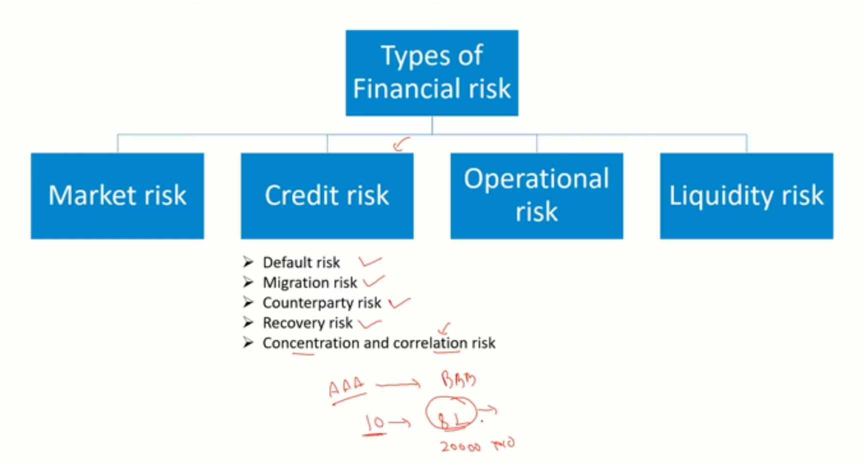
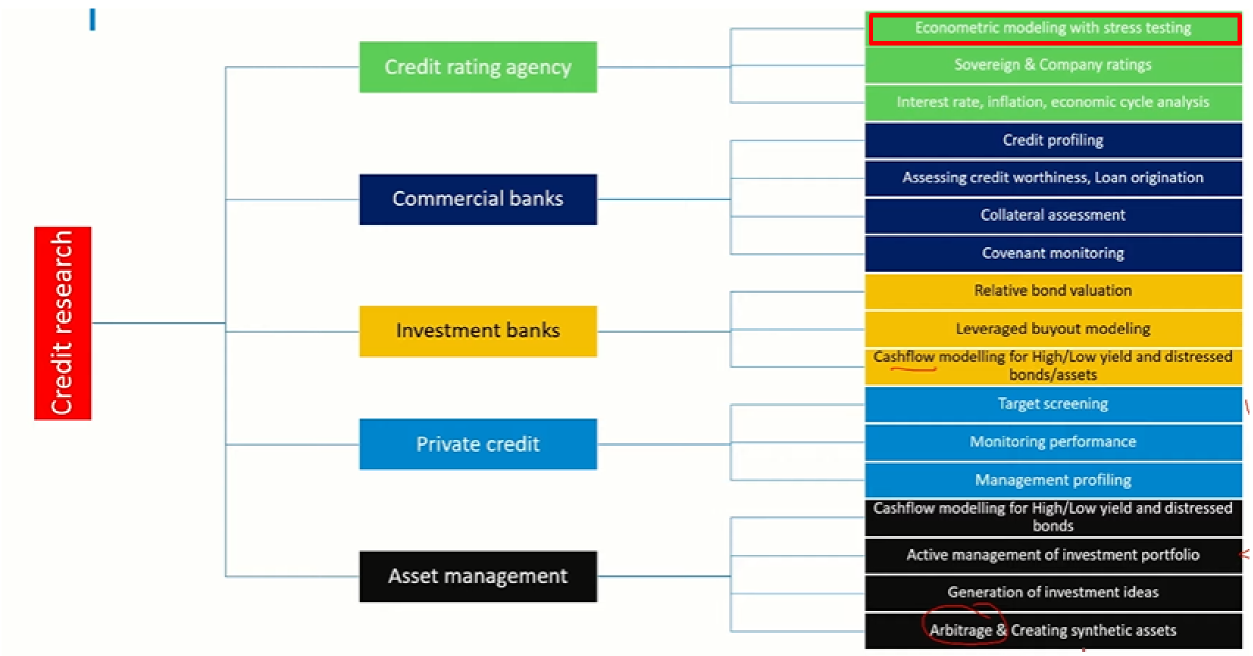




Credit Scorecards:

Credit or application scorecards can be excellent tools for both lender and borrower to work out the debt servicing capability of the borrower. For lenders, scorecards can help them assess the creditworthiness of the borrower and maintain a healthy portfolio – which will eventually influence the economy as a whole. Additionally to the borrower, they can provide valuable information such as 45% of people with her socio-economic background have struggled to keep up with the EMI commitment.

At a very high level, credit scorecards have their roots in the classification problem in statistics & data mining. The classification problems present an extremely broad methodology/thought process that has multiple business applications.

Including:
1. Classification problem and sampling
2. Variable selection and coarse classing
3. Predictive Models
4. Logistic regression and scorecards
5. Model validation
6. Application and business process integration

<u>For more details about score cards read the book which was downloaded at another PC</u>

Credit Scorecards – Classification Problem

In the case of credit scorecards, the problem statement is to distinguish analytically between the good and bad borrowers. Hence, the first task is to define a good and a bad borrower. For most loan products, good and bad credit is defined in the following way

1. Good loan: never or once missed on the EMI payment

2. Bad loan: ever missed 3 consecutive EMIs in a row (i.e. 90 days-past-due)

Additionally, for tagging someone good or bad, you need to observe his or her behaviour for a significant length of time. This length of time varies from product to product based on the tenor of the loan. For home loans, with a tenor of 20 years, 2-3 years is a reasonable observation period.

Sampling Strategy for Credit Scorecards

Black swan problem.
Example: You may need to visit every single planet to rule our possibility of an intelligent form of life.

For credit scorecard development, the accepted rule of thumb for sample size is at least 1000 records of both good and bad loans.

Variables Selection [redacted]

$$GDP = C + I + G + (Ex - Im)$$

$$C = \textit{collective spending by consumers}$$

$$I = \textit{cumulative investment by businesses}$$

$$G = \textit{total spending by government}$$

$$(Ex - Im) = \textit{net exports (exports − imports)}$$

Gross Domestic Product

As per the formulae, the above-mentioned parameters are the best fit to calculate GWP. But there could be better formulae which may include more parameters to calculate GWP as India is too diversified.
The idea is to select the right variables to build your model!

the variables selection process is performed through statistical significance – a reasonably automated process.

Data Collection:
Application forms are a major source of data collection regarding the borrower. However, nobody wants to fill out a lengthy form hence an optimal size of the form ensures accurate information provided by the borrower. The idea is to select the right variable and ensure accurate measurement.

There are several aspects regarding variables but I will mention just one of them here (coarse classing).

[redacted]

Coarse Classing as a technique during model development.

Quite a few academicians & practitioners for a good reason believe that coarse classing results in loss of information. However, in my opinion, coarse classing has the following advantage over using raw measurement for a variable.

1. It reduces random noise that exists in raw variables – similar to averaging and yes, you lose some information here.

[redacted]

2. It handles extreme events – on two extremes of a variable – much better where you have thin data.

3. It handles the non-linear relationship between dependent and independent variables without a lot of effort of variable transformation from the analyst.

PCA could be better, or AIC or variable importance plot

	Original Data				Coarse Classes					
Age Group	Total Number of loans	Number of Bad loans	Number of Good Loans	% Bad loans	Age Group	Total Number of loans	Number of Bad loans	Number of Good Loans	% Bad loans	Name of Coarse Groups
21-24	310	14	296	4.5%	21-30	4821	206	4615	4.3%	G_1
24-27	511	20	491	3.9%						
27-30	4000	172	3828	4.3%						
30-33	4568	169	4399	3.7%	30-36	10266	357	9909	3.5%	G_2
33-36	5698	188	5510	3.3%						
36-39	8209	197	8012	2.4%	36-48	32926	776	32150	2.4%	G_3
39-42	8117	211	7906	2.6%						
42-45	9000	216	8784	2.4%						
45-48	7600	152	7448	2.0%						
48-51	6000	84	5916	1.4%	48-60	12788	183	12605	1.4%	G_4
51-54	4000	64	3936	1.6%						
54-57	2000	26	1974	1.3%						
57-60	788	9	779	1.1%						

Table 1 – Coarse Class

MODELING:

A model is defined as a simplified representation of reality.

Data warehouse, Business Intelligence and Advanced Analytics

Analytics has received a massive boost because of the emergence of information technology. We are living in the era of big data. A plethora of data collected at every stage of the business process had created a need to extract knowledge out of the information. This overall process has three aspects to it

1. **Data warehouse or data marts:** transactional data is extracted-transformed and loaded (ETL) into a data model / schema for the purpose of analysis

2. **Business Intelligence or dashboards:** “as is” business reports

3. **Predictive Analytics or Advanced Analytics:** high-end statistical and data mining exercise

As the quantum of data is exponentially increasing, **Hadoop and big data** technologies are replacing the data warehouses.

Credit Scoring Models

Credit scorecards are models to predict the probability of a borrower default on his/her loan. The following is a simplified version of credit score with three variables

Credit Score = Age + Loan to Value Ratio (LTV) + Installment (EMI) to Income Ratio (IIR)

Points Table					
Age of the applicant (in years)		**Loan to Value Ratio (LTV)**		**Instalment to income ratio(IIR)**	
Below 32 years	10	0 to 50	100	Below 20	140
32 to 50 year	50	50 to 80	50	20 to 50	75
Above 50 years	20	80 to 100	10	Above 50	5

Score Points wise Risk					
Below 100	High Risk	100 to 180	Medium Risk	Above 180	Low Risk

A 28-year-old man with the LTV of 75 and the IIR of 60 will have the score of 10+50+5 =65 and hence is a high credit risk.

