# Credit Scoring

## Introduction

* Banks have lots of data about customers
* Banks also have lots of data about customer performance
* It is now the aim of credit scoring to analyze both sources of data into more detail and come up with a statistically based decision model which allows to score future credit applications and decide which ones to accept or reject.
* A key assumption which is made when building a credit scoring model is that the future resembles the past.
* To summarize the goal of credit scoring is to estimate whether an applicant will successfully repay the loan based on applicant characteristics such as age, income, employment status, time at address, etc
* Credit bureaus are data pooling organizations that gather default information from various financial institutions information such as delinquency history. Bureau checks and bureau scores can be very useful for credit scoring.
* There are two main approaches to assessing credit risk: the judgmental approach and the statistical approach.
  + The judgmental approach is a qualitative expert based approach whereby based on business experience and common sense the credit expert or credit committee, which is a group of credit experts, will make a decision about the credit risk.
  + Usually this is done based upon inspecting the five C's of the applicant and loan. The five C's being the character of the applicant, the capital of the loan, whether there was any collateral provided yes or no, the payment capacity of the applicants, and the conditions of the loan.
  + The statistical approach is based upon statistically analyzing historical data to find the optimal multivariate relationship between a customer's characteristics and the binary good/bad target variable.
  + It is less subjective than the judgmental approach since it is not tied to a particular credit expert's background knowledge and experience.
* Note that both approaches assume that the future will resemble the past

## Retail Credit Scoring

### Application Scoring

* The purpose of application scoring is to come up with a credit score which reflects upon the default risk of a customer at the moment of loan application
* In order to build an application scorecard one first needs to define a concept of default.
  + in the earlier days of credit scoring: a customer was more than three months of payment arrears.
  + Basel capital Accords: 90 days in payment arrears
  + In the United States in retail credit for residential mortgages, the default definition is 180 days
  + For qualifying revolving exposures, it's also 180 days and for other retail exposures, it's 120 days.
* application variables
  + date of birth or age, gender, income marital status, years living at current address, employment status etc.
  + can be complemented by bureau variables obtained from credit bureaus which are external to the bank.
* Credit bureau
  + an organization that assembles and aggregates credit information from various financial institutions or banks.
  + It can collect both positive or negative credit information, depending upon the country in which it operates.
  + two sources of information.
    - raw bureau data
    - total amounts of credit that a customer has outstanding, whether the customer has already been late with some payments at other financial institutions (yes or no), number of credit checks that have been made for a particular customer at other financial institutions, etc.
    - Using this raw bureau data credit bureaus can now build bureau credit scores.
    - These bureau scores can then be sold to interested counterparties and used into their application scoring models.
  + In the US, you have very popular bureaus like Experian, Equifax and TransUnion. which each cover their own geographical region.
  + All three provide a FICO score which ranges between 300 to 850 with higher scores reflecting better credit quality.
  + In Australia, there is BayCorp advantage,
  + Germany has the Schufa,
  + Netherlands BKR,
  + Belgium CKP.
  + Dun and Bradstreet is a popular credit bureau targeting the mid corp and small and medium-sized enterprise market.
* application scorecard
  + Each category has points assigned to it.
  + The more points the higher the credit quality.
  + points represent the total credit quality of the customer.
  + This now needs to be compared against the cutoff, the minimum required criticality set by the bank.
* decisions that need to be made during the scorecard development process.
  + how do we select the characteristics age, known customer and salary?
  + Why don't we include other ones like employment status, number of years living at current address, etc?
  + Why do we categorize age into four categories?
  + How do we decide upon the points assigned to each category?
  + Why do we set the cutoff at 500?
* When building an application scorecard, you are actually taking two snapshots of customer behavior.
  + The first snapshot is taken at loan origination where you will gather both the application and credit bureau data.
  + The second snapshot is taken at some later point during the loan at which the default behavior will be determined.
  + Empirical analysis has shown that the majority of customers default in the first 18 months.
  + Hence, many firms take the second snapshot 18 months after the loan origination to see where our customers defaulted or not.
* Finally note that application scoring aims at ranking customers in terms of their default risk.
  + Hence application scores can take on any value and no calibrated probabilities of default bounded between zero and one are needed for application scoring.

### Behavioral Scoring

* Behavioural scoring is another statistical credit scoring approach whereby we're going to look at our existing credit customers.
* That recent information could be my checking account behavior summarized by the average of my checking account balance, the maximum or minimum there off, or the trends during the previous 12 months.
* Other interesting information could be delinquency information like if I already incurred payment delays, etc Also changes in job status or home address could be considered.
* Behavioral scoring models are typically constructed using a 24 month timeframe.
* 12 months are taken to measure and quantify all the information which will be used as predictors and a subsequent 12 months to determine the default status.
* Behavioral scoring is dynamic since it summarises the behavior into various dynamic variables such as average checking account balance, maximum checking account balance, trend in checking account balance, etc
* Behavioral scorecards have different types of usage.
  + First, they can be used for debt provisioning and profit scoring.
    - As such they will prove to be a very important and valuable input for Basel capital calculation as we will discuss later.
    - Behavioral scores can also be used to authorize accounts to go in excess of their credit limit.
  + Behavioral scores can also be used to do cross selling of other products.
    - selling insurance products to a customer having a good behavioral score on his or her mortgage.
  + Finally they can be used for proactive debt collection.
* Behavioral scoring data sets typically have a few hundred of variables to consider.
* Just as with application scoring the aim of behavioral scoring is to provide a score which is as explained earlier for application scores a relative credit assessment allowing to rank our customers from low risk to high risk in terms of their default likelihood

## Corporate Credit Scoring

### Prediction Approach

* Assumes:
  + historical data is available about firms and their bankruptcy status
  + This data can then be analyzed by statistical techniques to predict corporate bankruptcy.
  + Usually accounting information such as balance sheet and financial statement ratios and stock price behavior if the firm is publicly listed.
* Altman's z model for manufacturing firms
  + built in the late 60s using a statistical technique called discriminant analysis.
  + Separate versions exist for public and private industrial companies.
  + the z-score is a linear combination of five accounting ratios:
    - working capital divided by total assets,
    - retained earnings divided by total assets,
    - earnings before interest in taxes divided by total assets,
    - market or book value of equity divided by total liabilities,
    - net sales divided by total assets.
  + A higher z-score reflects a more healthy firm and thus lower bankruptcy risk
    - Public industrial: gt 3, good. Lt 1.8, bad
    - Private industrial: gt 2.6 good. Lt 1.1 bad.
  + Extensions of the original z-score model have been provided for privately held and non manufacturing firms.
  + The z-score can be used by a bank as its internal bankruptcy prediction model. It can also be used to benchmark other bankruptcy prediction models.

### Expert-based Approach

* in the absence of data the expert based approach is also used in corporate credit risk modeling.
* builds a scorecard in a qualitative way using the business experience intuition and common sense of one or more credit experts.
* Expert based scorecards are often written down as a set of if-then business rules.
* Although they might seem inferior to statistical based scorecards at first sight they are still quite commonly used in the industry for specific corporate portfolios where no historical data is available.

### Agency Rating Approach

* agency ratings approach is an approach that can be adopted if none of these is available.
* Rating agencies are interesting partners to collaborate with in this case since they provide credit ratings for almost any type of corporate exposure.
* These ratings typically vary from AAA which represents excellent credit quality to AA, A and up to D which represents the default status.
* The ratings also come with default rates measured across different time horizons such as one, two, three or even five years.
* Banks can then purchase these credit ratings to score their corporate exposures.
* The most popular rating agencies are
  + Moody's Investors Service
  + Standard & Poor's
  + Fitch.
* ratings for
  + companies (both private and public),
  + countries and governments (sovereign ratings),
  + local authorities
  + banks.
* Retail exposures are typically not covered by the rating agencies.
* The methodology behind the rating assignment is obviously not disclosed but it is based upon a combination of both quantitative and qualitative modeling

### Shadow Ratings Approach

* starts from a set of ratings for a particular set of corporate obligors.
* next step: information will be collected which might have an influence on the rating.
  + accounting ratios
  + firm characteristics
  + stock price behavior
* then: use all this information as predictors in a statistical regression model to predict the ratings.
* Advantages
  + one obtains a white box understandable model which clearly indicates how the various characteristics of an obligor contribute to the rating.
  + gives clear advice to corporates on how to improve their rating furthermore in the long term
  + allows the bank to become independent from the rating agency since the internal statistical model can now be used to rate any obligor given its characteristics
* summarize: the shadow' ratings approach aims at building a statistical model mimicking the external ratings provided by a rating agency

## Quiz

## Discussion

In the US, three popular credit bureaus are Experian, Equifax and TransUnion that each cover their own geographical region. All three provide a FICO credit score which ranges between 300 to 850, with higher scores reflecting better credit quality.

Find out which five data sources are used to calculate FICO scores.

1. payment history (35%)
2. amounts owed (30%)
3. length of credit history (15%)
4. new credit (10%)
5. credit mix (10%)

Besides banks, what other businesses make use of FICO scores in the US?

Lenders, either installment lenders or branded card companies certainly do. Employers might?

## Discussion

Find out which of the following variables can be used for application scoring in your country of residence:

* ~~Age~~
* ~~Gender~~
* Income
* ~~Marital status~~
* Employment status

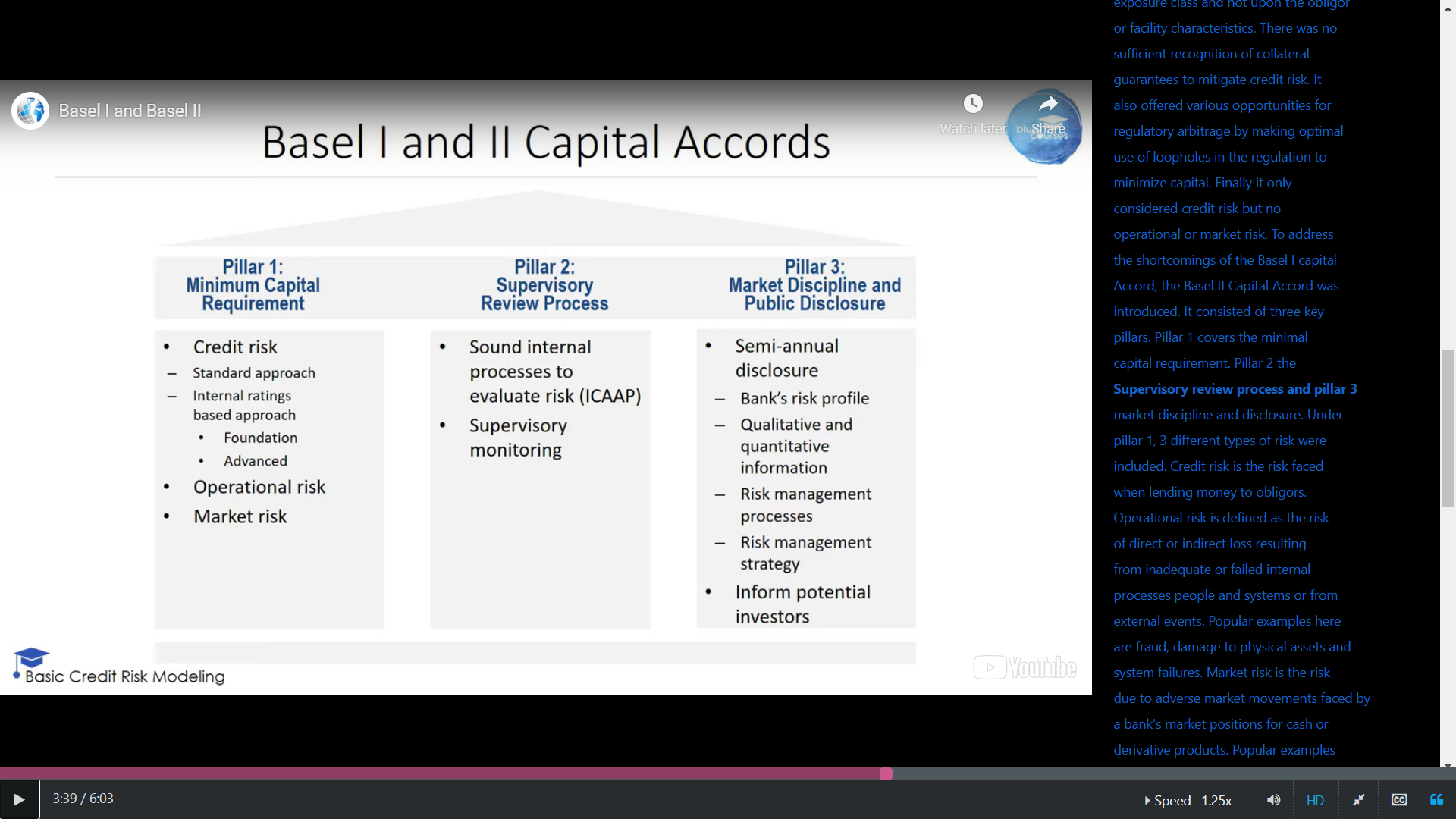
# Basel Accords / IFRS 9 / CECL

## Regulatory vs Economic Capital

* Bank cash and flow sources
  + Liability / equity
    - savings depositors who buy savings products from the bank.
      * savings accounts
      * term accounts
      * pension funds
    - shareholders or investors.
      * They buy shares which gives them an ownership relationship with the bank.
      * If the firm makes profit then part of it can be paid back to the shareholders as dividends.
  + Assets
    - the money obtained by making various investments.
      * Lending
      * various market securities such as bonds, stocks options, etc
* Note that these investments always have a risk associated with them.
  + Given the societal impact of banks in any economic system they need to be well protected against the risks they are exposed to.
  + the risks which banks take on their asset size should be compensated by appropriate liabilities.
  + To safeguard their savings depositors these people should be guaranteed to always get their savings money back whenever they wanted.
* Hence a bank should forsee enough shareholder capital as a buffer against losses.
  + a well capitalized bank has a sufficient amount of equity to protect itself against its various risks.
  + there should be a direct relationship between risk and equity.
* Usually this relationship is quantified in two steps.
  + First the amount of risk on the asset side is quantified into a specific number.
  + This number is then plugged into some kind of formula which precisely calculates the corresponding equity and thus capital buffer required.
* There are two views on finding both this risk number and the formula to be used.
  + The first view is a regulatory view whereby a regulation such as Basel I, Basel II and Basel III have been introduced to precisely define how to calculate the risk number and what formula to use.
    - Regulatory capital is then the amount of capital a bank should have according to a regulation.
  + Banks can use their own risk modeling methodologies to come up with a risk number and use their own formula to calculate the buffer capital.
* economic capital: the amount of capital a bank has based on internal modeling strategy and policy.
* actual capital: the amount of capital a bank actually holds.
* types of capital
  + Tier 1 capital typically consists of common stock, preferred stock and retained earnings.
  + Tier 2 capital is of somewhat less quality and is made up of subordinated loans, revaluation reserves, undisclosed reserves and general provisions.
  + The Basel I capital accord also included tier 3 capital which consisted out of short-term support subordinated debt but as we will discuss later this has been abandoned in the more recent Basel III capital afterwards.

## Basel I and II Capital Accords

* The Basel Accords <- Basel Committee on Banking Supervision, founded in 1974 by the Board of Governors of the G10 central banks
  + counts 27 members
  + meet approximately every three months at the Bank for International Settlements (BIS)in Basel Switzerland
* Basel I capital accord (1988)
  + Aim: minimum regulatory capital requirements in order to ensure that banks are able at all times to return depositors funds.
  + predominantly focused on credit risk
    - Cooke Ratio — named for Peter Cooke of the Bank of England: the ratio of commitments (assets weighed by the risk of default) to total assets. Also known as the solvency ratio, Basel ratio, and capital ratio, it was first established by the Basel Commission in its 1988 accord.
    - Cook ratio which is the ratio of the available buffer capital and the risk-weighted assets.
  + It put a lower limit on this ratio of 8%.
    - In other words the capital should be bigger than 8% of the risk weighted assets.
    - capital both tier 1 and tier 2
  + introduced fixed risk weights depending upon the exposure class.
    - cash exposures: risk weight was 0%
    - mortgages: 50%
    - other commercial exposures:100%
  + drawbacks
    - the solvency of the debtor was not properly taken into account since the risk weights only dependend upon the exposure class and not upon the obligor or facility characteristics.
    - There was no sufficient recognition of collateral guarantees to mitigate credit risk.
    - It also offered various opportunities for regulatory arbitrage by making optimal use of loopholes in the regulation to minimize capital.
    - Finally it only considered credit risk but no operational or market risk.
* Basel II
  + To address the shortcomings of the Basel I
  + It consisted of three key pillars.
    - Pillar 1 covers the minimal capital requirement.
    - Pillar 2 the Supervisory review process and
    - pillar 3 market discipline and disclosure.
  + Pillar 1
    - 3 different types of risk were included
      * Credit risk is the risk faced when lending money to obligors.
      * Operational risk is defined as the risk of direct or indirect loss resulting from inadequate or failed internal processes people and systems or from external events.
        + Popular examples here are fraud, damage to physical assets and system failures.
      * Market risk is the risk due to adverse market movements faced by a bank's market positions for cash or derivative products.
        + Popular examples here are equity risk, currency risk, commodity risk and interest rate risk.
    - credit risk
      * The standard approach
      * the foundation internal ratings based approach
      * the advanced internal ratings based approach
    - All quantitative models built under pillar 1 will need to be reviewed by overseeing supervisors.
  + Pillar 2
    - the introduction of sound processes to evaluate risk
      * the internal capital adequacy assessment process (ICAAP)
      * the Supervisory monitoring.
    - once all quantitative risk models have been approved they can be disclosed to the market.
  + Pillar 3
    - bank will typically semi-annually disclose its risk profile
    - provide qualitative and quantitative information about its risk management processes and strategies to the market.
    - The objective here is to inform the investors and convince them that the bank has a sound and solid risk management strategy which hopefully will result into a favorable rating for the bank such that the bank can attract funds at cheap rates.



## Basel III

* Introduced because of the 2008 credit crisis
* standards took effect between January 2013 and January 2019
* builds further upon the Basel II Accord but tries to further strengthen global capital standards
  + a bigger focus on tangible equity capital since this is the component with the greatest loss absorbing capacity
  + reduced the reliance on models developed internally by the bank and ratings obtained from external rating agencies
  + It also put a greater focus on stress testing for systemically important banks
* It puts a greater focus on tier 1 capital consisting of shares and retained earnings by abolishing the tier 3 capital introduced in Basel II as it was deemed of insufficient quality to absorb losses
* It introduced a risk insensitive leverage ratio as a backstop to address model risk
* includes some facilities to deal with procyclicality whereby due to a too cyclical nature of capital economic downturns are further amplified
* introduced a liquidity coverage and net stable funding ratio to satisfy liquidity requirements

The tier 1 capital ratio was 4% of the risk-weighted assets in the Basel II capital Accord. It was increased to 6% in Basel 3 to be implemented by 2015.

The common tier 1 capital ratio whereby common tier 1 consists of common equity, which is common stock and retained earnings but no preferred stock, was 2% of the risk weighted assets in Basel II and 4.5 percent of the risk weighted assets in Basel 3 which had to be implemented by 2015.

A new capital conservation buffer was introduced which was set to 2.5 percent of the risk weighted assets to be covered by common equity by 2019.

Also a counter-cyclical capital buffer was added ranging between 0 to 2.5 percent of the risk weighted assets which had to be implemented by 2019.

As already mentioned a non-risk based leverage ratio was introduced which should be at least 3 percent of the assets and covered by tier 1 capital.

Very important to note here is that we look at the assets and not risk weighted assets as with the previous ratios. The assets also include off-balance sheet exposures and derivatives. The idea here is to add this ratio as a supplementary safety on top of the risk-based ratios in Basel II. The total capital ratio is the sum of the tier 1, tier 2, and tier 3 capital ratio. In Basel III, the capital ratio then becomes the sum of the tier 1 capital ratios, the capital conservation buffer, the counter-cyclical capital buffer and if relevant an additional capital ratio for systemically important banks.

## Basel IV

Basel IV is the informal name for a set of proposed banking reforms building on the international banking accords known as Basel I, Basel II, and Basel III. It is also referred to as Basel 3.1. It is scheduled to begin implementation on Jan. 1, 2023

It started from the following observation of a G20 report published in 2014: the studies confirmed that there are material variances in bank's regulatory capital ratios that arise from factors other than differences in the riskiness of bank's portfolios. These variances undermine confidence in capital ratios.

The accords aim to strengthen the international banking system by standardizing rules from country to country, including those relating to risk.

## Basel Approaches to Model Credit Risk

### Standardized Approach

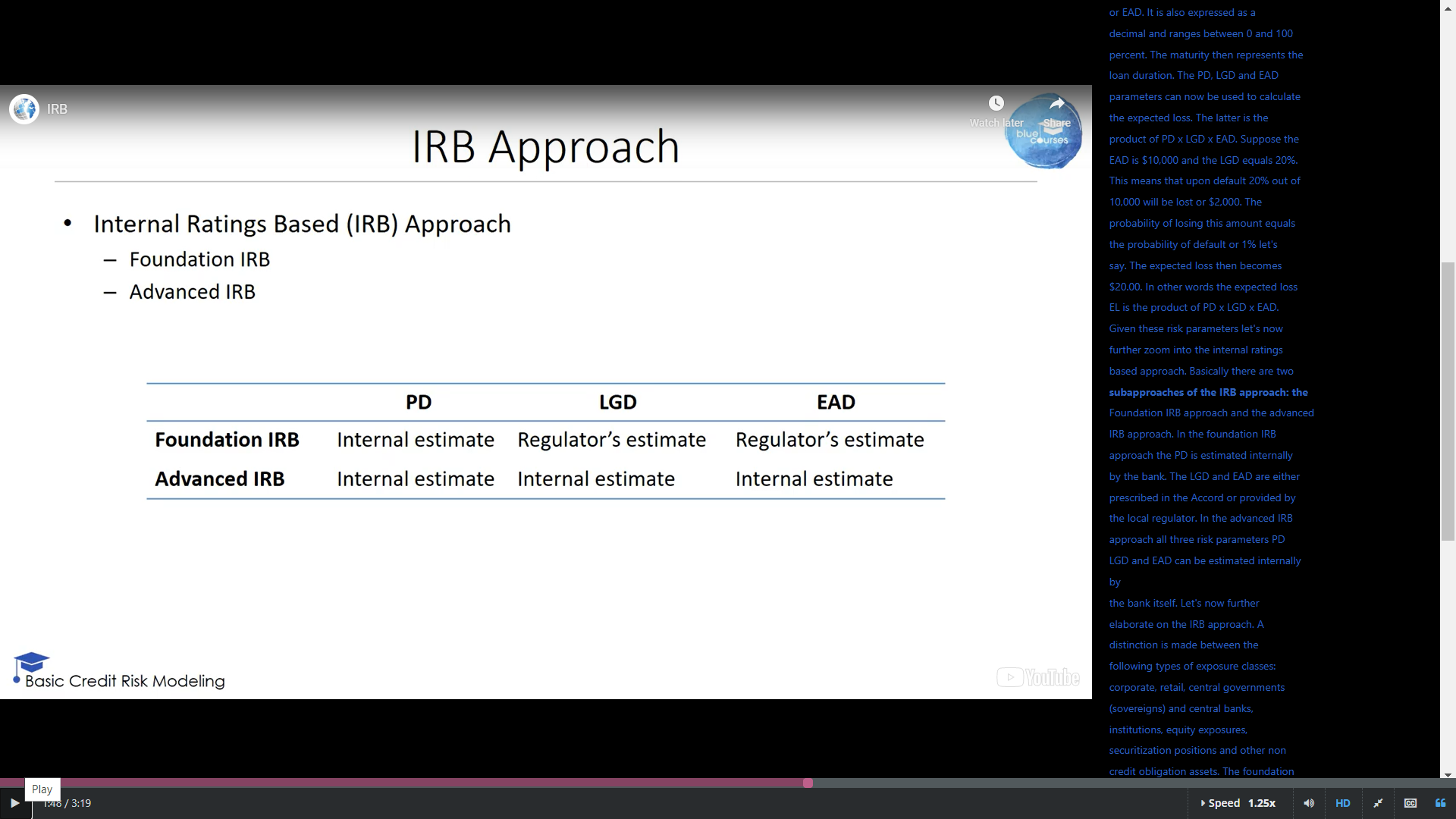
Let's first discuss the standardized approach for non retail exposures. This approach relies on external credit assessment institutions or ECAIs to provide credit ratings.

Popular examples of ECAIs ar Moody's, Standard & Poor's and Fitch.

