding to analysts, model developers, model validators and line of business in the risk organization on the model design and architecture of such an LGD modeling framework. The paper does not include a discussion on model segmentation, model estimation approaches of ordinary least square, Tobit or two-stage LGD estimation as there are several academic papers and industry literature available on these topics.

LGD is defined as the ratio of net loss amount and loan balance at the time of default measured as gross charge-off. The model objective is to predict the LGD percentage using macroeconomic, time hazard index and loan-level attributes as explanatory variables.

Net Loss = Loan Balance at Default – Total Recovery Amount

Choice of Data Structure

One of the biggest challenges of modeling LGD is to get reliable account level recovery information, which can be used for macroeconomic LGD model estimation. Once the reliability of the data is ascertained through reconciliation with Collections, General Ledger and Finance reports,the model developer decides on the data structure to be used for modeling.

The choice of data structure for modeling time hazard LGD can be categorized into the following options

1) Flat panel data structure constructed using the loan level information at the time of default
2) Triangular data structure constructed using information prior to default
    I. All records prior to the time of default are included
    II. Single record prior to the time of default included through random sampling

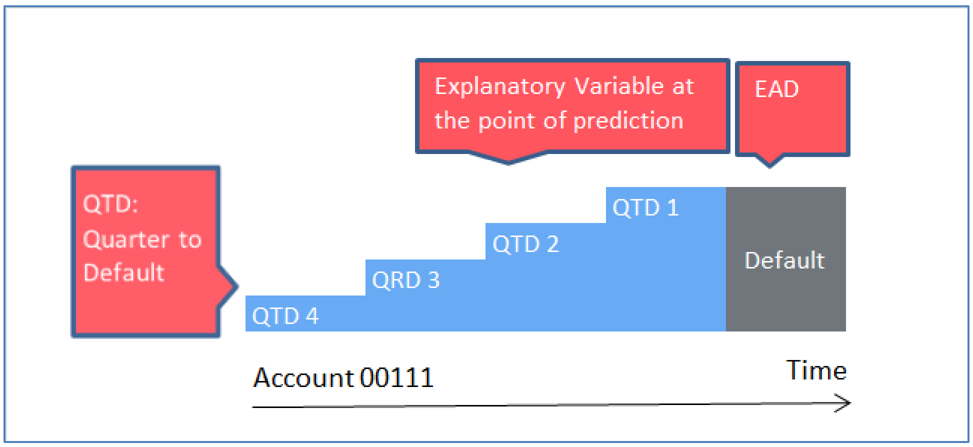
Flat panel data structure of the defaulted accounts: This data structure includes the loan level information at the time of default to model loss given default. LGD is calculated at the account level based on historical recovery proportion analysis to determine a recovery window.

The flat panel data structure however simplistic has several limitations. Loan level attributes are worst at or near the time of default and are thus are not very useful as explanatory variables or segmentation driver in modeling. For example as the loans age towards default their FICO score worsens. Thus most of the accounts closer to default will have poor FICO score which will not serve as differentiator of risk either as a modeling variable or segmentation. Flat panel data misses on the critical timing component of the time hazard LGD model, which makes it incompatible with time hazard suite of PD and EAD models.

Triangular data structure on the other hand includes the loan level characteristics prior to the point of default to model LGD along with the time index in the model through quarter to default variable. Loan level information prior to default (i.e. data implies the model is unconditional on default) will suit the model building process considering usage of the model for forecasting LGD on observation period. Triangular data structure has the option of either randomly selecting a single record prior to the point of default or including all the records prior which thickens the data for modeling and has shown the empirical evidence of improved modeling accuracy.

However, the limitation of triangular data is that there are less and less observations as one moves away from the point of default and the prediction accuracy decreases for higher "quarters since observation"in forecasting.

