Multifamily Loan Credit Risk

Robbie Walsh

11/28/2022

# Executive Summary

When people think of housing in the United States, they often think of single-family homes, but close to 30% of housing in the US today is multifamily. Multifamily housing is defined as apartment buildings, condominiums, townhouses, and mixed-use developments. While small in comparison to the single-family mortgage market, the market for multifamily loans in nonetheless impressive. According to the Mortgage Bankers Association, total multifamily debt originated in 2021 totaled $487 billion. Whenever any type of loan is made, there is a risk that the borrower will not repay the loan and default, this risk is referred to in the financial industry as credit risk. Lenders need to be aware of how much they stand to lose on loans that they have originated, both for business purposes, and for regulatory purposes.

One crucial factor in determining expected credit losses is the broader economic environment. Characteristics of the borrower and the loan terms impact the credit risk associated with a loan, but the lender knows this information up front, the credit risk associated with the broader economy is more uncertain to the lender. When considering credit risk, it is important to consider both borrower/loan characteristics and the economy.

This study will explore how borrower and loan characteristics as well as the broader economic environment impact multifamily loan defaults, allowing for a quantification of credit risk. The goal of this study will be to produce a statistical model that can forecast multifamily default rates, at the loan level, given certain assumptions about the economy. Such a model could be used to provide expected credit loss estimates on multifamily loan portfolios that would be of interest to multifamily lenders, particularly those without sufficient credit history to build their own model.

The results of the study show that both borrower/loan characteristics and the broader economy are important predictors of credit losses, however, as banks have little control over the broader economy, adequate cash buffers should be maintained as credit losses rise significantly during recessionary periods.

# Project Description and Research Questions

The client is a mid-sized bank looking to break into the multifamily lending market. As the bank does not have any historical performance data on multifamily loans, they are looking for a statistical model that will forecast loan defaults on typical multifamily loans, given a few key inputs and assumptions about the economy. The bank is interested in explaining performance using just several key variables, while maintaining maximum explainability for regulatory purposes. This model will ultimately be used for forecasting profit and loss in the bank’s emerging multifamily portfolio, as well as assisting with key estimates that need to be produced for several regulations including CCAR/DFAST and CECL. CCAR/DFAST are regulations that require banks to assess how their loan portfolios will perform in a recession and CECL is an accounting standard specifying how banks treat their expectations for credit losses.

Since data was not available from the bank, industry data was sourced from publicly available sources. This project will use the Fannie Mae Multifamily Loan Performance Data (MFLPD), which is a publicly available performance dataset on Fannie Mae’s Delegated Underwriting Service (DUS) for multifamily loans. The dataset shows active and liquidated multifamily loans acquired by Fannie Mae from January 2000 or when loan products were initiated. Publicly available economic data from FRED (Federal Reserve Economic Database) maintained by the Federal Reserve Bank of St. Louis will also be incorporated. The data contains monthly observations from 2000m1 through 2022m3 on 61,048 loans, which makes for a dataset that contains 4,078,435 observations.

This project will address the following questions:

*• What are the key borrower/loan characteristics that influence the default rate on multifamily loans and what is their directional impact? This can be answered with a regression analysis and graphical exploratory data analysis.*

*• What macroeconomic variables can be used to explain the impact of the economy on the multifamily loan market and what is their relative impact? This can be answered with a regression analysis.*

*• What is the model error that can be expected in any month? This can be answered by calculating error metrics from the predictions of the statistical model.*

# Variables of Interest

## Fannie Mae Data

The dataset used contains many variables, the important variables considered in this study are outlined below.

**Loan Number** - This variable is unique for each loan, which is observed every month. This variable can be thought of as a subject identifier variable and allows us to keep track of individual loans.

**Maturity Date at Acquisition** - The date at which the loan is scheduled to mature when Fannie Mae acquired the loan. This can be used (along with the term length) to infer the origination date of the loan.

