ASYMMETRIX

Asymmetrix IRB Estimation, Validation & Calibration Tool

User Manual

Nov-2013

V 1.0

Table of Contents

Abbreviations

EVC	Asymmetrix – IRB Estimation, Validation & Calibration Tool
PD	Probability of Default
CCF	Credit Conversion Factor
LGD	Loss Given Default
AR	Accuracy Ration or Gini
IV	Information Value
KS Statistic	Kolmogorov-Smirnov Statistic
FAR	False Alarm Rate
ROC	Receiver Operating Characteristics
AUROC	Area Under ROC
HR	Hit Rate
CIER	Conditional Information Entropy Ratio
KL	Kullback-Leibler
CAP	Cumulative Accuracy Profile
SPGH	Spiegelhalter Test
MSE	Mean Square Error
CLAR	Cumulative LGD Accuracy Ratio
CEAR	Cumulative EAD Accuracy Ratio

Document History

Version	Date	Author	Changed Sections	Description of Changes
1.0	Sep-12	Asymmetrix		First version

1. Introduction

PD, LGD and EAD estimates are required for Regulatory and Economic capital estimation as well as credit risk pricing, sanctioning, provisioning, limit setting and monitoring. Risk Rating systems need to be continuously validated in terms of their discriminatory power, calibration and stability, to ensure reliable output.

Asymmetrix IRB Estimation, Validation & Calibration Tool helps financial institutions in meeting Basel II IRB minimum requirements and validating the health of its internal risk rating systems based on industry best practices.

This document explains the data flow (input & output) of the solution along with output interpretation.

2. Master Data

The following Master data points are required to be maintained

a. Rating Source Master

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_RATING_SRC_CD	Rating Source Code	This column stores the rating model code for which the rating code has been provided. Ex. CRISIL, LC, SME, etc	String	30	Yes	No
V_RATING_SRC_DESC	Rating Source Description	This column stores the rating model code for which the rating code has been provided. Ex. CRISIL, Large Corporate, Small and Medium Enterprises, etc	String	100	No	No
N_LWR_CORR	Correlation for highest PD	The asset correlation function is built of two limit correlations of 12% and 24% for very high and very low PDs (100% and 0%, respectively). Source: An Explanatory Note on the Basel II IRB Risk Weight Functions July 2005 by BIS.	Number	38,25	No	No
N_UPR_CORR	Correlation for lowest PD	The asset correlation function is built of two limit correlations of 12% and 24% for very high and very low PDs (100% and 0%, respectively). Source: An Explanatory Note on the Basel II IRB Risk Weight Functions July 2005 by BIS.	Number	38,25	No	No
N_SLOPE	Slope or K-factor	The exponential function decreases rather fast; its pace is determined by the so-called “k-factor”, which is set at 50 for corporate exposures. Source: An Explanatory Note on the Basel II IRB Risk Weight Functions July 2005 by BIS.	Number	38,25	No	No
N_FIRM_SIZE	Firm Size	In addition to the exponentially decreasing function of PD, correlations are adjusted to firm size, which is measured by annual sales. Source: An Explanatory Note on the Basel II IRB Risk Weight Functions July 2005 by BIS.	Number	10	No	No
F_INT_EXT_FLAG	External or Internal Flag	This column stores the flag to indicate whether the rating model is internal model (IB) or an external rating agency (EB)	String	3	No	No

b. Rating Code Master

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_RATING_CD	Rating Code	This column stores the rating nomenclature used, ex. AAA, AA+,BBB, etc	String	30	Yes	No
V_RATING_SRC_CD	Rating Source Code	This column stores the rating model code for which the rating code has been provided. Ex. CRISIL, LC, SME, etc	String	30	Yes	No
N_RATING_ORDER	Rating Order	This column stores the rating rank order, for ex AAA code has rank 1, AA+ has rank 2 and so on	Number	10	No	No
V_RATING_TYPE	Rating Type	This column stores the type of rating, ex. BORROWER-STD for non-default borrower ratings, BORROWER-DEF for default borrower ratings	String	30	No	No
F_MODIFIER_FLAG	Modifier Flag	This flag indicates, if the rating code used has a modifier attached or not, ex. A+, AA-, etc have a modifier, while A, AA, etc don’t. The value should be saved as ‘Y’/’N’.	String	3	No	No

Note: The two master tables are joined on the basis of rating source code.

3. Rating & Default Data

The following data points are required at monthly/quarterly/annual intervals.

a. Rating Data

This table stores the applicable rating of a borrower on any given date.

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_RATING_SRC_CD	Rating Source Code	This column stores the internal rating model code under which the borrower has been rated. Ex. LC, SME, etc	String	30	Yes	No
V_BANK_CUST_CD	Bank Customer Code	This column stores the customer code for which self-cure data is being provided.	String	30	Yes	No
D_RATING_DATE	Rating Date	This column stores the date on which the borrower was rated	Date		Yes	No
V_RATING_CD	Rating Code	This column stores the borrower rating assigned on the given rating date	String	30	Yes	No
N_PARAM_SCORE1	Score for Parameter 1	This column stores the score assigned to the borrower for parameter 1	Number	35,25	No	Yes
N_PARAM_SCORE2	Score for Parameter 2	This column stores the score assigned to the borrower for parameter 2	Number	35,25	No	Yes
N_PARAM_SCORE3	Score for Parameter 3	This column stores the score assigned to the borrower for parameter 3	Number	35,25	No	Yes
N_PARAM_SCORE4	Score for Parameter 4	This column stores the score assigned to the borrower for parameter 4	Number	35,25	No	Yes
N_PARAM_SCORE5	Score for Parameter 5	This column stores the score assigned to the borrower for parameter 5	Number	35,25	No	Yes
N_PARAM_SCORE6	Score for Parameter 6	This column stores the score assigned to the borrower for parameter 6	Number	35,25	No	Yes
N_PARAM_SCORE7	Score for Parameter 7	This column stores the score assigned to the borrower for parameter 7	Number	35,25	No	Yes
N_PARAM_SCORE8	Score for Parameter 8	This column stores the score assigned to the borrower for parameter 8	Number	35,25	No	Yes
N_PARAM_SCORE9	Score for Parameter 9	This column stores the score assigned to the borrower for parameter 9	Number	35,25	No	Yes
N_PARAM_SCORE10	Score for Parameter 10	This column stores the score assigned to the borrower for parameter 10	Number	35,25	No	Yes
N_PARAM_VALUE1	Value for Parameter 1	This column stores the Value assigned to the borrower for parameter 1	Number	35,25	No	Yes
N_PARAM_VALUE2	Value for Parameter 2	This column stores the Value assigned to the borrower for parameter 2	Number	35,25	No	Yes
N_PARAM_VALUE3	Value for Parameter 3	This column stores the Value assigned to the borrower for parameter 3	Number	35,25	No	Yes
N_PARAM_VALUE4	Value for Parameter 4	This column stores the Value assigned to the borrower for parameter 4	Number	35,25	No	Yes
N_PARAM_VALUE5	Value for Parameter 5	This column stores the Value assigned to the borrower for parameter 5	Number	35,25	No	Yes
N_PARAM_VALUE6	Value for Parameter 6	This column stores the Value assigned to the borrower for parameter 6	Number	35,25	No	Yes
N_PARAM_VALUE7	Value for Parameter 7	This column stores the Value assigned to the borrower for parameter 7	Number	35,25	No	Yes

N_PARAM_VALUE8	Value for Parameter 8	This column stores the Value assigned to the borrower for parameter 8	Number	35,25	No	Yes
N_PARAM_VALUE9	Value for Parameter 9	This column stores the Value assigned to the borrower for parameter 9	Number	35,25	No	Yes
N_PARAM_VALUE10	Value for Parameter 10	This column stores the Value assigned to the borrower for parameter 10	Number	35,25	No	Yes
V_SECTION_RATING 1	Rating for Section 1	This column stores the rating assigned to the borrower for section 1	Varchar2	30	No	Yes
V_SECTION_RATING 2	Rating for Section 2	This column stores the rating assigned to the borrower for section 2	Varchar2	30	No	Yes
V_SECTION_RATING 3	Rating for Section 3	This column stores the rating assigned to the borrower for section 3	Varchar2	30	No	Yes
V_SECTION_RATING 4	Rating for Section 4	This column stores the rating assigned to the borrower for section 4	Varchar2	30	No	Yes
V_SECTION_RATING 5	Rating for Section 5	This column stores the rating assigned to the borrower for section 5	Varchar2	30	No	Yes
V_RATING_TYPE	Rating Type	This column stores the rating type, i.e. whether the rating is a borrower rating or a facility rating. Deafult values are 'BORROWER'/'FACILITY'.	Varchar2	30	No	No

Note: users can provide up to 200 parameter scores, 200 parameter values, 15 sectional scores, 15 sectional values and 15 sectional ratings.

b. Default Data

This table stores the IRB default data, i.e. data relating to restructured and NPA borrowers.