Following diagram illustrates the triangular data structure of time hazard LGD modeling –



Model development data structure –

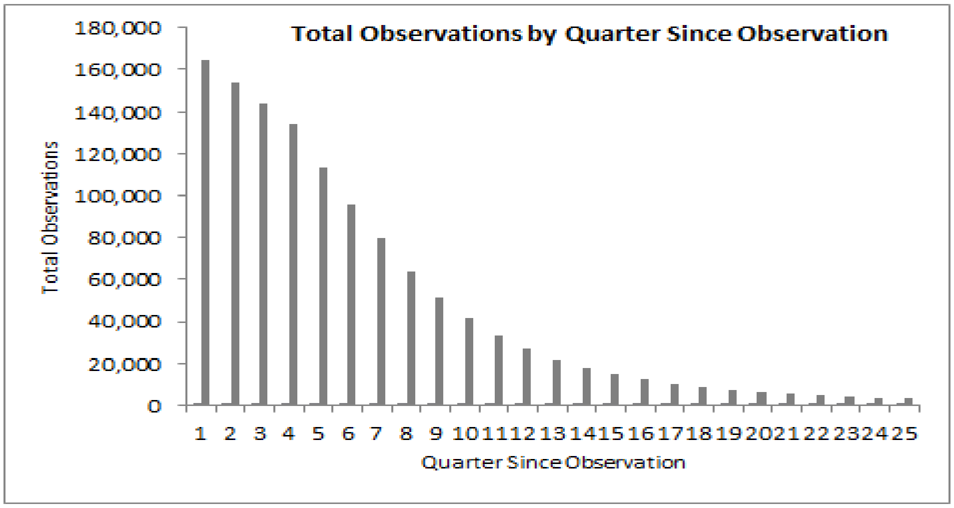
Account	Obs_Qtr	PrinBal	Orig_FICO	Current_FICO	QOB_Obs	QOB_Obs	QTD	Default Flag	Unemployment	LGD	
00111	30-Jun-12	7,500	650	650	9	15	6	0	7.8	89%	
00111	30-Sep-12	7,000	650	620	10	15	5	0	7.8	89%	
00111	31-Dec-12	6,500	650	620	11	15	4	0	7.8	89%	
00111	31-Mar-13	6,000	650	610	12	15	3	0	7.8	89%	
00111	30-Jun-13	5,500	650	580	13	15	2	0	7.8	89%	
00111	30-Sep-13	5,000	650	580	14	15	1	0	7.8	89%	
00111	31-Dec-13	4,500	650	560	15	15	0	1	7.8	89%	**Default**

Loan level characteristics at the time of prediction

Quarter to Default as a time index variable

Macroeconomic variable at the time of performance

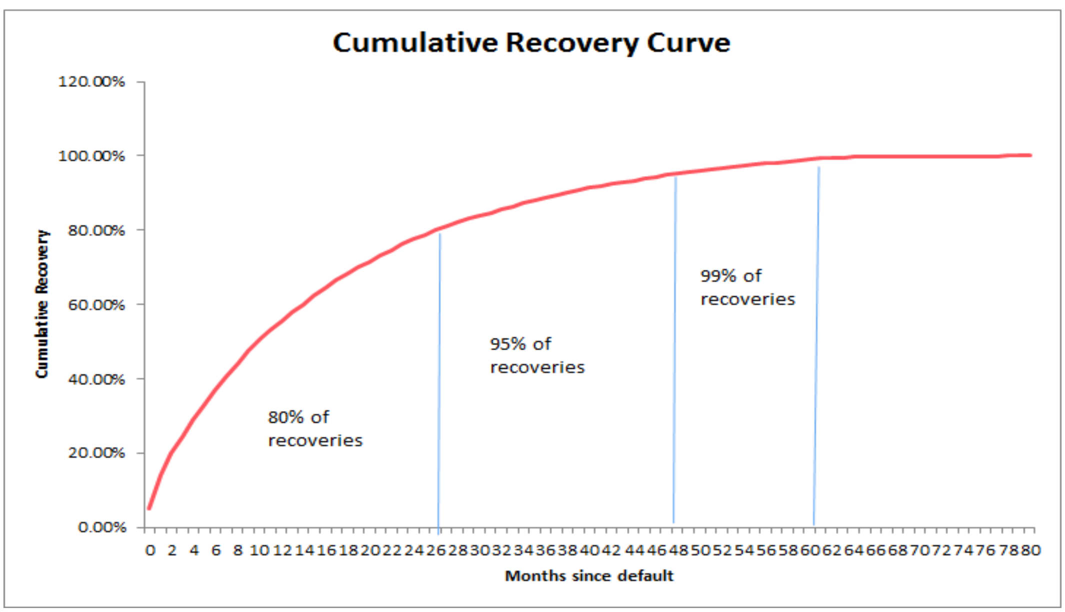
The below chart shows the limitation of triangular data structure in time hazard LGD modeling with decreasing number of observation over which the model gets trained on.



