Notes - CRR Models

Thursday, July 30, 2020 2:32 PM

Creditor ----(% interest)--- Lender, banks make profit out of this interest rate

What is a credit risk?

- The likelihood that a borrower would not repay their loan to lender
- When a borrower fails to replay loan
- The lender will have to sustain substantial cost in an effort to recover outstanding debt - cost

Risk based pricing

- Based on credit risk associated with the particular client, banks decide the % interest rat collateral
- Lending to borrowers with high probability default is one of the main reasons of financia 2008,
  - Factor - high default rate of sub-prime mortgages; low interest rate; loans were pr 100% or more value of the home; demand for homes increased -> selling rice incre
  - Mortgage backed securities lost value
  - Big banks holding this instruments got affected - Le Mans Brother

Regulatory rule

- Lender must assess credit risk associated with each borrower
- Lenders know certain amount of credit risks is always associated with each/every client
- It is important to estimate expected loss - amt of money a lender lose by lending to a bo
- There are many models but established credit risk model has 3 components
  - PD (Probability of default)
  - LGD (Loss given default)
  - EAD (Exposure at default)

PD (Probability of default):

- The borrower inability to repay his/her debt in full or on-time
- Estimate of a borrower to likelihood the borrower will default

Loss given default (LGD)

- Loss of an asset due to borrowers default called collection

e and size of the

l crisis. Example

ovided with
eased

rrower.

- The proportion of the total exposure that cannot be recovered by the lender once a defa

EAD (Exposure at default)

- Total value that a lender is exposed to when a borrower defaults i.e. maximum amount t lose if a borrower defaults

Example: A borrower wants to buy a house
Price = $500,000
(Agreement) Bank funds 80% (loan to value) = $400,000
Borrower repaid = $40,000
Outstanding balance = $400,000 - $40,000 = $360,000
When a borrower defaults, total value that a lender is exposed to is $360,000. This is **Exposure**
Assume that the banks can sell the house is at $342,000, then
Loss given default = (EAD - $342,000)/EAD

- Assume that there is an empirical evidence that one in four homeowners have defaulted
  PD = 1/4
- Expected Loss (EL) = PD * LGD * EAD, i.e. 25% * 5% * $360,000 = $4,500

<u>Capital Adequacy/Regulations and Basel III accord</u>

- Govt. regulators impose certain requirements for banks to make sure bank conduct their without risking the stability of the economy system
- Regulators set rules
  - Regulate bank operations & hence reduce bank risky behaviour
  - Guarantee to the public that banking system is in good health
- Capital Adequacy: Banks require to hold sufficient capital to absorb capital losses from d obligation is called "**Capital Requirement**"
- CAR(Capital Adequacy ratio) = Capital / Risk-weighted assets (loans)
- Basel III accord - primary objective is to ensure capital allocation. Greater the risk the bar greater amount of capital it needs to hold
  - Three pillars of Basel III -
    1. Market discipline
    2. Supervisory review
    3. **Minimum Capital Requirement**
       Two approaches
       - Standardized approach
       - Internal ratings based approach (IRB)
         - ◇ Foundation internal rating based (F-IRB)
         - ◇ Advanced internal rating based (A-IRB)

ult has occurred

that a bank can

at default.

r business

efault. This

nk exposed to

- Basel III accord prescribes regulators should allow banks to choose from 3 different appr calculate/modelling credit risk i.e. calculating & modelling each of the three components loss
  - Standardized approach
  - Foundation internal rating based (F-IRB)
  - Advanced internal rating based (A-IRB)
- Capital requirements are calculated differently in 3 approaches

	SA	F-IRB	A-IRB
PD	External data	Internally calculated	Internally calculated
LGD	External data	External data	Internally calculated
EAD	External data	External data	Internally calculated

- External data comes from credit rating agencies such as FICO, S&P, Moody's, Fitch (for fi
- Two major credit reporting agency that collects data from FICO
  - Equifax
  - TransUnion
- Credit score ranges from 300 to 800
- For firms, credit cards & consumer loans, banks should hold around 75% of the total exp (Capital adequacy ratio)
- For secured residential property, banks should hold around 35% of the total exposure * (adequacy ratio)

- The more precise banks estimate expected -- Less capital is needed to hold -- more new can generate

- IRB approaches
  - Allows bank to establish their own credit rating
  - Precise calculations about the held capital for each individual exposure
  - Allocate resources to cover losses
  - Be more profitable

Characteristics of the data for individual clients

Biological info	External data	Characteristic of data
Age Sex Marital status Education	Credit rating No. of recent enquiries	Fix term consumer loan Life-span of product Purpose of the loan Interest rate

oaches to
s of expected

rms)

osure * CAR

CAR (Capital

business bank

Income Zip-code		

- Credit limit utilization - the proportion of money spent on the credit card

Data availability
- Most of the variables are available at time of application
- Many of the variable/information is available after loan is granted & behaviour of the loa be obtained for long sufficient

Application models	Behaviour models
Most of the variables are available at time of application	- Most of the variables are available at time of applicati - Many of the variable/information is available after loa behaviour of the loan borrower can be obtained for lo

- Behaviour model is used to calculate probability of default or expected loss after loan is
- If a customer holds credit card, banks can use behaviour model to grant or reject loan ap

Understanding the data
Dependent variables
Independent variables - predictor/features

PD model : Logistic regression
- Non-statistically savvy user example: front office workers, present in a simplified manne cards

LGD model: Beta regression
- How much loan has been recovered after the client defaulted

EAD model: Beta regression

Different data types
- Discrete
- Continuous

Based on data types data-preprocessing technique varies

Distinctive feature of PD model is --- all the independent variables have to be categorical - the is much easier to present a model in a simplified form and turn into score card - we transform independent variable into categorical variable/dummy variable an borrower can

ion an is granted & ong sufficient

granted
plication

r such as score

reason is that it
all the

Fine Classing: We slice the independent variable into equally sized intervals or classes
Coarse classing: we find how well each of the intervals discriminate b/w defaulted and non-de
adjacent discrete variables discriminate each other very well, then we can merge them.

