THE SECURITISATION & STRUCTURED FINANCE HANDBOOK
2018

APSA
Asia-Pacific Structured Finance Association

Estimating loss distribution for a securitisation transaction

by Vishal Saxena and Dilip Mohan, IFMR Capital

IFMR Capital has structured, arranged and invested in more than 400 securitisation transactions over the last eight years which has helped large and small originators to access capital at affordable cost. These originators are from different asset classes like microfinance, affordable housing finance, small business loans and commercial vehicles. To give additional comfort to the investors, IFMR Capital retains skin-in-the-game by regularly investing in the senior or subordinated tranches of the securitisation deal it structures.

One of the key questions any investor would have on a securitisation transaction is what could be the potential loss in it. Based on the loss distribution, the credit enhancement and the price of the transaction are determined. Rating agencies also evaluate a securitisation transaction based on the underlying pool of assets, estimated losses and the extent of loss protection that is provided to the investors through various forms of credit enhancement.

Hence, for a securitisation transaction it is crucial to come up with a suitable method to estimate the loss distribution.

From a large database of historic performance of loans, along with access to loan portfolio data of some originators, we have built different models for estimating loss distribution for different asset classes also taking into account the vintage of the asset class and the originator.

The losses on a loan portfolio data of an originator can be estimated with the Transition Matrix & Loss Given Default (TM-LGD) model (discussed in Securitisation & Structured Finance Handbook 2016, *Commercial vehicle securitisation: Loss given default estimation using Transition Matrix (TM-LGD)* by Vaibhav Anand and Amit Mandhanya, IFMR Capital). The TM-LGD model takes the entire loan portfolio data of the originator as its input and generates a set of


Vishal Saxena Dilip Mohan

Vishal Saxena, Associate Director - Risk Analytics and Modelling

tel: +91 95 6620 5899

email: vishal.saxena@ifmr.co.in

Dilip Mohan, Senior Associate - Risk Analytics and Modelling

tel: +91 75 5000 9981

email: dilip.mohan@ifmr.co.in

Transition Matrices (which we will be discussing in this article). The losses are then estimated from these TMs (Transition Matrices). A limitation of the TM-LGD model is that it is very difficult to generate a set of TMs for the loan portfolio from a given set of peer TMs. The model depends on the availability of the loan portfolio data.

We have built an alternative model which estimates the losses of a securitisation transaction based on a few historical parameters. We can calculate these parameters based on some historical data of the originator or a peer. These parameters can then be notched up or down based on the riskiness of the originator's loan portfolio.

This model uses a limiting form for the portfolio loss distribution with a systemic factor (using a conditional independence framework) derived by Oldrich Vasicek, a Czech mathematician and quantitative analyst, in 1991. But, one of the assumptions in Vasicek's work was that the loans in the portfolio were homogenous. In reality, however, there are wide variations in the repayment patterns and behaviours for different asset classes – like housing finance, vehicle finance, microfinance, small business loans, etc. Even within a specific asset class, the loan repayment behaviour varies between subgroups. For example in vehicle finance, the characteristic of a light commercial vehicle loan is completely different from that of a heavy commercial vehicle loan.

We have come up with an approach that extends Vasicek's model to take care of non-homogenous loans in the portfolio, so that it can be used to determine the loss distribution of securitisation transactions.

The limiting distribution of portfolio losses

By using Merton's approach Vasicek assumes that a loan will default at its maturity T, if at T its assets are lower than its obligations payable B.

Let A_i be the value of the i^{th} borrower's assets, described by the process:

$$dA_i = \mu_i A_i dt + \sigma_i A_i dX_i$$

The asset value at T can be obtained by integration:

$$\log A_i(T) = \log A_i + \mu_i T - \tfrac{1}{2}\sigma_i^2 T + \sigma_i \sqrt{T} X_i \quad (1)$$

Where $X_i \sim N(0,1)$ is a standard normal variable.

The probability of default of the i^{th} loan is given by:

$$p_i = P[A_i(T) < B_i] = P[\log A_i(T) < \log B_i] = P[X_i < c_i] \quad (2)$$

$$\text{Where, } c_i = \frac{\log B_i - \log A_i - \mu_i T + \tfrac{1}{2}\sigma_i^2 T}{\sigma_i \sqrt{T}}$$

c_i represents the default threshold.

Consider a portfolio of n loans of equal amount, equal probability of default p, same maturity T and a flat correlation coefficient ρ between the asset values of the two observations.