Considering the model usage for CCAR, IFR 9 and CECL, the triangular data structure is best suited for the time hazard suite of PD and EAD models. This data structure allows for changing macro scenarios, loan level characteristics at the time of prediction and moving time index variables like quarters since observation and quarters on book.

Recovery Curve

Prior to estimating the LGD model, a consistent recovery window needs to be selected so all defaults would have the same length of observed time available for recovery calculation regardless of how recently the default occurred. A recovery window is selected so as to capture the vast majority of recoveries.



Based on the review of cumulative recovery curve along with the inputs from the line of business and nature of a the product structure, a cutoff is selected to provide a balance between capturing a high percentage of recoveries while not losing too much data, given that loans without the specified recovery window will be removed from the modeling dataset.

It should be noted here that an LGD model based on fixed recovery window may be conservative, and thus a preferred approach for stress testing use case; however, it can be a less preferred assumption for allowance calculation in case of IFRS 9 and CECL. Thus, it is recommended that model developers normalize the cumulative recovery proportion to 100% post the choice of the recovery window based on the data. Banks should make a note of this assumption for LGD model,0-while tweaking their exiting CCAR stress test models for IFRS 9 and CECL use cases.

Discounting

Another key aspect that model developers should take into consideration while tweaking their existing stress test models (e.g. CCAR models) for IFRS 9 and CECL use is the discounting of the recoveries using the effective interest rate at the account level. While discounting of recoveries is not required for CCAR, it is required for the allowance calculations in IFRS 9. Interestingly, CECL does not mandate that banks do discounting for the sake of reducing complexity. However, institutions would be tempted to follow the route of discounting as it would reduce their allowances especially in the context of life of the loan calculation in the absence of any staging in CECL. Banks can develop a simpler discounting approach based on the level of granularity of the existing model where discounting can be applied at group level using historical timing curve analysis. Weighted interest rate can be used if the discounting is performed at the group level.

Model Developers should also perform recovery curve analysis by default year and analyze both the sensitivity of recovery rates and speed of recovery by macroeconomic conditions. This is especially relevant in the context of macro sensitive LGD model requirements under CCAR, IFRS 9 and CECL.

Model Testing

Banks needs to follow a holistic performance evaluation framework to test for accuracy of the time hazard LGD model for CCAR and allowance use cases of IFRS 9 and CECL. The key aspects of the performance evaluation framework that banks should consider include

- Performance metrics – Varied performance metrics including Mean Absolute Error, Mean Absolute Percentage Error, Weighted Absolute Percentage Error is recommended to be used
- Time index over which the performance metrics are evaluated – Evaluation of model performance over quarters on book and quarters since observation is recommended
- Forecast window for each of the use case – Stress testing evaluation is performed over 9 quarters, allowance testing for 4 quarters and life of the loan under new paradigm
- Model fit over the observation quarters
- Model scoring evaluation over historical portfolio snapshots to assess performance deterioration in the stress vs benign period. It is also critical to note that in a time hazard LGD

model, the actual vs predicted accuracy of accounts is evaluated for the quarter in which it defaults and not the prior quarters leading up to default.

Conclusion

Account level time hazard LGD model developed using loan level characteristics and macroeconomic drivers are well suited for stress testing use case in CCAR and allowance calculation under IFRS 9 and CECL. Such a modeling approach can serve as a unified framework that can be leveraged for multiple use cases including CCAR as well as longer horizon forecast under IFRS 9 and CECL, without having the need to develop separate models. This reduces the number of models in model inventory, lessens the burden of model governance and helps better manage, control and mitigate model risk due to a unified modeling framework. In addition, leveraging these synergies also helps banks reduce costs and realize a better return on investment.