UNC CHARLOTTE

Mortgage Default Scorecard

A report submitted to Dr. Thomas Mayock

Cross Section and Time Series Econometrics

Fall 2018

Author: Akshay D Patil

Abstract:

The goal of the project is to build up a “scorecard” to predict mortgage defaults. The task is to build up an econometric model that predicts the probability that given loan defaults. In this case, a logistic regression model is used since the dependent variable, BAD_OVER_48_60 is a categorical variable. After selecting the model, independent variables are stepwise selected and added to the model to increase the predictive performance of the model. After developing the final model, an analysis is performed on out-of-sample data and is compared against in-sample data.

Table of Contents

1 Introduction

The Great Recession of 2007 resulted in dramatic economic and political changes and its effects are still being felt all over the world. Since then mortgage lending companies are meticulously evaluating the underlying default risk to avoid a future recession. It is in this sense important to comprehend the fundamental drivers in the market and how they are connected to the risk of default. Considering this situation our client who is a mortgage lender wishes to engage in risk-based pricing for its mortgage loans. The first step for this is to construct a mortgage scorecard using a loan-level dataset that will predict the probability that a loan defaults.

2 Data Description

The data used is from the *Freddie Mac Single-Family Loan-Level Dataset.* The loans included in the dataset are fully amortizing and fixed-rate Single Family mortgages. To create a simple random sample, loans were selected at random, so that each loan in the full Dataset had an equal probability of being selected [1]. The dataset is split in calendar quarters and for each quarter there is loan-level origination data and monthly loan performance data. Each loan in origination data is assigned with unique Loan Sequence Number and this is used to merge the two transformed datasets. No other exterior data is used while developing the model.

2.1 Variable Description:

This segment talks about the attributes of the variables in the Loan-level dataset and the thinking to incorporate them in the model. The information for each variable was obtained from the User Guide and the command codebook in STATA. The variables described are the independent variables.

2.1.1 LTV_Combined_Orig:

The ratio is obtained by dividing the original mortgage loan amount on the note date plus any secondary mortgage loan amount disclosed by the Seller by the lesser of the mortgaged appraised value on the note date or its purchase price [2].

2.1.1.1 Missing Observations Treatment:

There are 2 missing observations for LTV_Combined_Orig which are not reported. These observations can be deleted without losing the attributes of the dataset.

2.1.1.2 Grouping:

As seen from the histogram plot 1, uniform groups of LTV_Combined_Orig cannot be formed since the data is not uniformly distributed. So, LTV_Combined_Orig is divided into 15 groups starting from 30 to 105 with a difference of 5 percentage points in combined LTV and then each group is assigned to a new variable LTV_Group enumerated from 1 to 15. The data is then set to be a time series data. It can be clearly seen from the TS 1 graph that a non-linear relationship is present. The graph is analyzed to find appropriate groups that will help understand the relationship between combined LTV and default better and also the impact of combined LTV on the likelihood of default. The adjusted groups of LTV_Group are mentioned in Table 1.

2.1.2 CreditScore:

A credit score is a statistical number that assesses a purchaser's financial creditworthiness and depends on credit history. Credit score ranges from 300-850, and the higher score, indicates that the borrower has timely repaid his obligations in past. A credit score plays a vital role in loan specialist's choice to offer credit.

2.1.2.1 Missing Observations Treatment:

There are 29 missing observations for which CreditScore are not reported which records to under 0.1% of the aggregate number of observations. These observations can be deleted without losing the attributes of the dataset.

2.1.2.2 Grouping:

A new group Credit_Group is created and similar grouping procedure was carried out as mentioned before, for CreditScore also. The TS 2 graph is analyzed to find appropriate groups that will help understand the relationship between CreditScore and default better and also the impact of credit scores on the likelihood of default. The adjusted groups of credit scores are mentioned in Table 2.

2.1.3 DTI_Backened_Orig:

Debt to Income ratio is defined as the ratio of the total debt owed by the purchaser with the total income of the purchaser. It is referred to as DTI_Backened_Orig in our dataset.

2.1.3.1 Missing Observations:

There are 1224 missing observations for DTI_Backened_Orig which records to less than 2% of the of the aggregate number of observations. These observations can be deleted without losing the attributes of the dataset.

2.1.3.2 Grouping:

A new group DTI_Group is created and similar grouping procedure was carried out as mentioned before, for DTI_Backened_Orig also. The TS 3 graph is analyzed to find appropriate groups that will help understand the relationship between DTI ratio and Default better and also the impact of DTI ratio on the likelihood of default. The adjusted groups of DTI_Group are mentioned in Table 3.

2.1.4 Rate_Orig:

The Interest rate or note rate referred to as Rate_orig in our data is rate offered to the borrower as indicated on the mortgage rate. This rate determines the interest the borrower has to pay on their loan balance.

2.1.4.1 Missing Observations and Outliers Treatment:

There are no missing observations for Rate_Orig. So, no treatment is required. Also, no outliers were present.

2.1.4.2 Grouping:

A new group Rate_Group is created and similar grouping procedure was carried out as mentioned before, for Rate_Orig also. The TS 4 graph is analyzed to find appropriate groups that will help understand the relationship between Interest Rate and Default better and also the impact of Interest rate on the likelihood of default. The adjusted groups of Rate_Group are mentioned in Table 4.

2.1.5 NumBorrowers:

NumBorrowers represent the number of borrowers who are obliged to pay a particular loan. A new binary variable NumBorrower_2 is generated and it is set to 0 if the number of borrowers is one and 1 otherwise.

2.1.5.1 Missing Observations Treatment:

There are 23 missing observations for NumBorrowers. These observations can be deleted without losing the attributes of the dataset.

2.1.6 FirstTimeBuyer:

FirstTimeBuyer, as the name suggests, is a self-explanatory categorical variable. According to the data, a borrower who is purchasing a mortgaged property to reside in that property as a residence and has not bought any other property during the three-year period before the purchase of mortgaged property [1] will be considered as the first-time buyer.