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_BANK_CUST_CD	Rating Source Code	This column stores the rating model code for which the rating code has been provided. Ex. CRISIL, LC, SME, etc	String	30	Yes	No
D_RESTR_DATE	Restructured Date	This column stores the date on which the borrower was restructured	Date		No	No
D_DEFAULT_DATE	Default Date	This column stores the date on which the borrower defaulted	Date		No	No
V_RATING_SRC_CD	Rating Source Code	This column stores the internal rating model code under which the borrower has been rated. Ex. LC, SME, etc	String	30	Yes	No
V_BORR_STATUS_CD	Borrower Status Code	This column stores the borrower status, NPA for defaulted borrowers and RESTR for restructured borrowers.	String	30	Yes	No

c. Cure Data

This table stores the details of the cured borrowers after a particular default/ restructuring

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_BANK_CUST_CD	Bank Customer Code	This column stores the customer code for which self-cure data is being provided.	String	30	Yes	No
D_CURE_DATE	Cure Date	This column stores the date on which the borrower got cured	Date		Yes	No

V_RATING_SRC_CD	Rating Source Code	This column stores the internal rating model code under which the borrower has been rated. Ex. LC, SME, etc	String	30	Yes	No

All these tables are linked through Rating Source Code and Borrower Code.

4. Setup Tables

a. Parameter Mapping Configuration Table

This table stores the details of the parameters for which the data has been provided. The reference table is STG_RATING_DATA.

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_PARAM_COLUMN	Parameter Name	This column stores the parameter name	String	100	Yes	No
COLUMN_NAME	Column Name	This column stores the column name in the stg_rating_data table in which the given parameter is stored	String	30	Yes	No
PARAM_DATA_TYPE	Parameter Data Type	This column stores the parameter data type. i.e. Discreet or Continuous	String	30	Yes	No
V_PARAM_TYPE	Parameter Type	This column stores the parameter type, i.e. Score, Value, Rating	String	30	Yes	No
N_SORT_ORDER	Sort Order of the Parameter	This column stores the sort order of the parameter. 1 - sort by descending order 2 - sort by ascending order 3 - sort positive values by ascending order negative values in descending order (positive first) 4 - sort positive values by descending order negative values in ascending order (positive first) 5 - sort positive values by ascending order negative values in descending order (negative first) 6 - sort positive values by descending order negative values in ascending order (negative first) 7 - sort positive values by descending order negative values in descending order (negative first) 8 - sort positive values by ascending order negative values in ascending order (positive first) 9 - sort positive values by ascending order negative values in one bin (negative first) 10 - sort positive values by descending order negative values in one bin (negative first)	Number	10	Yes	No
N_BANDS	Number of Bins	This column stores the number of bins to be formed for continous data.	Number	10	No	Yes
COL_TYP	Column Type	This column stores the type of data stored in the column, Number or String	String	30	Yes	No
V_SECTION_NAME	Section Name	This column stores the section name to which the parameter belongs	String	100	Yes	No
V_THRESHOLD_TYPE	Threshold Type	This column stores the threshold type for calculating certainty rating	String	30	Yes	No
V_RATING_SRC_CD	Rating Source Code	This column stores the rating model code for which the rating code has been provided. Ex. CRISIL, LC, SME, etc	String	30	Yes	No

b. Test Threshold Master

This table stores the details of the various test (AR, AUROC, etc) and their related thresholds. This table is used to assign certainty rating to the test values derived.

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_RATING_SRC_CD	Rating Source Code	This column stores the internal rating model code for which the validation is being done. Ex. LC, SME, etc	String	30	Yes	No
V_THRESHOLD_TYPE	Threshold Type	This column stores the threshold types, i.e. whether the threshold provided is at the model level (Model), or for a particular section (Section1, Section2) or for a parameter type (Parameter1, Parameter2), etc	String	30	Yes	No
V_COLOR_BAND_CD	Color Band	This column stores the color band for which the threshold is being provided. Red/Amber	String	30	Yes	No
N_THRESHOLD_LEVEL	Threshold Value	This column stores the threshold value (between 0 and 1)	Number	35,25	Yes	No
F_MODIFIER_FLAG	Modifier Flag	This flag indicates, if the rating code used has a modifier attached or not, ex. A+, AA-, etc have a modifier, while A, AA, etc don't. The value should be saved as 'Y'/'N'.	String	3	No	No

c. Assigned PD Master

This table stores the assigned PD details for each model and is used for calibration of the model.

Column Name	Column Short Description	Column Long Description	Data Type	Length	Is PK?	Is Null?
V_RATING_CD	Rating Code	This column stores the rating nomenclature used, ex. AAA, AA+,BBB, etc	String	30	Yes	No
V_RATING_SRC_CD	Rating Source Code	This column stores the internal rating model code for which the calibration is being done. Ex. LC, SME, etc	String	30	Yes	No
N_RATING_ORDER	Rating Order	This column stores the rating rank order, for ex AAA code has rank 1, AA+ has rank 2 and so on	Number	10	No	No
N_ASSIGNED_PD	Assigned PD	This column stores the assigned pd for the rating grade	Number	35,25	Yes	No

5. Adding/Modifying a new model

To add/ modify a new model, changes will have to be done to MST_RATING_SOURCE

V_RATING_SRC_CD	V_RATING_SRC_DESC	N_LWR_CORR	N_UPR_CORR	N_SLOPE	N_FIRM_SIZE	F_INT_EXT_FLAG
PD_Model	PD_Model	0.12	0.24	0.5	25	IB

Note: If a new model is added then additional changes will have to be done to MST_RATING_CODE, MST_PD_DETAILS & MST_TEST_THRESHOLDS.

6. Adding/Modifying rating grades

To add/ modify rating grades of a given model changes will have to be done to MST_RATING_CODE

V_RATING_CD	V_RATING_SRC_CD	N_RATING_ORDER	V_RATING_TYPE
AAA	PD_Model	1	BORROWER-STD
AA	PD_Model	2	BORROWER-STD
A	PD_Model	3	BORROWER-STD
BBB	PD_Model	4	BORROWER-STD
BB	PD_Model	5	BORROWER-STD
B	PD_Model	6	BORROWER-STD
C	PD_Model	7	BORROWER-STD
D	PD_Model	8	BORROWER-DEF

7. Note: If a new rating grades are added then, additional changes will have to be done to MST_PD_DETAILS.

8. Adding/Modifying a new section/parameter to the rating data

To add a new parameter changes will have to be done to STG_RATING_DATA file and CNFG_PARA_MAPPING file.

The new section/parameter can be added (as the last column) in the rating file and the .ktr file modified after mapping the new parameter to a column in STG_RATING_DATA.

The details of this mapping will also have to be provided in the cnfg_para_mapping file as new row.

For ex. If the initial rating file has only 1 parameter.

V_BANK_CUST_CD	V_RATING_SRC_CD	D_RATING_DATE	V_RATING_CD
XYZ	Model1	31-03-2012	AA

Now, if the user now wants to add a new parameter, viz. Rating Score then the rating file will have to be modified to have following columns.

V_BANK_CUST_CD	V_RATING_SRC_CD	D_RATING_DATE	V_RATING_CD	N_RATING_SCORE
XYZ	Model1	31-03-2012	AA	85

For the system to recognize this as a new parameter, the details of this parameter will have to be added to CNFG_PARA_MAPPING table.

NAME	VALUES (V_RATING_CD)	VALUES (N_RATING_SCORE)
V_PARAM_COLUMN	V_RATING_CD	N_RATING_SCORE
COLUMN_NAME	V_RATING_CD	N_MODULE_SCORE_1
QUERY_COL	Discrete	Continuous
V_PARAM_TYPE	Rating	Value
V_AUDIT_CD	SYS	SYS
N_SORT_ORDER	1	2

N_BANDS		10
COL_TYP	RATING	NUMBER
V_SECTION_NAME	Model	Model
V_THRESHOLD_TYPE	Model	Model
V_RATING_SRC_CD	Large Corporate Model	Large Corporate Model
N_REPORT_ORDER	1	3
F_MODIFIER_FLAG	N	

9. User Input Screen

Field Name	Details
Report Name	Unique name to save a set of user inputs
Cohort Data available	Select Yes if Rating grade/Pool wise count of Total and Defaults is available. Select No if the system has to create cohort and count of Total and Defaults from rating data.
Model	Select Internal Rating Model to be validated
Method	Select Cohort Approach to form static pools at discrete points in time, ignoring transitions during cohort interval. Select Hazard Rate Approach for considering transitions during the interval.
Start Date	Cohort Approach: Start date of the first cohort. Hazard Rate Approach: Start date of the transition window
End Date	Cohort Approach: End date of the last cohort. Hazard Rate Approach: End date of the transition window.
Horizon	PD estimation Horizon for which transitions are to be generated.
Frequency	Cohort Formation Frequency. Cohort Approach: Frequency at which cohorts are to be formed, beginning from the start date. Hazard Rate Approach: Frequency at which sub-cohorts are to be generated.
Self-cure Treatment	Select "Self-cures as default" for "Ever Default" definition i.e. defaults are considered irrespective of cure. Select "Cured within 12M as standard", if defaults that got cured within 12M of default date are to be considered as Non-Defaults Select "Cured before cohort end date as standard", if defaults that got cured prior to chort end date are to be considered as Non-Defaults
Iterations	Number of iterations for validating rating-grade wise PD calibration using Monte Carlo Simulation
Description	Detailed description for the report

10. PD Validation

a. CAP/Gini/Somer's-D

Non-technical Introduction – The Cumulative Accuracy Profile (CAP) curve assesses the quality of the rating system to rank order defaulters to worst rating grade. Gini coefficient quantifies the ability of the rating system to discriminate between defaulters and non-defaulters.