Let, L_i be the gross loss on i^{th} loan. No recovery is assumed which means $L_i = 1$ when a loan default and $L_i = 0$ when the loan does not default. So the gross loss L of the portfolio is:

$$L = \sum_{i=1}^{n} L_i$$

If the default of loans in the portfolio is independent of each other, then by central limit theorem the portfolio loss distribution would converge to a normal distribution as the portfolio size increases. But the event of defaults is not independent and hence the conditions for central limit theorem are not satisfied and L is therefore not normally distributed.

The variables X_i in Equation (1) are jointly standard normal with equal pair-wise correlations ρ, and can therefore be represented as:

$$X_i = \sqrt{\rho}\, Y + \sqrt{1-\rho}\, Z_i \quad (3)$$

Where, Y, Z_1, Z_2, ... , Z_n are mutually independent standard normal variables.

Y can be interpreted as a common risk factor for the portfolio, such as GDP over the interval (0, T), $\sqrt{\rho}Y$ can be interpreted as the contracts exposure to the common risk factor and $\sqrt{1-\rho}Z_i$ represents company specific risk.

When the common factor Y is fixed, the conditional probability of loss on any one loan is:

By using equation (2) and (3):

$$p(Y) = p\ (L_i = 1\ |Y) = p(X_i < Z_i|Y)$$

$$p(Y) = N\left(\frac{N^{-1}(p) - \sqrt{\rho}Y}{\sqrt{1-\rho}}\right) \qquad (4)$$

By the law of large numbers the portfolio loss conditional on Y converges to its expectation p(Y).

As $n \to \infty$: $L(Y) \to p(Y)$

Then:

$$P(L \leq x) = P(p(Y) \leq x) = P(Y \geq p^{-1}(x)) = N(-p^{-1}(x))$$

By using equation (4):

$$P(L \leq x) = N\left(\frac{\sqrt{1-\rho}N^{-1}(x) - N^{-1}(p)}{\sqrt{\rho}}\right)$$

This result is given in Vasicek (1991).

This limiting distribution also holds true in case of a portfolio with contracts of different ticket size, as long as the portfolio is not dominated by a few loans which are much larger than the others.

Extension of the model to handle non homogeneity in loan repayment behaviour and estimation of transaction loss

Let us take an example of a securitisation transaction of a diversified vehicle finance portfolio which has exposure to light commercial vehicles (LCV), medium commercial vehicles (MCV), heavy commercial vehicles (HCV), construction equipment (CE) and cars.

Loans in different vehicle types are not homogeneous, they are of different maturities, ticket size and loan behaviour. For example, HCV and CE loans have larger ticket sizes and longer maturity. Their loan repayment depends heavily on economic cycle. LCV and car loans have relatively smaller ticket sizes and shorter maturity. In LCVs, the loan repayment depends on local factors like route profitability and in cars the loan repayment depends on individual borrower characteristics.

Also, we have observed that loan behaviour among subgroups may be correlated. For example, a reduction in mining can negatively impact the cashflows for HCV as well as CE. So, for estimating the total portfolio loss, the inter group correlation in the subgroups needs to be incorporated in the model as well.

To handle the non-homogeneity, we divide the total loan portfolio into different homogeneous sub-groups and then estimate subgroup-wise parameters for Vasicek'slimiting form distribution i.e. correlation (p) and probability of default (p). This division can be done either based on expert judgement or by using some statistical technique like clustering. We also estimate a correlation matrix C, of the subgroups, and arrive at the inter subgroup correlations from this. These parameters are estimated based on historical loan repayment data.

3

By using Cholesky Decomposition, the correlation matrix C is decomposed as:

$C = L*L^T$

Using this matrix L we can generate correlated random numbers $\bar{Z}$ from uncorrelated numbers $\bar{R}$ by multiplying them with matrix L.