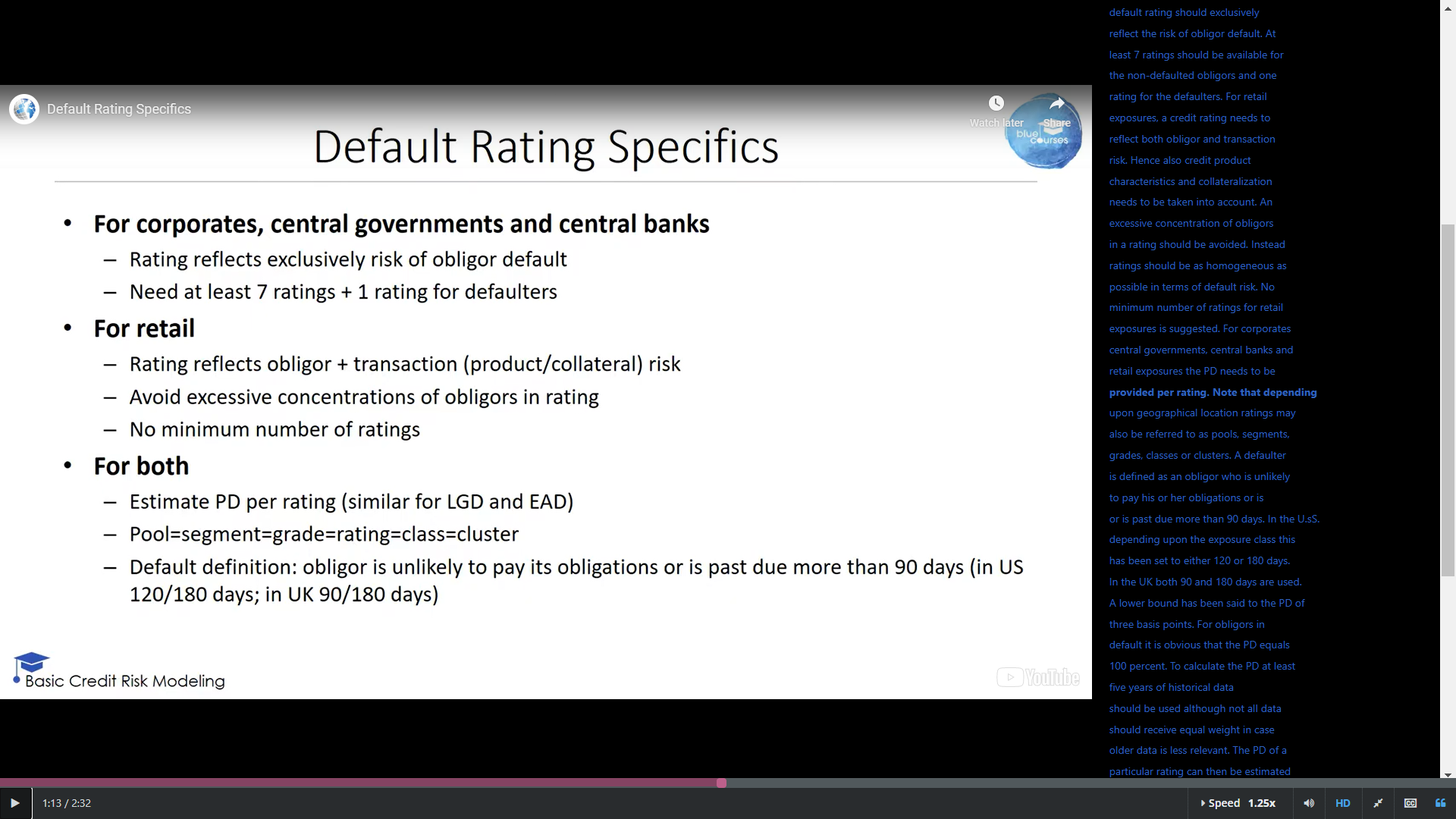
Risk weights are provided for sovereigns, banks, corporates and other exposures. The capital itself is then calculated using the formula capital equals eight percent of the risk weighted assets. Capital = RWA \* 8%.

retail exposures are only discriminated in terms of mortgage or non mortgage. a more detailed categorization is highly desirable here. Ideally, every obligor should have his or her own risk profile whereby not only default risk is considered but also loss and exposure risk as measured by LGD (Loss Given Default) and EAD (Exposure at Default).

