

# Default Rate on Auto Loans

## II. Introduction

In completing this model, we want to view a portfolio of loans at a bank and be able to predict the default rate on an auto loan. We also want to be able to look at an individual loan and predict the default probability taking into consideration interest rate, term, number of days late, times late, original loan amount, collateral value, LTV and credit score. The data will assist in forecasting default rates in future years and in projecting income and loss for a bank. We predict at the end of the model, credit score will be the main indicator of default rate on any given loan.

Our data source is the underwriting system for Commercial Bank of Texas, or CBTx, which consolidates all of a person's financial information into a usable format for decision making. We chose these variables based our predictions of what would affect the default probability and based on data availability.



Variable	Mean ( $\mu$ )	Std.Dev. ( $\sigma$ )	Min.	Max.	Count (n)
Default	0.0573	0.2325	0	1	3,751.00
Interest Rate	0.0841	0.0339	0	0.18	3,751.00
Term Number	58.2306	9.9581	1	79	3,751.00
Original Loan Amount	\$20,359	\$8,546	\$2,500	\$63,383	3,751.00
Times Late 1-10 Days	12.39	15.2871	0	75	3,751.00
Times Late 11-30 Days	7.5665	13.387	0	72	3,751.00
Times Late 31-60 Days	3.6822	9.8196	0	63	3,751.00
Collateral Value	\$21,386	\$11,790	\$1,745	\$338,083	3,751.00
LTV	1.044	0.4754	0.0207	8.9238	3,751.00
CREDIT Score	727.089	59.756	450	860	3,751.00

	Def.	IR	Term #	Loan Amt	1-10 Days	11-30 Days	31-60 Days	CV	LTV	CS
Default	1.00									
Interest Rate	0.18	1.00								
Term Number	0.03	-0.11	1.00							
Original Loan Amount	-0.04	-0.25	0.44	1.00						
Times Late 1-10 Days	0.44	0.16	0.19	0.07	1.00					
Times Late 11-30 Days	0.59	0.30	0.11	0.01	0.87	1.00				
Times Late 31-60 Days	0.72	0.44	0.00	-0.09	0.54	0.72	1.00			
Collateral Value	-0.12	-0.29	0.16	0.62	-0.13	-0.18	-0.20	1.00		
LTV	0.15	0.17	0.17	0.09	0.37	0.41	0.31	-0.35	1.00	
CREDIT Score	-0.27	-0.26	-0.06	0.03	-0.28	-0.32	-0.28	0.11	0.17	1.00

## Executive Summary

Our regression model is based on the probability of default on an automobile loan at Commercial Bank of Texas with the following independent variables: Interest Rate, Term, Original Loan amount, Times Late 1-10 Days, Times Late 11-30 Days, Times Late 31-60 days, Collateral Value, Loan to Value, and Credit Score. Our sample size is 3,751 loans from the indirect portfolio. This sample size is sufficiently large for us to assume that the sample is distributed normally and thus allows us to continue with our analysis based on Central Limit Theorem 2. Our model would assist in portfolio forecasting and creating the profit plan for the indirect department. A default is technically a payment that is 30 days or more past due, but for the purposes of this regression model we defined a default as repossession of the vehicle. Repossession of the vehicle occurs after the customer becomes 90 days or more past due. Our regression model captured 56.73% of the variation in the Y variable, which is the default rate on the indirect loan portfolio. We predicted that the credit score would be the most telling independent variable, when in fact, credit score turned out to be completely insignificant based on a P-Value of 0.7372. This was also surprising considering Interest Rate is significant at the 99% confidence level. Credit score and interest rate are closely related since the credit score determines the interest rate. We checked for a multicollinearity problem, but found none in our excel data. The largest factor affecting the chances for a default is the interest rate. For every point that the interest rate goes up, your chances of default decrease by 1.2801 on average and holding all other variables constant. The coefficient sign does not make sense if you consider that the lower the interest rate, the higher the credit score and thus the lower the default history for that customer. This means that our model has a flaw somewhere.

Regression Statistics	
Multiple R	0.754
R Square	0.568
Adjusted R Square	0.567
Standard Error	0.153
Observations	3751

## ANOVA

	df	SS	MS	F	Sig. F
Regression	9	115.197	12.800	547.369	0
Residual	3741	87.480	0.023		
Total	3750	202.677			

	Coef	Std. Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.143	0.046	3.100	0.002	0.053	0.233
Interest Rate	-1.280	0.095	-13.475	0.000	-1.466	-1.094
Term Number	0.001	0.000	2.266	0.024	0.000	0.001
Original Loan Amount	0.000	0.000	0.093	0.926	0.000	0.000
Times Late 1-10 Days	-0.002	0.000	-4.582	0.000	-0.002	-0.001
Times Late 11-30 Days	0.005	0.001	9.608	0.000	0.004	0.006
Times Late 31-60 Days	0.017	0.000	41.284	0.000	0.016	0.017
Collateral Value	0.000	0.000	-2.805	0.005	0.000	0.000
LTV	-0.058	0.007	-8.818	0.000	-0.071	-0.045
CREDIT Score	0.000	0.000	-0.336	0.737	0.000	0.000

## VII. Conclusion

Even though the Credit Score was an insignificant variable, we do not believe that the credit score is irrelevant. Theory and common sense tells us that the credit score must play a part in the future behavior of a customer. Our results were run through Probit for clarification on accuracy because we violated a rule by having 0,1 variables for our default.

We feel that our data was not a very good representation of the information required to forecast the default rate of a population. If we could have used Debt to Income, Income, Length of time at address, Length of time at job, Rent or Own, prior default or bankruptcy, and age, we feel that we could have come up with a more accurate model for predicting if a loan would go bad before making the loan. Where we are now, we can only predict the probability that a loan will go bad once it is already on the books. This does not help to prevent loss of income, but to predict loss of income; which is not quite as helpful. We did try to get this information in data, but there was not enough default information for the data to be useful.

**All data for this project was given courtesy of Commercial Bank of Texas, N.A. and their lending staff. All personal information of the customer was omitted for the purpose of this research.**