$\bar{Z} = \bar{R}*L$

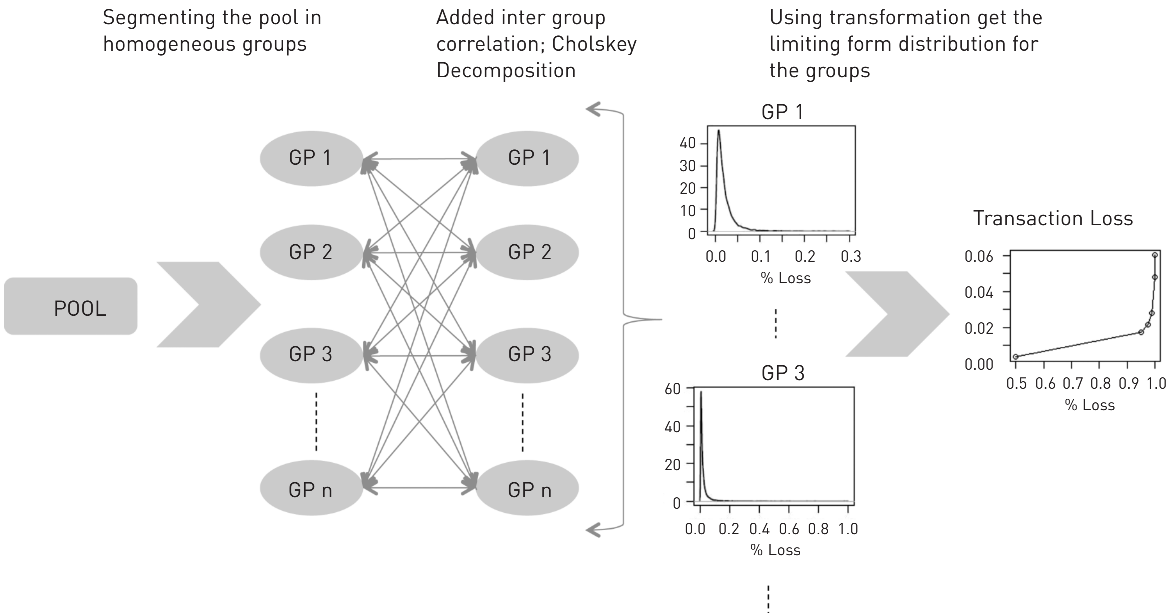
Monte Carlo simulation is used to generate the portfolio loss distribution given a set of marginal loss CDFs of different subgroups, correlation matrix C and a covariance matrix V amongst the subgroups.

Let $X=(X_1, X_2, ..., X_n)$ be a random vector. We already have the marginal loss distributions of different subgroups $F_{xi}(x)$, the correlation matrix C and the covariance matrix V.

Assuming that the marginal distributions and the correlation matrix are consistent, which means that there is a joint CDF with the desired marginal distributions and correlation matrix.

Let $(Z_1, Z_2, ..., Z_n) \sim MN\ (0,V)$ where V is a covariance matrix with 1's on the diagonal.

Transaction loss model schema **Exhibit 1**



Source: IFMR Capital

$Z_i \sim N(0,1)$ for $i = 1,2,\ldots,n$; $\phi(.)$ And $\varphi(.)$ be the PDF and CDF of the standard normal variable.

Then $X_i = F^{-1}_{xi}(\varphi(Z_i))$ has the desired marginal distribution $F_{xi}(.)$

Since Z_i's are correlated, $\varphi(Z_i)$'s are correlated and hence X_i's are also correlated.

The total loss for the portfolio is obtained by aggregating all the subgroups' losses in every simulation.

$$\textit{Total portfolio loss} = \sum_{i=1}^{k} X_i$$

where k is the total number of subgroups in the pool.

The loss distribution of the loan portfolio is obtained from multiple simulations.

Parameter estimation

Two parameters in the loss model need to be estimated. They are (i) the probability of default of the subgroups; and (ii) correlation within a subgroup. For estimating the inter group correlation, the model also requires a correlation matrix among various subgroups.

Identification of instance of default

The first step in estimating the probability of default is identifying the instance of default for the subgroups. For this, we need the "Transition Matrices" of loan delinquencies in the respective subgroup. A Transition Matrix is a *nxn* matrix with rows denoting the initial delinquency states and the columns denoting the future delinquency states after a transition period. Each cell in the transition matrix gives the probability of loans moving from one delinquency state to another. An example of transition matrix is shown in Exhibit 2.