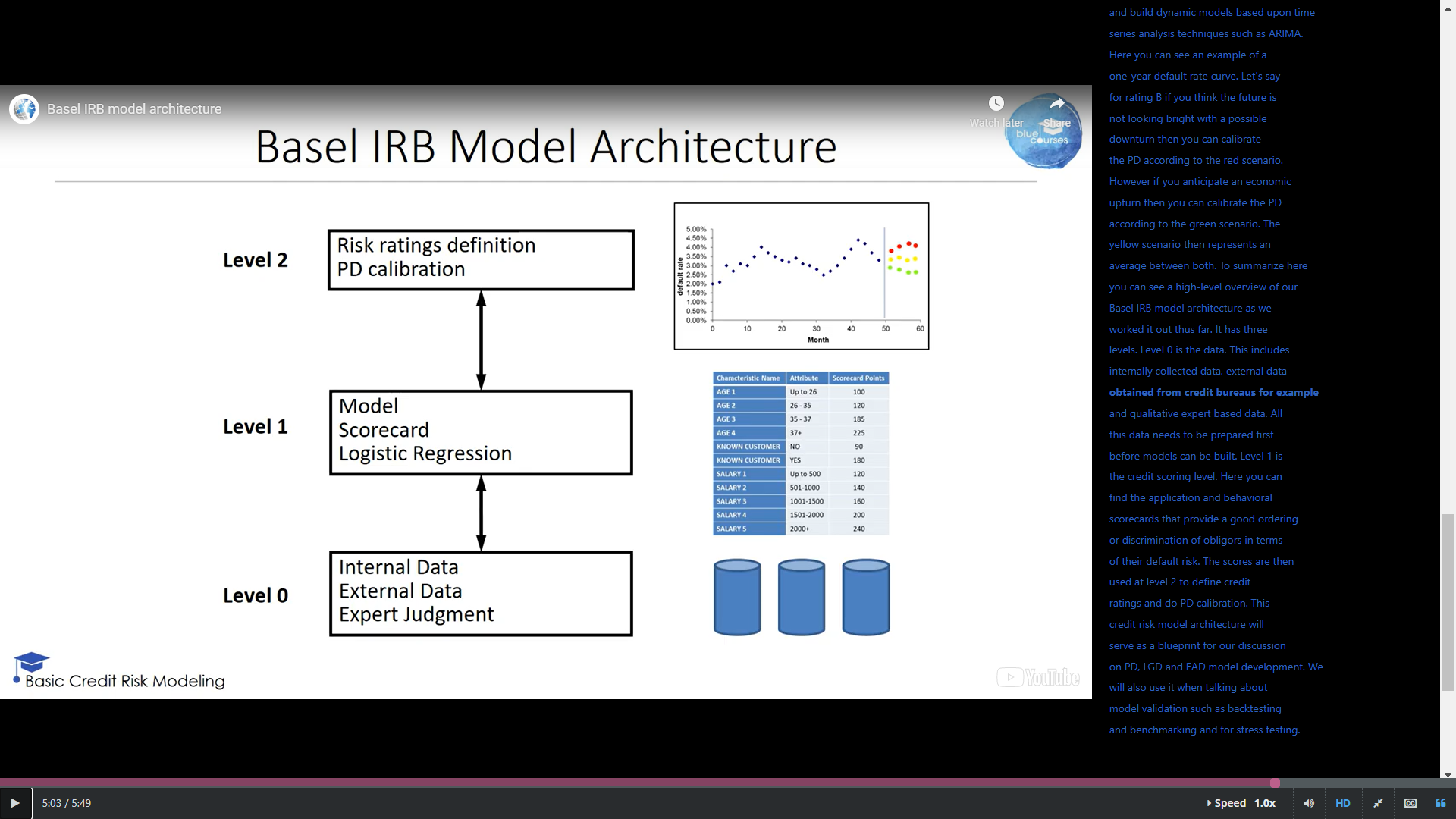
### IRB Approach



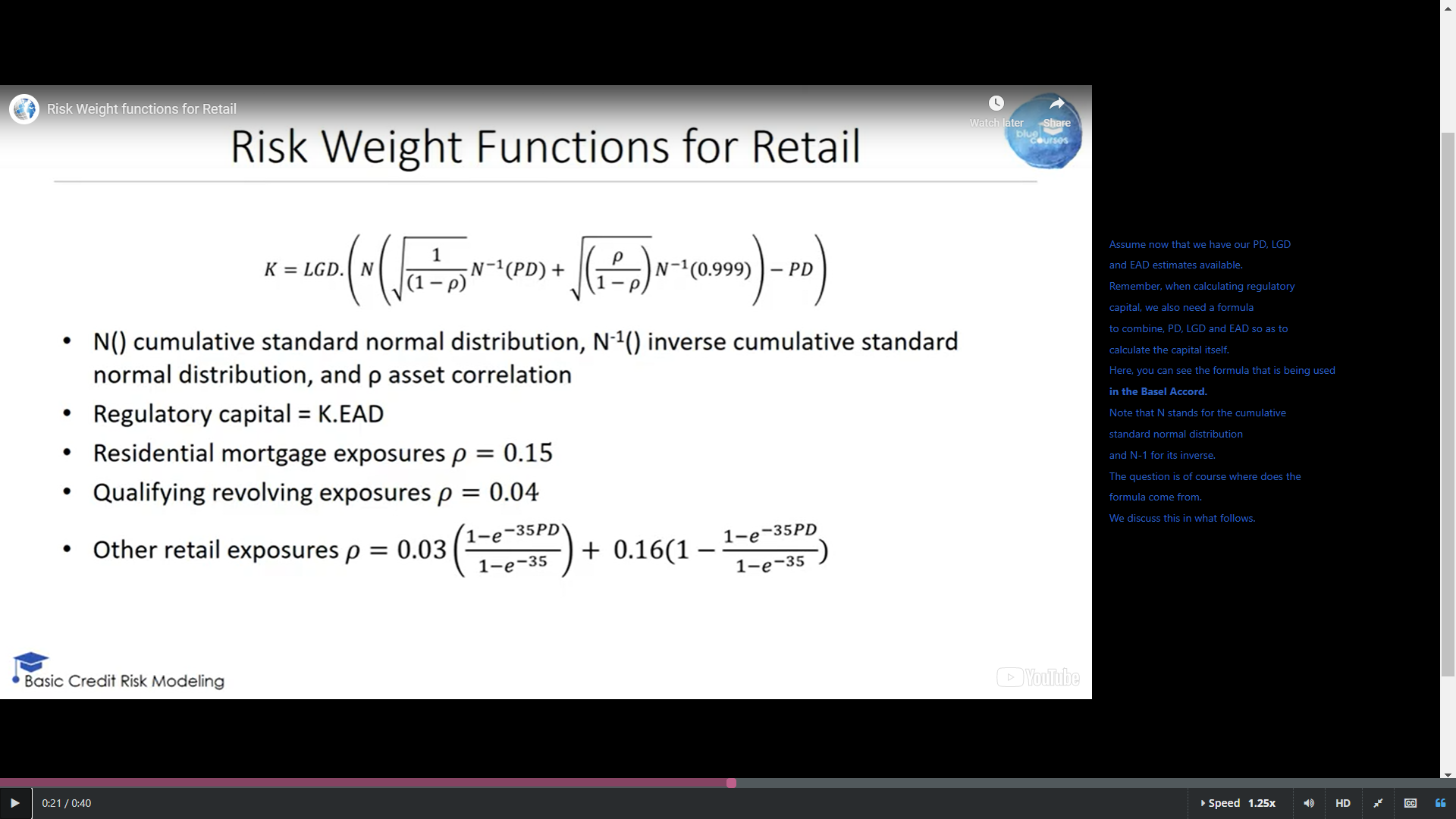
#### Default Rating Specifics



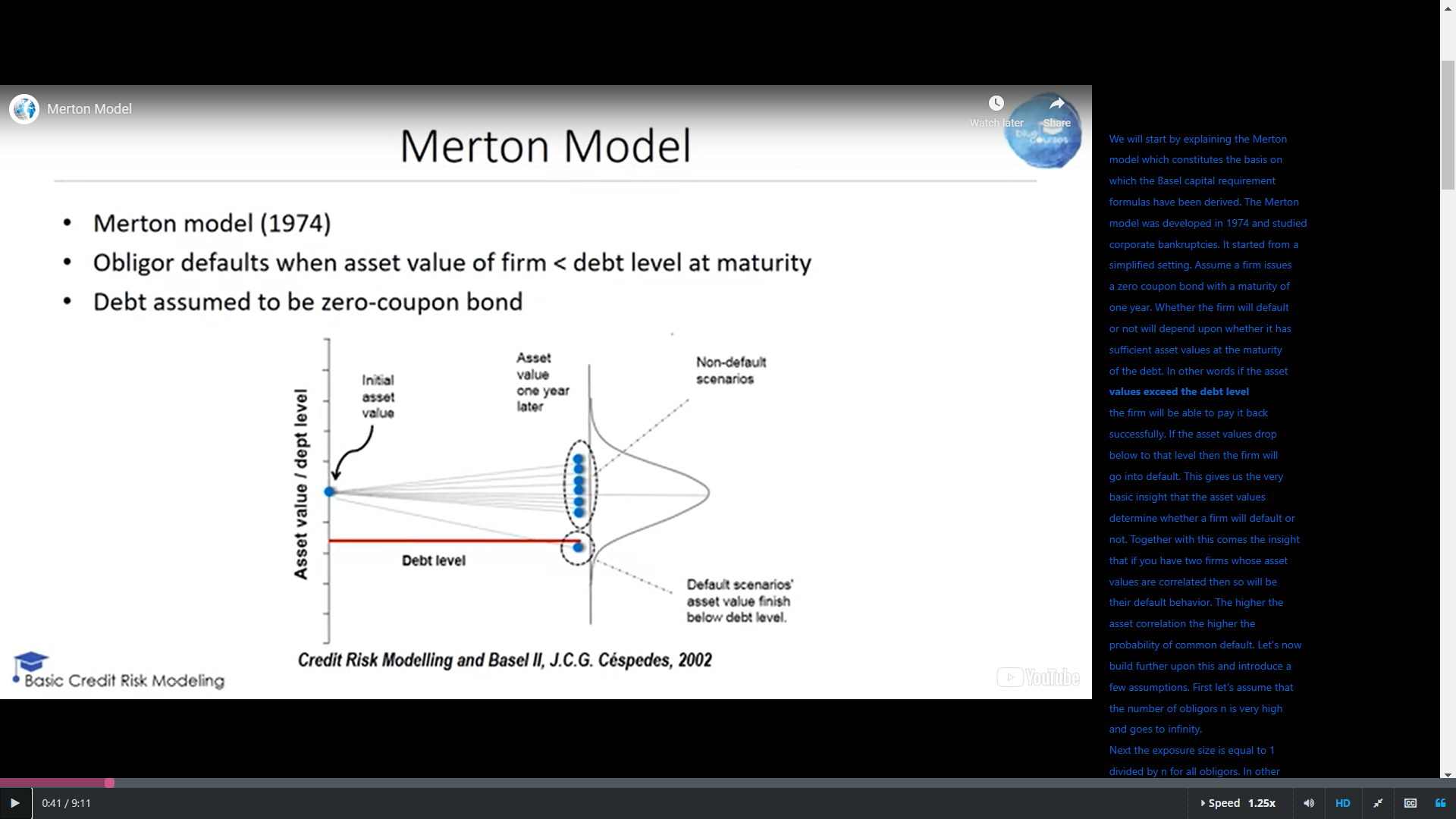
#### Basel IRB Model Architecture

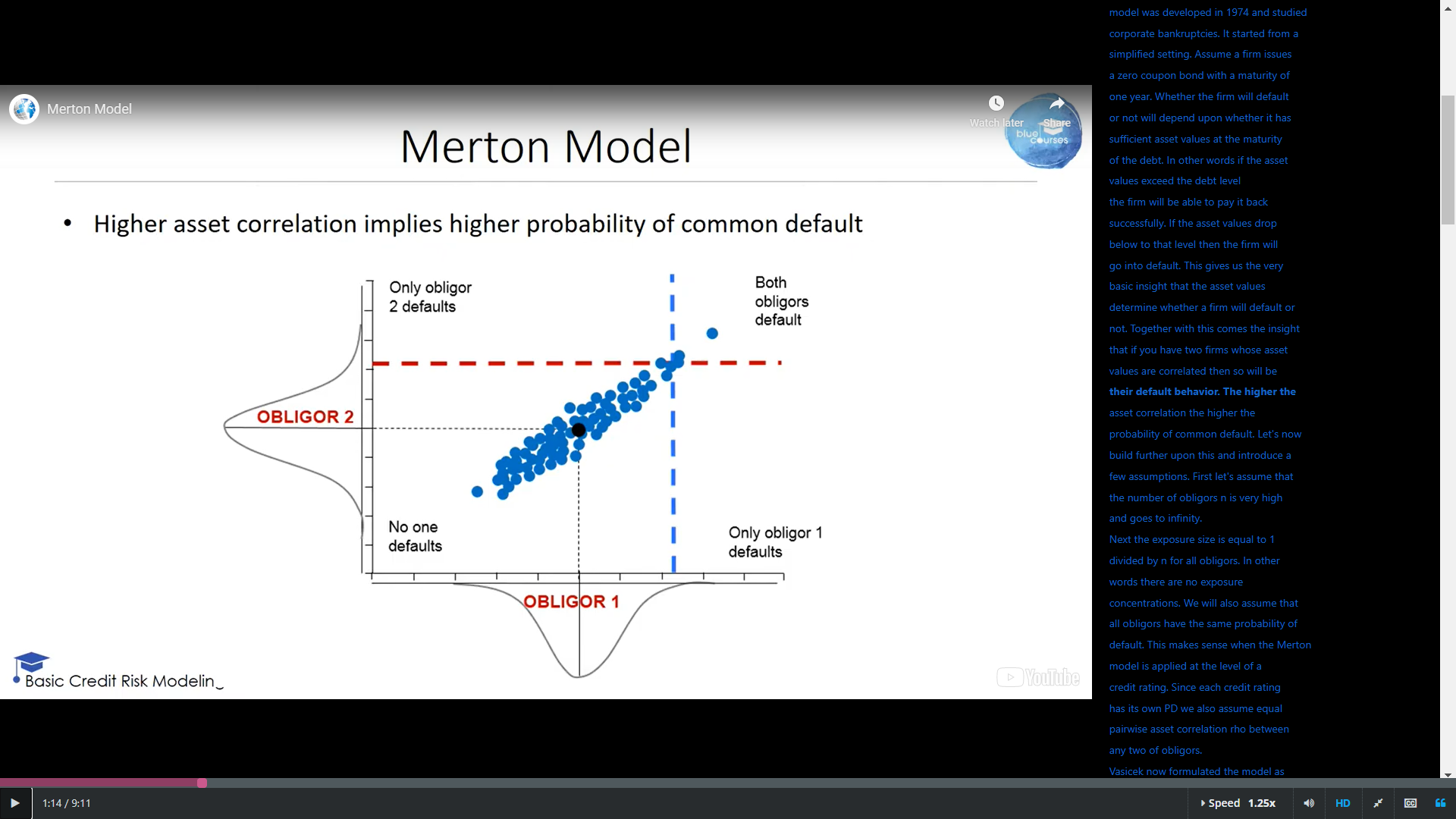


#### Risk Weight Functions for Retail

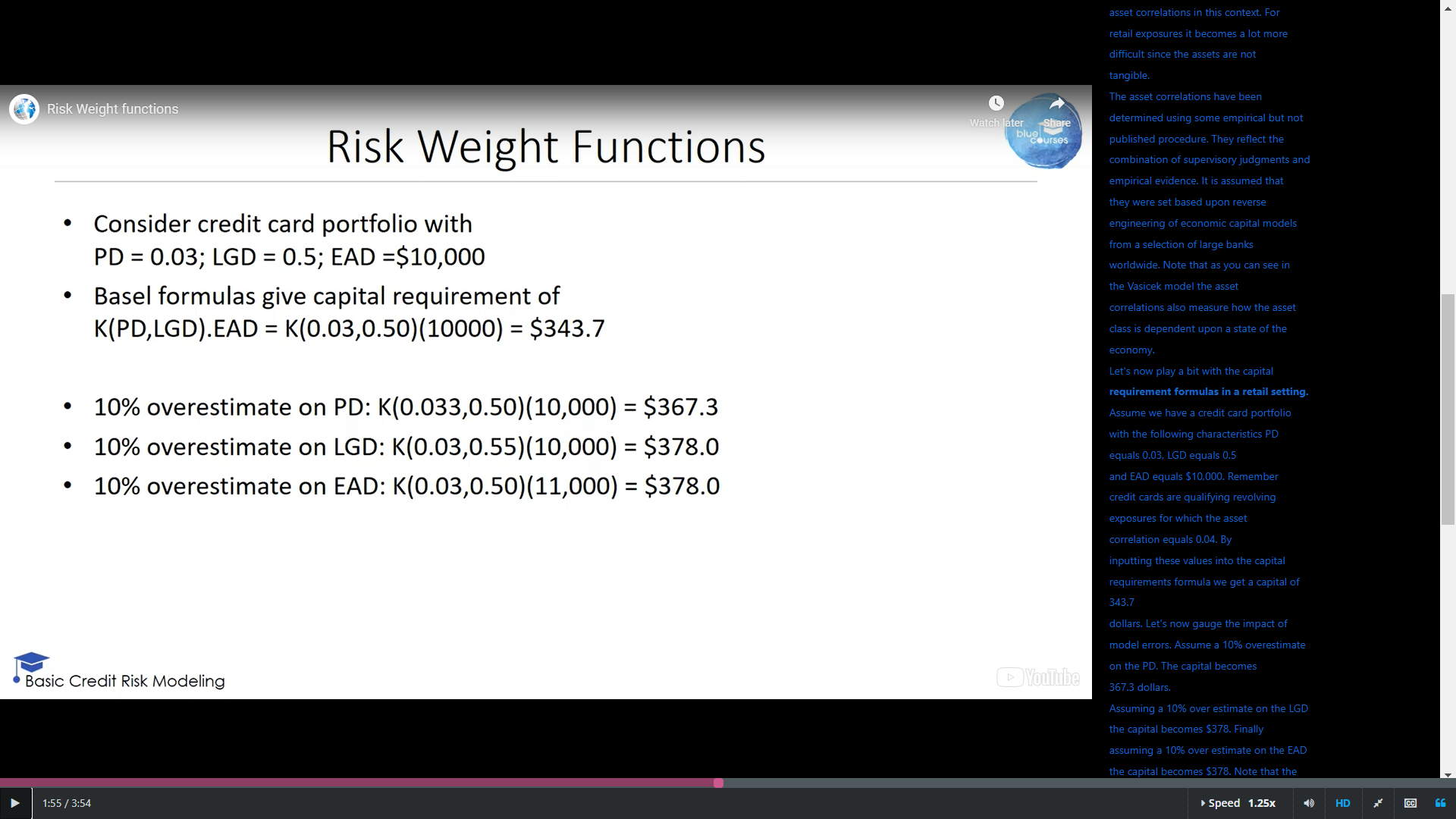


#### Merton Model

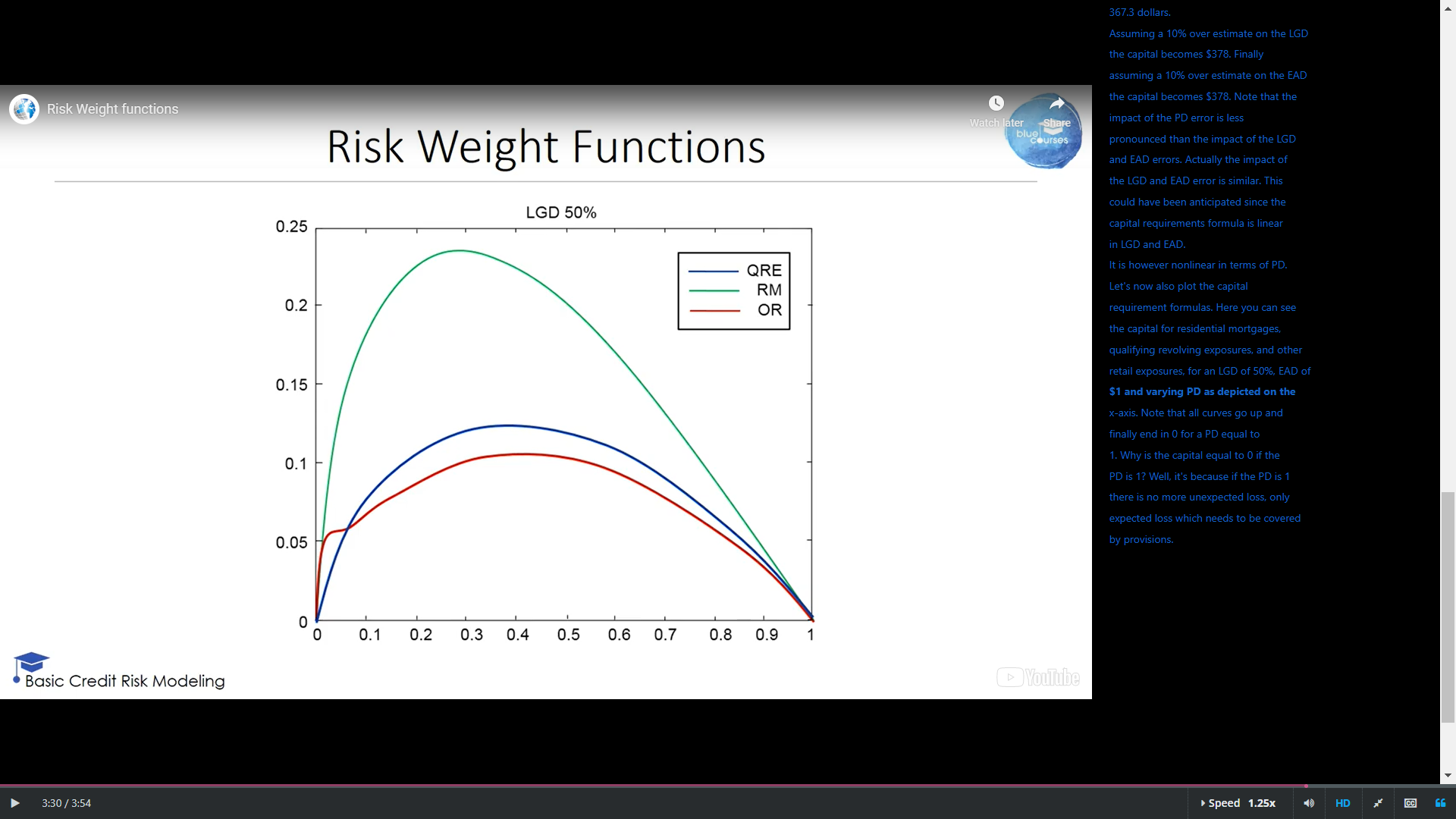




#### Risk Weight Functions



Because of the way the risk weight formulas work, they are "linear" with respect to LGD and EAD, but non-linear with respect to PD.