2.1.6.1 Grouping:

A new variable NEWBUYER is created which contains a value 1 if the purchaser is first time buyer for a particular observation and 0 otherwise.

2.1.7 STATECODE:

The variable STATECODE is created using metropolitan statistical area (MSA). The value of STATECODE is 1 if the borrower belongs to that particular state otherwise zero for the rest of the states.

2.1.1 Summary Statistics of Categorical Variable:

Categorical Variable	Missing Observation	Tabulation	
NumBorrowers	23	Value=1- 21886 obs	Value=2- 25654 obs
NEWBUYER	3330	Value=Y- 38804 obs	Value=N- 5429 obs

Obs - Observations

2.1.2 Summary Statistics of Continuous Variable:

Continuous Variables	**Missing**	**Mean**	**Median**	**Std. Dev**	**Min**	**Max**
CreditScore	29	724.887	731	57.7894	333	844
LTV_Combined_Orig	2	74.0047	80	18.8969	6	115
DTI_Backend_Orig	1224	36.6369	37	12.4546	1	65
Rate_Orig	0	6.36524	6.375	.426158	3.375	8.875

2.2 Data Partition:

The data used is divided into two parts i.e. Development (70%) and Holdout (30%), using random uniform distribution. The Development sample is used to build the model and Holdout sample is used to validate the performance of the model built [2].

3 Model Description:

3.1 Logistic Regression Description:

The dependent variable of the model is BAD_OVER_48_60. This variable is a categorical variable which takes value 1 that is the loan is considered as default when it is 60 days-past-due or entering foreclosure at any point within 4 years of origination or value 0 that is the purchaser of the loan does not default.

Categorical Variable	Missing Observation	Tabulation	
BAD_OVER_48_60	0	Value=1- 4889 obs	Value=0- 42674 obs

Obs denotes Observations

So, for this dataset logistic regression model is better suitable since the dependent variable, BAD_OVER_48_60 is a categorical variable. Approximately 10% of the borrowers default in our sample.

3.2 Purpose of the model:

The purpose of the model is to build a Mortgage default scorecard to predict the probability that the loan defaults at the time of origination [2].

3.3 Expected Economic Relationship of Independent variables with Dependent Variable

Variable	Expected Relation	Economic Explanation
CreditScore	Negative	A person with higher credit score timely repays his obligations and hence is less likely to default.
LTV_Combined_Orig	Positive	If liabilities of the borrower are more than his assets, he is more likely to skip the payments and hence more likely to default.
DTI_Backend_Orig	Positive	A borrower with high income and less debt is more likely to pay his mortgage payments and hence less likely to default.
Rate_Orig	Positive	Higher interest rate causes higher mortgage payments, which in turn may lead the borrower to skip the payment and increase his chances to default.
NumBorrowers	Negative	The income stream is expected to be more consistent as the number of borrowers goes up and hence they are less likely to default.
NEWBUYER	Positive	A borrower with more than one property has to pay multiple mortgage payments and hence is more likely to skip the payment.

Positive denotes as the value of independent variables increases the probability that the loan default increases and negative denotes inverse relationship.

3.4 Initial Model Specification:

The Logit model was selected since the dependent variable was a categorical variable. Next step was to carefully select independent variables. At the inception stage, CreditScore, LTV_Combined_Orig, DTI_Backened_Orig were considered as independent variables and the logistic regression was ran using these independent variables. KS Statistics was noted. Later, other independent variables were stepwise added to model to increase the predictive accuracy. In each step, the variables were tested for significance. Even though some variables such as the number of units, loan purpose etc. increased the KS statistics, they were dropped since they were found insignificant. Groups were made for certain variables to better understand the effect of variables on the probability of default.

4 Model Output

4.1 Logistic Regression Output

```
                                                   LR chi2(80)       =    3891.75
                                                   Prob > chi2       =     0.0000
Log likelihood = -8852.1381                        Pseudo R2         =     0.1802
```

BAD_OVER_48_60	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.NEWBUYER	-.3487702	.0671723	-5.19	0.000	-.4804255	-.2171149
LTV_Group						
2	.5586081	.1053921	5.30	0.000	.3520434	.7651729
3	1.045069	.0996455	10.49	0.000	.8497679	1.240371
4	1.263453	.1055089	11.97	0.000	1.056659	1.470246
5	1.224007	.0976542	12.53	0.000	1.032608	1.415406
6	1.456165	.118349	12.30	0.000	1.224206	1.688125
7	1.538756	.1044093	14.74	0.000	1.334117	1.743394
8	1.570543	.102275	15.36	0.000	1.370088	1.770998
Credit_Group						
2	.716624	.1166723	6.14	0.000	.4879505	.9452976
3	1.021519	.1458134	7.01	0.000	.7357301	1.307308
4	1.453467	.1223714	11.88	0.000	1.213623	1.69331
5	1.379328	.136473	10.11	0.000	1.111846	1.646811
6	1.800896	.1157235	15.56	0.000	1.574082	2.02771
7	2.148142	.1207421	17.79	0.000	1.911492	2.384792
8	2.328102	.1357144	17.15	0.000	2.062107	2.594097
9	2.760305	.1232957	22.39	0.000	2.51865	3.00196
10	2.846663	.1528144	18.63	0.000	2.547152	3.146173
11	3.111465	.1316321	23.64	0.000	2.853471	3.369459
Rate_Group						
2	.7806827	.6174531	1.26	0.206	-.4295032	1.990869
3	1.007474	.6162365	1.63	0.102	-.200327	2.215276
4	1.35389	.6196301	2.18	0.029	.1394373	2.568343
DTI_Group						
2	.2327686	.0676039	3.44	0.001	.1002674	.3652697
3	.5559271	.0762825	7.29	0.000	.406416	.7054381
4	.6978661	.0786309	8.88	0.000	.5437524	.8519798
5	.5575902	.0901235	6.19	0.000	.3809514	.734229
6	.6636928	.1007613	6.59	0.000	.4662042	.8611813
7	.8385467	.1117454	7.50	0.000	.6195298	1.057564
NumBorrowers_2	-.5563816	.0402613	-13.82	0.000	-.6352923	-.4774709
STATECODE AK	-.8734223	.8051104	-1.08	0.278	-2.45141	.7045651