Output Interpretation - Higher the CAP curve from random curve, more discriminating the rating tool is, as it indicates that the default cases were assigned poor rating grades prior to default and non-default cases were assigned better rating grades prior to default. Similarly, higher Gini Coefficient is preferable[2].

Technical Details – Cumulative Accuracy Profile (CAP) curve is constructed by plotting cumulative frequencies of defaulters on Y axis and total number of borrowers along the X axis beginning from worst rating grade. Therefore, a perfect rating model will assign the worst grade to defaulters. In this case, CAP will be increasing linearly and staying at one. Similarly, a random model without any discriminatory power will be a diagonal. Real rating systems will be somewhere in between these two extremes.

Accuracy Ratio (AR) is given as follows[1]:

$AR = a_r / a_p$

Where, a_r the area between the CAP of the rating model being validated and the CAP of the random model

And a_p is the area between the CAP of the perfect rating model and the CAP of the random model

Calculation Methodology – Construction of CAP curve and calculation of Gini coefficient is done as follows:

- Cumulative frequencies of total borrowers (%) and cumulative frequencies of defaulters (%) are calculated from the data, starting from worst rating grade to best rating grade.
- After the frequencies have been calculated, we will have N number of two-dimensional coordinates for N number of rating grades.
- We plot these coordinates and generate a curve (CAP curve).
- Once the CAP curve is constructed, we calculate Gini Coefficient using the Trapezoidal Rule.

Regulatory References – [1] The quality of a rating system is measured by the Accuracy Ratio *AR*. It is defined as the ratio of the area a_r between the CAP of the rating model being validated and the CAP of the random model, and the area a_p between the CAP of the perfect rating model and the CAP of the random model, i.e.

$AR = a_r / a_p$

Thus, the rating method is the better the closer *AR* is to one.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems

[2] Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische Nationalbank (OeNB)

b. ROC/Mann-Whitney U

Non-technical Introduction – The Receiver Operating Characteristics (ROC) curve is a visual tool to assess the ability of the rating system to rank order defaulters to worst rating grades.

Output Interpretation - Greater the Area below the ROC curve, better is the ability of the rating system to discriminate between defaulters and non-defaulters.

Technical Details – ROC curve is constructed by plotting the cumulative frequencies of bad cases as points on the y axis and the cumulative frequencies of good cases along the x axis. Each section of the ROC curve corresponds to a rating class, beginning at the left with the worst class[1]. ROC curve plots Hit rate (percentage of bad correctly identified as bad) and False Alarm Rate (percentage of good identified as bad) at different rating scores/grades. Area Under ROC (AUROC) quantifies the discriminatory power of the rating system.

Calculation Methodology - Construction of ROC curve is done as follows:

- Cumulative frequencies of non-defaulters (%) and cumulative frequencies of defaulters (%) are calculated from the data, starting from worst rating grade to best rating grade.
- After the frequencies have been calculated, we will have N number of two-dimensional coordinates for N number of rating grades.
- We plot these coordinates and generate a curve (ROC curve).

Regulatory References – [1] One common way of depicting the discriminatory power of rating procedures is the ROC Curve,80 which is constructed by plotting the cumulative frequencies of bad cases as points on the y axis and the cumulative frequencies of good cases along the x axis. Each section of the ROC curve corresponds to a rating class, beginning at the left with the worst class.

References - Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische Nationalbank (OeNB)

c. Mean Difference

Non-technical Introduction – Mean Difference is the standardized difference in the mean rating scores of defaulters and non-defaulters in the portfolio.

Output Interpretation - The higher the difference, greater is the discriminating power of the rating tool.

Technical Details – The formula for calculating mean difference is as follows:

Mean Difference = ABS (Mean Score of Defaulters – Mean Score of Non-defaulters) / Pooled Std.Dev.

Calculation Methodology – Mean difference can be calculated as follows:

If mean score of defaulters and non-defaulters is 20 and 50 respectively, and if the standard deviation of their pooled scores is 20, then

Mean Difference = ABS (20-50)/20 = 1.5

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems

d. FAR at 50% Hit Rate

Non-technical Introduction – A good rating tool that is a good discriminator between defaulters and non-defaulters will correctly predict at least 50% of defaulters while classifying minimum number of non-defaulters wrongly as defaulters. The definitions of False Alarm Rate (FAR) and Hit Rate are as follows:

		Actual	
Rating Model Prediction		**Default**	**Non-default**
	Default	Hit rate (Correct prediction)	False Alarm Rate (Type-I error): Loss of Business
	Non-Default	Type II error (1-Hit rate)	Correct prediction

Output Interpretation – The lower the False Alarm Rate at 50% Hit Rate, better is the discriminatory power of the rating tool.

Technical Details – The formula for calculating hit rate and false alarm rate is as follows[1]:

$HR(C) = H(C) / N_D$

Where *H(C)* is the number of defaulters predicted correctly with the cut-off value *C*, and *ND* is the total number of defaulters in the sample. This means that the hit rate is the fraction of defaulters that was classified correctly for a given cut-off value *C*.

$FAR(C) = F(C) / N_{ND}$

Where F(C) is the number of false alarms, i.e. the number of non-defaulters that were classified incorrectly as defaulters by using the cut-off value C. The total number of non-defaulters in the sample is denoted by N_{ND}.

Calculation Methodology – If at a cut-off score of rating score of 45, 35 out 50 defaulters were classified correctly as defaulters while 10 out 50 non-defaulters were wrongly classified as defaulters then
Hit Rate (45) = 35 / 50 = 70%
False Alarm Rate (45) = 10 / 50 = 20%

Regulatory References – [1] we define the hit rate *HR(C)* as
$HR(C) = H(C)/N_D$
Where *H(C)* is the number of defaulters predicted correctly with the cut-off value *C*, and N_D is the total number of defaulters in the sample. This means that the hit rate is the fraction of defaulters that was classified correctly for a given cut-off value *C*. The false alarm rate *FAR(C)* is defined as
$FAR(C) = F(C)/N_{ND}$
Where *F(C)* is the number of false alarms, i.e. the number of non-defaulters that were classified incorrectly as defaulters by using the cut-off value *C*. The total number of non-defaulters in the sample is denoted by N_{ND}.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische Nationalbank (OeNB)

e. KS & Pietra Index

Non-technical Introduction – Kolmogorov-Smirnov (KS) statistic is used to examine if two distributions are similar or not. Pietra Index is defined as twice the area of the largest triangle which can be drawn between the diagonal and the ROC Curve[1].

Output Interpretation – The greater the values of KS statistic and Pietra index, better is the discriminatory power of the rating system.

Technical Details – Pietra Index can be defined as the maximum area a triangle can obtain that is inscribed between the ROC curve and the diagonal of the unit square. Equivalently, the Pietra Index can be seen as half the maximum distance of ROC curve and diagonal. Interpretation of the Pietra Index as a distance leads to the representation

$$\text{Pietra Index} = \sqrt{2}/4 \max_C |HR(C) - FAR(C)|.$$

The sup-term on the right-hand side of this equation is just the well-known Kolmogorov-Smirnov (KS) test statistic[2].

Calculation Methodology – First, we calculate cumulative probabilities of defaulters and non-defaulters at each rating grade. Then, we take the absolute difference between cumulative probabilities of defaulters and non-defaulters at each rating grade. The maximum of these values is simply KS value.

Regulatory References – [1] the Pietra Index as a Measure of Discriminatory Power. Another one-dimensional measure of discriminatory power which can be derived from the ROC curve is the Pietra Index. In geometric terms, the Pietra Index is defined as twice the area of the largest triangle which can be drawn between the diagonal and the ROC curve. In the case of a concave ROC curve, the area of this triangle can also be

calculated as the product of the diagonal's length and the largest distance between the ROC curve and the diagonal.
[2] Geometrically, the Pietra Index can be defined as the maximum area a triangle can obtain that is inscribed between the ROC curve and the diagonal of the unit square.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische Nationalbank (OeNB)

f. Information Value (IV)

Non-technical Introduction - IV measures the difference between the distribution of defaulters and that of non-defaulters across obligor grades (or scores).

Output Interpretation – The higher the IV, the more is the separation of the distributions, and the better is the discriminatory power of a rating model.