There are several statistical & data mining techniques that could help us achieve our object such as

1. Decision tree

2. Neural Networks

3. Support Vector Machines

4. Probit Regression

5. Linear discriminant analysis

6. Logistic Regression

Logistic regression is the most commonly used technique for the purpose. We will explore more about logistic regression in the next article.

Applications of logistic regression:
• Good and Bad borrowers
• Fraud and genuine cases
• Buyers and non-buyers

Out-of-sample test: where we have divided our sample into the training and the test sample. The first level of testing happens on the holdout or test sample. The test sample needs to perform as well as the training sample.

Out-of-time sample test: since the model was built on a sample of the portfolio with reasonable vintage, the analyst would like to test the performance of a more recent portfolio.

On-field test: this is where the test of the pudding is; **the analyst needs to be completely aware of any credit policy changes that the bank has gone through** since the scorecard is developed and more importantly, the impact the changes will have on the scorecard.

Performance Tests for Model Validation

There are several ways to test the performance of the scorecard such as c**onfusion matrix, KS statistics, Gini and area under ROC curve (AUROC)** etc. The KS statistics is a widely used metric in scorecard development. However, I personally prefer the AUROC to the others. I must add the Gini is a variant of the AUROC. The reason for my liking of the AUROC could be my formal training in Physics and engineering. I think it is a more holistic measure

and lets the analyst visually analyze the model performance. I prefer graph and visual statistics any day to raw numbers.
GINI Helps to tell the separation between 1 and 0. More separation better the model.
KS also almost tells the same.

BANKING ANALYTICS

Financial Ratios

When corporate analysts try to analyze financials of a company they often work with several financial ratios. Working with ratios has a definite advantage over working with plain vanilla variables. Combined variables often convey much higher information. Seasoned analysts understand this really well. Moreover, variable creation is a creative exercise that requires sound domain knowledge. For credit analysis, the ratio of the sum of obligations to income is highly informative since this provides an insight about percentage disposable income for the borrower.

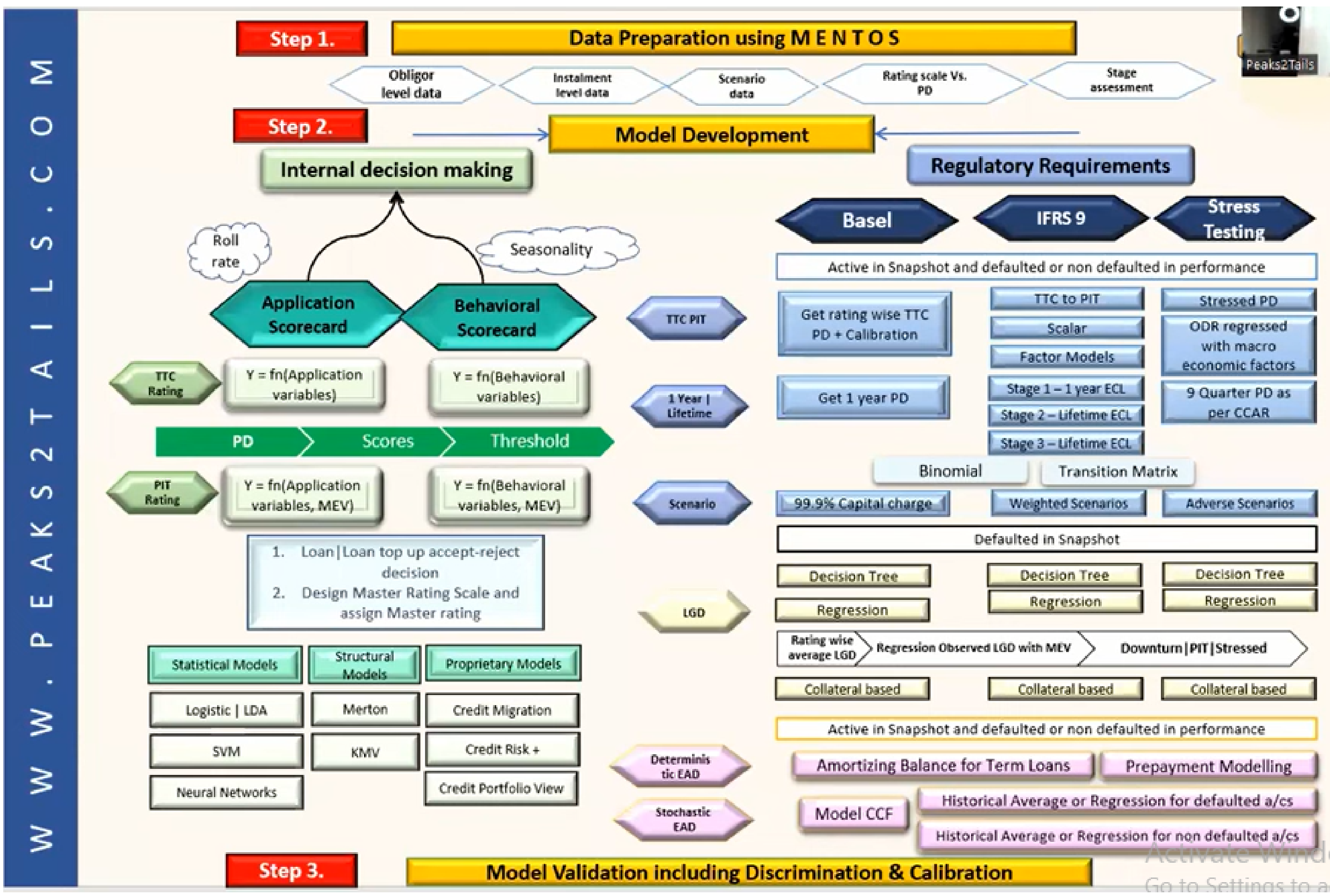
Example Links:
https://github.com/brunokatekawa/credit_risk/blob/master/Credit_Risk.ipynb

Credit Risk Analysis:

Steps for modelling:
DP MENTOS
D- data analysis
P- partition of data train test
M- missing value imputation
E - Encoding
NT - Normal Transformation
O- Outlier detection and correction
S- Scaling

Linear Discriminate Analysis: Required normally distributed data. Hence data transformation is required.

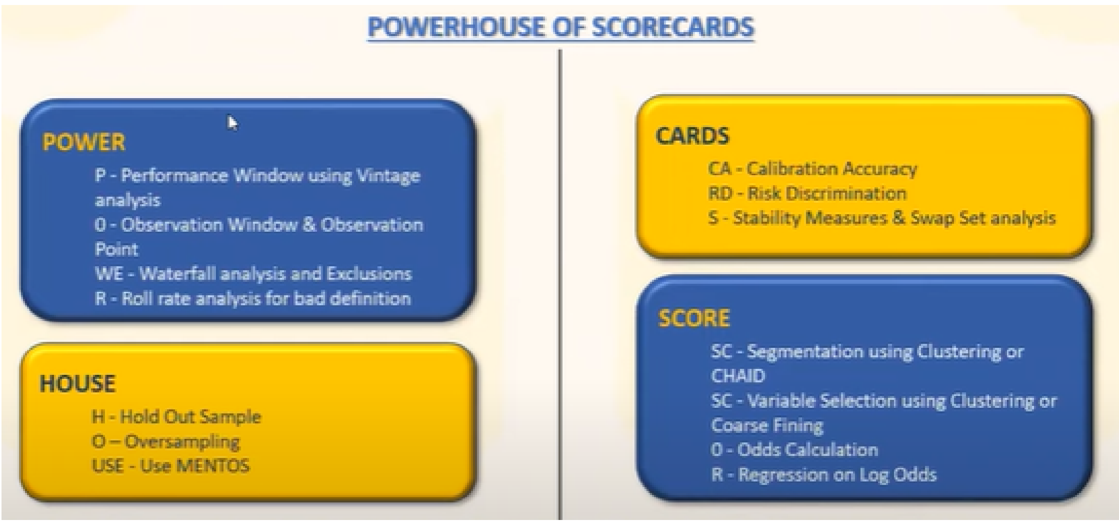
Overview:

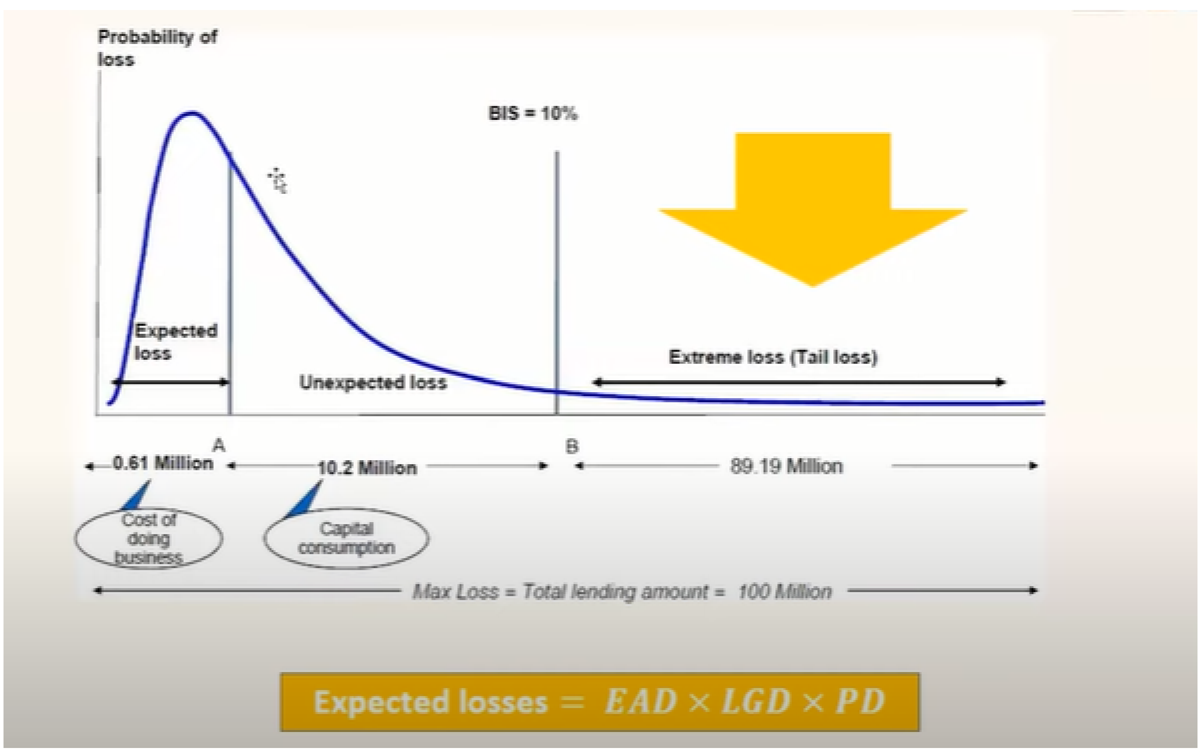


Expertise in Scorecards: Power House of Score Cards

DPD: days past due, ex: 30+ DPD 60+DPD 90+DPD
It's a time span in which default was observed.
DPD is identified based on roll rate analysis.

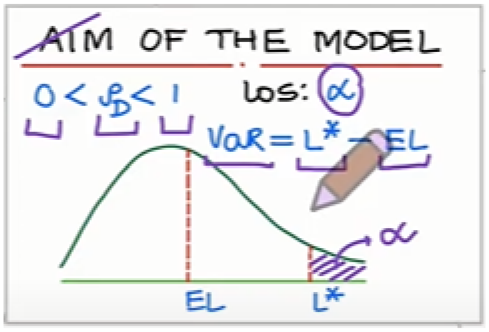
Segmentation: The accounts are separated and for each segmented group, a separate model is created, this allows for the model to train the characteristic behaviour of the particular segment group and predict better results.





B is the worst-case loss, which can be calculated as per Basel.

Where expected loss can be calculated using IFRS provisions, also the bank considers 2% as the default expected loss.



Whereas the unexpected loss is calculated using VaR. **Unexpected loss is the actual loss in a year above the expected loss. So-called Capital is required for unexpected loss.**
Whereas the L* is the worst-case loss here which is derived using monte carlo simulations.

QUANTIFICATION OF RISK COMPONENTS

- **Measuring expected loss**
  - Probability of default
  - Loss given default
  - Exposure at default

- **Measuring unexpected loss**
  - Value at risk (VaR)
  - Estimated shortfall
  - Stress test and reverse stress test
  - Scenario analysis

Extreme Loss can be known using stress testing.

In credit risk analysis BASEL, IRFS9 and Stress testing are all interlinked. Each act has an important section.

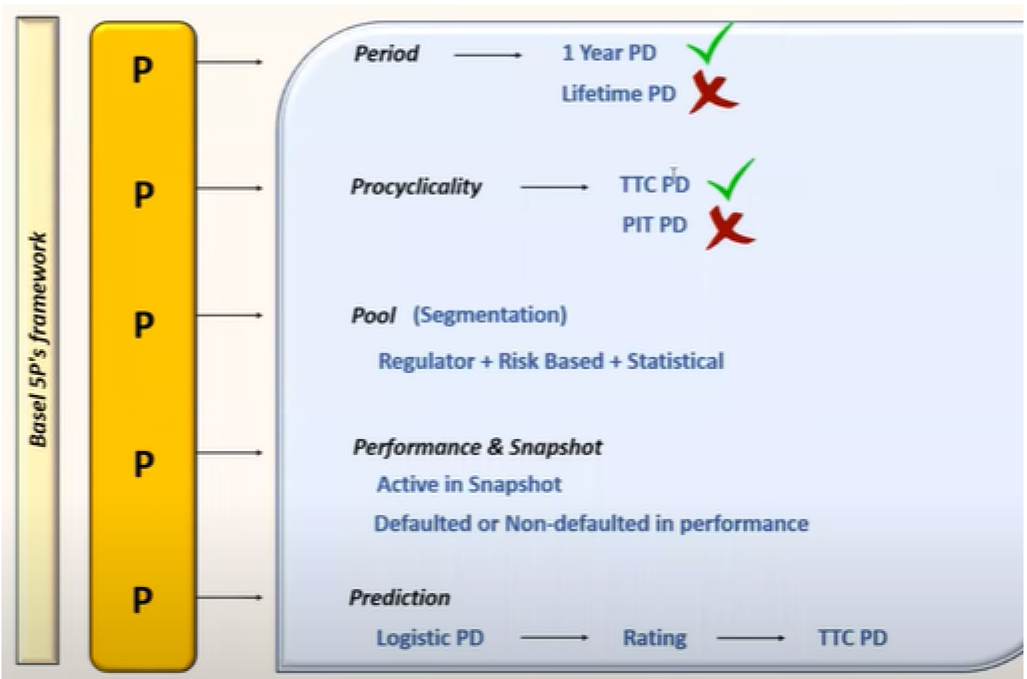
The primary components of RWAs under the Advanced Approach include

Credit Risk RWAs: Reflect the risk of loss associated with a borrower or counterparty default (failure to meet obligations in accordance with agreed-upon terms), are presented by exposure type including wholesale credit risk, retail credit risk, counterparty credit risk, securitisation credit risk, equity credit risk, and other exposures

Market Risk RWAs: Reflect the risk of possible economic loss from adverse changes in market risk factors such as interest rates, credit spreads, foreign exchange rates, equity and commodity prices, and the risk of possible loss due to counterparty exposure.

Operational Risk RWAs: Reflects the risk resulting from inadequate or failed internal processes, people and systems, or from external events.

Basel 5P’s framework : (for worst-case loss calculation)



2 types of models can be created.

1. Though the cycle - taking over the period data (average of many cycles)
2. Point in time - on time (recent cycle)

PD will be higher at a point in time if macroeconomic conditions are not good.
To calculate worst-case loss will take TTC PD.

Pool (segmentation)
Taking decile, PD for the segmented group. This shows the stability of the model.
PD should be closed for each group.

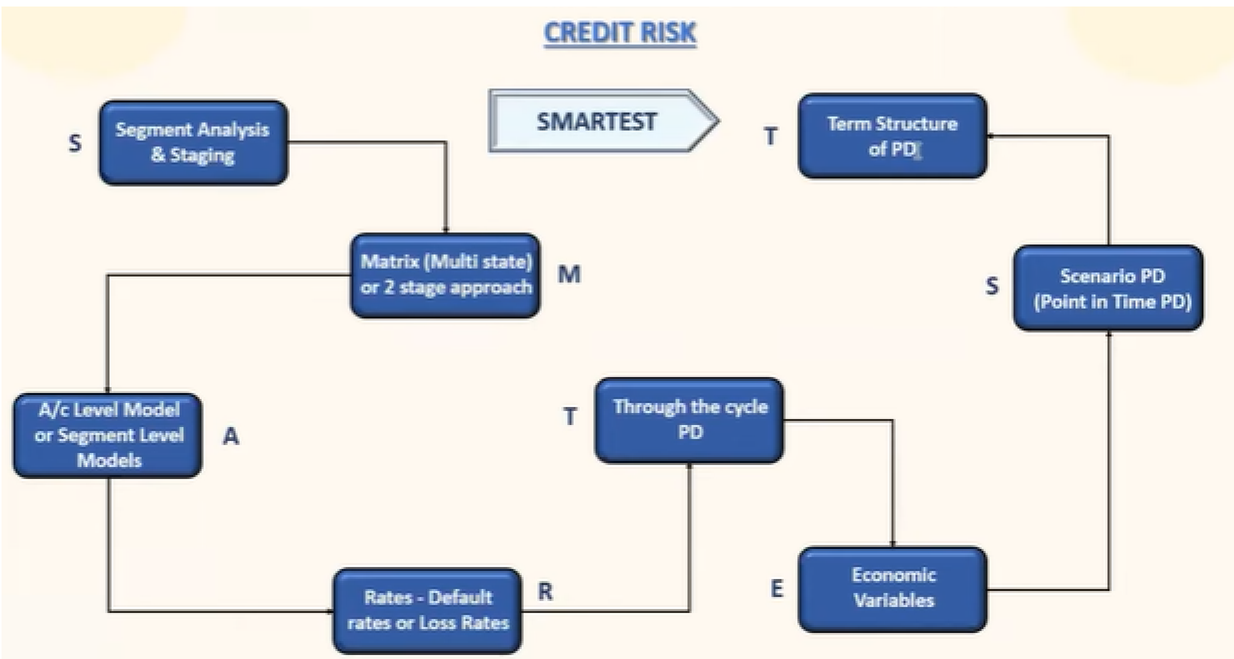
As per BASEL, **Worst-case loss (K)**: (known as the heart of credit risk)

$$K = EAD \cdot LGD \cdot \left(N\left(\frac{N^{-1}(PD) + q_{\alpha}(Z) \cdot \sqrt{\rho}}{\sqrt{1-\rho}} \right) - PD \right)$$

Z - worst case macroeconomic variable defined as per BASEL
N - threshold, Rho - correlation

Unexpected loss = worst case loss - expected loss
This unexpected loss is prevented by keeping sufficient capital. This capital maintained comes under Basel Standards.