4.2 Average Marginal Effects

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.NEWBUYER	-.0254663	.0045045	-5.65	0.000	-.0342948	-.0166377
LTV_Group						
2	.0250031	.0045701	5.47	0.000	.0160459	.0339603
3	.0565023	.0049242	11.47	0.000	.046851	.0661536
4	.0741711	.0060676	12.22	0.000	.0622788	.0860633
5	.0708052	.0048557	14.58	0.000	.0612881	.0803222
6	.0917629	.0082861	11.07	0.000	.0755224	.1080033
7	.0998994	.0064615	15.46	0.000	.087235	.1125638
8	.1031284	.0060572	17.03	0.000	.0912566	.1150003
Credit_Group						
2	.0244384	.0033946	7.20	0.000	.0177851	.0310916
3	.0401821	.0063429	6.33	0.000	.0277502	.0526139
4	.0698614	.0054666	12.78	0.000	.0591471	.0805757
5	.0640769	.0069745	9.19	0.000	.0504071	.0777467
6	.1012477	.0052103	19.43	0.000	.0910358	.1114597
7	.1402964	.0074351	18.87	0.000	.1257239	.1548689
8	.1637575	.0117199	13.97	0.000	.1407869	.186728
9	.2292327	.0104674	21.90	0.000	.208717	.2497484
10	.2438157	.0189061	12.90	0.000	.2067604	.280871
11	.2913889	.0145261	20.06	0.000	.2629183	.3198595
Rate_Group						
2	.0420905	.0247963	1.70	0.090	-.0065093	.0906903
3	.0591213	.024655	2.40	0.016	.0107984	.1074441
4	.0900469	.0255222	3.53	0.000	.0400243	.1400694
DTI_Group						
2	.015607	.0043607	3.58	0.000	.0070601	.0241539
3	.041528	.0056078	7.41	0.000	.0305369	.0525191
4	.0545962	.0061609	8.86	0.000	.0425211	.0666713
5	.041675	.0070393	5.92	0.000	.0278783	.0554717
6	.0513525	.0085006	6.04	0.000	.0346916	.0680133
7	.0686129	.0104713	6.55	0.000	.0480896	.0891362
NumBorrowers_2	-.0440391	.0031804	-13.85	0.000	-.0502727	-.0378056
STATECODE_AK	-.0691337	.0637282	-1.08	0.278	-.1940387	.0557712
STATECODE_AL	.0049668	.0441744	0.11	0.910	-.0816134	.091547
STATECODE_AR	.0164592	.0455506	0.36	0.718	-.0728184	.1057367
STATECODE_AZ	.1314255	.0426851	3.08	0.002	.0477642	.2150868
STATECODE_CA	.1322582	.0423481	3.12	0.002	.0492575	.215259

4.3 Interpretation of regression Output:

The output of logistic regression was as expected. All the variables are significant even at 1% level of confidence. The coefficients of logistic regression are difficult to interpret so the coefficients are interpreted by calculating the marginal effects.

- **NEWBUYER**

The non-first-time buyer is the reference group. The marginal effect for NEWBUYER is negative which explains that a first-time home buyer is 2.5% less likely to default as compared to a purchaser who is not a first-time buyer.

- **LTV_Group**

Group 1, for which the combined LTV is less than equal to 50% is the reference group. The marginal effect is positive which explains that the purchaser is more likely to default as his Loan-to-Value ratio increases. If a purchaser's combined LTV lies between 50% and 65% (group 2) then he is 2.5% more likely to default as compared to a purchaser whose Combined LTV is less than 50%. Similarly, the marginal effect of six other groups can be interpreted with group 1 (combined LTV < 50%) as the reference group.

- **Credit_Group**

Group 1, for which the credit score is more than 790 is the reference group. The marginal effect is positive which explains that as the purchaser's credit score decreases he is more likely to default. A person with a credit less than 610 (group 11) is approximately 29% more likely to default than a person whose credit score is more than 790. Similarly, the marginal effect of 9 other groups can be interpreted with group 1 as the reference group.

- **Rate_Group**

Group 1, for which the interest rate is less than equal to 5% is the reference group. The marginal effect is positive which explains that as the interest rate on a loan increases the purchaser is more likely to default. A loan with an interest rate of more than 7% is 9% more likely to default than a loan with an interest rate of less than 5%. Similarly, the marginal effect of two other groups can be interpreted with group 1 as the reference group.

- **DTI_Group**

Group 1, for which the DTI rate is less than equal to 25% is the reference group. The marginal effect is positive which explains that as the Debt to Income ratio increases the purchaser is more likely to default. A borrower whose DTI rate lies between 60% to 65% is approximately 7% more likely to default than a borrower whose DTI rate is less than 25%. Similarly, the marginal effect of five other groups can be interpreted with group 1 as the reference group.

- **NumBorrowers_2**

The Number of borrowers equal to 1 is the reference group. The marginal effect is negative which explains that if more than one borrower is obligated to pay back for a particular loan then they are 4.4% less likely to default as compared to the loan being purchased by a single borrower.

❖ STATECODE

The reference group for STATECODE changes randomly for each particular state. Thus, a borrower from Alabama State is 0.4% more likely to default than a borrower from the reference state. A Similar interpretation can be done for the rest of the states.

5 Model Performance Testing

5.1 KS of Development vs Holdout

The Kolmogorov-Smirnov (K-S) statistic for the development sample is 0.4603 and the K-S statistic for the holdout sample is 0.4548. There is a very small difference between this two values and hence can be neglected.