Technical Details – Information Value is simply addition of Sub IV of each grade. Sub IV is calculated as
Sub IV = (% of Defaulters - % of Non-defaulters) * log (% of Defaulters / % of Non-defaulters)

Calculation Methodology – First, we calculate the percentage of defaulters and percentage of non-defaulters within each rating grade. Then we calculate Sub IV using the formula
Sub IV = (% of Defaulters - % of Non-defaulters) * log (% of Defaulters / % of Non-defaulters)
Finally, we add the Sub IVs of each rating grade to compute Information Value.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische Nationalbank (OeNB)

g. Entropy Measures

Non-technical Introduction – These measures assess the information gained by using the rating model. In this context, information is defined as a value which is measurable in absolute terms and which equals the level of knowledge about a future event[1]. These measures include Kullback-Leibler Distance and Conditional Information Entropy Ratio (CIER).

Output Interpretation – The greater the Kullback-Leibler Distance and CIER, better is the discriminatory power of the rating model.

Technical Details – Unconditional Entropy is calculated using the following formula
$H_0 = -(PD * \log_2(PD) + (1-PD) * \log_2(1-PD))$
For each rating class c, conditional entropy is H_c:
$h_c = -(PDc * \log_2(PD_c) + (1-PD_c) * \log_2(1-PD_c))$
Where PD_c is % of defaulters within rating class c.
Across all rating classes in a model, the conditional entropy H1 (averaged using the observed frequencies of the individual rating classes' p_c) is defined as:
$H_1 = -\sum p_c * h_c$

Calculation Methodology – First, we calculate H_0 (Unconditional Entropy) from the overall default probability. Then, we calculate h_c for each rating grade to arrive at H_1 (Conditional Entropy).
Kullback-Leibler Distance is calculated as

Kullback-Leibler Distance = $H_0 - H_1$
Conditional Information Entropy Ratio (CIER) is calculated as:
CIER = $(H_0 - H_1)/ H_0$

Regulatory References - [1] Entropy-Based Measures of Discriminatory Power
These measures of discriminatory power assess the information gained by using the rating model. In this context, information is defined as a value which is measurable in absolute terms and which equals the level of knowledge about a future event.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische Nationalbank (OeNB)

h. Chi-square test

Non-technical Introduction – Chi-square test serves as a measure of discriminatory power of the rating system.

Output Interpretation – The higher the value of Chi-square test statistic, the better the discriminatory power of the rating system. The closer the p-value of chi-square test to zero, the better the rating system is in discriminating between good and bad borrowers.

Technical Details – The statistic is calculated as [1]

$T_k = \sum$ (Default Rate (i) – Portfolio Default Rate)^2

With Default Rate (i) = Number of realized defaults in rating grade i as a percentage of total number of borrowers in the portfolio.

Calculation Methodology – First, we simply calculate the number of realized defaults for each rating grade as a percentage of total number of borrowers in the portfolio. Then, we calculate deviations of default rate in each rating grade from portfolio level default rate. Sum of all the squared deviations is Chi-squared value.. P-value of this statistic serves as a measure of discriminatory power of the rating system

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische National bank (OeNB)

i. Bayesian Error rate

Non-technical Introduction – Bayesian Error Rate specifies minimum probability of error if the rating system under consideration is used for a yes/no decision whether a borrower will default or not[1].

Output Interpretation – The lower the Bayesian Error Rate, better is the rating system in differentiating between defaulters and non-defaulters.

Technical Details – The logic behind Bayesian Error rate is quite simple. If we plot the probability density functions of defaulters and non-defaulters, then the Bayesian Error Rate is simply the area underneath where both curves overlap. Therefore, if the distributions of defaulters and non-defaulters are similar to each other, the Bayesian Error Rate will be close to 100%.

Calculation Methodology – The Bayesian error rate is calculated as follows:

Error Rate = min ($P_D(1 - HR(C)) + (1 - P_D) FAR(C)$)

Where Hit Rate (C) and False Alarm Rate (C) are defined as follows:

		Actual	
Model Rating Prediction		Default	Non-default
	Default	Hit rate (Correct prediction)	False Alarm Rate (Type-I error): Loss of Business
	Non-Default	Type II error (1-Hit rate)	Correct prediction

Regulatory References – [1] The **Bayesian error rate** (or classification error or minimum error) specifies the minimum probability of error if the rating system or score function under consideration is used for a yes/no decision whether a borrower will default or not.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische Nationalbank (OeNB)

j. Visual Tests

Non-technical Introduction – Visual Tests help in assessing the ability of the rating tool to differentiate between defaulters and non-defaulters visually. Some of the Visual Tests are Cumulative Accuracy Profile (CAP) curve, Receiver Operating Characteristics (ROC) curve, Probability Density Functions of Defaulters and Non-defaulters, Cumulative Density Functions of defaulters and non-defaulters etc.

Output Interpretation – One can interpret the output just by looking at the plots. E.g. In case of CAP and ROC, the greater the distance between CAP/ROC curves and diagonal, better is the rating system in discriminating between defaulters and non-defaulters.

11. PD calibration

a. Chi Square Test

Non-technical Introduction – Chi-square test serves as a measure of accuracy of the estimated default probabilities.

Output Interpretation – The lesser the value of Chi-square test statistic, the better is the accuracy of the rating system in predicting default probabilities.

Technical Details – The statistic is calculated as [1]

$$T_k = \sum \frac{(n_i p_i - \Theta_i)}{n_i p_i (1 - p_i)}$$

With n_i = number of debtors with rating *i* and Θ_i = number of defaulted debtors with rating i.

Calculation Methodology – First, we simply calculate expected defaults for each rating grade as overall default probability multiplied by the number of borrowers in a particular grade. Then, we calculate deviations of actual defaults from expected defaults. Sum of all the squared deviations is Chi-squared value. Then we compute the statistic using expected defaults and actual defaults. P-value of this statistic serves as a measure of accuracy of expected default probabilities.

Regulatory References – [1] $T_k = \frac{\sum \; (n_i p_i - \Theta_i)}{n_i p_i (1 - p_i)}$

The p-value of Chi-square test could serve as a measure of the accuracy of the estimated default probabilities: the closer the p-value is to zero, the worse the estimation is.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Guidelines on Credit Risk Management, Rating models and validation, Oesterreichische National bank (OeNB)

b. Spiegelhalter Test (SPGH)

Non –Technical Introduction – It is used in the calibration of rating system. It compares forecast and realized default probabilities on an individual level.

Output Interpretation - SPGH test statistic value greater than the critical value indicates poor overall calibration.

Technical Details – The starting point of SPGH is the Mean Square Error which is a measure of accuracy of the rating system in predicting default probabilities. SPGH test statistics is calculated by subtracting Expected Value of MSE from MSE and then dividing it by the variance of MSE. This test statistic follows a standard normal distribution and the familiar steps coming to a test decision have to be conducted.

Calculation Methodology – First, calculate MSE, Expected value of MSE and variance of MSE. Then compute the test statistic which is given as follows:

$$Z_S = \frac{MSE - E(MSE_{\pi_i = \hat{\pi}_i})}{Var(MSE_{\pi_i = \hat{\pi}_i})^{0.5}} = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \pi_i)^2 - \frac{1}{N}\sum_{i=1}^{N}\pi_i \cdot (1 - \pi_i)}{\sqrt{\frac{1}{N^2}\sum_{i=1}^{N}(1 - 2\pi_i)^2 \cdot \pi_i \cdot (1 - \pi_i)}}$$

Once the test statistic is calculated, we compare it with the critical value. If Test Statistic is greater than the critical value, then it indicates poor calibration of the rating system.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
The Basel II Risk Parameters – Engelmann, Rauhmeier

c. Brier Score / Mean Square Error (MSE)

Non –Technical Introduction – Brier Score is a measure for accuracy of the rating system in predicting default probabilities.

Output Interpretation – Brier Score quantifies deviations of the forecasts and observed default probabilities. Therefore, the higher the accuracy of forecasts of rating system, the smaller the Mean Square Error (MSE)[1].

Technical Details – It is an exploratory method. It calculates average quadratic deviation of the forecasted PD and the realized default rates. The resulting score between zero and one is called brier score. MSE can be interpreted as the weighted average of independent Bernoulli distributed random variables.

Computational Methodology – Mean Square Error is calculated as follows:

$MSE = 1/N \sum (y_i - \pi_i)$^2

Representing the squared difference of the default ($y_i = 1$) and non-default ($y_i = 0$) indicators, respectively, and the corresponding default probability forecast π_i averaged across all obligors.

Regulatory References – [1] obviously the MSE gets small, if the forecast PD assigned to defaults is high and the forecast PD assigned to non-defaults is low. Generally speaking, a small value of MSE indicates a good rating system. The higher the MSE the worse is the performance of the rating system (keeping other things equal).

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
The Basel II Risk Parameters – Engelmann, Rauhmeier

d. Geometric Mean Probability

Non –Technical Introduction – Geometric Mean Probability assesses the ability of the rating system in estimating the magnitude of the PDs correctly, not just their rank. Geometric Mean Probability tells us if the rating system is getting the probabilities of default right.

