MA 477

PROBABILITY OF DEFAULT AND HISTORY OF DEFAULTS

Submitted By: **Ankur Ingale (170123007)**
Garvit Mehta (170123018),
Tanya Chauhan (170123052)

Probability of Default

The probability of default (PD) is the probability of a borrower or debtor defaulting on loan repayments. Within financial markets, an asset's probability of default is the probability that the asset yields no return to its holder over its lifetime and the asset price goes to zero. Investors use the probability of default to calculate the expected loss from an investment.

The market's view of an asset's probability of default influences the asset's price in the market. Therefore, if the market expects a specific asset to default, its price in the market will fall (everyone would be trying to sell the asset).

Creditors typically want a higher interest rate to compensate for bearing higher default risk. Financial metrics—such as cash flows relative to debt, revenues or operating margin trends, and the use of leverage—are common considerations when evaluating the risk. A company's ability to execute a business plan and a borrower's willingness to pay are sometimes factored into the analysis as well.

For businesses, a probability of default is implied by their credit rating. PDs may also be estimated using historical data and statistical techniques. PD is used along with "loss given default" (LDG) and "exposure at default" (EAD) in a variety of risk management models to estimate possible losses faced by lenders. Generally, the higher the default probability, the higher the interest rate the lender will charge the borrower.

People sometimes encounter the concept of default probability when they purchase a residence. When a homebuyer applies for a mortgage on a piece of real estate, the lender makes an assessment of the buyer's default risk, based on their credit score and financial resources. The higher the estimated probability of default, the greater the interest rate that will be offered to the borrower. For consumers, a FICO score implies a particular probability of default.

Credit Rating	FICO Score	The Percent of Population (%)	The Probability of Default (%)	Interest Rate
C1	800 or more	13	1	5.99
C2	750–799	27	1	5.99
C3	700–749	18	4.4	6.21
C4	650–699	15	8.9	6.49
C5	600–649	12	15.8	7.30
C6	550–599	8	22.5	8.94
C7	500–549	5	28.4	9.56
C8	Less than 499	2	41	-

For businesses, a probability of default is implied by their credit rating. PDs may also be estimated using historical data and statistical techniques.

Rating grade	2009			2010			2011		
	Rated	Def.	DR	Rated	Def.	DR	Rated	Def.	DR
AAA	81	0	0.00	72	0	0.00	51	0	0.00
AA+	37	0	0.00	25	0	0.00	36	0	0.00
AA	188	0	0.00	143	0	0.00	120	0	0.00
AA−	245	0	0.00	209	0	0.00	207	0	0.00
A+	340	1	0.29	353	0	0.00	357	0	0.00
A	510	2	0.39	474	0	0.00	470	0	0.00
A−	546	0	0.00	528	0	0.00	560	0	0.00
BBB+	498	2	0.40	457	0	0.00	473	0	0.00
BBB	541	1	0.18	583	0	0.00	549	0	0.00
BBB−	459	5	1.09	430	0	0.00	508	1	0.20
BB+	266	0	0.00	254	2	0.79	260	0	0.00
BB	295	3	1.02	276	1	0.36	319	0	0.00
BB−	441	4	0.91	379	2	0.53	403	0	0.00
B+	438	24	5.48	393	0	0.00	509	2	0.39
B	482	48	9.96	436	3	0.69	586	7	1.19
B−	303	52	17.16	290	6	2.07	301	12	3.99
CCC–C	190	92	48.42	220	49	22.27	138	22	15.94
All	5860	234	3.99	5522	63	1.14	5847	44	0.75

Sources: Standard & Poor's (2010, Tables 51–53), Standard & Poor's (2011, Tables 50–52), Standard & Poor's (2012, Tables 50–52).