5.2 Results from Out of Sample testing

The K-S statistic for the out of sample data is 0.4583 which is very close to the K-S statistic for the development sample. Thus, the model is working well on out of sample data in terms of distinguishing between default and no-default.

5.3 Predictive accuracy assessment table

The table of deciles with Hosmer Lemeshow test for development and holdout sample is as follows.

Development Sample

Logistic model for BAD OVER 48 60, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0119	17	23.8	3231	3224.2	3248
2	0.0208	42	52.9	3206	3195.1	3248
3	0.0310	65	83.0	3183	3165.0	3248
4	0.0438	129	120.0	3118	3127.0	3247
5	0.0603	151	167.2	3097	3080.8	3248
6	0.0826	240	231.0	3009	3018.0	3249
7	0.1144	338	317.8	2909	2929.2	3247
8	0.1642	455	447.3	2792	2799.7	3247
9	0.2561	718	664.9	2530	2583.1	3248
10	0.8452	1203	1250.1	2044	1996.9	3247

number of observations = 32477
number of groups = 10
Hosmer-Lemeshow chi2(8) = 20.81
Prob > chi2 = 0.0077

Holdout Sample

Logistic model for BAD OVER 48 60, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0126	7	10.6	1374	1370.4	1381
2	0.0213	10	23.2	1370	1356.8	1380
3	0.0314	32	36.1	1349	1344.9	1381
4	0.0441	45	51.9	1335	1328.1	1380
5	0.0603	70	71.7	1311	1309.3	1381
6	0.0824	105	97.6	1275	1282.4	1380
7	0.1151	147	135.0	1234	1246.0	1381
8	0.1629	194	187.7	1186	1192.3	1380
9	0.2530	292	278.4	1089	1102.6	1381
10	0.7874	483	526.0	897	854.0	1380

number of observations = 13805
number of groups = 10
Hosmer-Lemeshow chi2(10) = 18.93
Prob > chi2 = 0.0412

The null hypothesis for development sample can be rejected at 1% level and the null hypothesis for holdout sample can be rejected at 5% level. Rejecting the null hypothesis for the HL test shows that there are significant differences between the fitted and observed values. But since our dataset is large, there is a trivial difference between the number of observed good's and expected good's for both the sample. Also, the difference between the number of observed bad's and expected bad's for both the sample is not that significant.

6 Conclusion

A logistic model has the adaptability of incorporating both the qualitative and quantitative variables and hence is better suited to build the Mortgage Default Scorecard. The independent variables used to build the model are statistically significant. The logit models are non-linear and hence Maximum Likelihood Estimation (MLE) is used to estimate the parameters. Interpreting the output of Logit model is complicated, so it's better to calculate the marginal effects and then interpret the coefficients. The K-S Statistic of Development sample and the Out of sample data is almost similar which suggests that the model is working well in distinguishing whether the loan defaults or does not default. The results of the model are economically meaningful.

7 Reference

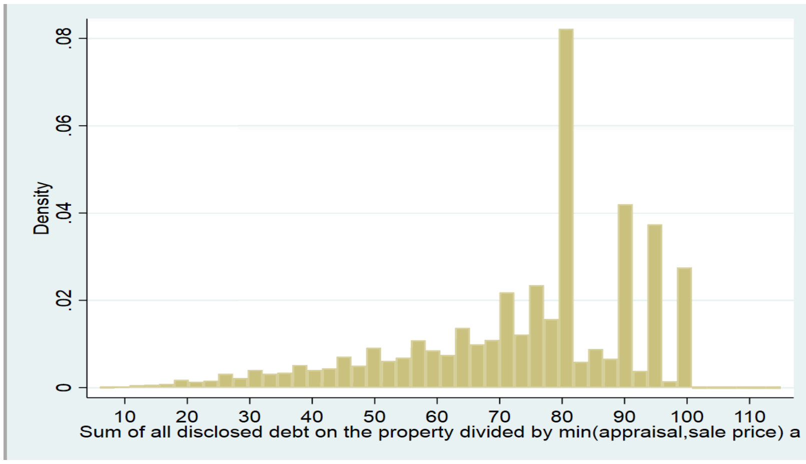
[1] http://www.freddiemac.com/research/pdf/faq.pdf

[2] http://www.freddiemac.com/research/pdf/release_notes.pdf

8 Appendix

8.1 LTV_Group

8.1.1 Histogram Plot 1



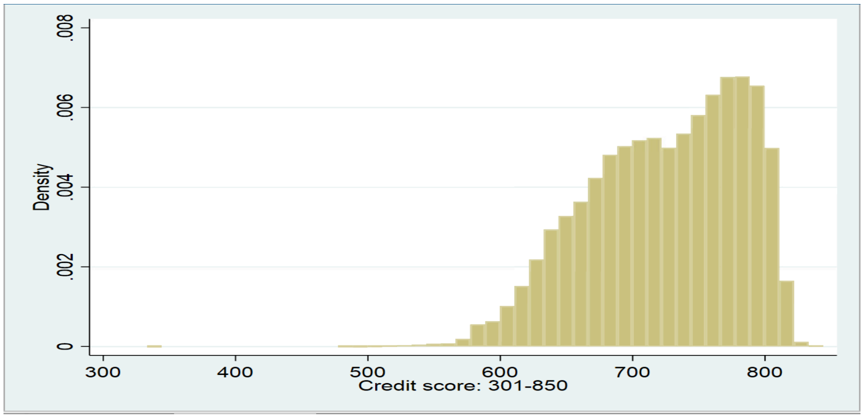
8.1.2 Adjusted Groups of LTV_Group

LTV_Group	Combined LTV Range (%)
Group 1	<50
Group 2	50-65
Group 3	65-75
Group 4	75-80
Group 5	80-85
Group 6	85-90
Group 7	90-95
Group 8	>=90