**Amortization Term** - The number of months over which the loan’s unpaid balance is scheduled to be amortized. This gives the number of months that a loan is scheduled to last for, barring default, prepayment, or modifications to the loan.

**Original Interest Rate** - The original interest rate on the loan, specified in percentage (int\_rate\_o).

**Issue Date** - The first day of the month in which the the security was issued.

**Underwritten DSCR** - A ratio of underwritten net cash flow to an actual or calculated principal and interest payments (underwritten\_dscr, debt service coverage ratio).

**Liquidation/Prepayment Code** - Indicates the type of scheduled or unscheduled principal collection or liquidation, will be used to determine when loans enter into default.

**Reporting Period Date** - The month and year that the loan is observed in, this will be used as the time variable.

**Loan Payment Status** - Indicates the status of the loan payment, whether it is current or is in a stage of delinquency. This variable will also be used.

**LTV at Acquisition** - The loan to value ratio (ratio of the balance of the loan to the value of the home collateralizing the loan) when Fannie Mae acquired the loan (ltv\_acquisition).

The variable we will be modeling will be an indicator variable indicating whether the loan entered into default. Default will be defined as the first date the loan reaches 90+ DPD (using the loan payment status variable) or the loan enters into foreclosure (using the liquidation/prepayment code variable). DPD stands for days past due, which is the number of days late that a borrower is late on their loan payment.

## Economic Variables

The economic variables were sourced from the FRED (Federal Reserve Economic Database), which is maintained by the Federal Reserve Bank of St Louis and contains hundreds of economic series at a variety of frequencies. Several economic variables that are known to impact residential loan performance were selected as candidate variables. All variables are measured at the national level, meaning that all quantities measured are measured on the entire United States. The variables are listed as follows, all on a monthly frequency:

**Disposable income** - The income level of the average American household after taxes and other deductions, adjusted for inflation (i.e., real disposable income) and seasonally adjusted (dispinc\_t).

**Mortgage interest rate** - The average interest rate paid on a 30-year fixed-rate mortgage, not seasonally adjusted (mort\_rate\_t).

**HPI (house price index)** - The median sale price of existing homes, not seasonally adjusted (medhpi\_t).

**Unemployment rate** - The percentage of the workforce (people actively looking for work) not employed, seasonally adjusted (ump\_t).

**Normalized, seasonally adjusted GDP (gross domestic product)** - This is the gross domestic product (sum of the value of all goods and services produced in the economy) seasonally adjusted, with the trend removed. The removal of the trend ensures stationarity and accounts for shifts in GDP as opposed to the level (gdp\_t).

## Created Variables

Additional created variables are defined as follows:

**Loan Age** - The number of months that the loan has existed since origination. This variable is calculated using the reporting date variable and the issue date variable and using datetime math (loan\_age).

**Loan Age Percentage** - The percentage of a loan’s amortization term that it has progressed through at a particular time. This is calculated by taking the loan age and dividing it by the amortization term.

**Unemployment change since origination** - The change in the unemployment rate between the observation date and the origination date of the loan. This difference is taken to better incorporate changing economic condition’s impact on loans and to ensure stationarity (ump\_vt).

**HPI percent change since origination** - The percentage change in the median home price between the observation date and the origination date of the loan, expressed as a percentage. The percentage change is taken to better incorporate the impact of changing home equity on loan performance and to ensure stationarity (hpi\_vt).

**Mortgage interest rate change since origination** - The change in the mortgage interest rate between the observation date and the origination date of the loan. The difference is taken to better incorporate changing interest rate’s impact on a particular loand and to ensure stationarity (mort\_rate\_vt).

## Additional Data Cleaning

Not much data cleaning is needed, there are a few loans that do not have an origination date listed, but we are able to figure out what it is based on the amortization term and maturity date of the loan. A ceiling was applied to the loan age percentage variable at 1, as values above 1 should not be possible (loans are modified which is what causes this) and loan behavior is volatile beyond that point. A ceiling is also applied to underwritten dscr at 3, as there were few loans with values above that point and behavior was volatile above that point.