Under Basel II, a default event on a debt obligation is said to have occurred if

- it is unlikely that the obligor will be able to repay its debt to the bank without giving up any pledged collateral
- the obligor is more than 90 days past due on a material credit obligation

MERTON MODEL

Merton developed a structural model based on the Black Scholes option pricing model. The model can both be used for equity valuation and credit risk management. The intuition behind the use of option pricing for equity valuation in the Merton model is simple. Equity holders are the residual holders of the company.The value of the assets above debt K, will be paid out to them.

Therefore Merton argues that the equity holders have a call option on the company's assets with strike price K. If the value of assets is larger than K, then we get V(t)-K. If not, then they get nothing.

In contrast, the bondholders are argued to own a zero-coupon bond with par-value of K. They retrieve the value of K when the assets are worth more. If the assets are worth less, then they only receive the asset value V(t). Bondholders are thus short a put option.

The formula below values the equity in function of the value of assets corrected for the value of debt. The additional parameters required are the risk free rate, r, the volatility of assets, sigma(v), and the time to maturity T.

$$E = V_t \cdot N(d_1) - K \cdot e^{-r \cdot \Delta T} \cdot N(d_2)$$

$$d_1 = \frac{ln\frac{V_t}{K} + (r + \frac{\sigma_v^2}{2}) \cdot \Delta T}{\sigma_v \cdot \sqrt{\Delta T}}$$

$$d_2 = d_1 - \sigma_v \cdot \sqrt{\Delta T}$$

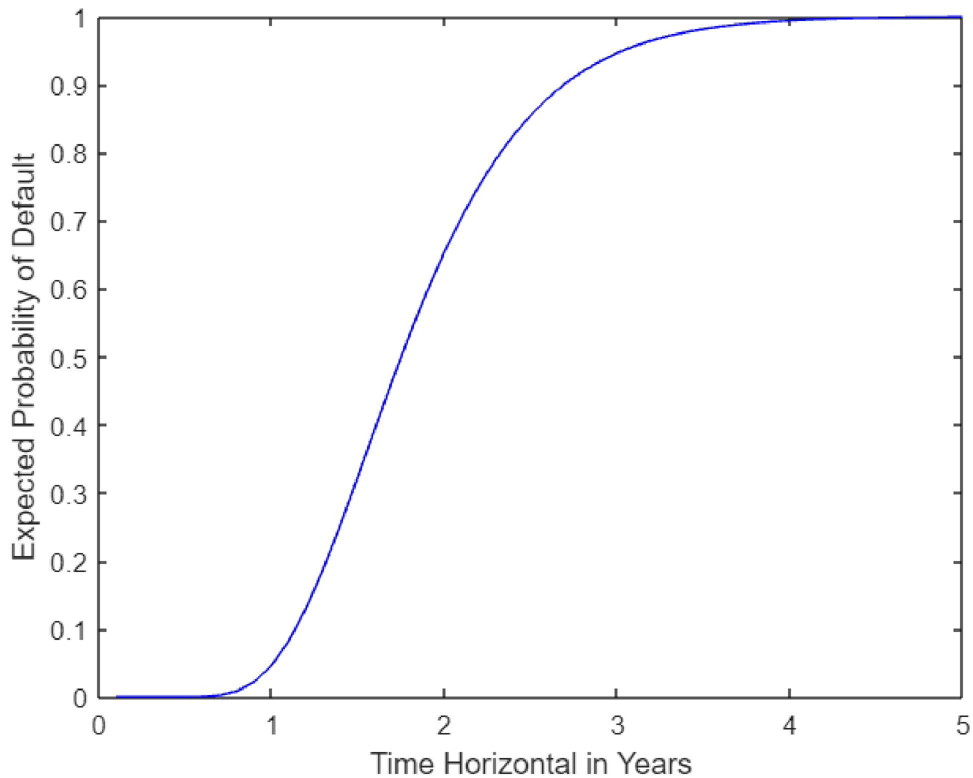
The Merton model also allows us to estimate a probability of default. This indicates the chance that a company will not be able to repay its debt considering a certain time frame, in many cases 1 year.