PD

capital

## IFRS 9

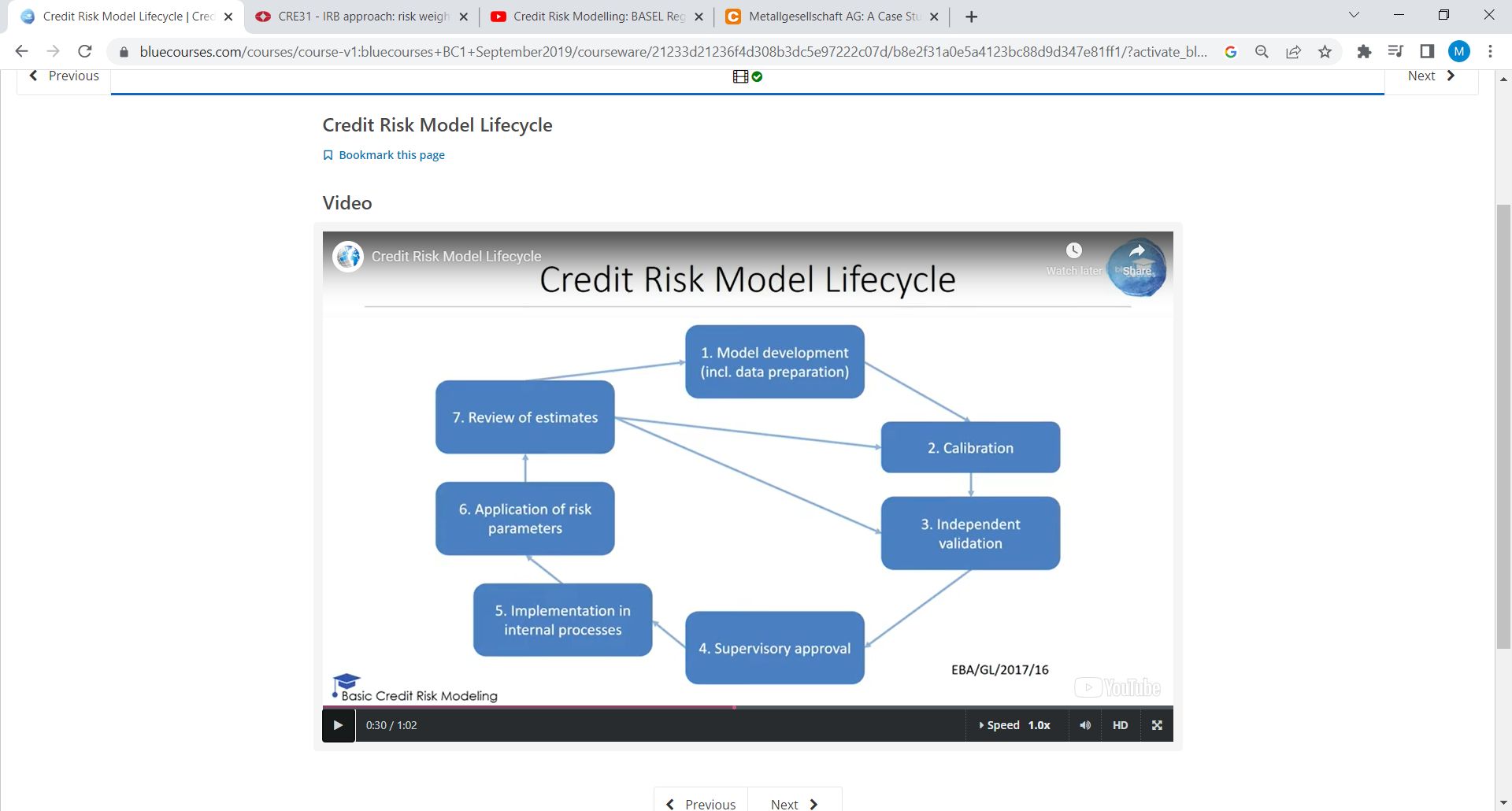
IFRS 9 recognizes three stages of credit risk

1. credit risk not increased significantly since initial recognition; 12 month PD
2. credit risk has increased significantly since initial recognition; lifetime PD
3. credit is impaired; lifefime PD

Basel vs IFRS 9

|  |  |
| --- | --- |
| Basel | IFRS 9 |
| Expected + Unexpected Loss | Expected Loss |
| Default definition (90/180 days in arrears | No default definition |
| One-year PD | I year for stage 1, lifetime for stages 2 & 3 |
| TTC rating philosophy (long-run avg PD) | PIT rating philosophy (foucs on reporting date) |
| Downturn LGD (direct + indirect costs) | Best estimate LGD (direct costs) |
| Downturn EAD | Best estimate EAD |
| EL = PD\*LGD\*EAD | EL = PD\*PV of cash shortfallss |
| Conservative calibration |  |
| Regulatory PD/LGD floors |  |

## Credit Risk Model Lifecycle



# Data Preprocessing

## Motivation

Because data can be crap

## Types of data

This and that

## Denormalizing Data

You start with a normalized database, and turn it into an analytic table. Denormalized.

## Sampling

a key requirement for a good sample is it should be representative for the future entities on which the credit risk model will be run.

The sample should also be taken from an average business period to get an accurate picture of the target population

the performance window is long enough to stabilize the default rate

In application scoring one commonly adopts an eighteen-month performance window.

Various methods for reject inference have been suggested.

1. A first method to do reject inference is to classify all rejects as bad payers.
2. A softer version can also be adopted whereby only a subsample of the rejects is classified as bads based upon expert knowledge.
3. Another way to get more information about the rejects is via the credit bureau
4. You can also analyze the rejects with the analytical model you built on the accepts as we discuss later

you will see that you also have withdrawals in there

The synthetic minority oversampling technique or SMOTE is another interesting approach to deal with skewed class distributions. It oversamples the minority class by creating synthetic examples.

1. In step one of SMOTE, for each minority class observation the k nearest neighbors are determined in Euclidean sense for example. Let's say we set k equal to one and look at the nearest neighbor.
2. Step two then generates the synthetic examples as follows.
   1. Take the difference between the variables of the current minority sample and those of its nearest neighbor.
   2. Multiply this difference with a random number between zero and one and add it to the sample.

Throughout our research we have found that this SMOTE method works really good when dealing with skewed class distributions

## Visual Data Exploration

## Descriptive Statistics

## Missing Values

keep, delete and replace

## Outliers

Too far away is then quantified as more than 1.5 times the interquartile range IQR which equals Q3 minus Q1.

A practical rule of thumb then defines outliers when the absolute value of the z-score is bigger or equal than three.

A popular scheme is truncation, capping or winsorizing

## Categorization

Categorization is also known as coarse classification, classing, binning and grouping

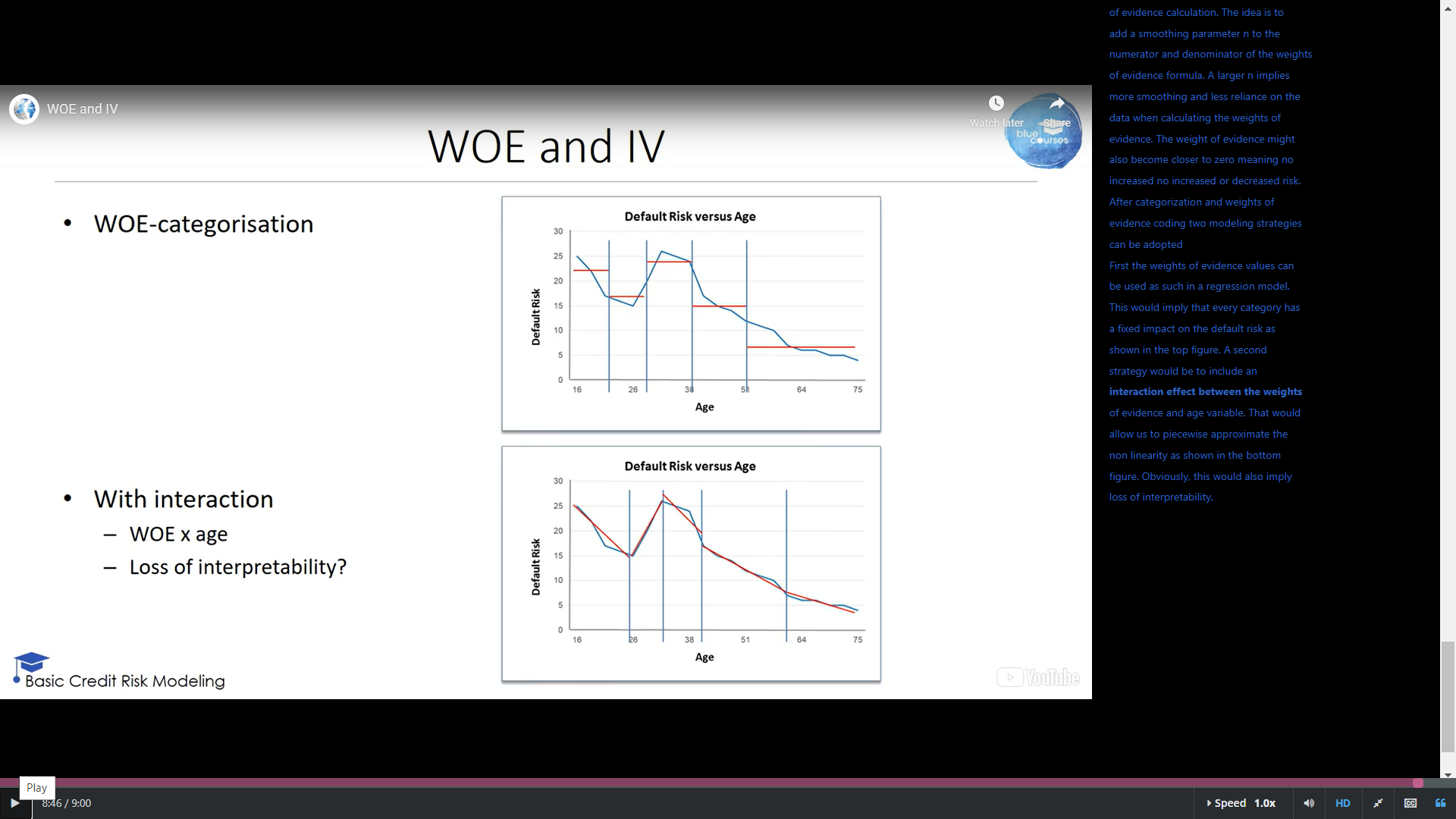
The goal of using categorization for categorical variables is to reduce the number of categories to a manageable size.

five popular methods for categorization:

1. Binning
   1. Equal range-sized bins
   2. Equal frequency bins
2. pivot tables
   1. We can now perform categorization by grouping purpose values with similar odds
3. chi squared analysis
   1. A more sophisticated method for doing coarse classification is based on chi squared analysis.
   2. Take a potential grouping, and get the goods and the bads and the total
   3. Calculate what the goods and bads would be if equal to global odds
   4. Add up a squared diff over the actuals. What a mess
   5. we want to have the empirical frequencies as different as possible from the independence frequencies
   6. The bigger this value (the sum of part d above), the bigger the dissimilarity between both tables and thus the more dependence there is between the good/bad variable and the residential status variable
4. business knowledge
5. decision trees

## WOE and IV

1. The weights of evidence is defined as the logarithm of the proportion of goods in a category divided by the proportion of bads within the same category
2. if the weights of evidence is bigger than zero it means that the input to the logarithmic transformation is bigger than 1 or that there are more goods than bads within the category
3. The information value is the sum across all categories of the product of the difference between the distribution of goods and bads multiplied by the weights of evidence
   1. the product tells us something about the absolute difference between the distribution of goods and bads in a particular category
   2. Hence the higher the value of the information value the more predictive the variable is since it gives us more information about the target
4. Interpretation
   1. if the information value is less than 0.02 the variable is considered unpredictive
   2. If the information value is between 0.02 and 0.1 the variable has weak predictive power
   3. If the information value is between 0.1 and 0.3 the variable has medium predictive power
   4. If the information value is above 0.3 the variable has strong predictive power
5. After categorization and weights of evidence coding two modeling strategies can be adopted
   1. First the weights of evidence values can be used as such in a regression model. This would imply that every category has a fixed impact on the default risk as shown in the top figure.
   2. A second strategy would be to include an interaction effect between the weights of evidence and age variable. That would allow us to piecewise approximate the non linearity as shown in the bottom figure. Obviously, this would also imply loss of interpretability.