IRFS9 is the <u>SMARTEST</u> regulation.



IFRS: Smartest regulation in credit risk. Recently came in Jan 2018.

- In different segments, different models are created. Depends if the account level model or segment level model.

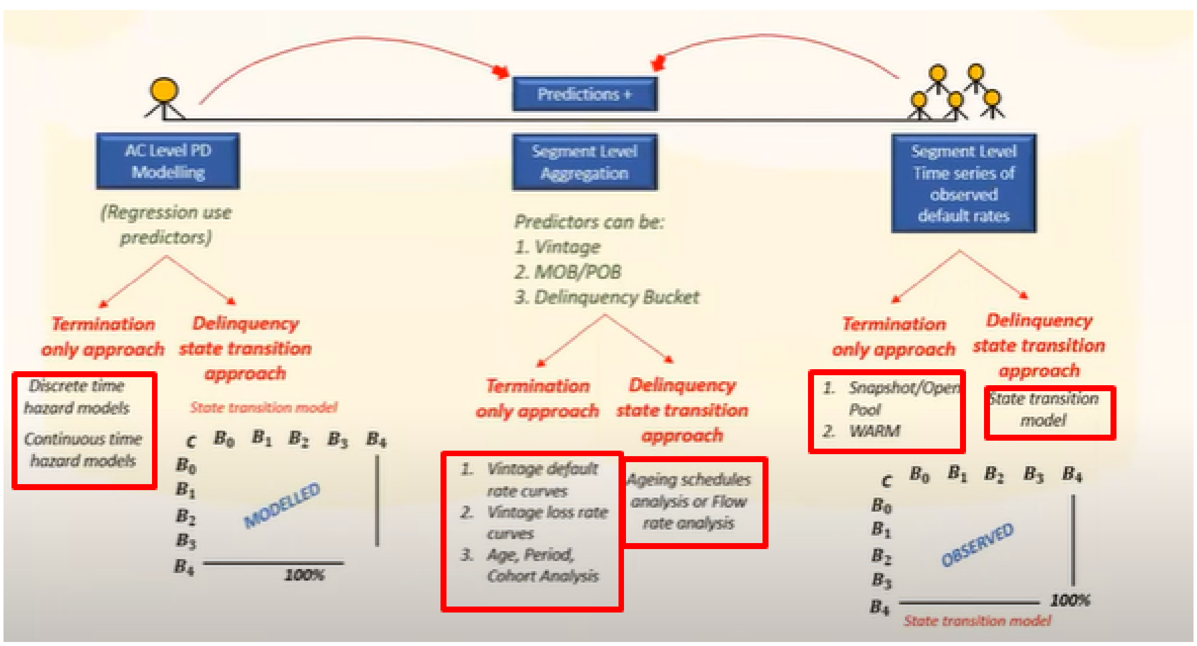
- Multi-State (delinquency state transition): complex but more granular as compared to simple binary models (default and non-default), 30DPD moving to 60DPD, 60DPD moving to 90DPD.

- Stage 1: 1-year PD
  Stage 2: 10-year PD (if the loan is for 10 years)
  Stage 3: Defaulted accounts (separate analysis)

IFRS9 vs BASEL

In IFRS9 we take TTC PD and use a macroeconomic variable to convert it to point-in-time PD and then define it different term structure.
Whereas a macroeconomic variable is decided based on different macroeconomic scenarios in future. Base Scenario, Adverse Scenario, Severe Adverse Scenario.
In Basel, we take a worst-case macroeconomic variable, whereas in IFRS9 we take a macroeconomic variable as per the macroeconomic scenario.
In Basel, we take 1-year PD, Whereas in IFRS9 we calculate term structure PD, based on stages i.e 1 year, 10 years, and Lifetime.

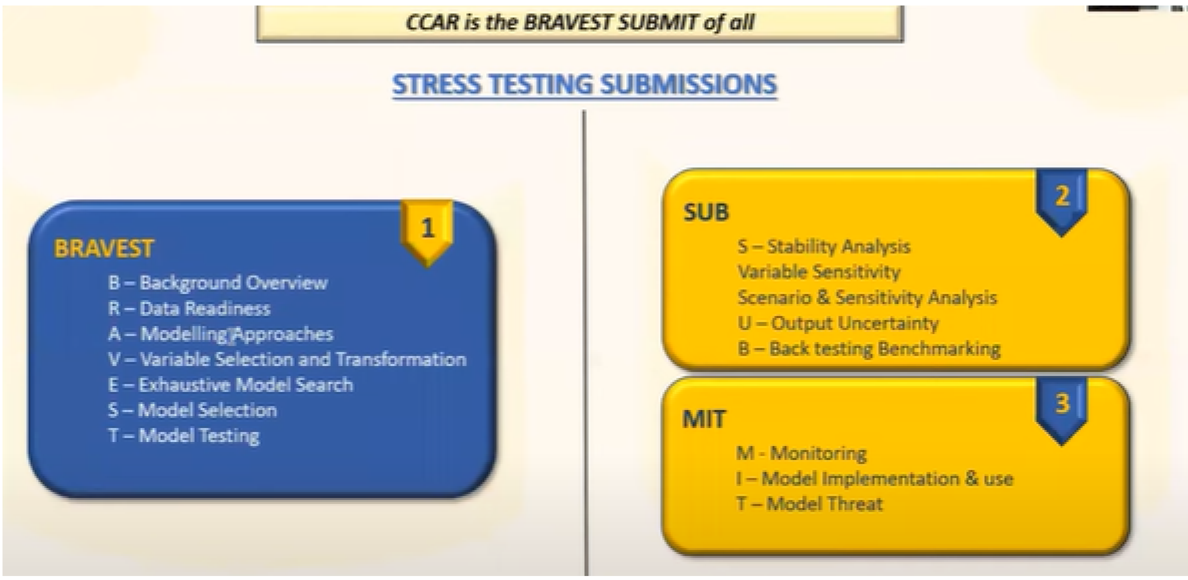
10+ approaches in modelling, which depend from case to case.
Hence IFRS9 has the highest potential for making money in the industry. The act is new and people don't know much about the different approaches followed in it.
Different models are highlighted in red below:



In IFRS we make a model through the cycle and then convert it to a point in time using macroeconomic variables using the equation of worst-case loss (the above one)

In Basel, we take the worst but in IFRS Instead of the worst case we take macroeconomic scenario baseline, severe, adverse.

Stree testing: (Framework - CCAR and DFAST)
Establishing the relationship between macroeconomic variables and default rate.
We stress macroeconomic variables to project worst-case losses.

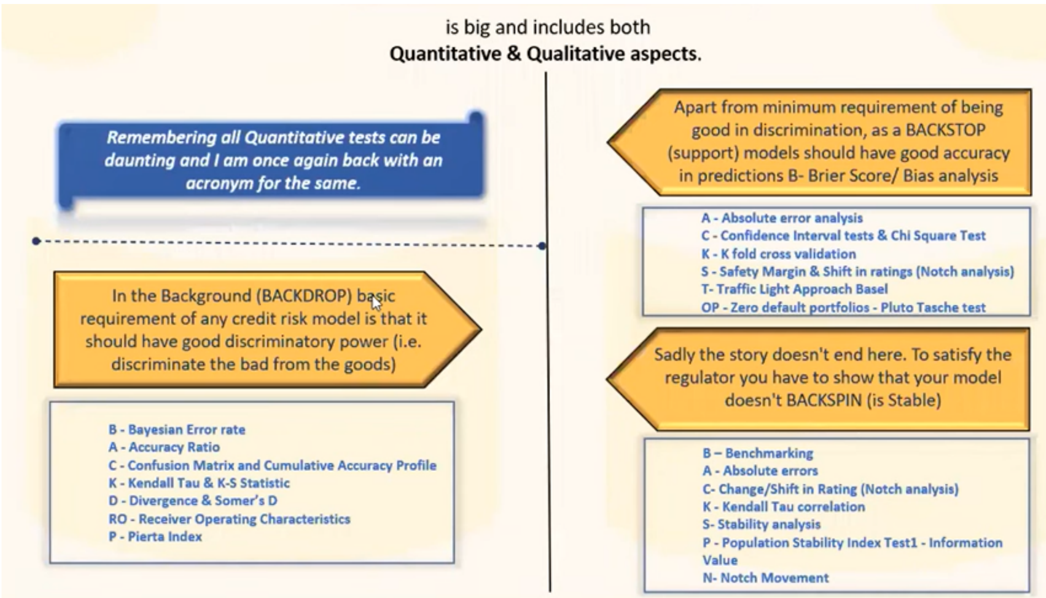


Model Approaches: Regression and time series.