The Merton model allows to calculate a risk-neutral probability of default for a certain company.

The risk neutral PD indicates the chance at maturity the value of assets will be worth less than the debt.

Just like for option, it indicates the probability that the option ends up out-of-the-money. The risk neutral probability of default is calculated as follows:

$$PD = N(-d_2)$$



Probability of Default using Transition Matrices

Credit market for bonds issued by corporations and government entities as well as for asset-backed securities (ABS). The three major global credit rating agencies are Moody's Investors Service, Standard & Poor's, and Fitch Ratings. Each provides quality ratings for issuers as well as specific issues. These are ordinal ratings focusing on the probability of default.

The credit rating agencies consider the expected loss given default by means of notching, which is an adjustment to the issuer rating to reflect the priority of claim for specific debt issues of that issuer and to reflect any subordination. The issuer rating is typically for senior unsecured debt. The rating on subordinated debt is then adjusted, or "notched," by lowering it one or two levels — for instance, from A+ down to A or further down to A–. This inclusion of loss given default in addition to the probability of default explains why they are called "credit ratings" and not just "default ratings."

- Notching is when a credit rating agency bumps up or down the credit rating on an issuer's specific debts or obligations.
- Because certain types of debt—for instance, subordinated debts—are riskier than senior debts, the rating on junior debts can be notched lower.
- Similarly, those debts from the issuer that are senior and secured by collateral may be notched higher.

The rating agencies report transition matrices based on their historical data

We can obtain the Standard & Poor's Average One-Year Transition Rates For Global Corporates using historical data from 1981–2019.

The following table is the Standard & Poor's 2019 transition matrix. It shows the probabilities of a particular rating transitioning to another over the course of the following year. An A rated issuer has an 78.88% probability of remaining at that level, a 0.03% probability of moving up to AAA; a 0.22% probability of moving up to AA; an 0.86% probability of moving down to BBB; 0.10% down to BB; 0.02% to B, 0.01% to CCC, CC, or C; and 0.05% to D, where it is in default.

Average One-Year Transition Rates For Global Corporates By Rating Modifier (1981-2019) (%)

	From/to	AAA	AA+	AA	AA-	A+	A	A-	BBB+	BBB	BBB-	BB+	BB	BB-	B+	B	B-	CCC	D
0	AAA	87.03	5.89	2.51	0.69	0.16	0.24	0.13	0.00	0.05	0.00	0.03	0.05	0.03	0.00	0.03	0.00	0.05	0.00
2	AA+	2.31	78.94	10.91	3.54	0.71	0.33	0.19	0.05	0.09	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	AA	0.42	1.31	80.76	8.53	2.72	1.15	0.36	0.39	0.13	0.08	0.05	0.03	0.02	0.02	0.00	0.02	0.05	0.02
6	AA-	0.04	0.11	3.77	78.80	9.68	2.19	0.60	0.25	0.15	0.07	0.03	0.00	0.00	0.03	0.08	0.00	0.00	0.03
8	A+	0.00	0.06	0.44	4.44	78.38	8.73	2.15	0.61	0.34	0.09	0.06	0.09	0.01	0.07	0.03	0.00	0.00	0.05
10	A	0.03	0.04	0.22	0.41	5.32	78.88	6.74	2.38	0.86	0.27	0.10	0.10	0.06	0.08	0.02	0.00	0.01	0.05
12	A-	0.04	0.01	0.06	0.15	0.42	6.49	78.12	7.23	1.98	0.57	0.13	0.13	0.11	0.10	0.02	0.01	0.03	0.06
14	BBB+	0.00	0.01	0.05	0.06	0.20	0.74	7.13	75.83	7.98	1.56	0.36	0.29	0.13	0.15	0.10	0.02	0.06	0.10
16	BBB	0.01	0.01	0.04	0.03	0.10	0.31	1.00	7.73	76.00	6.11	1.34	0.58	0.27	0.22	0.11	0.03	0.05	0.16
18	BBB-	0.01	0.01	0.02	0.04	0.06	0.14	0.25	1.17	9.31	72.40	5.47	2.08	0.83	0.36	0.22	0.16	0.21	0.25
20	BB+	0.04	0.00	0.00	0.03	0.03	0.08	0.08	0.41	1.59	11.33	65.29	7.42	2.61	0.95	0.53	0.24	0.36	0.31
22	BB	0.00	0.00	0.03	0.01	0.00	0.06	0.05	0.16	0.47	2.00	9.44	65.41	8.46	2.22	1.02	0.31	0.52	0.51
24	BB-	0.00	0.00	0.00	0.01	0.01	0.01	0.05	0.09	0.23	0.35	1.69	9.57	63.71	8.42	3.04	0.81	0.66	0.91
26	B+	0.00	0.01	0.00	0.03	0.00	0.03	0.06	0.04	0.05	0.10	0.31	1.42	8.17	62.91	9.20	2.51	1.71	1.98
28	B	0.00	0.00	0.01	0.01	0.00	0.03	0.04	0.02	0.05	0.03	0.11	0.23	1.09	7.38	62.00	9.32	3.85	3.20
30	B-	0.00	0.00	0.00	0.00	0.02	0.03	0.00	0.06	0.05	0.10	0.08	0.13	0.46	2.18	10.06	54.63	11.70	6.49
32	CCC/C	0.00	0.00	0.00	0.00	0.03	0.00	0.08	0.05	0.08	0.05	0.03	0.16	0.40	0.98	2.57	9.41	43.64	27.08

