# CR – IRB – Retail × Corporates portfolio
(Summary)



- Retail : PD, EAD, LGD
- Corp : PD

Outline:

Ⅰ PD
1. Popu. data collection
2. Seg. × sampling
   DD Development
3. Model
   i. Develop
      - Data treatment
      - Uni. ana.
      - Transform data
      - Multi. ana
   ii. Validate
   iii. Calibrate

long list → short list → final shortlist

Ⅱ EAD
1. Data + sampling
2. CCF modeling
3. EAD estimate
4. Backtest

Ⅲ LGD
1. Data + sampling
2. Recoveries modeling
3. Backtest

## Data. Seg. DD. For Retail × Corp:

- Data : . at least 1 business cycle (cover all stage)
  corp : > 7y . Retail : > 5y (Basel II)
  - Train / test ~ 7/3 ; O-O-sample : 1 mth after
- Segmentation: Corp : usually : industry, size
- Default drivers:

| Corp | Retail |
|---|---|
| • Financial drivers: Progit, debt ratio ... | |
| • Non-Fin : | |
| . level × quality of mgt | • Customer info (gender, age, Edu, marital status ...) |
| . cic info | |
| . internal environment | |
| . Product info (length ...) | • Product info |
| . credit history, pmt history | • credit history, pmt —— |

DETAIL Ⅰ PD. %

## 2. Seg × sampling

+ Seg : segments of similar Risk characteristic/profile based on:
  - + Sta ana : . clustering (K-mean, Agg hie. clustering)
    . decision tree
  - + Expert perspective, business sense

+ Sample : + Stratified
+ Imbalance dataset → solution :
  - . oversampling / undersampling
  - . ensemble
  - . use Class weight = 'balanced'
  - . metric : $F_1$, Recall, Precision

S.t GIB ~ 4:1 or 3:1

+ Missing value : (M.V)
  - + Remove if : . M.V Random
    . Remove is not biased for model
  - + Replace by mean / median for continous variable
  - + Group in 1 group : for discrete variable.
- + Outlier
  - + true outlier : [ > $\bar{X} \pm 3\sigma$ ; $\bar{X} \pm 3\sigma$ → ≠ value from business sense ]
  - + data quality issue : treat as M.V

## 3. Model

i. Develop

Data treat + Exclusions : . borrower exclusion, if missing > 20% for Retail ; > 50% for corp.
. facility ——
. factor (variable) —— : ——

a. Univariate ana.

{ predictive power
  stability
  business sense
by: { Statistical ana
    { expert opinions

Sta ana.
- WOE : far from 0 → Good ig
- IV : 0.1 – 0.5 ↑
- Gini ↑
- for uni. vana : 5 – 10
- For model : 30 – 50%

+ Stepwise: mix both method:
{ 1 step fw
  — backward
  Repeat

+ Model selection based on { sta tests
                           { credit expert opinions

Sta ana:
- |ρ| < 0.7
- VIF < 5
- p < 0.05
- gini 30–50% or > 55%

- testing: seeing μ of G & B samples are statistic signif. ≠
  (p < α)
  $H_0: \bar{\mu}_X = \bar{\mu}_Y$

PD: G & B sample
[ (M) of 1 variable )
EAD : ≠ btw CCF
LGD LGD

b. Transform
variables with ≠ ranges → bad Regression

How? num variable : [ normalization → range(0,1)
                    [ std $x_{std} = \frac{x-\mu}{\sigma}$
                    $x_{new} \sim Dist(\mu=0; \sigma=1)$
dis/categori : score adjustment

c. Multi. ana
{ • develop some models by variable combination
{ • Select some optimized ——

How? + Corr ana [ Pearson : corr (value) : continous variable
                [ Spearman : corr (Rank) cont & dis —

if |ρ| > 0.7 or 0.6–0.8:
[ Keep 1 : the one is better
[ Keep both: but not together in 1 model
[ discuss

+ multicolinearity: ∃ linear Rela btw $(X_i)$s
check VIF = $\frac{1}{1-R^2}$ ← $R^2$ ← $X_1 = b_1 + b_2 X_2 + \dots$
VIF > 5 → remove $X_1$

+ Stepwise Regression
+ backward : all variables at beginning
  • Remove till all vari. meet Requirement
+ Forward : add 1 vari. at a time
  till no variable meet requirement
+ Stepwise

• Condition to add/drop:
most / least signi. variable in the model

**Validate** assess model performance

- model stability : accuray (if the imbalanced issue is treated already)
- confusion matrix

other:
- Gini : overall performance
- KS peak — : 0.2-0.4
- Lift : Relative — at a cutoff point
- Hosmer lemeshow test
- Area under ROC . > 80%
  = (Gini+1)/2

**Calibrate**

add offset factor

when: if default rate in sample not ≈ in population

as time goes & more new data

---

II **EAD** $

EAD = current outstandig + CCF (commited $ − current outstandig $)

estimate : additional drawing up to default time

CCF: credit conversion factor . ∈ [0,1]

III **LGD** %

$$\%LGD = 1 - \frac{\sum PV(\text{Recoveries}) - PV(\text{cost})}{EAD}$$

---

⊛ Decision making PD:

Calculate PD for a customer → create score card → PD → set cutoffs

Some formulas

**WOE, IV**

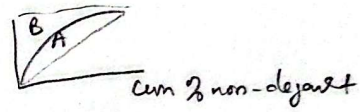
$$WOE = \ln\left(\frac{\%G}{\%B}\right)$$

$$IV = \sum [(\%G - \%B)\, WOE]$$

i.e:

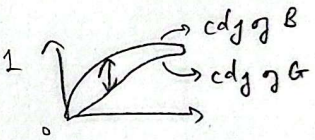
| category | #G | #B | %G | %B |
|---|---|---|---|---|
| Higher Edu | | | | |
| no education | | | | |
| Σ | | | | |

**Gini**




$$Gini = \frac{A}{A+B}$$

**KS**




KS = max distance btw cdf of G & B curves

∈ [0,1]