8.1.3 Distribution of LTV_Group with respect to the ratio of defaults to the total number of defaults (TS 1)

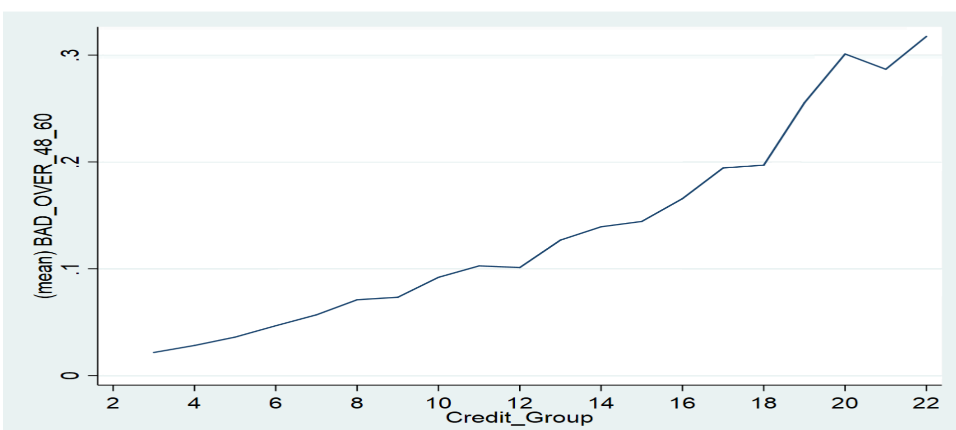
8.2 Credit_Group

8.2.1 Histogram Plot 2



8.2.2 Adjusted Groups of Credit_Group

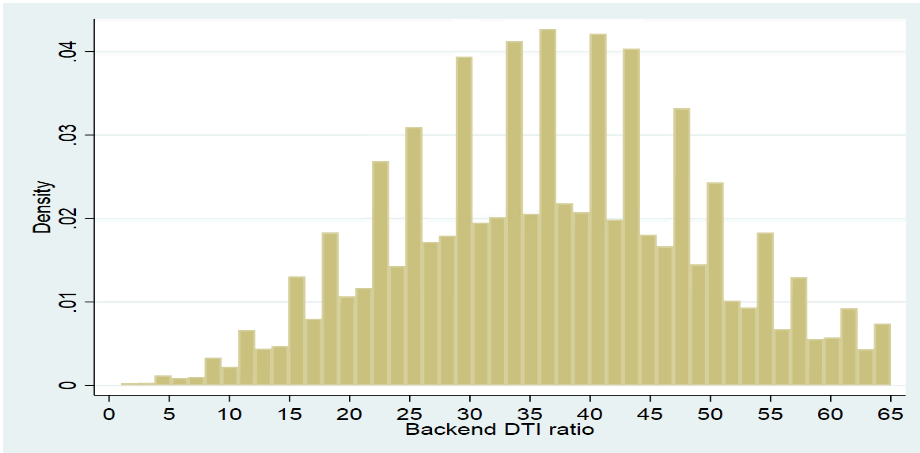
Credit_Group	CreditScore Range
Group 1	>790
Group 2	740-790
Group 3	730-740
Group 4	710-730
Group 5	700-710
Group 6	670-700
Group 7	650-670
Group 8	640-650
Group 9	620-640
Group 10	610-620
Group 11	<=610



8.2.3 Distribution of Credit_Group with respect to the ratio of defaults to the total number of defaults (TS 2)

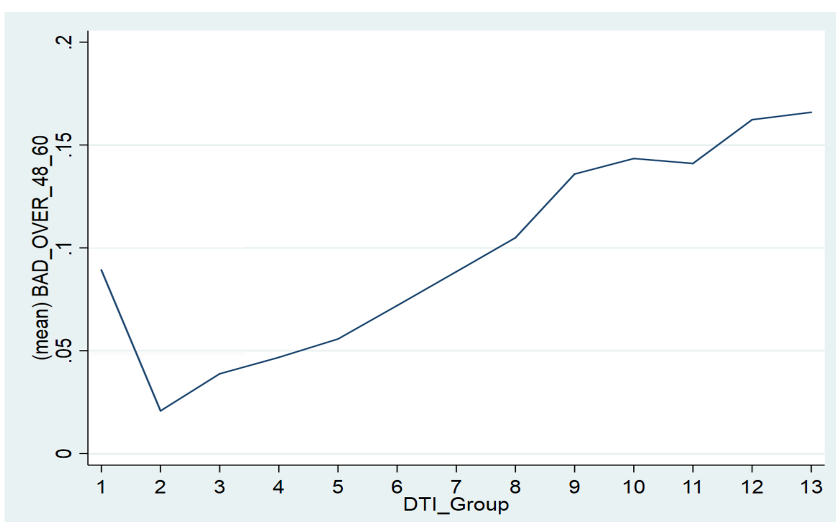
8.3 DTI_Group

8.3.1 Histogram Plot 3



8.3.2 Adjusted Groups of DTI_Group

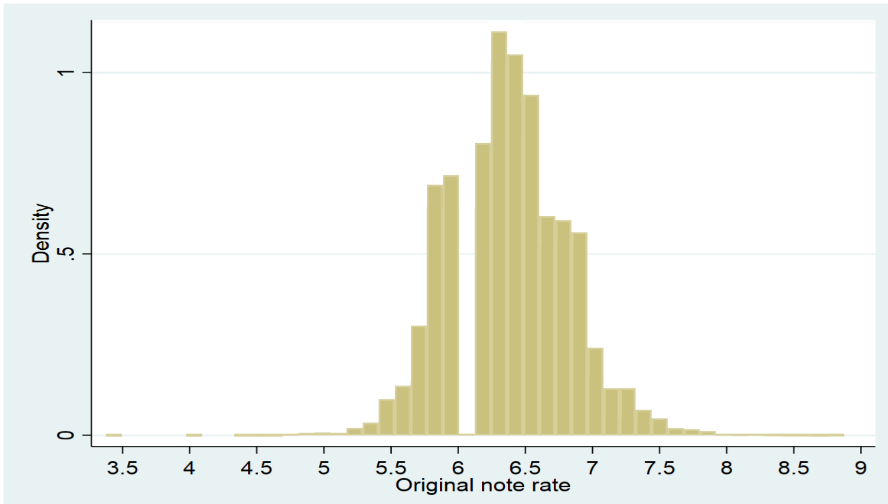
DTI_Group	DTI Range (%)
Group 1	<=25
Group 2	25-40
Group 3	40-45
Group 4	45-50
Group 5	50-55
Group6	55-60
Group 7	60-65