PD Model

- Established dependent variable
- Binary
  - 0 = Bad loan
  - 1 = Good loan
- "Default definition": If a borrower is more than 90 days past due on a loan
- Statistical methodology to model PD is a logistic regression where dependent variable is customer is defaulter or not
- Logistic regression estimates the relationships b/w 2 things
  - Odds of an outcome (dependent variable) --- linear combination (Independent ∑va predictors
- Logistic regression
  - $\ln(\text{Non-default/default}) = \sum \beta j X_j = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots \beta_n X_n$

- "**Weight of evidence**" - The ability of each category to predict the dependent variable.
  - To what extent an independent variable would predict a dependent variable
  - ln(%Good/%Bad)

Variable Categories	Good	Bad	Proportion of good	Proportion of bad	Weight of evidence
Higher education	4000	600	4000/16000 =25%	600/4000 = 15%	ln(25/15) = 0.51
No Higher education	12000	3400	12000/16000 = 75%	3400/4000 =85%	ln(75/85) = -0.13
Total	16000	4000			

- "**Information value**" - shows how much information the original independent variable br explaining the dependent variable

Range 0 - 1	Predictive power
IV < 0.02	No predictive power
0.02 <= IV =< 0.1	Weak predictive power
0.1 <= IV =< 0.3	Medium predictive power

faulted. If two

whether a

ariable) of

rings with to

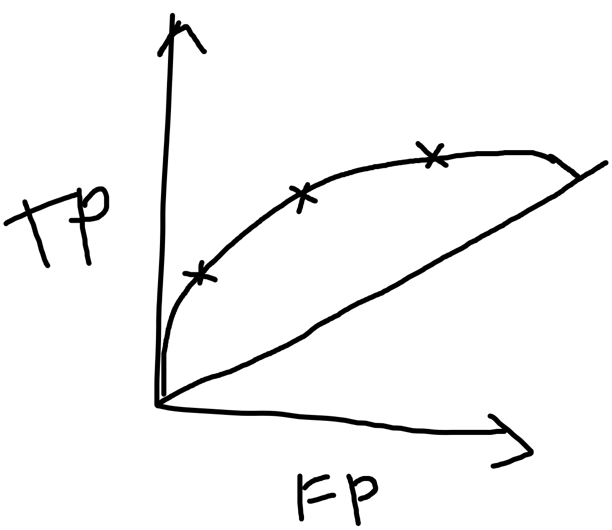
0.3 <= IV =< 0.5	Strong predictive power
0.05 < IV	Suspiciously high, too good to be true

Variable Categories	Good	Bad	Proportion of good	Proportion of bad	Weight of evidence	%good-% bad	Infor
Higher education	4000	600	4000/16000 =25%	600/4000 = 15%	ln(25/15) = 0.51	0.25-0.15 = 0.1	0.51 =0.0
No Higher education	12000	3400	12000/16000 = 75%	3400/4000 =85%	ln(75/85) = -0.13	0.75-0.85 = -0.1	0.13 =0.0
Total	16000	4000					0.06 pred

Overfitting: A substantial issue we might face when statistical model has focused on a pa so much that it missed the point

Underfitting: is under fitting when model fails to capture the underlying logic of data tha learn well so it doesn't know what to do and therefore it provides inaccurate answers

ROC - Receiver operating curve



fpr, tpr, thresholds = roc_curve(df_actual_predicted_probs['loan_data_targets_test'], df_actu

rmation Value
*0.1 511
*0.1 13
4 = weak ictive power

articular dataset

at is it didn’t

ual predicted probs['y hat test proba'])

fpr, tpr, thresholds = roc_curve(df_actual_predicted_probs['loan_data_targets_test'], df_actu
Here we store each of the three arrays in a separate variable.

Interpretation	Area under the curve
Bad	50-60%
Poor	60-70%
Fair	70-80%
Good	80-90%
Excellent	90-100%

from sklearn.metrics import roc_curve, roc_auc_score
AUROC = roc_auc_score(df_actual_predicted_probs['loan_data_targets_test'], df_actual_pre
Gini = AUROC *2 -1

Population Stability Index

Values of PSI: 0-1	Population difference
PSI = 0	No difference
PSI < 0.1	Little no difference
0.1 > PSI > 0.25	Little difference (no action is taken)
PSI > 0.25	Big difference (Action is taken)
PSI = 1	Absolute difference

LGD and EAD model

Dependent variables
RECOVERY RATE = amt_recovered/total_fund_amt
CREDIT CONVERSION FACTOR = (total_fund_amt - total_rec_principal) / total_fund_amt

al_predicted_probs['y_hat_test_proba'])

dicted_probs['y_hat_test_proba'])