Output Interpretation – A Geometric Mean Probability of less than 0.5 is not acceptable. A Geometric Mean Probability of the rating model should be greater than the Geometric Mean Probability of the Naïve Model which will assign a PD equal to the overall portfolio level default rate to all the borrowers. Generally, this difference should be at least 50 to 100 bps. The larger the difference, better the performance of the rating model[1].

Technical Details – Geometric mean probability tells us if the rating system is getting the probabilities of default correct. It is given by the equation:

$\text{Log(likelihood)} = \sum \log [y_i\, p(x_i) + (1-y_i)(1-p(x_i))]$

The log likelihood described above is a negative number which ranges from negative infinity to zero. This becomes difficult to interpret. Therefore, the geometric mean of the observed outcomes is

GMP = e^1/N(log_likelihood)

Computational Methodology – First, the log-likelihood measure is calculated for the overall portfolio of borrowers. Since, the log-likelihood measure is difficult to interpret. It is then converted into Geometric Mean Probability which is given as:

GMP = e^1/N(log_likelihood)

Regulatory References – [1]A typical performance comparison would be against a “naïve” model. A typical naïve model can be the following: A model that correctly predicts the average default rate for the portfolio as a whole but is unable to distinguish among the obligors.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Basel II Implementation – Ozdemir, Miu

12. PD Benchmarking

a. Hit-Miss-Match (Similarity Matrix)

Non-technical Introduction - It provides an intuitive way to analyze how many consensus ratings exactly match the ratings provided by a rating agency (E.g. Moody's, Fitch, S&P). It helps to indicate whether there is a bias in the rating provided by the rating agency or not.

Output Interpretation - After comparing the consensus ratings with the rating agencies' ratings, we classify the proportion of deviations in the ratings (between consensus and rating agencies') into different buckets, e.g. -3, -2, -1, 0, 1, 2, 3.
Higher the percentage in the '0' bucket, more appropriate the ratings are with respect to the consensus ratings.
Proportion of ratings per rating class deviation between the consensus ratings and the origin ratings provided by the big three rating agencies Fitch, Moody's and Standard&Poor's is described below (data only for example purpose):

Agency	-4	-3	-2	-1	0	1	2	3
Fitch	0.000	0.000	0.008	0.154	0.724	0.109	0.004	0.000
Moody's	0.000	0.000	0.017	0.027	0.276	0.544	0.115	0.020
S&P	0.003	0.030	0.426	0.533	0.008	0.000	0.000	0.003

We may infer from the matrix above that Fitch's ratings are mostly in line with the consensus ratings and S&P's ratings are most deviated from the consensus ratings and also seem to be pretty pessimistic.

Technical Details –

Computational Methodology - Map the different rating agencies' ratings each to the consensus ratings and form a matrix. Now form buckets depicting deviations from the consensus rating. E.g. -3, -2, -1, 0, 1, 2, 3. Classify the proportional deviations for each agency into the buckets and infer from the resultant matrix about the accuracy of the rating agencies' ratings in accordance with the consensus ratings.
Mapping for one of the rating agency's ratings to consensus ratings is depicted below (data only for example purpose):

Consensus rating	Fitch rating										
	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+
AAA	0	0	0	0	0	0	0	0	0	0	0
AA+	0	33	0	0	0	0	0	0	0	0	0
AA	6	52	124	17	0	0	0	0	0	0	0
AA-	0	0	21	157	44	14	0	0	0	0	0
A+	0	0	3	19	148	50	0	0	0	0	0
A	0	0	0	0	34	166	33	0	0	0	0
A-	0	0	0	0	0	13	308	82	3	0	0
BBB+	0	0	0	0	0	0	69	350	93	0	0
BBB	0	0	0	0	0	0	0	22	218	4	0
BBB-	0	0	0	0	0	0	0	0	1	26	2
BB+	0	0	0	0	0	0	0	0	0	0	0

References - Deriving Consensus Ratings of the Big Three Rating Agencies, Working paper – Grun, Hofmarcher, Hornik, Leitner and Pichler.

b. Kendall's Tau

Non-technical Introduction - It depicts the agreement between ratings from two different sources (for instance, Internal and External ratings) as to how similar the internally assigned ratings are in comparison with the ratings assigned by the external rating agency

Output Interpretation - Values of Tau range from −1 to 1

- If tau = 1, it indicates perfect agreement between the ratings from two sources
- If tau = -1, it indicates absolute disagreement between the ratings from two sources
- If tau = 0, there is a no correlation between the ratings from two sources.

Technical Details - Any pair of observations (Xi, Yi) and (Xj, Yj) are said to be concordant if ranks of both the elements agree i.e. Xi>Xj and Yi>Yj or if both Xi<Xj and Yi<Yj; and are said to be discordant otherwise.
If Xi=Xj or Yi=Yj, the pair is neither concordant nor discordant and the pair is said to be tied. When tied pairs arise, tau-b coefficient is calculated.
The Kendall tau-b coefficient is termed as:

$$\tau b = (C-D)/\sqrt{[(n(n-1)/2-t)(n(n-1)/2-u)]}$$

Where, n is the sample observations; C is the number of concordant pairs; D is the number of discordant pairs; t is the number of tied X values and u is the number of tied Y values

Computational Methodology: - Take a sample of ratings from internal and external sources and make pairs of both internal as well as external ratings each by arranging them in a rank-wise order. Then find out the number of concordant and discordant pairs in the sample and the number of tied pairs. Post this, calculate the tau-b rank correlation coefficient and infer about the association of the internal ratings with external ratings from the value of the tau-b coefficient obtained.

References - Pearson versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds - Sorana-Daniela and Lorentz, 2006

c. Spearman Rank Correlation Coefficient

Non-technical Introduction - It signifies the agreement between ratings from two different sources (for instance, Internal and External ratings) as to how similar the internally assigned ratings are in comparison with the ratings assigned by the external rating agency

Output Interpretation - Spearman coefficient varies from -1 to 1.

- If Spearman coefficient is 1, it indicates perfect agreement between the ratings from two sources
- If Spearman coefficient is -1, it indicates absolute disagreement between the ratings from two sources
- If Spearman coefficient is zero, it indicates there is no correlation between the ratings from two sources

Technical Details - It is a non-parametric measure of correlation between variables which assess how well an arbitrary monotonic function could describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables. It is denoted by rho (ϱ).

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

where, x_i and y_i are the ranked variables; $\bar{x}$ and $\bar{y}$ is the average of the ranked variables respectively.
Tied values are assigned a rank equal to the average of their positions in the ascending order of the values.

Computational Methodology - Sort the sample ratings from two raters into two columns Xi(internal) and Yi(external) and then arrange both the columns in a rank-wise order, from AAA to D. Assign a rank to each rating grade from both the raters. Calculate the mean of both the ranked columns. Calculate the Spearman coefficient using the formula above and comment on the association between ratings assigned by the two raters. For example, the sample data can be arranged in a rank-wise order as follows:

Internal rating , Xi	**External rating , Yi**	**Rank , x_i**	**Rank, y_i**
AAA	AAA	1	1
AA	AA	(2+3)/2 =2.5	2
AA	BBB	(2+3)/2 =2.5	(3+4)/2 = 3.5
BBB	BBB	4	(3+4)/2 = 3.5
C	C	5	5
		$\bar{x} = 3$	$\bar{y} = 3$

References - Pearson versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds - Sorana-Daniela and Lorentz, 2006

d. Kruskal Gamma

Non-technical Introduction - It signifies the agreement between ratings from two different sources (for instance, Internal and External ratings) as to how similar the internally assigned ratings are in comparison with the ratings assigned by the external rating agency

Output Interpretation - Values of Gamma range from −1 to 1

- If gamma = 1, it indicates it indicates perfect agreement between the ratings from two sources
- If gamma = -1, it indicates it indicates absolute disagreement between the ratings from two sources
- If gamma = 0, there is no correlation between the ratings from two sources.

Technical Details - Any pair of observations (Xi, Yi) and (Xj, Yj) are said to be concordant if ranks of both the elements agree i.e. Xi>Xj and Yi>Yj or if both $X_i<X_j$ and Yi<Yj; and are said to be discordant otherwise.
Gamma is the surplus of concordant pairs over discordant pairs, as a percentage of all pairs, ignoring ties. Under statistical independence, Gamma will be zero, but it can be zero also when concordant and discordant pairs are equal.

$$G = (C-D)/(C+D)$$

Where, C is the number of concordant pairs and D is the number of discordant pairs of variables.

Computational Methodology - **From** a sample of ratings from two raters, we make rating pairs according to the rank and calculate the number of concordant and discordant pairs of ratings therein. Thereafter, we calculate the Kruskal's Gamma with and infer about the correlation between the ratings assigned by two different sources.

References - Non- parametric Measures of Bivariate Relationships – UNESCO

e. Cohen's Kappa

Non-technical Introduction - It signifies the agreement between ratings from two different sources (for instance, Internal and External ratings) as to how similar the internally assigned ratings are in comparison with the ratings assigned by the external rating agency. However, it doesn't account for the divergence from the observed ratings.