8.3.3 Distribution of DTI_Group with respect to the ratio of defaults to the total number of defaults (TS 3)

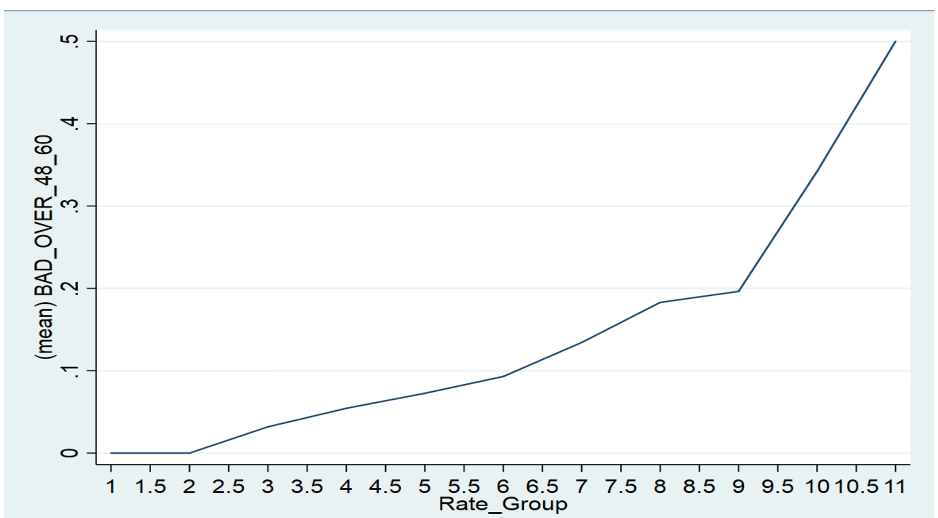
8.4 Rate_Group

8.4.1 Histogram Plot 3



8.4.2 Adjusted Groups of Rate_Group

Rate_Group	Rate Range (%)
Group 1	<=5
Group 2	5-6
Group 3	6-7
Group 4	>7



8.4.3 Distribution of Rate_Group with respect to the ratio of defaults to the total number of defaults (TS 4)

8.5 Results

8.5.1 Logistic Regression Output

	LR chi2(80)	=	3891.75
	Prob > chi2	=	0.0000
Log likelihood = -8852.1381	Pseudo R2	=	0.1802

BAD_OVER_48_60	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.NEWBUYER	-.3487702	.0671723	-5.19	0.000	-.4804255	-.2171149
LTV_Group						
2	.5586081	.1053921	5.30	0.000	.3520434	.7651729
3	1.045069	.0996455	10.49	0.000	.8497679	1.240371
4	1.263453	.1055089	11.97	0.000	1.056659	1.470246
5	1.224007	.0976542	12.53	0.000	1.032608	1.415406
6	1.456165	.118349	12.30	0.000	1.224206	1.688125
7	1.538756	.1044093	14.74	0.000	1.334117	1.743394
8	1.570543	.102275	15.36	0.000	1.370088	1.770998
Credit_Group						
2	.716624	.1166723	6.14	0.000	.4879505	.9452976
3	1.021519	.1458134	7.01	0.000	.7357301	1.307308
4	1.453467	.1223714	11.88	0.000	1.213623	1.69331
5	1.379328	.136473	10.11	0.000	1.111846	1.646811
6	1.800896	.1157235	15.56	0.000	1.574082	2.02771
7	2.148142	.1207421	17.79	0.000	1.911492	2.384792
8	2.328102	.1357144	17.15	0.000	2.062107	2.594097
9	2.760305	.1232957	22.39	0.000	2.51865	3.00196
10	2.846663	.1528144	18.63	0.000	2.547152	3.146173
11	3.111465	.1316321	23.64	0.000	2.853471	3.369459
Rate_Group						
2	.7806827	.6174531	1.26	0.206	-.4295032	1.990869
3	1.007474	.6162365	1.63	0.102	-.200327	2.215276
4	1.35389	.6196301	2.18	0.029	.1394373	2.568343
DTI_Group						
2	.2327686	.0676039	3.44	0.001	.1002674	.3652697
3	.5559271	.0762825	7.29	0.000	.406416	.7054381
4	.6978661	.0786309	8.88	0.000	.5437524	.8519798
5	.5575902	.0901235	6.19	0.000	.3809514	.734229
6	.6636928	.1007613	6.59	0.000	.4662042	.8611813
7	.8385467	.1117454	7.50	0.000	.6195298	1.057564
NumBorrowers_2	-.5563816	.0402613	-13.82	0.000	-.6352923	-.4774709
STATECODE_AK	-.8734223	.8051104	-1.08	0.278	-2.45141	.7045651