Data after 2020m3 was excluded to remove the impact of the COVID19 pandemic on the model. The COVID19 pandemic caused the reported default rate to spike upwards, as many borrowers went 90+ DPD, but due to government the lender programs, many borrowers that would have otherwise defaulted did not. As this outcome is counter to the usual relationship we see between the default rate and traditionally used economic variables, this time period will be excluded so as not to complicate the analysis.

The data was also split into a development (also referred to as the in-sample dataset) and testing (sometimes referred to as out-of-time although this is not totally correct) dataset. First, the last year and a half of data (from 2018m9 to 2020m3) was taken as an out-of-time sample and saved. Next, a 70% sample of LoanIDs was taken from the remaining dataset to be used as the development dataset. The last 30% of LoanIDs was combined with the out-of-time sample to create a testing set. Note that the testing set contains both in-time but out-of-sample loans, and out-of-time observations.

# Exploratory Data Analysis

The table below gives some key summary stats for the data.

**Table 1: Overview of Development Data**

|  |  |
| --- | --- |
| **Field** | **Value** |
| Number of Loans | 36,548 |
| Number of Observations | 2,430,301 |
| Date Range | 2000m1 – 2018m9 |
| Origination Vintage Range | 1971m8 – 2018m9 |

Table 1 shows that the data contains many loans, and many observations per loan, which will allow the model we build to adequately represent typical multifamily loans. We can see that the origination vintage (origination date) of the loans goes as far back as the early 1970s, which means that the data contains loans that were originated in a variety of economic environments and contains observations on loans that are well into their amortization schedule. The observation date range encompassed in the data goes as far back as 2000, allowing us to capture the full span of the business cycle. There is data on time periods before a recession, during a recession (2007/08 financial crisis) as well as the recovery from a recession.

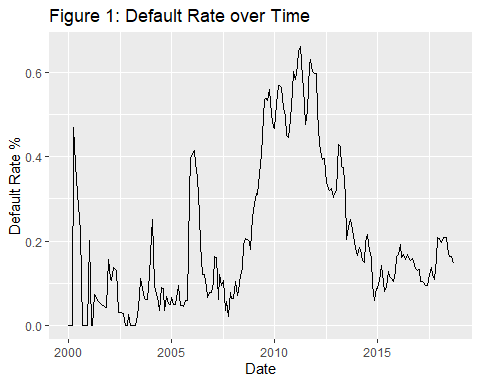
Table 2 presents summary statistics of key quantitative variables that will be considered as variables to predict the default rate.

**Table 2: Summary Statistics for Predictor Variables**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Missing** | **Minimum** | **Maximum** | **Mean** |
| Int\_rate\_o | 0 | 1.22 | 9.8 | 5.54 |
| Underwritten\_dscr | 0 | 0.39 | 3 | 1.57 |
| Ltv\_acquisition | 0 | 0 | 100 | 65.05 |
| Loan\_age | 0 | 0 | 353 | 44.94 |
| Dispinc\_t | 0 | 9,309.1 | 14,509.6 | 12,461.24 |
| Medhpi\_t | 0 | 160,100 | 343,400 | 261,522.91 |
| Mort\_rate\_t | 0 | 3.35 | 8.52 | 4.62 |
| Gdp\_t | 0 | 97.76 | 101.83 | 99.98 |
| Ump\_t | 0 | 3.7 | 10 | 6.3 |
| Hpi\_vt | 0 | -22.24 | 596.84 | 14.16 |
| Ump\_vt | 0 | -6.3 | 6.2 | -0.02 |
| Mort\_rate\_vt | 0 | -5.17 | 1.53 | -0.76 |

We can see from this table that all missing variables have been filled in (some of which occurred during data cleaning).

Figure 1 presents the average default rate (defaults / number of loans) across time, with time being at a monthly frequency and indicating the date of default of a loan. More univariate graphs are presented in the appendix.