The circled number A denotes that there is a 2.3% probability that a current contract (a contract with no overdue) will miss one instalment in the next period and move to '1 Instalment OD bucket'. The circled number B

Illustrative Transition Matrix **Exhibit 2**

	Fore Closed	Part Prepaid	Current	1 Inst OD	2 Inst OD	3 Inst OD	4 Inst OD	5 Inst OD	6 Inst OD	7 Inst OD	8 Inst OD
Fore Closed	100%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Part Prepaid	99.70%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Current	1.20%	5.0%	91.5%	2.3% A	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1 Inst OD	0.0%	1.6%	12.6%	66.8%	18.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2 Inst OD	0.0%	3.2%	3.1%	5.1%	49.8%	38.8%	0.0%	0.0%	0.0%	0.0%	0.0%
3 Inst OD	0.0%	1.8%	2.7%	0.7%	3.1% B	31.8%	60.0%	0.0%	0.0%	0.0%	0.0%
4 Inst OD	0.0%	2.4%	0.3%	0.2%	0.3%	1.6%	20.0%	75.2%	0.0%	0.0%	0.0%
5 Inst OD	0.0%	0.5%	1.1%	0.1%	0.1%	0.1%	0.8%	12.3%	84.9%	0.0%	0.0%
6 Inst OD	0.0%	0.3%	0.4%	0.0%	0.0%	0.0%	0.1%	0.4%	7.2% C	91.5%	0.0%
7 Inst OD	0.0%	0.0%	0.3%	0.0%	0.0%	0.1%	0.0%	0.1%	0.3%	7.6%	91.7%
8 Inst OD	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	99.3%

Source: IFMR Capital

denotes that there is a 3.1% probability that a contract which has missed three instalments till date will pay its next instalment due and will also pay one of its overdue instalment and hence the contract will move from '3 Instalment OD bucket' to '2 Instalment OD bucket'.

Historical transition matrices are generated for loans in the subgroups and they are analysed to identify the delinquency bucket from which the recovery is significantly low.

In the given example, the box C shows that the probability of any payment from '6 Instalments OD bucket' is less than 10%, which means that once a contract misses six instalments it is highly unlikely that the loan will further pay. Hence we will identify the instance of default as delinquency bucket 6.

Estimating probability of default '*p*' for the subgroup

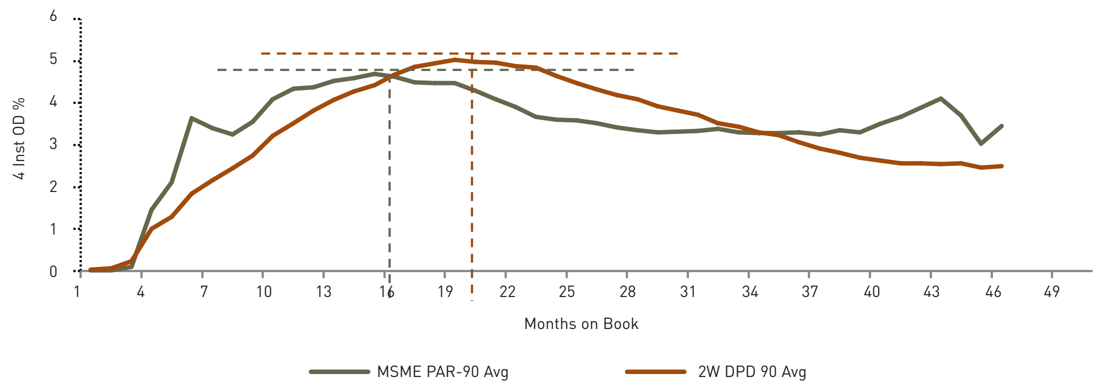
After identifying the instance of default for the subgroup, its probability of default (PD) can be estimated from the repayment history of similar loan contracts.

The loan contracts used for estimating PD should have a minimum vintage (or loan repayment history) that depends upon the instance of default. The minimum vintage for the loans can be estimated by using vintage charts of similar historical loans.

For calculating the PD of an MSME and a 2W portfolio of a multi-asset class originator, let us assume that from the TMs of the MSME and 2W loan portfolios, the instance of default in both of them is estimated to be '4 instalments overdue'. First a vintage chart of these portfolios are drawn, where the X-axis is the month on books for a loan and the Y-axis is the percentage of loans with four or more instalments overdue as shown in Exhibit 3.

As shown in the vintage chart, the percentage of four or more instalments OD loans is peaking at the 16th month for MSME and at the 20th month for 2W, and then it is going down. It means that a subset of MSME loans with a repayment history of 16 months and 2W loans with a repayment history of 20 months are sufficient to capture the movement of loans to and from '4 Instalments overdue' bucket.