### WOE and IV in R

install.packages('Information')

library(Information) # Information package author is Kim Larsen

hmeq <- read.csv("c:/temp/hmeq.csv")

IV <- create\_infotables(data=hmeq, y="BAD")

print(head(IV$Summary))

MultiPlot(IV, "LOAN")

### WOE and IV in Python

import numpy as np

import pandas as pd

hmeq = pd.read\_csv('c:/temp/hmeq.csv')

# Categorize variable in 10 bins based on distribution

hmeq["CLAGE\_cat"] = pd.qcut(hmeq.CLAGE, q=10, labels=False)

# Run calc\_IV function from

# https://www.kaggle.com/puremath86/iv-woe-starter-for-python#Calculate-Information-Value

calc\_iv(hmeq, "CLAGE\_cat", "BAD")

## Variable Transformation

Log

Box-cox

# Classification Techniques

## Classification

we discuss classification techniques that have practical relevance and fulfill the needs to build successful retail and corporate credit scoring models

## Linear Regression

* Linear regression estimates a linear relationship between the target variable Y representing the good/bad status and the predictor variables.
* In order to do so the target variable Y is represented in a numerical way: zero if it's a bad customer and one if it's a good customer.
* The beta parameters are then typically estimated using the idea of ordinary least-squares (OLS) whereby the parameters are estimated so as to minimize the sum of squared error terms.

Advantage: it usually works well and it's simple to interpret

Challenges:

* The first problem is that there is no guarantee that the target variable will have values between 0 and 1. In a classification problem such as credit scoring, it would be handy to have a target variable between 0 an 1 because we could then interpret it as a probability.
* A second problem is that the target variable and errors are not normally distributed but Bernoulli distributed which is also not that nice

Note however that despite both problems linear regression usually gives good performance for classification.

## Logistic Regression

* When combining both the linear regression with the bounding function one arrives at logistic regression
* Logistic regression is undoubtedly the most popular classification technique used for both retail as well as corporate credit scoring
* The odds are defined as the probability of good divided by the probability of bad
* the logistic regression model is almost linear in the odds
* To make it linear we just have to apply the logarithmic transformation
* In terms of the odds, the new odds are equal to the old odds multiplied by e to the power beta\_i. The latter is called the odds ratio. It is the multiplicative increase in the odds when xi increases with one unit assuming all other variables keep their values.
* Another way to analyze the impact of an individual variable is by calculating the doubling amount. This is the amount of change that is needed for doubling the outcome odds. It can be easily verified that this amount is equal to the logarithm of 2 divided by the beta coefficient of the variable.

## Nomograms

### Nomograms in R

 Bookmark this page

### Generate artificial data

# logit(pi) = 5 - 0.10 \* age - 0.05 \* income - 2 \* female + 0.005 \* debt

# Parameters:

n <- 3000   
beta0 <- 5  
beta1 <- -0.10  
beta2 <- -0.05  
beta3 <- -2  
beta4 <- 0.005

# Simulation:

library(dummies)  
library(dplyr)

age <- floor(runif(n, min = 25, max = 68))  
income <- round(runif(n, min = 9600, max = 104000)/1000,2) # in thousand €  
sex <- sample(x=c("male","female"),size=n, replace=TRUE, prob=c(1/2, 1/2))  
sex <- factor(sex,levels = c("male","female"))  
debt <- sample(c(round(runif(n\*0.8, min = 0, max = 500000)/1000,2),rep(0,n\*0.2))) # in thousand €  
linpred <- beta0 + beta1\*age + beta2\*income + beta3\*dummy(sex)[,-1] + beta4\*debt  
pi <- exp(linpred) / (1 + exp(linpred))    
y <- rbinom(n=n, size=1, prob=pi)    
mydata <- data.frame(x1=age, x2=income, x3=sex, x4=debt, y=y)  
colnames(mydata) <- c("age","income","sex","debt","default")

### Fit the logistic model

mod <- glm(default ~ age + income + sex + debt, family="binomial", data=mydata)

summary(mod)

# Visualize variables

library(sm)

mydata$default <- factor(mydata$default)  
sm.density.compare(mydata$age,mydata$default)  
legend("topright", levels(mydata$default), fill=2+(0:nlevels(mydata$default)))  
sm.density.compare(mydata$income,mydata$default)  
legend("topright", levels(mydata$default), fill=2+(0:nlevels(mydata$default)))  
sm.density.compare(mydata$debt,mydata$default)  
legend("topright", levels(mydata$default), fill=2+(0:nlevels(mydata$default)))  
barplot(table(mydata$default,mydata$sex), legend = rownames(table(mydata$default,mydata$sex)))

# Example:

# A man of age 42 with an income of 52,000 EUR and a debt of 6,000 EUR

example <- data.frame(42,52,"male",6,NA)  
colnames(example) <- c("age","income","sex","debt","default")

predict(mod, newdata = example, type = "response")

# Probability of default = 0.1763394

### Visualization with nomogram

library(VRPM)

colplot(mod, file="Colored nomogram", coloroptions = 1, zerolevel = "zero",  
        risklabel = "Probability of default", adverse = F)

# coloroptions:   
# If 1, the rainbow color map is used.  
# If 2, a sequential color map is used. (default)  
# If 3, a diverging color map is used.  
# If 4, a black-and-white color map is used.   
# If 5, the viridis color map is used.

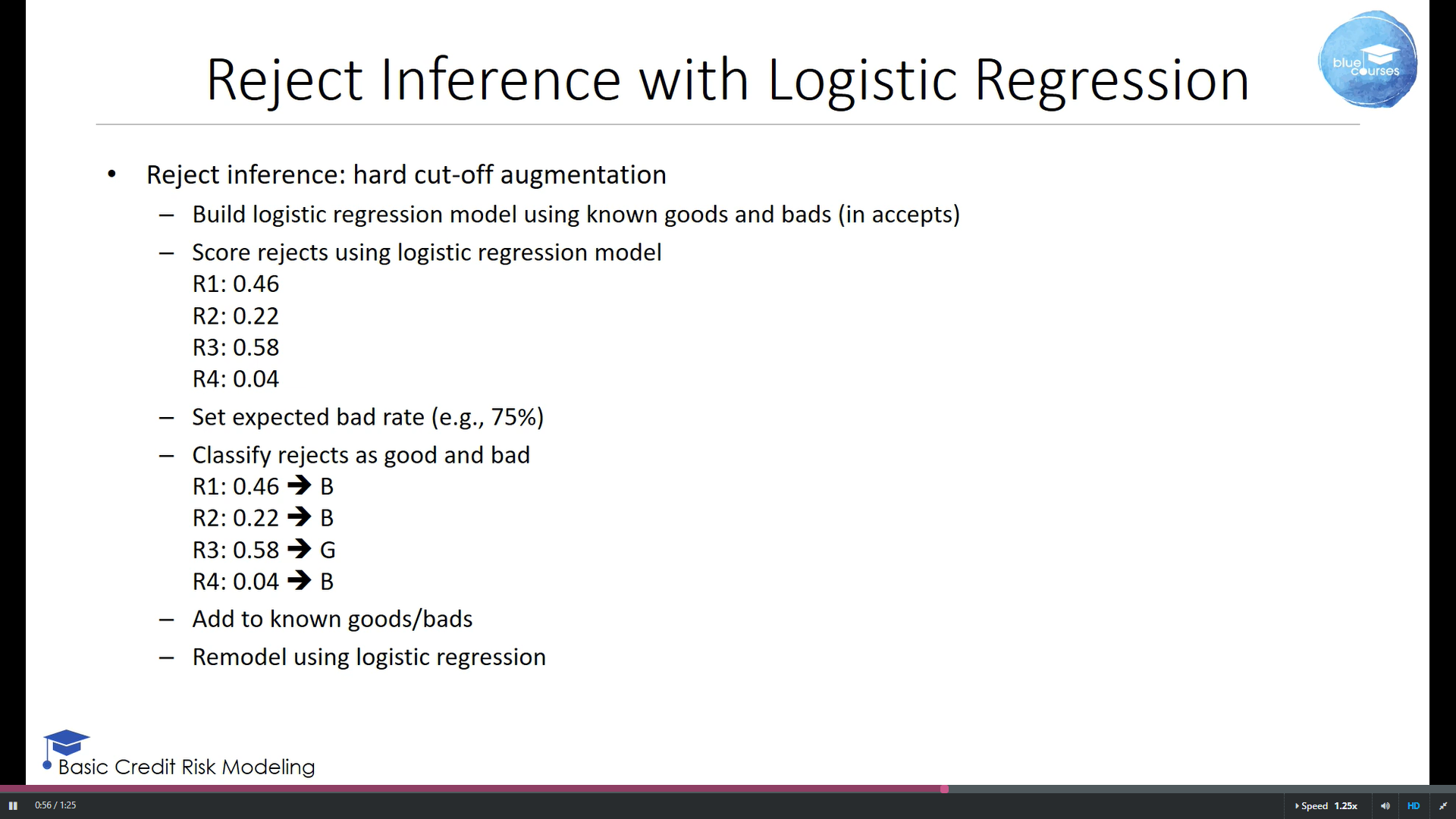
## Logistic Regression and WOE

Remember, higher weights of evidence means less risk and vice-versa.

This weights of evidence variable can then be directly used in the logistic regression model as illustrated here. This way less dummy variables need to be introduced hereby giving us a more robust model.

## Reject Inference with Logistic Regression

1. Assume a bad rate
2. Use logistic regression to rank order all of the rejects
3. The lower x% (the bad rate) are considered bad and the upper 1-x% are considered good



## Decision Trees

Decision trees are often referred to as recursive partitioning algorithms or RPAs because they recursively partition the data over and over again until a leaf node is reached where a classification is made

Various tree induction algorithms have been introduced in the scientific literature to induce decision trees from data. Amongst the most popular are

* C4.5, or its newer version C5, developed by Quinlan in 1993
* classification and regression trees abbreviated as CART developed by Breiman, Friedman, Olshen and Stone in 1984
* Chi-squared automatic interaction detection abbreviated as CHAID developed by Hartigan in 1975.

Decision trees can thus be used for both PD modeling where the target is binary, as well as LGD or EAD modeling where the target is continuous.

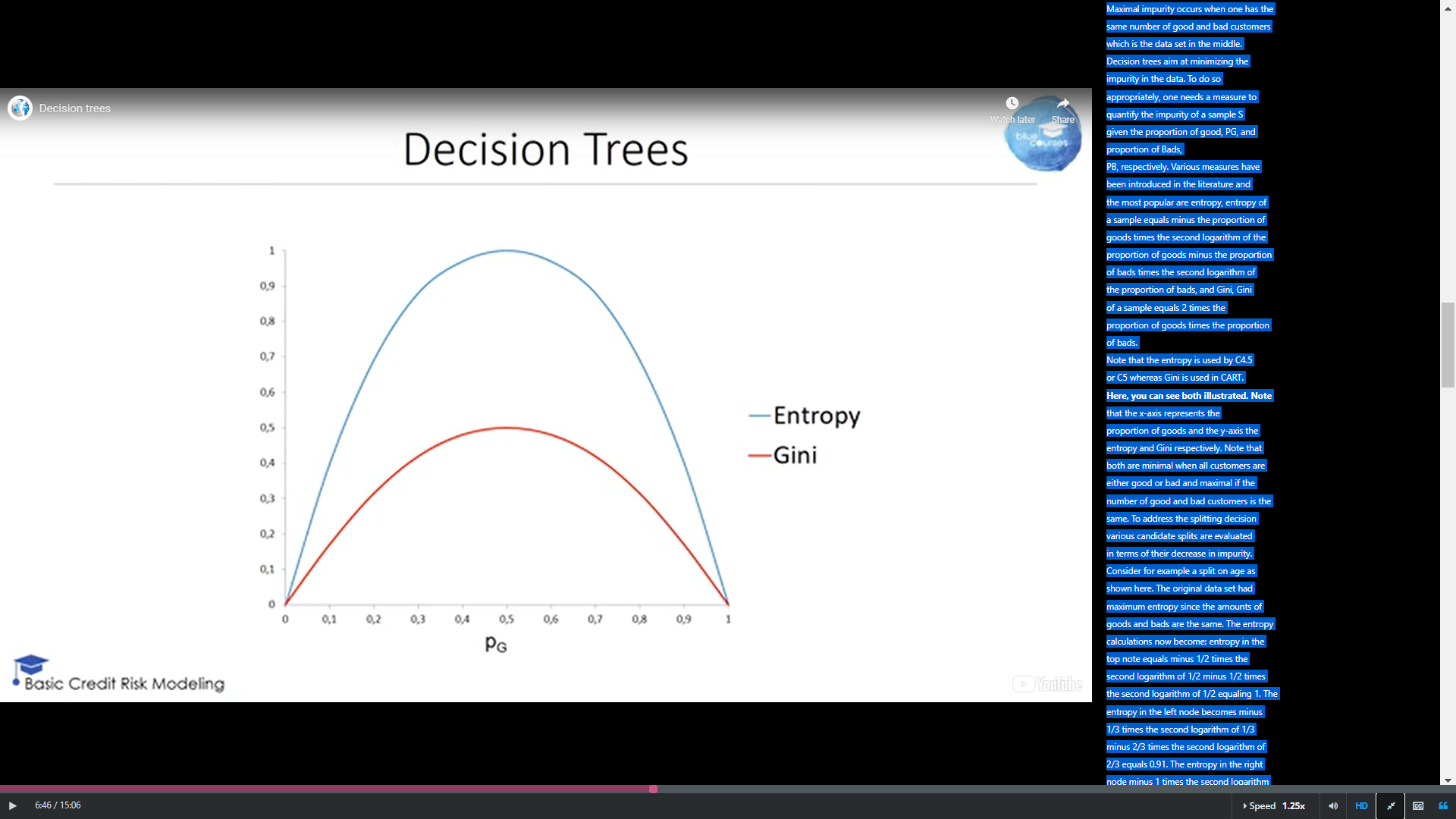
When estimating decision trees from data three decisions should be made:

* the splitting decision
* the stopping decision
* the assignment decision

we always assign a class in a leaf node according to the majority. This idea is also referred to as winner-take-all learning since the winner class makes the final assignment decision

Various impurity measures have been introduced in the literature

1. entropy: equals minus the proportion of goods times the second logarithm of the proportion of goods minus the proportion of bads times the second logarithm of the proportion of bads
2. Gini: equals 2 times the proportion of goods times the proportion of bads.



how they can be used in analytics

1. they can be used as an alternative to pivot tables **to perform categorization** as we discussed earlier in the section on data pre-processing. More specifically, you could calculate the gain for each categorization option and choose the option with the highest gain
2. can also be used to screen variables and do **variable selection**. Variables that appear at the top of the tree are the most predictive. In our example tree to the right you can see that income is the most predictive variable since it is the root node with the highest gain. In other words, you can very easily use the gain measure to compare various variables in terms of their predictive power
3. can also be used to perform **segmentation**. In this case, you could build a decision tree of only a few levels deep, for example two, and then build a second stage logistic regression model in each of the leaf nodes. In our example tree, that would mean that we build 4 additional logistic regression models in each of the leaf nodes. Our final model now consists of a decision tree and 4 logistic regression models.
4. you could also build a decision tree model and use that as **your final analytical model**. That would mean that we use our 4 leaf node decision tree as our final model and put it in production

Advantages

1. **easy to interpret** and understand assuming they are not too big. You can easily explain why a customer is classified as a good or bad payer. Note that especially in credit scoring being able to explain classifications is very important and required by various available regulations
2. nonparametric because no assumptions of normality, symmetric distributions or independence are needed. Hence, **no prior data transformations such as a logarithmic transformation are needed**
3. they're also very **robust with respect to outliers**

Disadvantages

1. very **sensitive to changes in the training data**. If you draw another sample from the same underlying data, you are likely to create another decision tree. A reason for this is that usually there are many competing good splits for the root node of the decision tree so because of data variations another sample might give another root node and thus a totally different tree. That's why decision trees are often referred to as unstable or weak classifiers

## Multiclass Classification

The target rating can have more than two values

### Cumulatie Logistic Regression

also called ordinal logistic regression

In cumulative logistic regression, the cumulative probabilities are modeled using a logit type of transformation. The transformation goes as follows: the probability that a firm with characteristics x has a rating lower than or equal to rating R is 1 divided by 1 plus e to the power minus theta\_R plus beta1 times x1 plus beta2 times x2 and so on.

mac: this is saying, if I want an ordinal regression on credit ratings high, medium, and low, I start with a logistic regression on high. Then, another on high and medium. Now, medium = high&medium – high. Then low is what's left.

#### Cumulative Logistic Regression in R

library(MASS)

ratings <- read.csv("c:/temp/ratings.csv")

ratings$rating <- factor(ratings$rating, ordered=T)

model\_logit <- polr(rating ~ COMMEQTA + LLPLOANS + COSTTOINCOME + ROE + LIQASSTA + SIZE,   
                    method="logistic", data = ratings, Hess=T)

summary(model\_logit)

## Multiclass Decision Trees

First, regarding the splitting decision, the entropy and gini measures are now calculated for all K classes as depicted here. The assignment decision in the leaf nodes is then based upon the class with the highest probability, again winner-take-all learning remember. And finally, the stopping decision is also made using a validation set and an early stopping plot whereby the error on the y-axis now represents the multiclass miss-classification error.

# Measuring the Performance of Credit Scoring Classification Models

## How to Measure Performance

First, the decision needs to be made about data set split up. In other words, on what data set are we going to measure performance. As we discuss, this will mainly depend upon the number of observations available.

Next, a decision needs to be made about the performance measure. Also here various choices are available both for binary as well as multi-class classification.

## Split Sample Method

You've got your training, your validation, and then (separate and apart) you testing. Don't overlap.

Maybe stratify. Your test bad percentage ought to equal your train bad percentage ought to enter your total pop bad percentage.

## Cross-Validation

In cross-validation, the data is split into K fold, for example 10 folds

A model is then trained on K minus 1 training folds and tested on the remaining validation fold

one may also choose to build one model on all observations and use that as the final model

The corresponding performance is then the performance as it comes out of the cross-validation procedure

## Single Sample Method

For small datasets, say less than 100 observations, the single sample method can be used

Here we will not split up the data into subsets but rather calculate the performance as a function of training error and model complexity

The idea is to penalize complex models since they are more likely to fit the noise in the data and thus overfit.

Popular examples here are the

1. Akaike information criterion (AIC) and the
2. Bayesian Information Criterion (BIC) also called the Schwarz Bayesian Criterion (SBC).

Each of them measures model complexity by looking at the number of estimated parameters.

Obviously, good models should have low AIC or BIC.

Note that is only meaningful to compare the AIC or BIC of models built on the same data set.

## Performance Measures for Binary Classification

### confusion matrix

#### classification accuracy

percent correctly predicted

#### classification error

percent incorrectly predicted

#### sensitivity

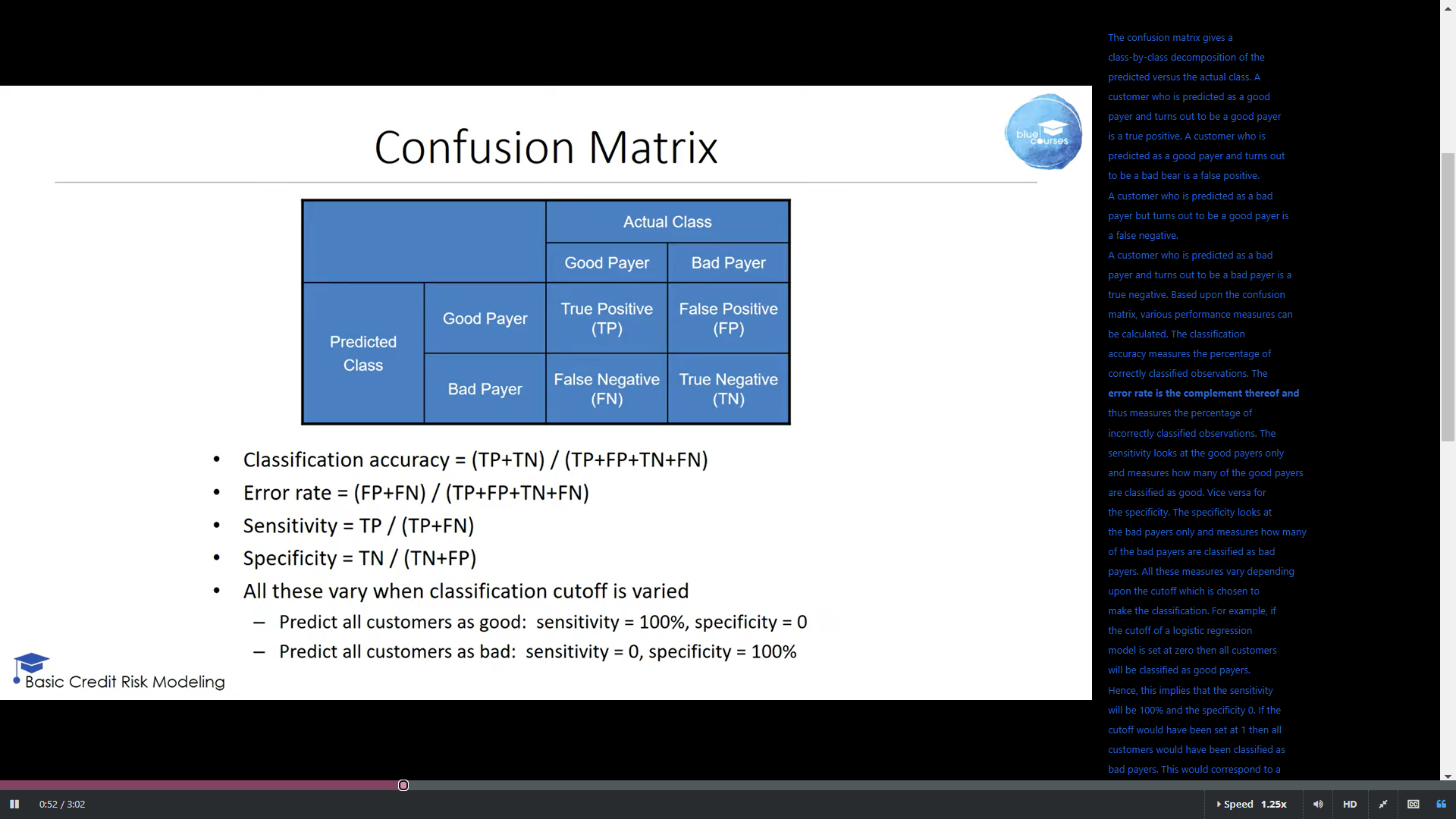
correctly predicted in the first column. In other words, of the goods, what percent were correctly predicted

#### specificity

correctly predicted in the second column. In other words, of the bads, what percent were correctly predicted

1 minus specificity

the percentage of bads predicted to be good



## the Receiver Operating Characteristic curve

The Receiver Operating Characteristic curve (ROC curve) then plots the sensitivity versus 1 - the specificity

When the cutoff is set at its minimum, zero in our case, then the sensitively becomes 1 and the specificity becomes zero.

When the cutoff is set at its maximum, 1 in our case, the sensitivity becomes zero and the specificity becomes 1.

A perfect model detects all the goods and all the bads and thus has a sensitivity of 1 and a specificity of 1

So: whichever model has the "highest" ROC curve is the best. But what if they intersect?

the area under the curve

The AUC is always bounded between 0 and 1 and can be interpreted as a probability

It represents the probability that a randomly chosen good payer gets a higher score than a randomly chosen bad payer

ROC Curve in R

hmeq <- read.csv("c:/temp/hmeq.csv")

install.packages("pROC")

library(pROC)

# Remove all missing values

hmeq.omit <- na.omit(hmeq)

# Convert JOB and REASON to factor

hmeq.omit$JOB <- as.factor(hmeq.omit$JOB)

hmeq.omit$REASON <- as.factor(hmeq.omit$REASON)

hmeq.full <- glm(BAD ~ ., data=hmeq.omit, family=binomial(link="logit"))

prob <- predict(hmeq.full, type=c("response"))

hmeq.omit$prob <- prob

g <- roc(BAD ~ prob, data = hmeq.omit)

plot(g)

auc(g)

ROC Curve in Python

import pandas as pd

import numpy as np

import statsmodels.api as sm

import matplotlib.pyplot as plt

from sklearn.metrics import roc\_curve, auc

hmeq = pd.read\_csv('c:/temp/hmeq.csv')

# Remove all missing values

hmeq = hmeq.dropna()

# Create dummies for JOB and REASON

cat\_vars = ['REASON', 'JOB']

for var in cat\_vars:

cat\_list = pd.get\_dummies(hmeq[var], prefix=var, drop\_first=True)

hmeq = hmeq.join(cat\_list)

hmeq.drop(columns=var, inplace=True)

Y = hmeq.loc[: , 'BAD']

X = hmeq.drop(columns='BAD')

logit\_model = sm.Logit(Y,X)

result = logit\_model.fit()

Ypred = result.predict(X)

fp\_rate, tp\_rate, thresholds = roc\_curve(Y, Ypred)

roc\_auc = auc(fp\_rate, tp\_rate)

plt.title('Receiver Operating Characteristic')

plt.plot(fp\_rate, tp\_rate, 'b', label = 'AUC = %0.2f' % roc\_auc)

plt.legend(loc = 'lower right')

plt.plot([0,1],[0,1],'r--')

plt.xlim([-0.1,1.2])

plt.ylim([-0.1,1.2])

plt.ylabel('True Positive Rate')

plt.xlabel('False Positive Rate')

plt.show()