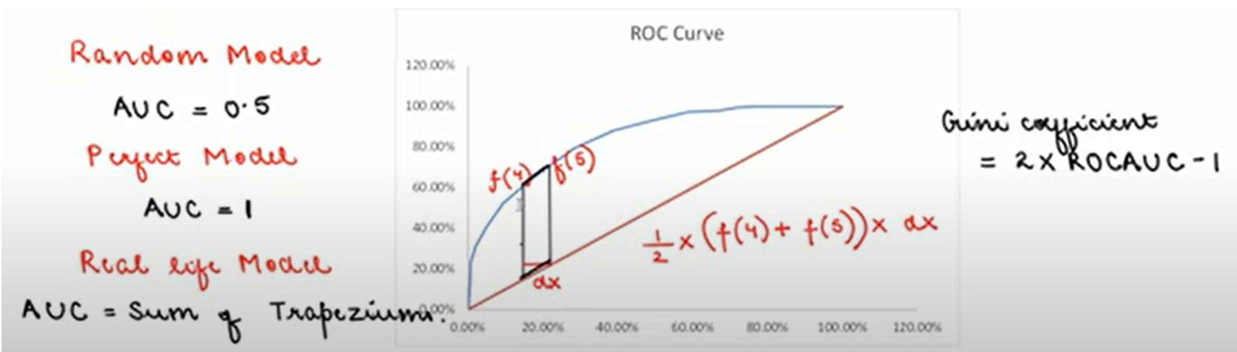
Exhaustive model search. (all types of models)
Selecting champ model and challenger model (2nd best)

Scenario testing - check for breaches under different stress, sensitive analysis.

Credit Risk Model Validation:



Model discrimination metric: AUC



Gini or accuracy ratio

CAPCAUC	0.72	AR
A_random	0.50	0.6813317
A_perfect	0.82	

Gini tells when compared to a random model which has a score of 0.5, how much better our prediction is.
Gini more than 0.6 is considered good.

Points to double the odds. Points in score need to double the odd value.

Factor	25	(Mid - Low)/10
Offset	550	Based on Mid Value of score band

$$\text{Score} = \alpha + \beta \times \ln(\text{Odds})$$

$$\text{Score} = \text{Offset} + \text{Factor} \times \ln\left(\frac{PD}{1-PD}\right)$$

$\uparrow PD,\ \uparrow \ln(\text{Odds})\ \ \uparrow \text{Score}$

But we want $\uparrow PD \rightarrow \downarrow \text{Score}$

$$\text{Score} = \text{Offset} + \text{Factor} \times \ln\left(\frac{1-PD}{PD}\right)$$

$\text{Odds} = \frac{PD}{1-PD}$, here $\text{Odds} = \frac{1-PD}{PD}$.

$PD = \frac{\text{Odds}}{1+\text{Odds}}$, here $PD = \frac{1}{1+\text{Odds}}$.

Example:

Logistic Regression - linking PD to Credit Score

In Credit Scoring Applications, score is linearly linked to Log (Odds)
Assume Score = Factor x ln (Odds) + Offset
First we decide a base score which corresponds to a base Odds
Let's say a Score 600 refers to an Odds ratio 50 : 1

Then 600 = Factor x ln(50) + Offset

Also, assume score increases by 20 points when the odds double

620 = Factor x ln(100) + Offset

Subtracting , 20 = Factor x ln(2)
Factor = 28.8539

Offset = 487.1229

Score = 28.8539 x ln ((1-PD)/PD) + 487.123

PD	**0.50%**
Score	**639.8554**

Check more on LGD Beta Regression:

CCAR: Stress PD, and Stress Losses to check to withstand capacity.

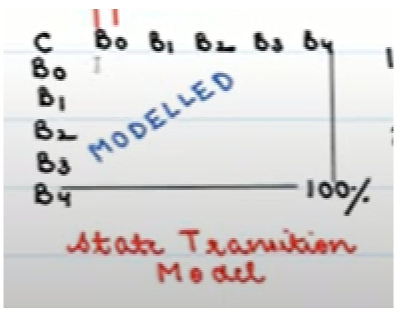
IFRS (1year point in time PD) and CECL (Lifetime PD)
Account-level PD modelling for big banks or Segment level (pool of customers) modelling for small banks.

Two types of approach:

1. Termination approach: default and non-default or continuous regression

2. Delinquency state transition approach.
   In this type, we try to predict the probability of moving from one state to the next state.
   Ex: 30DPD to 60DPD

Here multiple models are used each model predicts 1 state of transition.
Also, multi-nominal models can be used, but single logistic regression models are used for granularity. i.e making it more accurate but complex as well.



A transition matrix is created. Requires high computational power as this is done on the account level. So each account level matrix is created. Used only by bigger market cap banks. LGD and EAD need to be calculated separately.

The same matrix is created on a pool of customers, i.e segmented data by small banks

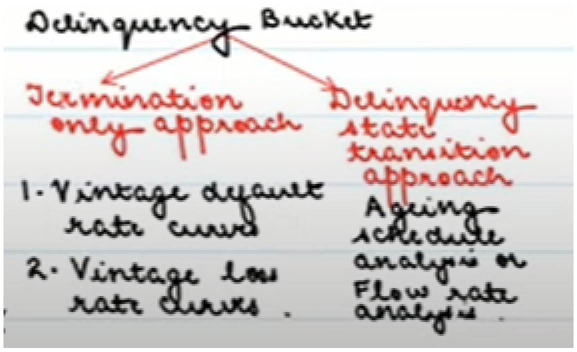
3. Mix Approach: predictors and segment level aggregation
   a. Vintage approach: macroeconomic situations and banking standards at the specific previous time period
   b. Month on books (MOB) or Period of books: Period of existing.

Age Period Cohort Analysis: combine vintage and MOB

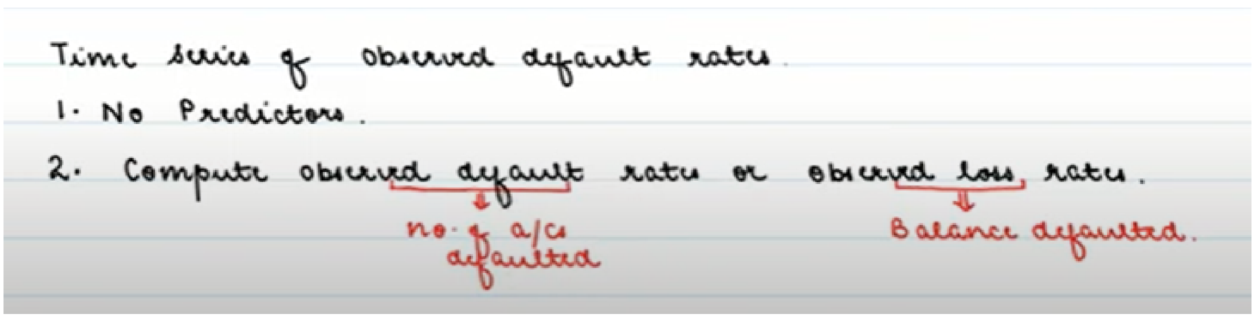
   c. Delinquency bucket:

Under this, we have again two types:

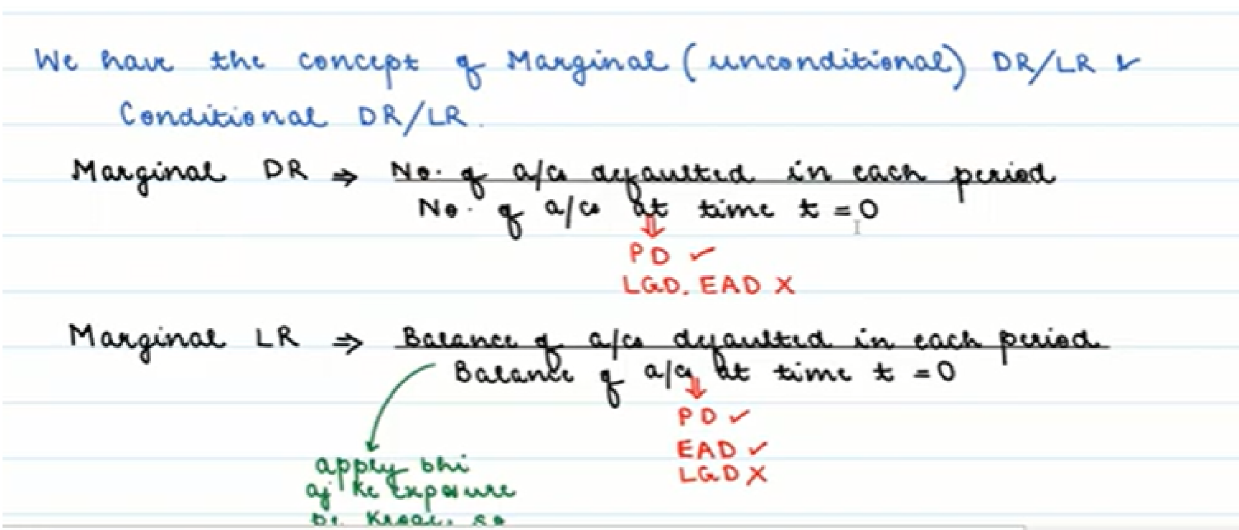
1. Termination only approach
2. **Delinquency state transition approach**



Vintage:



DR - Default rate
LR - Loss Rate



Unconditional: Starting at a fixed certain time till present
(in LR, PD and EAD are included when we calculate marginal LR)

1. Marginal rates decreases with time ∵ defaults / losses at later stage go down & denominator is fixed (time t = 0) → Static Balance

2. Because denominator is fixed you can add marginal rates to get cumulative rates (also cumulative rates saturates)

3. Can we take avg of Marginal loss rates across time
   - We can't ∵ Marginal loss rates are not stationary

Conditional DR ⇒ $\dfrac{\text{No. of a/cs defaulted in each period}}{\text{No. of a/cs at beginning of each period}}$

PD ✓
LGD, EAD ✗

Conditional LR ⇒ $\dfrac{\text{Balance of a/cs defaulted in each period}}{\text{Balance of a/cs at beginning of each period}}$

PD ✓
EAD ✗
LGD ✗

apply bhi har period ke beginning exposure pe krege, so there is a need to model EAD.