STATECODE_KY	.010201	.5606602	0.02	0.985	-1.088673	1.109075
STATECODE_LA	.3412477	.5590886	0.61	0.542	-.7545458	1.437041
STATECODE_MA	.2831578	.5512042	0.51	0.607	-.7971826	1.363498
STATECODE_MD	.8095642	.5431014	1.49	0.136	-.254895	1.874023
STATECODE_ME	.3232565	.606417	0.53	0.594	-.865299	1.511812
STATECODE_MI	.3356414	.5440161	0.62	0.537	-.7306106	1.401893
STATECODE_MN	.3677259	.5527344	0.67	0.506	-.7156136	1.451065
STATECODE_MO	-.0946493	.551101	-0.17	0.864	-1.174787	.9854888
STATECODE_MS	-.1514342	.6090066	-0.25	0.804	-1.345065	1.042197
STATECODE_MT	-.2017382	.6655508	-0.30	0.762	-1.506194	1.102717
STATECODE_NC	.0444966	.5436361	0.08	0.935	-1.02101	1.110004
STATECODE_ND	-.220613	.8059996	-0.27	0.784	-1.800343	1.359117
STATECODE_NE	-.2468813	.6217207	-0.40	0.691	-1.465431	.9716688
STATECODE_NH	.6046759	.5945881	1.02	0.309	-.5606953	1.770047
STATECODE_NJ	.9764537	.5421544	1.80	0.072	-.0861494	2.039057
STATECODE_NM	-.3558802	.6016098	-0.59	0.554	-1.535014	.8232534
STATECODE_NV	1.996425	.5513692	3.62	0.000	.9157614	3.077089
STATECODE_NY	.8412137	.5390441	1.56	0.119	-.2152933	1.897721
STATECODE_OH	-.088285	.5450045	-0.16	0.871	-1.156474	.9799041
STATECODE_OK	-.5739606	.5932478	-0.97	0.333	-1.736705	.5887836
STATECODE_OR	.4992035	.5556514	0.90	0.369	-.5898532	1.58826
STATECODE_PA	.0906745	.5429584	0.17	0.867	-.9735044	1.154853
STATECODE_PR	.9502994	.5814518	1.63	0.102	-.1893253	2.089924
STATECODE_RI	1.399026	.5880991	2.38	0.017	.2463725	2.551679
STATECODE_SC	.1830491	.5530196	0.33	0.741	-.9008495	1.266948
STATECODE_SD	-.6023902	.9071822	-0.66	0.507	-2.380435	1.175654
STATECODE_TN	.1327014	.5514303	0.24	0.810	-.9480822	1.213485
STATECODE_TX	-.3834994	.5402915	-0.71	0.478	-1.442451	.6754524
STATECODE_UT	.2485612	.5545422	0.45	0.654	-.8383215	1.335444
STATECODE_VA	.21524	.5483231	0.39	0.695	-.8594534	1.289934
STATECODE_VI	0	(omitted)				
STATECODE_VT	.0536085	.6858908	0.08	0.938	-1.290713	1.39793
STATECODE_WA	.4781146	.5442296	0.88	0.380	-.5885557	1.544785
STATECODE_WI	-.1736744	.5535155	-0.31	0.754	-1.258545	.911196
STATECODE_WV	.1233606	.6112299	0.20	0.840	-1.074628	1.321349
STATECODE_WY	0	(omitted)				
_cons	-6.473972	.825825	-7.84	0.000	-8.092559	-4.855384

8.5.2 Average Marginal Effects

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.NEWBUYER	-.0254663	.0045045	-5.65	0.000	-.0342948	-.0166377
LTV_Group						
2	.0250031	.0045701	5.47	0.000	.0160459	.0339603
3	.0565023	.0049242	11.47	0.000	.046851	.0661536
4	.0741711	.0060676	12.22	0.000	.0622788	.0860633
5	.0708052	.0048557	14.58	0.000	.0612881	.0803222
6	.0917629	.0082861	11.07	0.000	.0755224	.1080033
7	.0998994	.0064615	15.46	0.000	.087235	.1125638
8	.1031284	.0060572	17.03	0.000	.0912566	.1150003
Credit_Group						
2	.0244384	.0033946	7.20	0.000	.0177851	.0310916
3	.0401821	.0063429	6.33	0.000	.0277502	.0526139
4	.0698614	.0054666	12.78	0.000	.0591471	.0805757
5	.0640769	.0069745	9.19	0.000	.0504071	.0777467
6	.1012477	.0052103	19.43	0.000	.0910358	.1114597
7	.1402964	.0074351	18.87	0.000	.1257239	.1548689
8	.1637575	.0117199	13.97	0.000	.1407869	.186728
9	.2292327	.0104674	21.90	0.000	.208717	.2497484
10	.2438157	.0189061	12.90	0.000	.2067604	.280871
11	.2913889	.0145261	20.06	0.000	.2629183	.3198595
Rate_Group						
2	.0420905	.0247963	1.70	0.090	-.0065093	.0906903
3	.0591213	.024655	2.40	0.016	.0107984	.1074441
4	.0900469	.0255222	3.53	0.000	.0400243	.1400694
DTI_Group						
2	.015607	.0043607	3.58	0.000	.0070601	.0241539
3	.041528	.0056078	7.41	0.000	.0305369	.0525191
4	.0545962	.0061609	8.86	0.000	.0425211	.0666713
5	.041675	.0070393	5.92	0.000	.0278783	.0554717
6	.0513525	.0085006	6.04	0.000	.0346916	.0680133
7	.0686129	.0104713	6.55	0.000	.0480896	.0891362
NumBorrowers_2	-.0440391	.0031804	-13.85	0.000	-.0502727	-.0378056
STATECODE_AK	-.0691337	.0637282	-1.08	0.278	-.1940387	.0557712
STATECODE_AL	.0049668	.0441744	0.11	0.910	-.0816134	.091547
STATECODE_AR	.0164592	.0455506	0.36	0.718	-.0728184	.1057367
STATECODE_AZ	.1314255	.0426851	3.08	0.002	.0477642	.2150868
STATECODE_CA	.1322582	.0423481	3.12	0.002	.0492575	.215259