We can use this transition matrix to calculate the probability that a company with a given rating will default over the course of next year using the last column of this matrix.

Logistic Regression

Real world prediction involves using various machine learning methods to predict if the given entity who took the loan(person or company) will default or not. Here, we use Logistic Regression to predict if the person will default or not; that is, if he/she can pay the loan or not.

We have a database of people containing their loan amount, loan interest rate, annual income, home ownership, age and their grade(from A to G). The company labels their loan status as 0 if the person is able to pay the loan or 1 if he/she defaults.

	loan_status	loan_amnt	int_rate	grade	emp_length	home_ownership	annual_inc	age
1	0	5000	10.65	B	10	RENT	24000.00	33
2	0	2400	10.99	C	25	RENT	12252.00	31
3	0	10000	13.49	C	13	RENT	49200.00	24
4	0	5000	10.99	A	3	RENT	36000.00	39
5	0	3000	10.99	E	9	RENT	48000.00	24
6	0	12000	12.69	B	11	OWN	75000.00	28
7	1	9000	13.49	C	0	RENT	30000.00	22
8	0	3000	9.91	B	3	RENT	15000.00	22
9	1	10000	10.65	B	3	RENT	100000.00	28
10	0	1000	16.29	D	0	RENT	28000.00	22

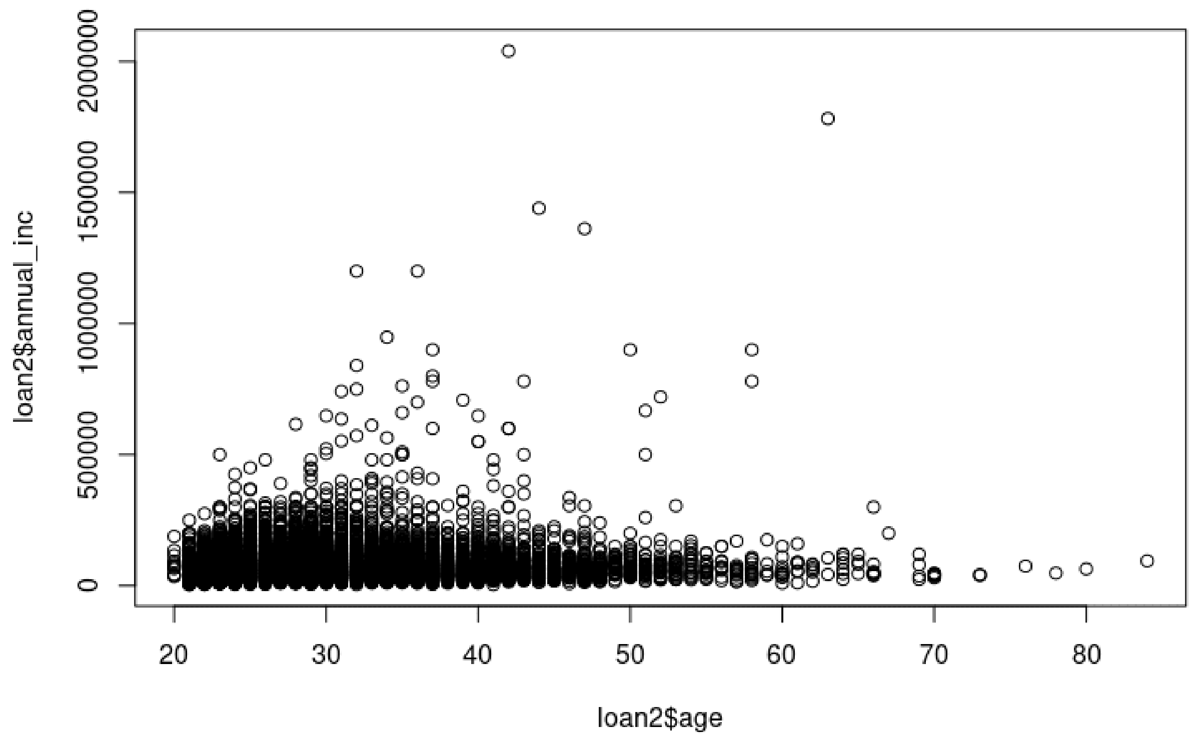
Snapshot of the database

```
  loan_status       loan_amnt        int_rate      grade      emp_length      home_ownership