Conditional: the beginning of each period
Ex: loss rate at each period: P1, P2,.....

	P_1	P_2	P_3	P_4	P_5
Bal at beginning	9374	10088	10914	10894	10369
Bal defaulted	32	33	50	42	31
Conditional loss rate	0.34%	0.33%	0.46%	0.39%	0.30%

Types of credit risk

Wholesale credit risk:

To calculate wholesale credit RWAs, the Company inputs its modeled risk parameters (PD, EAD, and LGD) and maturity (M) into the A-IRB risk weight formula, as specified by the Final Rule. PD is an estimate of the probability that an obligor will default over a one-year horizon. EAD is an estimate of the amount that would be owed to Wells Fargo if the obligor were to default. LGD is an estimate of the portion of the EAD that would be lost (including the economic cost of delayed recovery and the cost of collection) in a stressed environment with high default rates. M is the effective remaining maturity of the exposures. Additionally, modeled parameters may be supplemented with judgmental overlays to address model or data limitations and to help ensure conservatism where appropriate. The risk mitigating benefit of guarantees are reflected in the RWAs calculation by adjusting the PD or LGD. At March 31, 2021, $89.8 billion of wholesale exposures reflected the benefit of eligible guarantees. Table 5 provides the distribution of wholesale exposures and key parameter estimates by PD bands. The commercial loan portfolio comprises about half of the wholesale EAD and nearly 90% of the wholesale RWAs. The non-loan categories (identified in the bullet points at the start of the Wholesale Credit Risk section) add significant balances to the low-risk part of the portfolio.

(in millions, except ratios)					Exposure weighted average		
PD Range (percentage)	Balance Sheet Amount	Undrawn Commitments	Exposure at Default	Advanced Approach RWAs (2)	PD	LGD	Risk Weight
0.00 to < 0.05	$ 615,298	7,104	618,576	18,461	0.02 %	9.44	2.98
0.05 to < 0.25	208,386	302,682	320,010	95,679	0.13	34.56	29.90
0.25 to < 1.50	176,316	116,067	228,565	151,653	0.72	34.33	66.35
1.50 to < 5.00	46,219	31,586	59,443	53,938	2.20	32.30	90.74
5.00 to < 13.50	17,977	11,375	24,342	30,764	7.12	31.65	126.38
13.50 to < 100	4,107	912	4,516	8,506	16.60	35.68	188.35
100 (default)	4,444	162	4,643	4,830	100.00	40.39	104.03
Total Wholesale (3)	**$ 1,072,747**	**469,888**	**1,260,095**	**363,831**	**0.84 %**	**22.06**	**28.87**

(1) Loans made by the Company in connection with the Paycheck Protection Program (PPP) are not included in this table because those loans are guaranteed by the Small Business Administration (SBA) pursuant to the Coronavirus Aid, Relief, and Economic Security Act (CARES Act) and have zero risk weight.
(2) RWAs under Basel III Advanced Approach includes the 6.00% credit risk multiplier where applicable.
(3) Includes commercial loans, debt securities, deposits with (and other funds due from) banks/other institutions, plus other non-loan exposures. Beginning this quarter, wholesale exposure includes certain commitments extended on behalf of other parties that are to be reimbursed if funded.

Retail Credit Risk:

Retail segmentation is determined by portfolios which align with respective Basel categories: Residential Mortgage - First Lien, Residential Mortgage - Junior Lien, Residential Mortgage - Revolving, Qualifying Revolving Exposures, and Other Retail. The retail segmentation process uses various factors relevant to the credit risk of retail borrowers and groups those borrowers into pools for risk quantification purposes, after which the risk parameters are quantified at the pool level. The model development methodology selection incorporates expert judgment, business knowledge, account management, collection strategy, and risk management experience. PD and LGD are estimated separately for each retail segment, and EAD is estimated for each retail exposure.

(in millions, except ratios)					Exposure weighted average		
PD range (percentage)	Balance Sheet Amount	Undrawn Commitments	Exposure at Default	Advanced Approach RWAs (2)	PD (3)	LGD	Risk Weight
Residential mortgage - first lien:							
0.00 to < 0.10	$ 198,862	—	198,862	14,549	0.10 %	29.87	7.32
0.10 to < 0.25	15,708	16,776	29,042	3,911	0.22	29.96	13.47
0.25 to < 1.00	16,323	—	16,323	3,813	0.51	28.11	23.36
1.00 to < 5.00	16,247	93	16,340	7,614	1.79	24.77	46.60
5.00 to < 10.00	9,324	—	9,324	9,236	7.10	23.83	99.06
10.00 to < 100.00	3,937	116	4,054	4,730	31.04	22.71	116.67
100 (default)	14,610	—	14,610	7,429	100.00	20.88	50.85
Total residential mortgage - first lien	**275,011**	**16,985**	**288,555**	**51,282**	**5.95**	**28.74**	**17.77**
Residential mortgage - junior lien:							
0.00 to < 0.10	370	—	370	60	0.08	79.62	16.22
0.10 to < 0.25	41	—	41	9	0.19	54.13	21.95
0.25 to < 1.00	305	—	305	184	0.49	74.06	60.33
1.00 to < 5.00	280	—	280	425	2.31	66.40	151.79
5.00 to < 10.00	129	—	129	434	7.49	80.97	336.43
10.00 to < 100.00	29	—	29	119	29.87	73.21	410.34
100 (default)	97	—	97	99	100.00	70.70	102.06
Total residential mortgage - junior lien	**1,251**	**—**	**1,251**	**1,330**	**9.91**	**73.78**	**106.31**
Residential mortgage - revolving:							
0.00 to < 0.10	7,501	46,370	21,132	1,883	0.03	81.94	8.91
0.10 to < 0.25	13,319	3,970	13,930	4,442	0.17	81.70	31.89
0.25 to < 1.00	4,061	989	4,318	4,437	0.89	82.17	102.76
1.00 to < 5.00	1,801	72	1,854	3,813	2.76	82.65	205.66
5.00 to < 10.00	373	339	480	1,634	7.22	78.98	340.42
10.00 to < 100.00	382	30	394	1,908	24.00	83.45	484.26
100 (default)	1,193	71	1,260	1,336	100.00	77.29	106.03
Total residential mortgage - revolving	**28,630**	**51,841**	**43,368**	**19,453**	**3.48**	**81.76**	**44.86**
Qualifying revolving: (4)							
0.00 to < 0.25	3,440	95,548	20,780	1,371	0.11	96.26	6.60
0.25 to < 1.00	11,398	25,233	17,905	4,785	0.58	96.71	26.72
1.00 to < 2.50	8,228	6,054	10,860	6,744	1.67	96.97	62.10
2.50 to < 5.00	7,867	2,457	9,492	9,363	3.43	97.07	98.64
5.00 to < 10.00	2,957	441	3,376	5,197	6.72	97.31	153.94
10.00 to < 100.00	2,297	229	2,533	6,266	36.65	97.19	247.37
100 (default)	1	—	1	1	100.00	96.38	100.00
Total qualifying revolving	**36,188**	**129,962**	**64,947**	**33,727**	**2.76**	**96.71**	**51.93**
Other retail:							
0.00 to < 0.25	25,558	25,379	40,103	8,715	0.11	73.41	21.73
0.25 to < 1.00	23,705	3,963	27,074	15,214	0.57	65.01	56.19
1.00 to < 2.50	24,955	1,447	26,486	22,302	1.63	63.57	84.20
2.50 to < 5.00	7,507	1,021	8,428	8,625	3.70	68.59	102.34
5.00 to < 10.00	3,337	155	3,510	4,048	7.28	69.38	115.33
10.00 to < 100.00	2,983	22	3,103	4,827	24.83	69.67	155.56
100 (default)	678	14	691	636	100.00	51.98	92.04
Total other retail	**88,723**	**32,001**	**109,395**	**64,367**	**2.41**	**68.21**	**58.84**
Total Retail Exposures	**$ 429,803**	**230,789**	**507,516**	**170,159**	**4.58 %**	**50.58**	**33.53**

Securitization Credit Risk:

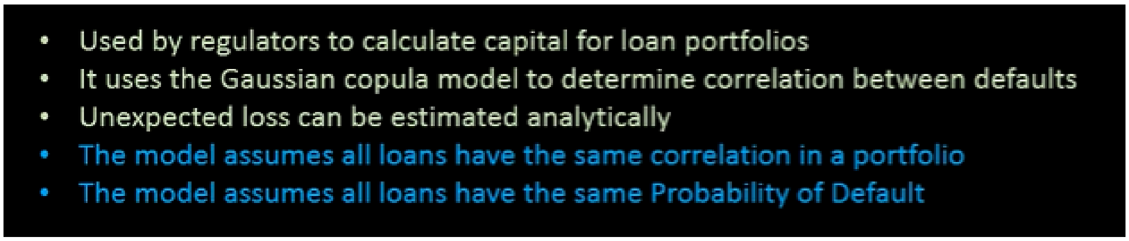
In securitization, a bank's exposure to credit risk is transferred into a Special Purpose Vehicle (SPV) that issues securities to a broad array of investors.