Vintage chart – 2W and MSME loans of a multi asset originator **Exhibit 3**



Source: IFMR Capital

After sub-setting the historical data, PD can be estimated as a function of % of loans in the instance of default in the subset portfolio.

$$PD = f\left(\frac{count\ (instance\ of\ default\ contracts}{count\ (live\ contracts\ )}\right)$$

The function *f* can be as simple as an average of historical periods or can be as complex as an equation combining min, max, average etc.

Let us assume *f* as a weighted average function where recent disbursements are given more weights than the older disbursements. At time '*t*', the % of Instance of defaults is calculated as:

$$Q_t = \frac{count\ (instance\ of\ default\ contracts)}{count\ (Live\ contracts\ )}$$

Month wise % of Instance of defaults Q_{ti}, is calculated from the monthly disbursements. PD is calculated as:

$$PD = w_{t-1}Q_{t-1} + w_{t-2}Q_{t-2} + \ldots\ldots\ldots\ w_{t-k-1}Q_{t-k-1} + w_{t-k}Q_{t-k}$$

The weights w_i will be in declining ratio i.e. $w_{t-i} > w_{t-i-1}$; which means that the recent group gets the highest weightage and the oldest group gets the lowest weightage, with intermediate weights reducing with time.

Estimating inter and intra group correlations

The Pearson's correlation coefficient can be obtained for the loans within a subgroup and for the loans across the subgroups. These will represent the estimate for intra-group and inter-group correlation for the model.

Model validation

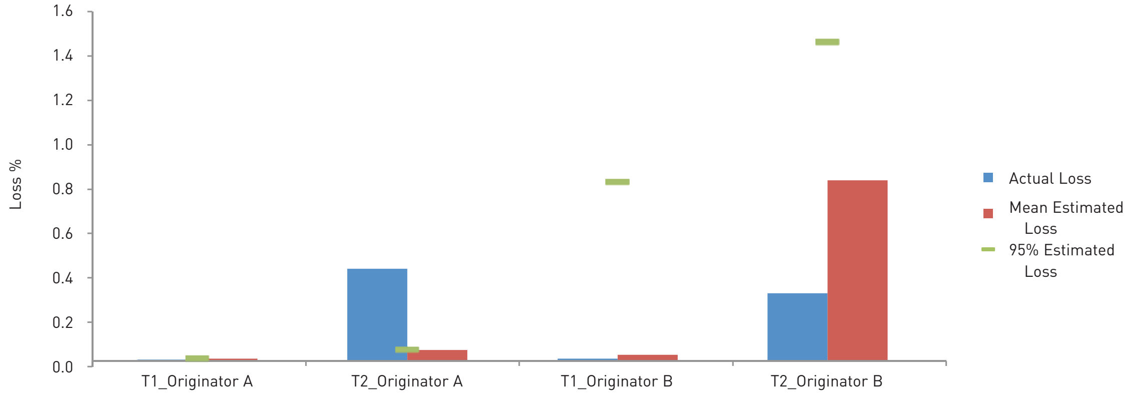
The methodology used to estimate parameters have to be validated and updated on a regular basis by back testing on existing data. Some of the results for microfinance and commercial vehicle transactions are shown below.

Microfinance loan portfolio

The model was used to estimate the loss for securitisation transactions of two microfinance originators. For these originators, we observed that the loan behaviour depends predominantly on the geographical factors and the defaults are generally a function of geographical events like natural disasters and/or socio-political events. Therefore, for estimating the loss in these transactions, we divided the pool into different geographical segments like District, City, Pin-Code, etc.

Microfinance – actual vs estimated

Exhibit 4



Source: IFMR Capital

From historical TMs (Transition Matrices), we identified that once a contract misses three instalments the chances of loan recovery gets very low, hence the instance of default is taken as '3 instalments OD'.

From the vintage charts we observed that we need to subset the loans with a minimum vintage of 12 months to sufficiently capture the loan behaviour.

For every geographical subgroup, the probability of default is estimated from previous four quarters' data. For a quarter *t*, the percentage of loans with three or more instalments overdue is:

$$Q_t = \frac{count\ (\geq 3\ instalment\ OD\ contracts)}{count\ (Live\ contracts\ )}$$

Probability of default at time t is estimated by using max function:

$$PD = \max\ (Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4})$$

Inter and intra geographical correlation is calculated from the subset data.

The result of the model validation is shown in Exhibit 4.

It can be seen from Exhibit 4 that the losses estimated by the model are ± 0.5% of the actual losses. Another key observation is that the tail of the model takes care of subgroup-wise concentration risk in the pool.