We can see that the average monthly default rate was fairly low before the financial crisis (with some exceptions), quickly rose during the financial crisis, and then fell as the economy recovered and remained fairly flat after 2015.

From looking at the univariate graphs presented in the appendix, we can also see that a loan’s default risk tends to rise as the loan progresses further through its amortization term, with a particularly large spike occurring as the loan gets close to its maturity date. LTV at acquisition appears to be positively related to default rate, as the default rate appears higher for high LTV loans. Interest rate at origination likewise appears to be positively related to the default rate, while underwritten DSCR appears to be negatively related (in the part of the distribution where many observations fall, see summary stats).

# Statistical Analysis

This section will give an overview of the statistical methodology used in this study, and will present the results.

## Model Methodology

### Theoretical Background

The model used in this study is a logistic regression model. Logistic regression is a commonly used statistical model for modeling a categorical response. Since the response variable in our model is a default flag (indicating if the loan entered default or not) we are modeling a categorical response and using logistic regression makes sense. Please see the appendix for more details on logistic regression.

The economic variables used in this model are time series variables, and as such they should be stationary when they enter the model. While economic variables are not always stationary in practice, taking transformations of the economic variables can make them theoretically stationary. An example of this would be taking the year over year % change in median house prices instead of using the level, as the % change has a fairly constant historical average while the level exhibits a long-run upward trend.

### Variable Transformation

See the appendix for a description of the transformations applied to variables in the dataset.

### Variable Selection

Variable selection occurs in an iterative manner. First, all considered quantitative variables that are not economic variables are put into the model and the model is evaluated. If there were any variables that did not exhibit a significant relationship, they were removed. Underwritten DSCR was eliminated in this way.

Once the non-economic variables were decided upon, all of the economic variables considered were tried in the model. Choosing variables that have an intuitive relationship with the default rate of multifamily loans was preferred, as this reduces the chances of identifying a spurious relationship because these variables will have some grounding in economic/financial theory. Variables such as the unemployment rate, GDP, and HPI satisfy these requirements as the unemployment rate and GDP are indicators of how well the economic is performing in general, and better economic conditions tend to lead to better loan performance. HPI has an intuitive relationship as higher house prices allow borrowers to access equity in their home to pay off existing debt or refinance easier. In addition, housing makes up a large portion of many American’s wealth, so incorporating house prices takes wealth effects into account as well.

Trying out several different combinations of economic variables, several economic variables were found to have statistically significant relationships with multifamily loan defaults that also have intuitive causal relationships and coefficients that are directionally consistent with what economic/financial theory would predict. The economic variables used in the model are as follows:

1. Change in unemployment since origination
2. % Change in median house price since origination
3. Normalized, seasonally adjusted GDP at time t

## Results

This section will give the final model specification, and present in-sample and out-of-sample graphical results. This section will also present performance metrics for the model.

### Final Model Specification

The final model specification is shown in the appendix, with the coefficient estimate, standard error, z-statistic, and p-value shown as well.

### In-Sample Univariate Fit

Please see the appendix for graphs of the model’s in-sample fit across several key dimensions.

We can see that the model tends to overpredict before the financial crisis, and underpredict in the last few years (2017/18) of the development timeframe, but that the model captures the peak stress observed and responds to changing economic conditions. The model appears to fit well across loan age percentage, interest rate, underwritten DSCR, and LTV.

### Out-of-Sample Fit

Please see the appendix for graphs of the model’s fit for the testing sample (out-of-sample/out-of-time hybrid) over several key dimensions.

We can see that the model fit on the testing sample is largely consistent with the in-sample fit, except for the time fit. The peak of the default rates observed in the financial crisis is not hit by the model. When using this model, care should be taken to ensure that the population the model is being applied to is consistent with the in-sample population. While this model attempts to control for differences in the population, not all factors that impact loan performance can be accounted for and so this model should be used with care and proper backtesting should be conducted if an untested set of loans is to be run through this model.

### Performance Evaluation

Two performance metrics will be used to evaluate the model. Please see the appendix for definitions of the performance metrics used.