 Min.   :0.0000   Min.   :  500   Min.   : 5.42   A:9649   Min.   : 0.000   MORTGAGE:12002
 1st Qu.:0.0000   1st Qu.: 5000   1st Qu.: 8.49   B:9329   1st Qu.: 2.000   OTHER   :   97
 Median :0.0000   Median : 8000   Median :10.99   C:5748   Median : 4.000   OWN     : 2301
 Mean   :0.1109   Mean   : 9594   Mean   :11.00   D:3231   Mean   : 6.145   RENT    :14692
 3rd Qu.:0.0000   3rd Qu.:12250   3rd Qu.:13.11   E: 868   3rd Qu.: 8.000
 Max.   :1.0000   Max.   :35000   Max.   :23.22   F: 211   Max.   :62.000
                                                  G:  56   NA's   :809
   annual_inc           age
 Min.   :   4000   Min.   : 20.0
 1st Qu.:  40000   1st Qu.: 23.0
 Median :  56424   Median : 26.0
 Mean   :  67169   Mean   : 27.7
 3rd Qu.:  80000   3rd Qu.: 30.0
 Max.   :6000000   Max.   :144.0
```

Total Observations in Table: 29092

	loan$loan_status		
loan$grade	0	1	Row Total
A	9084	565	9649
	0.941	0.059	0.332
B	8344	985	9329
	0.894	0.106	0.321
C	4904	844	5748
	0.853	0.147	0.198
D	2651	580	3231
	0.820	0.180	0.111
E	692	176	868
	0.797	0.203	0.030
F	155	56	211
	0.735	0.265	0.007
G	35	21	56
	0.625	0.375	0.002
Column Total	25865	3227	29092

Next, we removed the outliers in terms of annual income and age. The final annual income vs age graph looks as follows:

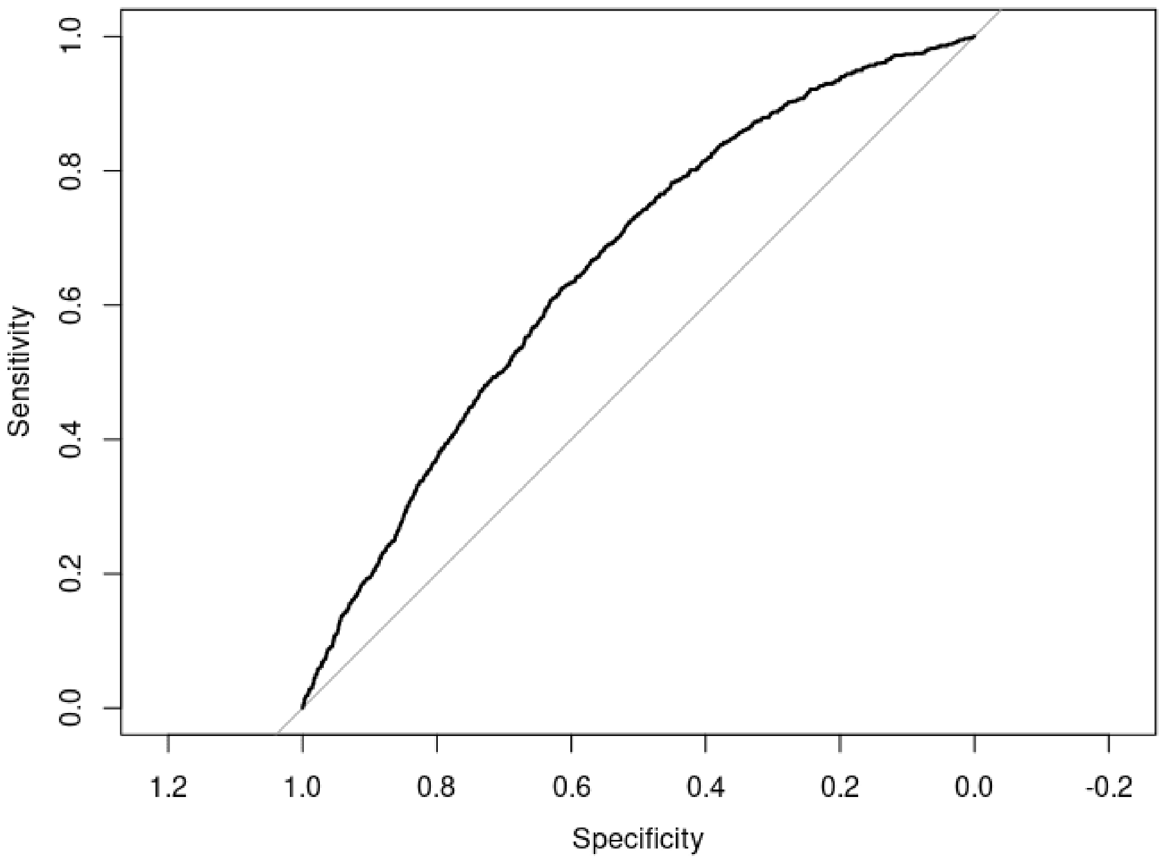


Next, we trained our model on two-thirds of data using Logistic Regression. We used the rest of the data to test the model. Next, we kept a threshold of 20%; that is, if default probability is less than 20%, it will default. We got the accuracy of 84.84%

Next, we plotted a **ROC curve** (**receiver operating characteristic curve**), which is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. This curve plots two parameters:

1. True Positive Rate (Sensitivity)
2. False Positive Rate (1 - Specificity)

Area Under the Curve(AUC) provides an aggregate measure of performance across all possible classification thresholds. AUC can be interpreted as the probability that the model ranks a random positive example more highly than a random negative example. The area under the curve for our model is 0.6581 (0 means 100% wrong predictions and 1 means 100% right predictions).


1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
Specificity