## the Cumulative Accuracy Profile curve and Accuracy Ratio

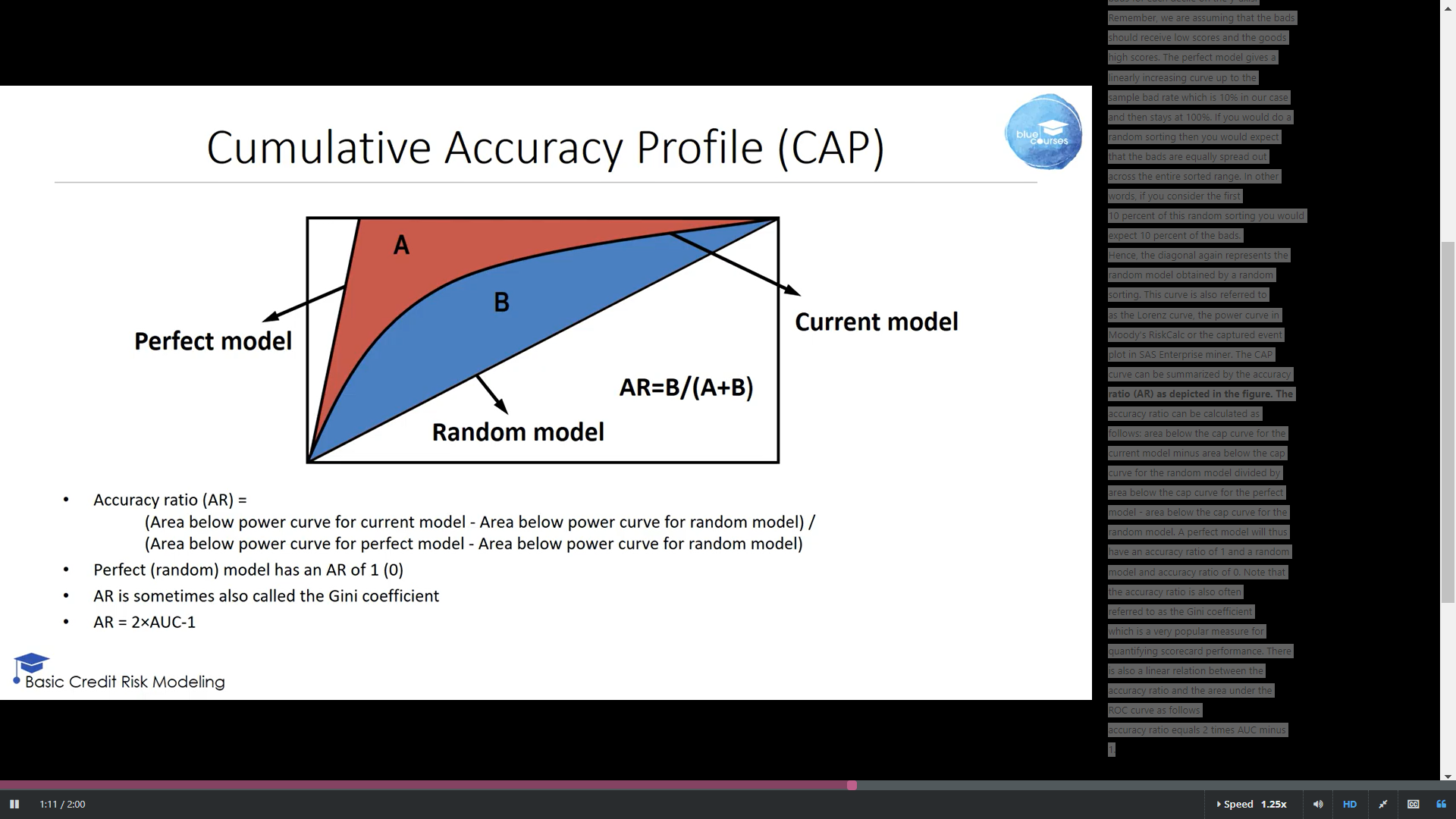
The cumulative accuracy profile or CAP curve starts by sorting the population from low score to high score and then measures the cumulative percentage of bads for each decile on the y-axis

The perfect model gives a linearly increasing curve up to the sample bad rate which is 10% in our case and then stays at 100%.

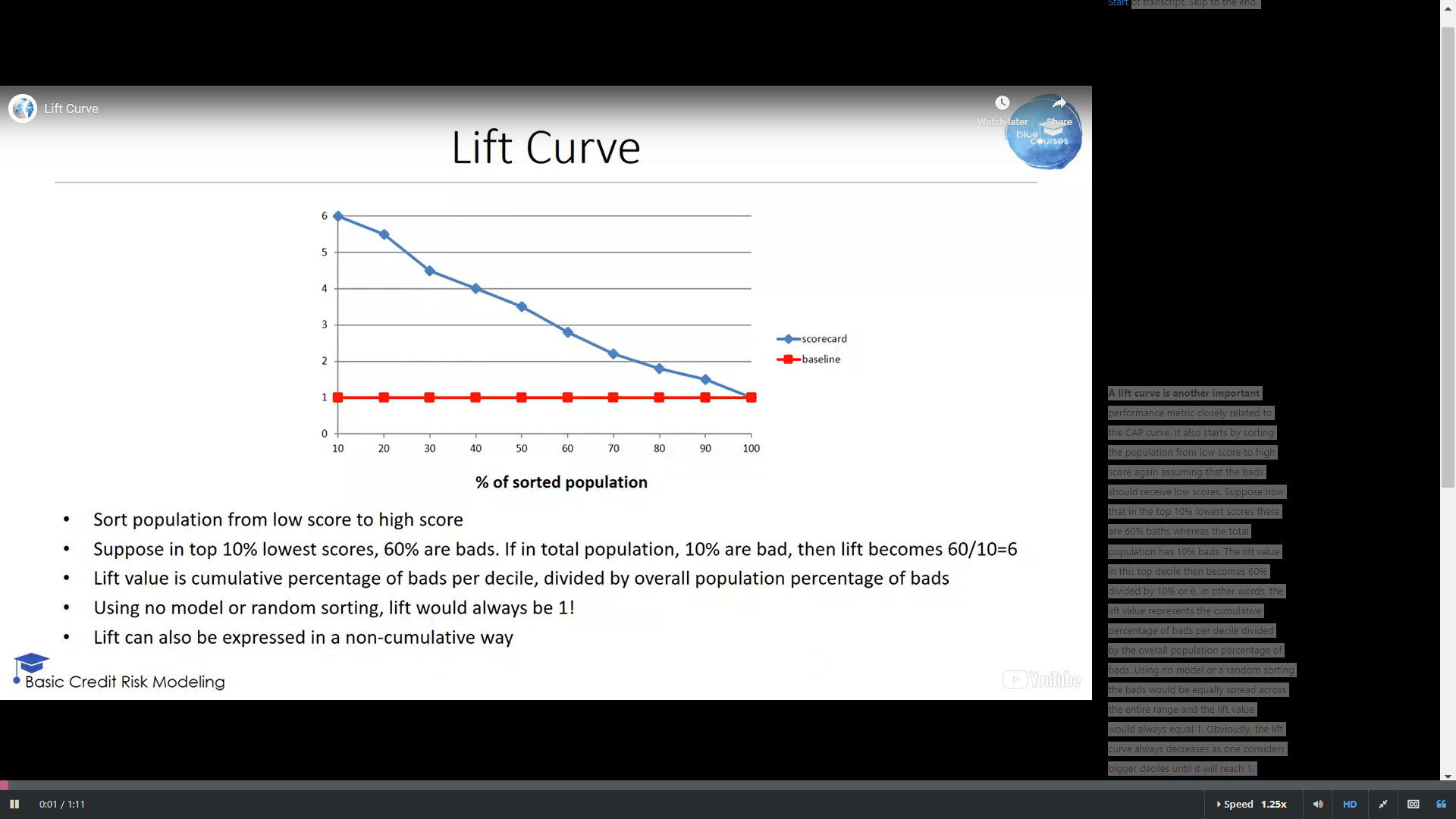
This curve is also referred to as the Lorenz curve, the power curve in Moody's RiskCalc or the captured event plot in SAS Enterprise miner.

Note that the accuracy ratio is also often referred to as the Gini coefficient which is a very popular measure for quantifying scorecard performance.

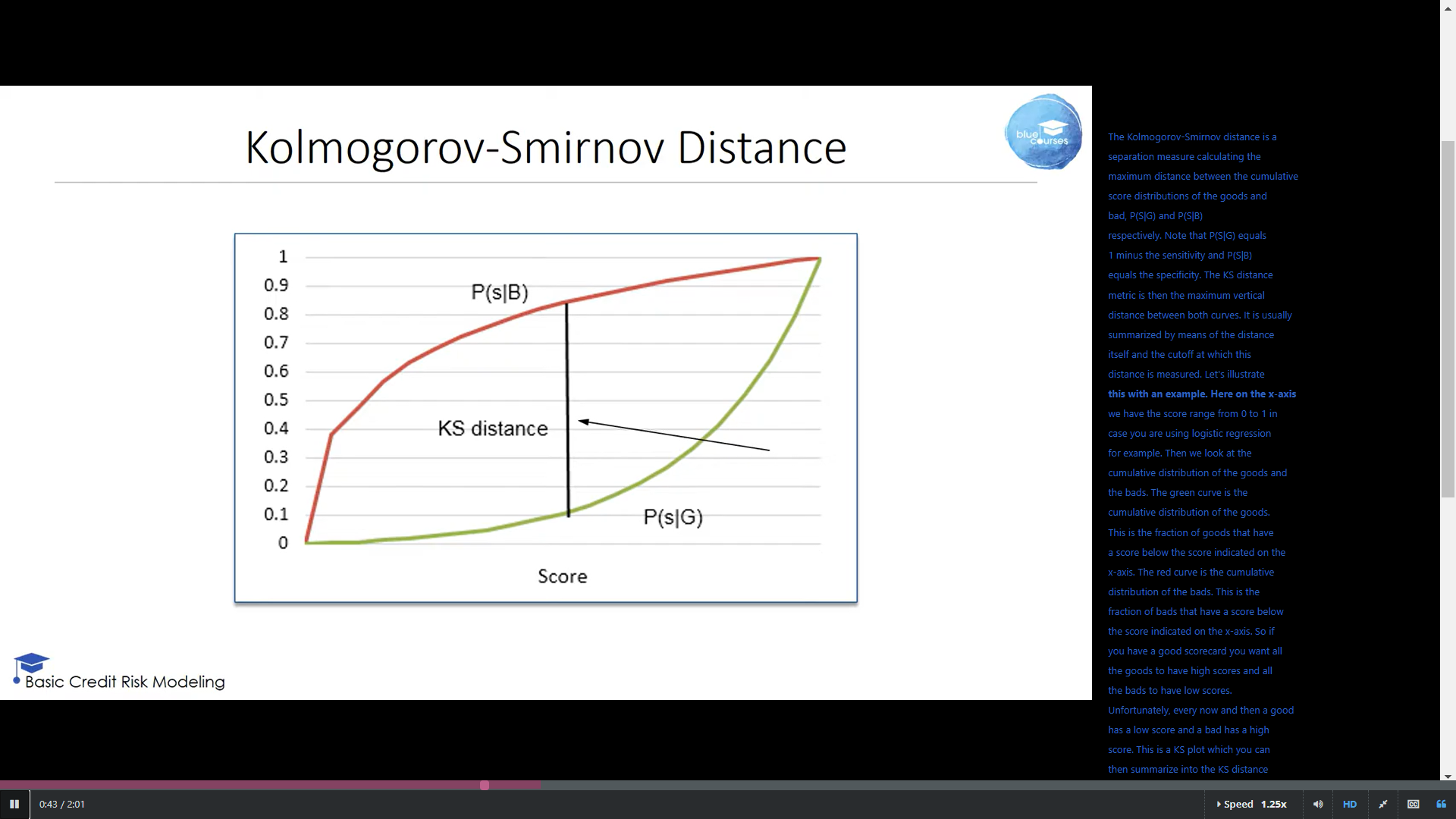
There is also a linear relation between the accuracy ratio and the area under the ROC curve as follows accuracy ratio equals 2 times AUC minus 1



## the lift curve



## the Kolmogorov-Smirnov distance



Red is the cumulative bad population.

Green is the cumulative good population

The KS distance metric typically comes with the distance itself as well as the score or cutoff at which this distance is reached.

It can also be easily measured on ROC plot as the maximum vertical distance between the ROC curve and the diagonal.

## the Mahalanobis distance

* the Mahalanobis distance defined as the difference between the mean scores of the goods and bads divided by the pooled standard deviation

It is better than the Euclidean distance because it takes the distribution of the scores into account by means of the standard deviation

* Closely related is the divergence metric which considers the squared difference of the means divided by the average of the sum of both variances

Higher is better in both cases

the Mahalanobis and divergence metric are seldom used