HHI index is a measure of concentration. High HHI index indicates more concentration. As shown in Exhibit 5, the transaction 'T1_Originator B' is having more

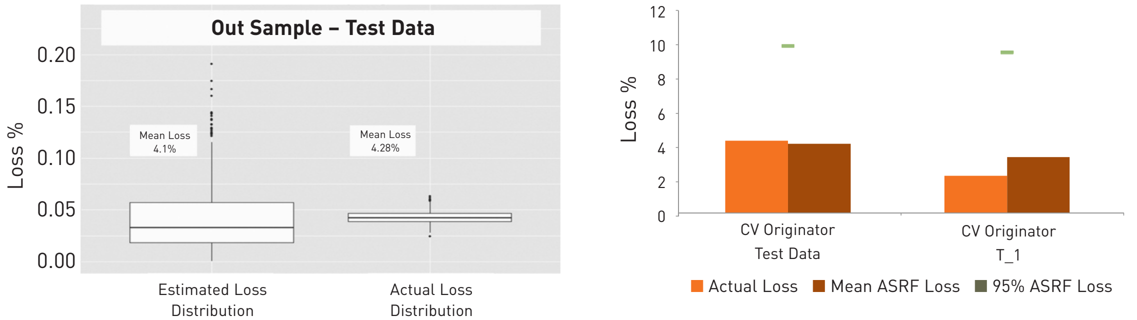
Concentration risk

Exhibit 5

	HHI Index	Estimated Mean Loss	Estimated 95% Loss
Tl_Originator B	0.1838	0.0293%	0.8197%
T2_OriginatorA	0.0610	0.0497%	0.0504%

Source: IFMR Capital

Commercial vehicle – actual vs estimated **Exhibit 6**



Source: IFMR Capital

concentration risk (0.1838) as compared to the transaction 'T2_Originator A'(0.0610). Even though the model estimated lower mean loss for 'T1_Originator B' due to its lower PD estimate, the tail losses i.e. the 95 percentile losses of 'T1_Originator B' is higher.

Commercial vehicle loan portfolio

We applied the model to estimate the loss of an out sample test portfolio of a commercial vehicle entity and a securitisation transaction of the same entity.

The loan portfolio is divided into different vehicle types like LCV (light commercial vehicle), HCV (heavy commercial vehicle), CE (construction equipment), cars, 3W (three wheelers) etc.

From the historical TMs, the instance of default is estimated to be '4 instalments OD'.

From the vintage charts we observed that we need to subset the loans with a minimum vintage of 12 months to sufficiently capture the loan behaviour.

Probability of default is estimated by using previous four quarters' data. For a quarter *t*, the percentage of loans with four or more instalments overdue is:

$$Q_t = \frac{\text{count } (\geq 4 \text{ instalment OD contracts})}{\text{count (Live contracts)}}$$

Probability of default at time t is estimated as:

$$PD = \max\,(Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4})$$

Inter and intra vehicle correlation is calculated from the subset data.

As shown in the output, the spread of the estimated loss distribution is greater than the actual loss distribution to cover the tail risks, however the mean losses for the estimated and actual loss distributions are similar.

In both the microfinance transaction as well as the commercial vehicle transaction, we could observe that the loss numbers provided by this model are similar to the actual losses, and moreover the differences are on the conservative side on most occasions. These evidences suggest that this model provides a better estimate of the loss distribution for securitisation transactions than heuristic assumptions being followed in many cases.

This approach, which we have discussed in this article, can also be used to estimate the loss distribution of a financial institution's entire portfolio, which can then be used to calculate the loan loss reserves, economic capital and value-at-risk.

References:

- Vasicek O, 1987, *Probability of Loss on Loan Portfolio*, KMV Corporation (available at kmv.com)

- Vasicek O, 1991, *Limiting Loan Loss Probability Distribution*, KMV Corporation (available at kmv.com)
- Bohdalová, Mária, and L'udomír Šlahor. "*Monte Carlo Simulations of the multivariate distributions with different marginal*".
- Vaibhav Anand and Amit Mandhanya, IFMR Capital, 2016, *Commercial vehicle securitisation: Loss given default estimation using Transition Matrix (TM-LGD)*, The Securitisation and Structured Finance Handbook.

Contact us:

IFMR Capital

10th Floor, IIT M Research Park,
1, Kanagam Village, Taramani, Chennai 600 113, India

tel: +91 44 6668 7377

web: http://capital.ifmr.co.in/