Equity Credit Risk:

Equity exposures that are subject to the equity credit risk capital rules include banking book equity exposures and trading book equity exposures not covered under the market risk capital rules.

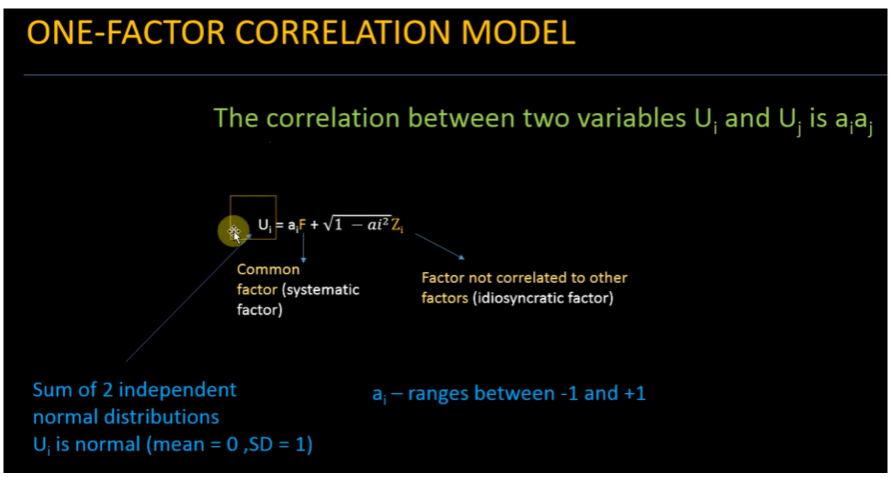
Process for TTC to PIT

Vasicek model

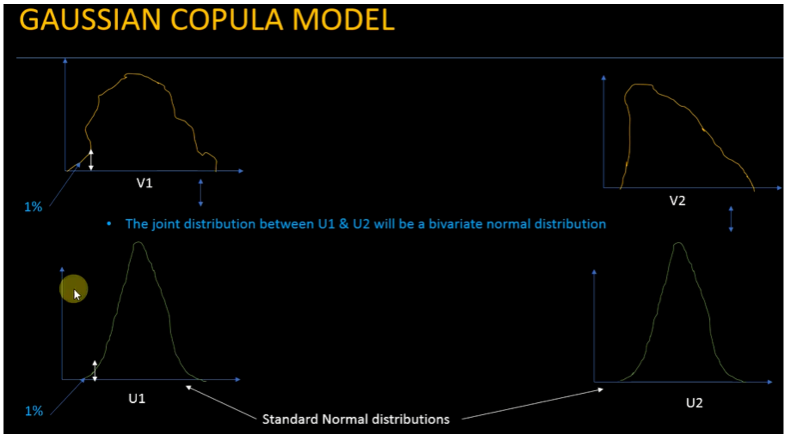


Where the Vasicek model has 3 components:

1. Common Factors (systematic factor)
2. Factors not correlated to other factors (idiosyncratic factor)
3. The equation to combine the above two factors:



Where Ui is derived from copula.


GAUSSIAN COPULA MODEL
V1
1%
V2
• The joint distribution between U1 & U2 will be a bivariate normal distribution
U1
1%
Standard Normal distributions
U2

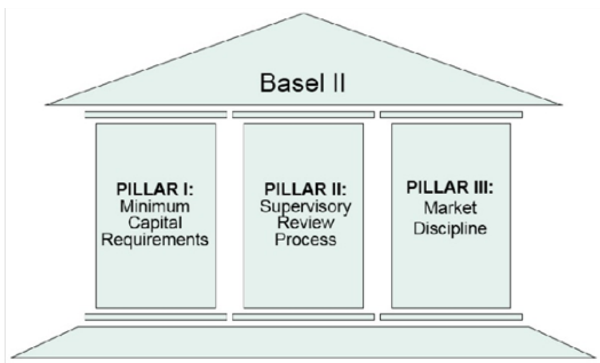
BASEL

Three Pillars were established as part of the Basel III capital adequacy framework

Pillar 1: **Minimum Capital Requirements** establishes capital requirements and prescribes rules for determining the regulatory capital treatment of capital instruments and calculating RWAs.

Pillar 2 requires banks to develop and maintain an Internal Capital Adequacy Assessment Process (ICAAP) to support the assessment of their capital adequacy. Pillar 2 also outlines principles of the **supervisory review process** to monitor banks' capital and evaluate banks' management of risks through the use of internal control processes.

Pillar 3 promotes **market discipline** through minimum requirements for qualitative and quantitative disclosures made available to the public to enable market participants to compare banks' disclosures of **RWAs** (Risk-Weighted Assets under Basel III) and improve the transparency of the internal model-based approaches that banks use to calculate minimum regulatory capital requirements.



Pillar 1 entitles banks to compute their credit risk capital in either of two ways:

1. Standardized Approach

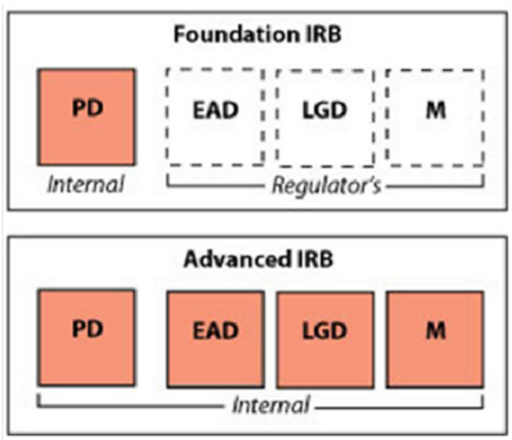
2. Internal Ratings-Based (IRB) Approach

    a. Foundation Approach
    b. Advanced Approach

Ratings:

Risk Grade	Default Probability (%)
AAA	1
AA	2
A	3
BBB	4
BB	6
B	10
C	15

AAA	Very low risk
AA	Marginal risk
A	Low risk
BBB	Moderate risk
BB	Fair risk
B	High risk
C	Very high risk
D	Default



The IRB approach is based on the following four key parameters:

1. Probability of Default (PD): the likelihood that a loan will not be repaid and will therefore fall into default **in the next 12 months.**

2. Loss Given Default (LGD): the estimated economic loss, expressed as a percentage of exposure, which will be incurred if an obligor goes into default - in other words, LGD equals: 1 minus the recovery rate.

3. Exposure At Default (EAD): a measure of monetary exposure should an obligor go into default.

4. Maturity (M): the length of time to the final payment date of a loan or other financial instrument

Expected loss:

Financial institutions expect a certain number of the loans they make to go into default; however they cannot identify in advance which loans will default. To account for this risk, a value for expected loss is priced into the products they offer.

From the parameters, PD, LGD and EAD, expected loss (EL) can be derived as follows:

$$EL = PD * LGD * EAD$$

For example, if PD = 2%, LGD = 40% and EAD = $10,000, then EL would equal $80. Expected Loss can also be measured as a percentage of EAD:

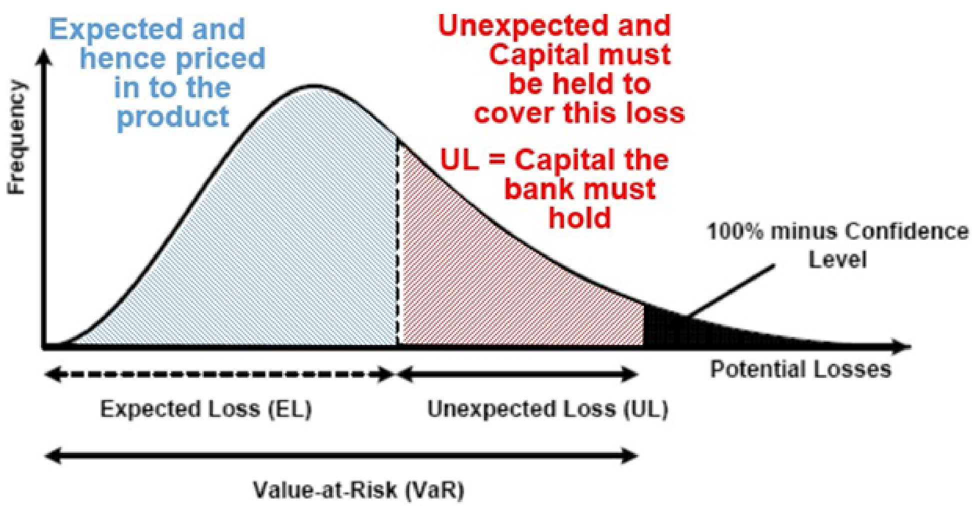
$$EL\% = PD * LGD$$

Expected loss as a percentage of EAD would be equal to $EL\% = 0.8\%$.

Unexpected Loss:

Unexpected loss is defined as any loss on a financial product that was not expected by a financial organization and therefore not factored into the price of the product.

Important: The purpose of the Basel regulations is to force banks to retain capital to cover the entire amount of the Value-at-Risk (VaR), which is a combination of this unexpected loss plus the expected loss.