The results are presented in the table below.

**Table 3: Model Performance Statistics**

|  |  |  |
| --- | --- | --- |
| **Dataset** | **AUC** | **MAPE** |
| In-Sample | 0.83 | 0.0995 |
| Testing | 0.86 | 0.1339 |

We can see that model performance is good for both the in-sample and testing dataset, according to both metrics, as the AUC values are above 0.8, which is a typical threshold for adequate performance for a credit risk model. The MAPE values are also very low (below 1), indicating that in a particular month we can expect the model to miss by less than a percentage point (not by less than 1 percent, less than one percentage point, this is a difference in percentages). This indicates that the loan level (AUC) and aggregated (MAPE) performance of the model is strong. Interestingly, the AUC of the model is higher for the testing sample, which is somewhat unexpected. The likely explanation for this is that in the testing sample there is less of an emphasis on time periods where economic conditions are changing. The testing sample contains more observations from the out of time portion, which was a benign economic period, as compared to the development sample, which is likely causing the model to perform better on the testing sample as compared to the development sample.

# Recommendations

This study set out to address the following questions:

*• What are the key borrower/loan characteristics that influence the default rate on multifamily loans and what is their directional impact? This can be answered with a regression analysis and graphical exploratory data analysis.*

*• What macroeconomic variables can be used to explain the impact of the economy on the multifamily loan market and what is their relative impact? This can be answered with a regression analysis.*

*• What is the model error that can be expected in any month? This can be answered by calculating error metrics from the predictions of the statistical model.*

The results of this model show that a variety of factors influence the behavior of multifamily loans. Particularly of interest is that the broader macroeconomy has a large impact on the rate at which multifamily loans default, with default rates more than doubling during economic recession. Thus, the recommendation to users of this model is to keep sufficient cash reserved to withstand an economic recession, and this model can be used to determine the impact that a recession of a particular strength would have on a typical multifamily loan portfolio.

Another recommendation would be to acquire the FICO score or other type of credit score at origination and an ongoing basis to evaluate the creditworthyness of borrowers. Once adequate observations have a FICO score assigned to them, the model will be able to predict default with significantly more precision and will be able to generalize to loans the model has not seen more robustly.

Another recommendation is to continue to monitor the performance of the model if it is used for loss forecasting. As we saw in the performance of the testing sample over time, model performance can differ between samples and different economic environments. If model performance begins to deteriorate, a model redevelopment or an overlay on top of model results to account for model error should be implemented to ensure that forecsted credit losses do not differ too much from reality.

# Considerations

One consideration is that the model developed on this data is industry data based on data from Fannie Mae. Since the client is a specific bank, the model may perform differently on a specific set of loans that contain different characteristics not controlled for by the model. The client should monitor performance of the model as they begin to track multifamily loans they have issued and make adjustments to the model as necessary to ensure that the model’s performance holds up over time.

This model does not contain any credit score, such as the FICO score, which is a key input into many models that forecast loan defaults. The model’s performance would likely be much better if we had access to a credit score. The bank will likely be able to develop a superior model if they are able to incorporate the impact of a credit score.

Another consideration is that the economic variables used in the creation of this model do not fully represent the complex interplay between the broader economic environment and loan defaults. While the variables in the model have been shown to have a statistically significant historical relationship with multifamily loan’s propensity to default, there is no guarantee that these relationships will hold into the future. For example, when the COVID-19 pandemic struck, the unemployment rate soared, which would usually imply high default rates. However, due to the high levels of government support to the economy, loan defaults decreased during COVID-19 despite the high levels of unemployment. As such, the forecasts and results from this model should be subject to critical scrutiny and the assumptions closely monitored.