Output Interpretation - **Higher** the value of Cohen's Kappa, greater is the similarity between the ratings from two different sources.

- If kappa =1, it indicates there is perfect agreement between the ratings from two sources
- If kappa= 0, it indicates there is no agreement between the ratings from two sources

Technical Details - Cohen's kappa measures the agreement between two raters who each classify a number of exposure items into different rating grades. It is given by:

$$K = \frac{Pr\ (a) - Pr\ (e)}{1 - Pr\ (e)}$$

Where, Pr(a) is the relative observed agreement among raters and Pr(e) is the hypothetical probability of chance agreement among the raters.

Computational Methodology - From a sample of ratings from two raters, we form a matrix which classifies each rating into a rating grade for each rater. Thereafter, we calculate the probabilities of observed agreement and chance agreement from the rows and columns in the matrix and calculate Cohen's Kappa to infer about the degree of association between the ratings of the two raters.

For example, consider a portfolio of 1000 loans. Externally 200 loans have been rated AAA and 800 loans have been rated BBB. Internally 100 loans have been rated AAA (which are part of 200 AAA rated loans by external rater) and 900 have been rated BBB.

Pr (a) = (100+800)/1000 = 0.9

Pr (e) = (100/1000*200/1000) + (800/1000*900/1000) = 0.74

K = (0.9 – 0.74)/ (1-0.74)

= 0.615

References - Assessing agreement on classification tasks: The kappa statistic – Carletta, Jean

f. Weighted Cohen's Kappa

Non-technical Introduction - It signifies the agreement between ratings from two different sources (for instance, Internal and External ratings) as to how similar the internally assigned ratings are in comparison with the ratings assigned by the external rating agency. It takes into account the effect of divergence from observed ratings as well. More divergence from the observed ratings is penalized more than the cases where less divergence is experienced.

Output Interpretation - Higher the value of Cohen's weighted Kappa, greater is the similarity between the ratings from two different sources.

- If kappa =1, it indicates there is perfect agreement between the ratings from two sources
- If kappa= 0, it indicates there is no agreement between the ratings from two sources

Technical Details - The rating matrix would also include some ratings which would differ for both the raters. So, we give the weights to such disagreements in the ratings of the two raters. Diagonal elements are the perfect agreement cells and hence are given weights 0. Cells one off the diagonal are weighted 1, those two off are weighted 2 and so on. We generate three matrices viz. weight, observed and expected. Then the Cohen's weighted Kappa is calculated as follows:

$$K = 1 - \frac{\sum \sum Wij\ Xij}{\sum \sum Wij\ Mij}$$

Where, I, j varies from 1 to k and K is the number of codes

Wij are the elements in the weight matrix

Xij are the elements in the observed matrix

Mij are the elements in the expected matrix

References - Assessing agreement on classification tasks: The kappa statistic – Carletta, Jean

13. LGD estimation

a. Workout LGD

Non-Technical Introduction – It states the quantified credit loss amount as a percentage of the exposure that is realized on a loan portfolio in case a default event occurs. It takes into consideration net recovered cash flows post default until the end of recovery process.

Output Interpretation - Lower the work-out LGD, better it is for the bank as it suggests lower amount of losses will be incurred in the event of a default.

Technical Details - **LGD** cannot be negative, hence work-out LGD is calculated as:

$$\text{Work-out LGD} = \text{Max}\left\{ \frac{1 - \sum_I R_i(r) - \sum_j P_j(r)}{EAD}, 0 \right\}$$ [1]

Where, R_i is each of the i discounted recoveries of the defaulted facility; P_j is each of the j discounted payments or costs during the recovery period; r is the discount rate.

Computational Methodology - Since LGD is given by (1-Recovery rate), we first calculate the recoveries as a percentage of Exposure at Default, where recoveries are calculated by taking the difference of discounted recoveries and discounted costs(direct and indirect). The discounting rate that is used should be risk – adjusted.

Regulatory references - [1] there are three main components for computing a workout loss: the recoveries (cash or noncash), the costs (direct and indirect) and the discount factor that will be fundamental to express all cash-flows in terms of monetary units at the date of default. When loss is calculated by setting all negative observations of loss to zero, as shown in equation (2), it is referred to as censoring the data.

$$\text{Work-out LGD} = \text{Max}\left\{ 1 - \sum_I R_i(r) - \sum_j P_j(r), 0 \right\}$$

EAD

Where Ri is each of the i discounted recoveries of the defaulted facility, Pj is each of the j discounted payments or costs during the recovery period and r represent a discount rate.

References - BCBS Working Paper No.14, Studies on the Validation of Internal Rating Systems

b. Market LGD

Non-Technical Introduction - It states the quantified credit loss amount as a percentage of the exposure that is realized on a loan portfolio in the event of a default. It takes into consideration the market price of the defaulted instrument (marketable bonds, loans) in the secondary market immediately after default (typically 30 days)[1]. It is mainly applicable to Large Corporates, Sovereigns and Banks.

Output Interpretation - Lower the Market LGD, better it is for the bank as it suggests lower amount of losses to be incurred in the event of a default.

Technical Details - Market LGD is calculated as one minus the recovery rate, where recovery rate is the percentage of the market price of the financial instrument in the secondary market immediately after default to the exposure at default. It is calculated as follows:

$$\text{Market LGD} = 1 - \frac{\text{Market Price of the Bonds/Loans after default}}{\text{EAD}}$$

Computational Methodology - Where there is limited recovery data, we observe the market price of the defaulted facility immediately post default and take it as a percentage of the total exposure on that facility and subtract it from 1.

Regulatory references - [1] Market LGD depends on the market price of a defaulted facility soon after the date of default (typically around 30 days). Most rating agency studies on recoveries use this approach.33
This method is useful since prices reflect the investor's assessment of the discounted value of recoveries. However, if markets are illiquid or they are driven by shocks unrelated to expected recoveries, this measure may not be appropriate. These concerns are particularly relevant for relatively new loan markets.

References - 1.BCBS Working Paper No.14, Studies on the Validation of Internal Rating Systems
2. Basel II implementation, Ozdemir and Miu

c. Implied Historical LGD

Non-technical Introduction - Loss Given Default is the amount of loss that the banks expect to lose on its portfolio in case a default event occurs. It takes into consideration historical total losses and PD estimates. It is applicable only to Retail portfolios, as specifically mentioned in the Basel IRB guidelines.

Output Interpretation - Lower the Implied Historical LGD, better it is for the bank, as it suggests lower amount of losses to be incurred in the event of a default.

Technical Details - LGD is calculated based on the equation: EL = PD*LGD

Computational Methodology - A historical loss data of the retail portfolios is accumulated and PD estimates are calculated. Thereafter, LGD is calculated by dividing the total expected losses by the PD estimate. [1]

Regulatory references - [1] The revised Framework considers an implicit method for obtaining LGDs for retail portfolios.27 This estimate of LGD uses the PD estimates and the experience of total losses in the portfolio to derive the implied LGD. In the following analysis, this method is called the implied historical LGD.

References - 1. BCBS Working Paper No.14, Studies on the Validation of Internal Rating Systems
2.Otto-Von-Guericke-University-Magdeburg, FEMM Working Paper No.25, A framework for LGD Validation of Retail portfolios.

d. Implied Market LGD

Non-Technical Introduction - Loss Given Default is the amount of loss that the banks expect to lose on its portfolio in case a default event occurs. It takes into consideration the credit spreads, which is nothing but the difference between the loan-specific interest rate and the risk free rate. It is mainly used for Large Corporate, Sovereign and Banks.[1]

Output Interpretation - Lower the Implied Market LGD, better it is for the bank, as it suggests lower amount of losses to be incurred in the event of a default.

Technical Details - Implied Market LGD is calculated for risky (but non-defaulted) loans like risky bonds, securitized loans/credits, etc. It is assumed that the spread between the loan-specific interest rate and the risk free interest rate equals the expected loss in percent. If the spread is known:

$$LGD = \frac{Spread}{PD} \quad ^{2}$$

It should be noted that the spread may contain two components viz. liquidity risk premium and Risk premium, hence it is necessary to determine each component separately.

Computational Methodology - The spread reflects the Expected Loss, thus both PD and LGD, as well as liquidity premium. So the risk premium is separated from the liquidity premium using a model [3]. Now the LGD is calculated by dividing the spread by the default probability.

Regulatory references - [1] the implied market LGD methods derive LGD from risky bond prices using an asset pricing model. However, rather than using defaulted facilities like the explicit market LGD, they use a reference data set that includes non-defaulted facilities and credit spreads instead of realized LGD, as basic inputs. The credit spreads on risky bonds reflect, among other things, the expected loss on the bonds. Recent models illustrate how to decompose this measure of expected loss into the PD and the LGD.