STATECODE_KS	-.0442891	.0484676	-0.91	0.361	-.1392838	.0507056
STATECODE_KY	.0008074	.0443778	0.02	0.985	-.0861714	.0877863
STATECODE_LA	.0270107	.0442533	0.61	0.542	-.0597242	.1137456
STATECODE_MA	.0224127	.0436291	0.51	0.607	-.0630987	.1079241
STATECODE_MD	.0640792	.0429864	1.49	0.136	-.0201726	.148331
STATECODE_ME	.0255866	.0479994	0.53	0.594	-.0684905	.1196638
STATECODE_MI	.0265669	.0430601	0.62	0.537	-.0578293	.1109632
STATECODE_MN	.0291065	.0437501	0.67	0.506	-.0566421	.1148551
STATECODE_MO	-.0074918	.0436212	-0.17	0.864	-.0929877	.0780042
STATECODE_MS	-.0119864	.0482045	-0.25	0.804	-.1064656	.0824927
STATECODE_MT	-.0159681	.0526803	-0.30	0.762	-.1192196	.0872833
STATECODE_NC	.003522	.0430302	0.08	0.935	-.0808157	.0878598
STATECODE_ND	-.0174621	.0637974	-0.27	0.784	-.1425027	.1075784
STATECODE_NE	-.0195413	.049211	-0.40	0.691	-.1159931	.0769105
STATECODE_NH	.0478617	.047063	1.02	0.309	-.0443801	.1401036
STATECODE_NJ	.077289	.0429105	1.80	0.072	-.006814	.1613919
STATECODE_NM	-.0281689	.0476192	-0.59	0.554	-.1215008	.065163
STATECODE_NV	.1580225	.0436257	3.62	0.000	.0725176	.2435274
STATECODE_NY	.0665843	.0426652	1.56	0.119	-.0170379	.1502066
STATECODE_OH	-.006988	.0431386	-0.16	0.871	-.0915381	.0775621
STATECODE_OK	-.0454305	.0469578	-0.97	0.333	-.1374661	.046605
STATECODE_OR	.0395133	.0439806	0.90	0.369	-.0466871	.1257137
STATECODE_PA	.0071771	.0429766	0.17	0.867	-.0770555	.0914098
STATECODE_PR	.0752188	.0460221	1.63	0.102	-.0149829	.1654204
STATECODE_RI	.1107367	.0465444	2.38	0.017	.0195112	.2019621
STATECODE_SC	.0144888	.0437729	0.33	0.741	-.0713045	.1002822
STATECODE_SD	-.0476808	.0718071	-0.66	0.507	-.1884202	.0930586
STATECODE_TN	.0105037	.0436472	0.24	0.810	-.0750432	.0960506
STATECODE_TX	-.030355	.0427657	-0.71	0.478	-.1141743	.0534643
STATECODE_UT	.0196743	.0438934	0.45	0.654	-.0663552	.1057038
STATECODE_VA	.0170368	.0434012	0.39	0.695	-.0680279	.1021015
STATECODE_VI	0	(omitted)				
STATECODE_VT	.0042433	.0542901	0.08	0.938	-.1021633	.1106499
STATECODE_WA	.0378441	.0430768	0.88	0.380	-.046585	.1222731
STATECODE_WI	-.0137468	.0438124	-0.31	0.754	-.0996174	.0721238
STATECODE_WV	.0097643	.0483805	0.20	0.840	-.0850597	.1045883
STATECODE_WY	0	(omitted)				

8.5.3 K-S Statistic of Development Sample

Two-sample Kolmogorov-Smirnov test for equality of distribution functions

Smaller group	D	P-value
0:	0.4603	0.000
1:	-0.0000	1.000
Combined K-S:	0.4603	0.000

8.5.4 K-S Statistic of Holdout Sample

```
Two-sample Kolmogorov-Smirnov test for equality of distribution functions

 Smaller group          D        P-value
 ----------------------------------------
 0:                     0.4548     0.000
 1:                     0.0000     1.000
 Combined K-S:          0.4548     0.000
```

8.5.5 K-S Statistic of Out of Sample data

```
Two-sample Kolmogorov-Smirnov test for equality of distribution functions

 Smaller group          D        P-value
 ----------------------------------------
 0:                     0.4583     0.000
 1:                    -0.0000     1.000
 Combined K-S:          0.4583     0.000
```

8.6 Predictive accuracy assessment table

8.6.1 Development Sample

Logistic model for BAD OVER 48 60, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0119	17	23.8	3231	3224.2	3248
2	0.0208	42	52.9	3206	3195.1	3248
3	0.0310	65	83.0	3183	3165.0	3248
4	0.0438	129	120.0	3118	3127.0	3247
5	0.0603	151	167.2	3097	3080.8	3248
6	0.0826	240	231.0	3009	3018.0	3249
7	0.1144	338	317.8	2909	2929.2	3247
8	0.1642	455	447.3	2792	2799.7	3247
9	0.2561	718	664.9	2530	2583.1	3248
10	0.8452	1203	1250.1	2044	1996.9	3247

```
      number of observations =      32477
            number of groups =         10
      Hosmer-Lemeshow chi2(8) =      20.81
                 Prob > chi2 =      0.0077
```

8.6.2 Holdout Sample

Logistic model for BAD OVER 48 60, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0126	7	10.6	1374	1370.4	1381
2	0.0213	10	23.2	1370	1356.8	1380
3	0.0314	32	36.1	1349	1344.9	1381
4	0.0441	45	51.9	1335	1328.1	1380
5	0.0603	70	71.7	1311	1309.3	1381
6	0.0824	105	97.6	1275	1282.4	1380
7	0.1151	147	135.0	1234	1246.0	1381
8	0.1629	194	187.7	1186	1192.3	1380
9	0.2530	292	278.4	1089	1102.6	1381
10	0.7874	483	526.0	897	854.0	1380

```
      number of observations =     13805
            number of groups =        10
     Hosmer-Lemeshow chi2(10) =        18.93
                 Prob > chi2 =         0.0412
```

==