Risk Weighted Assets (RWA) are the assets of the bank (money lent out to customers and businesses in the form of loans) accounted for by their riskiness. The RWA are a function of PD, LGD, EAD and M, where K is the capital requirement:

$$RWA = (12.5) * K * EAD$$

Important: Under the Basel capital regulations, all banks must declare their RWA

The capital requirement needs to ensure capital is no less than 8% of RWA, where:

The Capital Requirement (K) is defined as a function of PD, a correlation factor (R) and LGD

$$K = LGD \times \left(\phi\left(\sqrt{\frac{1}{1-R}}\phi^{-1}(PD) + \sqrt{\frac{R}{1-R}}\phi^{-1}(0.999) \right) - PD \right) \quad (1.4)$$

where *f* denotes the normal cumulative distribution function and *f* -1 denotes the inverse cumulative distribution function. The correlation factor (R) is determined based on the portfolio being assessed.

PD Calibration :

Data used to develop the model using its information from the past, may not answer the question of current condition, so it was born as a way to do Calibration (measuring something accurately).

PD Models:

1. **Logistic**
2. **GBM**
3. **Linear and Quadratic Discriminant Analysis**
4. **Random Forest**
5. **XGboost**
6. **Neural Network**

Typical examples of bad definitions are as follows:

1. 90 days delinquent – this is defined to have occurred where a customer has failed to make a payment for 90 days consecutively (within the observation period).
2. 2 x 30 days, 2 x 60 days, or 1 x 90 days – this is defined to have occurred where a customer has been either 30 days delinquent twice, 60 days delinquent twice, or 90 days delinquent once (within the observation period).

Types of scorecards:

1. Application scorecards : new to the business.
2. Behavioural scorecards: an existing account will turn bad

PD Model Reporting:

Reports include information such as model performance measures, development score and scorecard characteristics distributions, expected bad or approval rate charts, and the effects of the scorecard on key subpopulations. Reports facilitate operational decisions such as deciding the scorecard cutoff, designing account acquisition and management strategies, and monitoring scorecards.

LGD Models:

Note: When macroeconomic variables such as GDP growth are added as additional explanatory variables, they exhibit low explanatory power for the recovery rates. This indicates that in the prediction of the LGD (recovery rate) at an account level, macroeconomic variables do not add anything to the models which only incorporate individual loan-related variables derived from the data.

The best techniques to apply in the estimation of LGD, given its bimodal distribution: one model to first discriminate between zero-and-higher LGDs and a second model to estimate LGD for the subpopulation of non-zero LGDs.

Example: Logistic regression + OLS, B-OLS(Ordinary Least Squares with Beta Transformation), BR(Beta Regression), BCOLS(Ordinary Least Squares with Box-Cox Transformation), RT(Regression Trees), or ANN

Logistic regression to first estimate the probability of LGD ending up in the peak at 0 (LGD≤0) or to the right of it (LGD>0). A second-stage non-linear regression model is built using only the observations for which LGD>0. An LGD estimate is then produced by weighting the average LGD in the peak and the estimate produced by the second-stage model by their respective probabilities.

Metrics:

Metric	**Worst**	**Best**
RMSE	$+\infty$	0
MAE	$+\infty$	0
AUC	0.5	1
AOC	$+\infty$	0
R^2	0	1
r	0	1

Exposure at default:

in order to estimate EAD, for off-balance-sheet (unsecured) items such as example credit cards, one requires the committed but unused loan amount times a credit conversion factor (CCF).

Note: The term Loan Equivalency Factor (LEQ) can be used interchangeably with the term credit conversion factor (CCF) as CCF is referred to as LEQ in the U.S.

EAD is estimated as the current drawn amount, $E(t_r)$, plus the current undrawn amount (credit limit minus drawn amount), equivalency factor (LEQ): $L(t_r) - E(t_r)$, multiplied by a credit conversion factor, CCF or loan

$$EAD = E(t_r) + CCF * (L(t_r) - E(t_r))$$

The credit conversion factor can be defined as the percentage rate of **undrawn credit lines** (UCL) that have yet to be paid out but will be utilized by the borrower by the time the default occurs. (possibility of additional withdrawals when the limit allows)

However, a CCF calculation is not required for **secured loans such as mortgages.**

$$CCF_i = \frac{E(t_d) - E(t_r)}{L(t_r) - E(t_r)}$$

where $E(t_d)$ is the Exposure at the time of Default, $L(t_r)$ is the advised credit limit at the start of the time period, and $E(t_r)$ is the drawn amount at the start of the cohort.

Time Horizons for CCF

In order to initially calculate the CCF value, two-time points are required. The actual Exposure at Default (EAD) is measured at the time an account goes into default, but we also require a time point from which the drawn balance and risk drivers can be measured, Δt before default.

Three types of approach that can be used in the selection of the time period Δt for calculating the credit conversion factor:

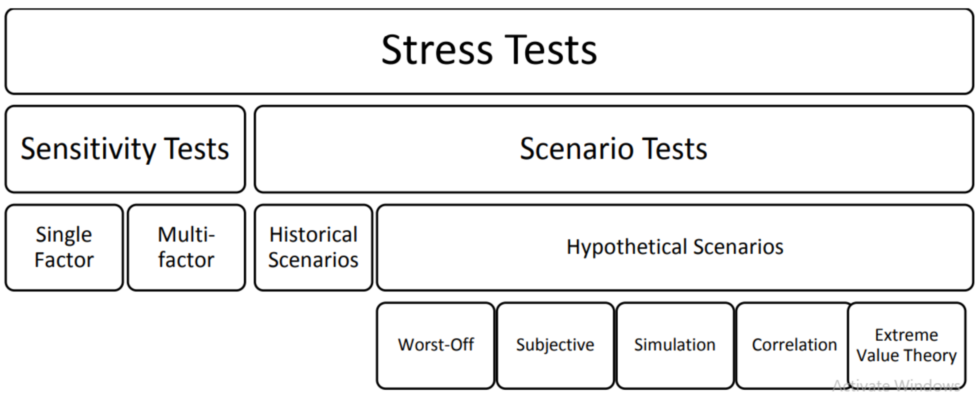
1. The Cohort Approach – This approach groups defaulted accounts into discrete calendar periods according to the date of default. A common length of time for these calendar periods is 12 months; however, shorter time periods may be more appropriate if a more conservative approach is required. The information for the risk drivers and drawn/undrawn amounts are then collected at the start of the calendar period along with the drawn amount at the actual time of default (EAD). With the separation of data into discrete cohorts, the data can then be pooled for estimation.

2. The Fixed-Horizon Approach – For this approach, information regarding risk drivers and drawn/undrawn amounts are collected at a fixed time period prior to the defaulting date of a facility as well as the drawn amount on the date of default. In practice, this period is usually set to 12 months unless other time periods are more appropriate or conservative.

3. The Variable Time Horizon Approach – This approach is a variation of the fixed-horizon approach by first fixing a range of horizon values (12 months) in which the CCF will be computed. Second, the CCF values are computed for each defaulted facility associated with a set of reference dates (1 month, 2 months, …, and 12 months before default). Through this process, a broader set of potential default dates are taken into consideration when estimating a suitable value for the CCF.

Stress Testing :

Figure 6.2: Stress Testing Methodologies



Sensitivity testing - Includes static approaches which do not take into account external (macroeconomic) information. Single factor sensitivity tests for credit risk can be conducted by multiple means:
1. Stressing the data, for example, testing the impact of a decrease in a portfolio of customers' income by 5%.
2. Stressing the PD scores, for example, testing the impact of behavioral scores falling by 15%.
3. Stressing rating grades, for example, testing the impact of an AAA rating grade decreases to an AA rating.

Multi-factor sensitivity tests seek to stress all potential factors by understanding the correlation between all of the factors. This type of sensitivity analysis is more synonymous with scenario-type testing.

Scenario stress testing: It is the method of taking historical or hypothetical scenario situations where the source of shock is well-defined as well as the parameter values that can be impacted. In most cases, these scenarios are either portfolio or event-driven and can take into account macroeconomic factors. Scenario testing can be broken down into two constituent parts:
1. Historical Scenarios –scenarios based upon actual events and therefore potential

variations in parameter values are known. How recent history can impact today's current portfolio?

2. Hypothetical Scenarios – requires expert judgment to assess potential threats and test against these. Hypothetical scenarios are much harder to conduct as stressing unknown conditions.
Market downturn – for financial institutions the most common hypothetical scenario is stressing adverse impacts on the market through macroeconomic factors. Correlation between PD and LGD values, the correlations defined in the past may not hold in times of stress.
Simulation scenario analysis: Simulations can be used to estimate potential outcomes based on stressed conditions. For example, if unemployment were to increase by 2%, 3% or 4%, forecast how this would impact portfolio-level loss.
Worst-case scenario analysis: this type of hypothetical scenario analysis stresses the most extreme movement in each risk parameter – This is the least plausible and also ignores correlations between risk factors. However, this is also one of the most common approaches.

Model Validation Categories
Model Performance: Ability of a model to discriminate
Model Stability: Track changes in the distribution of modelling and scoring data sets.
Model Calibration: Assess the accuracy of PD, LGD, and EAD models

Model Stability: It is used to create the PD report and the LGD reports.
System Stability Index (SSI): SSI monitors the score distribution over a time period. This can be calculated for both PD and LGD models.