References - BCBS Working Paper No.14, Studies on the Validation of Internal Rating Systems
[2] Bond prices, Default Probabilities and Risk Premiums – John Hull
[3] What do we know about Loss Given Default – Til Schuermann

e. DPD bucket roll rates

Debts which are not settled on or before due date are considered to be past due. Days Past Due (DPD) simply means number of days from the due date for which debt is outstanding. If an account is 90 DPD, it is considered to be in default. Thus, past due loans can be classified into various buckets. One common way of doing this is to form 1 to 30 DPD, 31 to 60 DPD, 61 to 90 DPD and 90+ DPD Buckets.
Once the buckets are formed, we can analyze the characteristics of each bucket by computing roll rates. These are classified as follows:
Stable Rate – It shows the number of accounts within one bucket staying in the same bucket after a certain time period. E.g. stable rate of 50% for 1 to 30 DPD means that 50% of 1 to 30 DPD remain delinquent.
Rollback Rate – It shows the number of accounts within a certain bucket that move to better buckets.
Forward Roll Rate – It shows the number of accounts that move to worse buckets.

f. Max Ever vs. Current DPD

Max Ever DPD indicates the maximum number of days after due date for which an account has remained delinquent.
Current DPD means that the account is not yet due.

g. Vintage Analysis of Cumulative Default Rates

A loan portfolio is segregated into different vintages on the basis of sanctioning date of each loan. Once the portfolio is segregated into vintages, we try to look at cumulative default rates for each vintage over time.

14. LGD Validation

a. Cumulative LGD Accuracy Ratio (CLAR)

Non-Technical Introduction – The Cumulative LGD Accuracy Ratio assesses the ability of a rating system to discriminate between different facility rating grades. It tests for the ordinal ranking of LGD ratings.

Output Interpretation – The CLAR coefficient can take a maximum value of 1 which indicates perfect discrimination. A higher value of CLAR coefficient indicates better discriminatory power of the facility rating system.

Technical Details – The CLAR curve compares predicted LGD bands with realized LGD bands for each rating grade. It tests the rank ordering of the rating system.

Computational Methodology – First step is to define LGD bands. These are termed as Predicted LGD bands. The realized LGDs are then sorted in descending order and grouped such that the number of realized LGDs in each of the bands is equal to the number of facilities being assigned the LGD rating in each of the predicted LGD bands. For example, assume the worst predicted LGD band has 10 facilities; the 10 worst (i.e., highest) realized LGDs are thus assigned to the equivalent realized LGD band[1]. Then we see how many observations in realized bands originated from corresponding predicted LGD bands. We do this on a cumulative basis. At the end, we will have cumulative predicted LGD observations and cumulative realized LGD observations at each grade. We will plot these points to generate a CLAR curve.
CLAR coefficient is calculated as
CLAR = 2 * Area under the CLAR curve

Regulatory References – [1] the first step is to determine the number of facilities being assigned the LGD rating (predicted LGD) in each LGD band as defined in previous section. These LGD bands are termed *predicted LGD bands*. The realized LGDs are then sorted in descending order and grouped such that the number of realized LGDs in each of the bands is equal to the number of facilities being assigned the LGD rating in each of the predicted LGD bands. For example, assume the worst predicted LGD band has 10 facilities; the 10 worst (i.e., highest) realized LGDs are thus assigned to the equivalent *realized LGD band*. Next, we record how many of the observations in the realized LGD bands in fact originated from the corresponding predicted LGD band.

References - References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Basel II Implementation – Ozdemir, Miu

b. Accuracy ratio

Non-Technical Introduction – Accuracy Ratio assesses the ability of the facility rating tool to discriminate between different facilities.

Output Interpretation - The Accuracy Ratio can take a maximum value of 1 which indicates perfect discrimination. A higher value of Accuracy Ratio indicates better discriminatory power of the facility rating system.

Technical Details – The Accuracy Ratio tests how many observations that were assigned a certain LGD band actually belonged to the same LGD band on a cumulative basis. E.g. if there were 20 observations for whom predicted LGD band was 60% to 100% and 10 out of those 20 observations had an LGD between 60% to 100% then my cumulative correctly assigned LGD observations will be 10 and cumulative predicted LGD observations will be 20.

Computational Methodology – First step is to define LGD bands. These are termed as Predicted LGD bands. The realized LGDs are set equal to corresponding predicted LGD bands. Then we see how many observations in realized bands originated from corresponding predicted LGD bands. We do this on a cumulative basis. At the end, we will have cumulative predicted LGD observations and cumulative realized LGD observations at each grade. We will plot these points to generate a curve.
Accuracy Ratio is calculated as
Accuracy Ratio = 2*Area under the curve

References - References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Basel II Implementation – Ozdemir, Miu

c. Predicted vs. Realized LGD

Non-Technical Introduction – Predicted vs. Realized LGD plots predicted and realized LGD for each facility. It is a visual test since one can interpret the results by simply looking at the chart.

Output Interpretation – The smaller the intercept, the better is the rating system. Similarly, the closer the trend line to the diagonal better is the rating tool since predicted and realized LGD will be very similar to each other.

Technical Details – Predicted vs. Realized LGD is a scatter plot of predicted LGD and realized LGD for each facility.

Computational Methodology – We plot predicted LGD and realized LGD for each facility with Predicted LGDs on X axis and realized LGDs on Y axis. Once the scatter plot is formed, we fit a trend line to it along with the equation. The idea here is to visually confirm the accuracy of the rating system.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Basel II Implementation – Ozdemir, Miu

d. Calibration of LGD

Non-Technical Introduction – One way to calibrate a facility rating tool is to calculate Mean Square Error (MSE). MSE is a measure for accuracy of the rating system in predicting LGD Bands.

Output Interpretation – Mean Square Error quantifies deviations of the predicted LGDs and realized LGDs. Therefore, the higher the accuracy of rating system in predicting LGD bands for a facility, the smaller the Mean Square Error (MSE).

Technical Details – It is an exploratory method. It calculates average quadratic deviation of the predicted LGD and the realized LGD. The resulting score between zero and one is called Mean Square Error.

Computational Methodology – Mean Square Error is calculated as follows:
MSE = Sum (Realized LGD – Expected LGD) ^2 / Number of observations

References - Basel II Implementation – Ozdemir, Miu

15. EAD Estimation

a. Fixed Horizon

Non-Technical Introduction – It is used to estimate Credit Conversion Factor (CCF) which is used in converting undrawn amounts into credit equivalent amounts given that a borrower defaults.

Output Interpretation – A higher Credit Conversion Factor will result into a higher Exposure at Default and vice versa.

Technical Details – A time horizon is first selected which is usually 1 year. Therefore, we need Exposure 1 year prior to default, Exposure at Default and Limit 1 year prior to default in order to calculate CCF for a particular facility. There are certain drawbacks of this method. They are[1]

- The fixed time horizon, T, is conventional.
- It is not possible to include directly defaulted facilities when the age of the facility at the date of default is less than T.

The advantages of this method are – 1. Dispersion of reference dates 2. The use of common horizon contributes to the homogeneity of the realized CCFs.

Computational Methodology – Realized CCF is calculated using following formula:

CCF (td – T) = [E (td) – E (td –T)] / [L (td –T) – E (td –T)]

Where CCF (td – T) is the Realized CCF
E (td) – E (td –T) is the increase in exposure amount over the selected time horizon e.g. 1 year
And L (td –T) – E (td –T) is the undrawn amount

Regulatory References – [1] Drawbacks - The fixed time horizon, T, is conventional.
It is not possible to include directly defaulted facilities when the age of the facility at the date of default is less than T.
Advantages – Dispersion of reference dates.
The use of common horizon contributes to the homogeneity of the realized CCFs.

References – Basel II Risk Parameters – Engelmann, Rauhmeier

b. Variable Horizon

Non-Technical Introduction - It is used to estimate Credit Conversion Factor (CCF) which is used in converting undrawn amounts into credit equivalent amounts given that a borrower defaults.

Output Interpretation - A higher Credit Conversion Factor will result into a higher Exposure at Default and vice versa.

Technical Details – The rationale for this method is to take into account a broader set of possible default dates than in the other approaches when estimating a suitable Credit Conversion Factor for a non-defaulted facility conditional on the default during the following year[1].

Computational Methodology - First, a range for horizon values (e.g. ne year) for which we are going to compute CCF is fixed. Second, for each defaulted facility, realized CCF associated with reference dates (e.g. 1m, 2m…12m) is calculated. Formula for calculating realized CCF is as follows:

CCF (td – j) = {[E (td) – E (td – j)] / [L (td – j) – E (td – j)], j = 1,…., 12 months}

Regulatory References – [1] the rationale for this method is to take into account a broader set of possible default dates than in the other approaches when estimating a suitable Credit Conversion Factor for a non-defaulted facility conditional on the default during the following year.

References – Basel II Risk Parameters – Engelmann, Rauhmeier

c. Cohort Approach

Non-Technical Introduction - It is used to estimate Credit Conversion Factor (CCF) which is used in converting undrawn amounts into credit equivalent amounts given that a borrower defaults.

Output Interpretation - A higher Credit Conversion Factor will result into a higher Exposure at Default and vice versa.

Technical Details – Cohort approach takes into account the possibility that current exposures can default at any moment during the following year.