# Appendix

## Theoretical Background

Logistic regression models the log odds (ratio of default to not default) as a linear combination of the predictor variables. One of the assumptions of logistic regression is that the log odds and the variables in the model have a linear relationship, while this is not true for all considered variables, transformations will be applied to certain variables to ensure that this assumption is satisfied. Another assumption is that each observation is independent, using panel data there can often be a dependence structure between observations of the same subject (loan in this case). This dependence does not typically exist in loan level credit data as time-dependent shocks are accounted for using economic variables, so conditional upon the variables used in the model, each observation is theoretically independent. While this assumption may not be convincing, the dependence structure does not bias the coefficients, only the standard errors, which are of secondary importance when creating a predictive model.

## Final Model Specification

The variable name nomenclature is described below:

(Intercept) = Intercept term

tseas# = accounts for seasonality, equal to the month of the observation (e.g. 2 = February, 12 = December)

sage## = piecewise linear spline for loan age percent, ## indicate which values the piece of the spline ranges between

age\_close = loan is 1 month before its maturity date

int\_rate\_o = interest rate at origination

ltv\_aquisition = LTV at acquisition

ump\_vt = change in the unemployment rate since origination

hpi\_vt = percent change in the median home price since origination

gdp\_t = normalized, seasonally adjusted GDP at reporting date

**Table 4: Regression Results**

term estimate std.error statistic p.value  
 (Intercept) -5.3207378843 1.825465647 -2.91472912 3.559974e-03  
 tseas2 0.0680709526 0.062307601 1.09249837 2.746141e-01  
 tseas3 0.0431442226 0.062557106 0.68967740 4.903971e-01  
 tseas4 0.0279076948 0.063111767 0.44219479 6.583483e-01  
 tseas5 0.0008823728 0.063422882 0.01391253 9.888998e-01  
 tseas6 -0.0566208383 0.063978848 -0.88499308 3.761604e-01  
 tseas7 -0.1062587586 0.064284349 -1.65294913 9.834122e-02  
 tseas8 -0.1245350474 0.064552917 -1.92919317 5.370689e-02  
 tseas9 -0.1132341118 0.064363098 -1.75930176 7.852627e-02  
 tseas10 -0.1829800875 0.065760088 -2.78254018 5.393519e-03  
 tseas11 -0.1024292445 0.065109125 -1.57319338 1.156740e-01  
 tseas12 0.0373341879 0.063714100 0.58596430 5.578995e-01  
 sagelt\_50 4.6724388490 0.130217934 35.88168481 5.897629e-282  
 sage\_50\_85 -1.1563301866 0.166124653 -6.96061763 3.387841e-12  
 sage\_85\_95 5.8512244939 0.921599791 6.34898635 2.167383e-10  
 sagegt\_95 59.7236402535 1.858994872 32.12684508 1.860218e-226  
 age\_close -1.1311663297 0.202404698 -5.58863674 2.288591e-08  
 int\_rate\_o 0.4045824255 0.017569065 23.02811328 2.437740e-117  
 ltv\_aquisition 0.0499476653 0.001420473 35.16270789 7.433246e-271  
 ump\_vt 0.0621719937 0.008965238 6.93478489 4.068397e-12  
 hpi\_vt -0.0238418219 0.001305170 -18.26722105 1.509442e-74  
 gdp\_t -0.0802007042 0.018342972 -4.37228505 1.229528e-05

## Performance Metric Definitions

### AUC

The first metric will be the AUC, which stands for area under the Reciever Operating Characteristic curve, and is a commonly used metric to evaluate classification models. This metric is calculated by evaluating the accuracy of the model given various thresholds for classifying the predictions as 0 or 1. This gives a metric of the discriminatory power of the model, ranging from 0.5 to 1, with higher numbers being favored, and 0.5 meaning the model is no better than a random guess.

### MAE

The second metric will be the MAE (mean absolute error). This metric will be calculated by aggregating the data, as it is not a metric used to evaluate classification models. The data are aggregated by reporting month (calculating the average of the default flag and predicted value at each month) and then taking the average of the absolute differences at each month. This metric gives a sense of the average model error in aggregate for a particular month, which is of particular importance for regulatory reporting and business use cases for credit risk analytics.