Computational Methodology – First, the observation period is divided into intervals of a fixed length (cohorts), e.g. One year intervals. Second, facilities are grouped into cohorts according to the interval that includes their default dates. Third, in order to compute the CCF of a facility, the starting point of the time interval that contains its default date is used as the reference date[1]. CCF is then computed as follows:

CCF (ti) = [E (td) – E (ti) / [L (ti) – E (ti)]

Regulatory References – [1] first, the observation period is divided into intervals of a fixed length (cohorts), e.g. One year intervals. Second, facilities are grouped into cohorts according to the interval that includes their default dates. Third, in order to compute the CCF of a facility, the starting point of the time interval that contains its default date is used as the reference date.

References – Basel II Risk Parameters – Engelmann, Rauhmeier

d. Momentum Approach

Non-Technical Introduction - It is used to estimate Credit Conversion Factor (CCF) which is used in converting undrawn amounts into credit equivalent amounts given that a borrower defaults.

Output Interpretation - A higher Credit Conversion Factor will result into a higher Exposure at Default and vice versa.

Technical Details - Momentum Approach estimates Loan Equivalent (LEQ) Factor. EAD from this approach will be realized LEQ factor multiplied by the total limit for a facility.

Computational Methodology – First, a time horizon is fixed. LEQ factor is calculated as:

LEQ (i) = EAD (i) / L (i)

Where LEQ (i) is the realized Loan Equivalent factor for a facility "i".
EAD is the Exposure at Default
And L (i) is the total limit for facility "i".

16. EAD Validation

a. Cumulative EAD Accuracy Ratio (CEAR)

Non-Technical Introduction – The Cumulative EAD Accuracy Ratio assesses the ability of a CCF model to discriminate between different facilities.

Output Interpretation – The CEAR coefficient can take a maximum value of 1 which indicates perfect discrimination. A higher value of CEAR coefficient indicates better discriminatory power of the CCF model.

Technical Details – The CEAR curve compares predicted CCF bands with realized CCF bands for each rating grade. It tests the rank ordering of the CCF Model

Computational Methodology – First step is to define CCF bands. These are termed as Predicted CCF bands. The realized CCFs are then sorted in descending order and grouped such that the number of realized CCFs in each of the bands is equal to the number of facilities being assigned the CCF in each of the predicted LGD bands. For example, assume the worst predicted CCF band has 10 facilities; the 10 worst (i.e., highest) realized CCFs are thus assigned to the equivalent realized CCF band[1]. Then we see how many observations in realized bands originated from corresponding predicted CCF bands. We do this on a cumulative basis. At the end, we will have cumulative predicted CCF observations and cumulative realized CCFs observations at each grade. We will plot these points to generate a CEAR curve.
CEAR coefficient is calculated as

CEAR = 2 * Area under the CEAR curve

Regulatory References – [1] the first step is to determine the number of facilities being assigned the CCF rating (predicted CCF) in each CCF band as defined in previous section. These CCF bands are termed predicted CCF bands. The realized CCFs are then sorted in descending order and grouped such that the number of realized CCFs in each of the bands is equal to the number of facilities being assigned the LGD rating in each of the predicted CCFs bands. For example, assume the worst predicted CCF band has 10 facilities; the 10 worst (i.e., highest) realized CCFs are thus assigned to the equivalent realized CCF band. Next, we record how many of the observations in the realized CCF bands in fact originated from the corresponding predicted CCF band.

References - References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Basel II Implementation – Ozdemir, Miu

b. Accuracy ratio

Non-Technical Introduction – Accuracy Ratio assesses the ability of the CCF model to assign an accurate Credit Conversion Factor for all the facilities.

Output Interpretation - The Accuracy Ratio can take a maximum value of 1 which indicates perfect discrimination. A higher value of Accuracy Ratio indicates better discriminatory power of the CCF model.

Technical Details – The Accuracy Ratio tests how many observations that were assigned a certain CCF band actually belonged to the same CCF band on a cumulative basis. E.g. if there were 20 observations for whom predicted CCF band was 60% to 100% and 10 out of those 20 observations had an CCF between 60% to 100% then my cumulative correctly assigned CCF observations will be 10 and cumulative predicted CCF observations will be 20.

Computational Methodology – First step is to define CCF bands. These are termed as Predicted CCF bands. The realized CCFs are set equal to corresponding predicted CCF bands. Then we see how many observations in realized bands originated from corresponding predicted CCF bands. We do this on a cumulative basis. At the end, we will have cumulative predicted CCF observations and cumulative realized CCF observations at node. We will plot these points to generate a curve.
Accuracy Ratio is calculated as
Accuracy Ratio = 2*Area under the curve

References - References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Basel II Implementation – Ozdemir, Miu

c. Predicted vs. Realized CCF

Non-Technical Introduction – Predicted vs. Realized CCF plots predicted and realized CCF for each facility. It is a visual test since one can interpret the results by simply looking at the chart.

Output Interpretation – The smaller the intercept, the better is the CCF model. Similarly, the closer the trend line to the diagonal better is the CCF model since predicted and realized CCF will be very similar to each other.

Technical Details – Predicted vs. Realized CCF is a scatter plot of predicted CCF and realized CCF for each facility.

Computational Methodology – We plot predicted CCF and realized CCF for each facility with Predicted CCFs on X axis and realized CCFs on Y axis. Once the scatter plot is formed, we fit a trend line to it along with the equation. The idea here is to visually confirm the accuracy of the CCF model.

References - BCBS Working Paper 14 - Studies on the Validation of Internal Rating Systems
Basel II Implementation – Ozdemir, Miu

17. Calibration of EAD

Non-Technical Introduction – One way to calibrate a CCF model is to calculate Mean Square Error (MSE). MSE is a measure for accuracy of the CCF mode in predicting CCF Bands.

Output Interpretation – Mean Square Error quantifies deviations of the predicted CCFs and realized CCFs. Therefore, the higher the accuracy of CCF model in predicting CCF bands for a facility, the smaller the Mean Square Error (MSE).

Technical Details – It is an exploratory method. It calculates average quadratic deviation of the predicted CCF and the realized CCF. The resulting score between zero and one is called Mean Square Error.

Computational Methodology – Mean Square Error is calculated as follows:
MSE = Sum (Realized CCF – Expected CCF) ^2 / Number of observations

References - Basel II Implementation – Ozdemir, Miu

18. Low Default Portfolios

a. Statistical PD estimation - Uncorrelated

Non-Technical Introduction – Statistical PD estimation allows us to estimate default probabilities for portfolios that have seen zero or very low defaults in the past.

Output Interpretation – A higher PD estimate indicates that there is a significant possibility that the borrower might default over a certain time period.

Technical Details – This method gives an upper bound PD estimate for each rating grade. It also assumes zero correlation i.e. it assumes that default events are independent. The estimates of default probabilities resulting from this approach are fairly conservative.

Computational Methodology – For cases where there are zero observed defaults, PD upper bounds are calculated as follows:

1 – Confidence level = (1-PD) ^N where N is the number of borrowers in the rating grade

Cases where there are one or more observed defaults, PD upper bounds can be found out by solving the following equation:

$$1 - confidence\ level = \sum_{i=0}^{r} {}^{N}C_r PD^r (1 - PD)^{N-r}$$

References - 'Estimating Probabilities of Default for Low Default Portfolios' by Pluto and Tasche

b. Statistical PD estimation - Correlated

Non-Technical Introduction – Statistical PD estimation allows us to estimate default probabilities for portfolios that have seen zero or very low defaults in the past. It assumes that default events are dependent.

Output Interpretation – A higher PD estimate indicates that there is a significant possibility that the borrower might default over a certain time period. Similarly, a higher correlation implies that there are greater chances that a borrower will default given some other borrower in the portfolio defaults.

Technical Details – This method gives an upper bound PD estimate for each rating grade. It also accounts correlation i.e. it assumes that the default events are dependent. The estimates of default probabilities resulting from this approach are fairly conservative.

Computational Methodology – PD estimates under this approach are calculated using the following equation:

$$1 - confidence\ level = \int_{-\alpha}^{\alpha} pdf\,(economic\ risk\ factor) \sum_{i=0}^{r} {}^{N}C_r\,(Basel\ IRB\,Worstcase\,PD)^r\,(1 - Basel\ IRB\,Worstcase\,PD)^{N-r}$$

Basel IRB Worst case PD is the downturn PD calculated using Basel II IRB equation, which assumes that all borrowers are dependent on single risk factor. PDF (economic risk factor) is the normal density function of the single risk factor. We solve the indefinite integral using Gauss Hermite integration

References - 'Estimating Probabilities of Default for Low Default Portfolios' by Pluto and Tasche

c. Likelihood approach

Non-Technical Introduction – Likelihood Approach allows us to estimate default probabilities for portfolio that have seen zero or very low defaults in the past.

Output Interpretation - A higher PD estimate indicates that there is a significant possibility that the borrower might default over a certain time period. Similarly, a higher correlation implies that there are greater chances that a borrower will default given some other borrower in the portfolio defaults.

ASYMMETRIX