## Variable Transformations

The following types of transformations were applied to variables in the dataset:

**Year over year % change:**

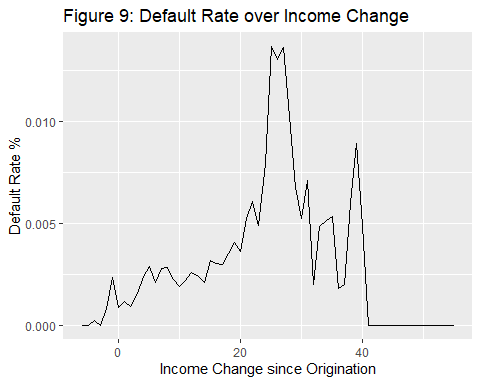
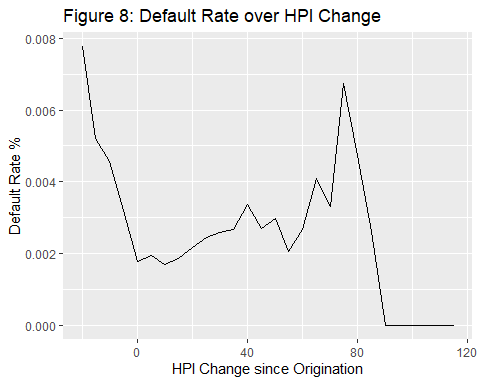
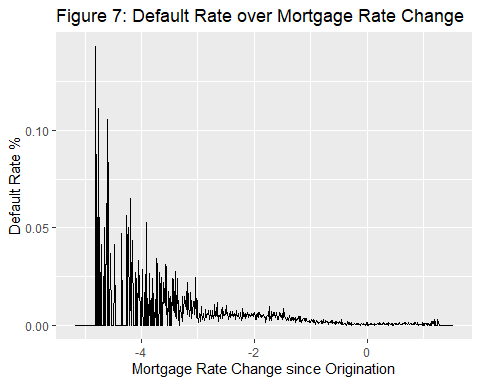
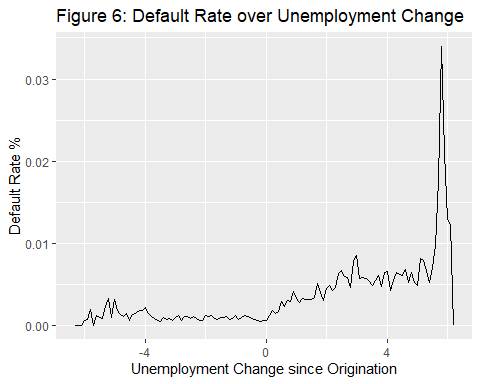
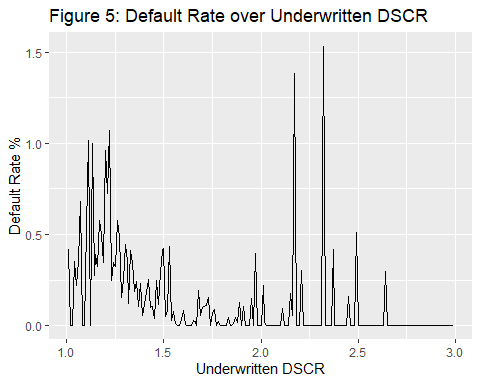
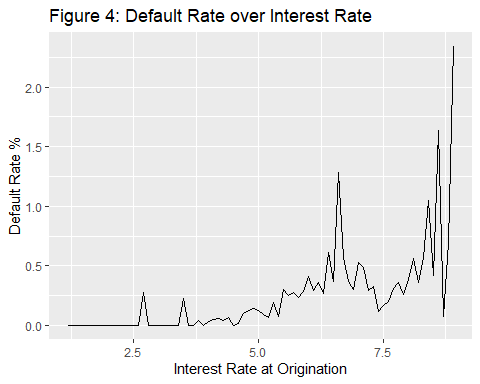
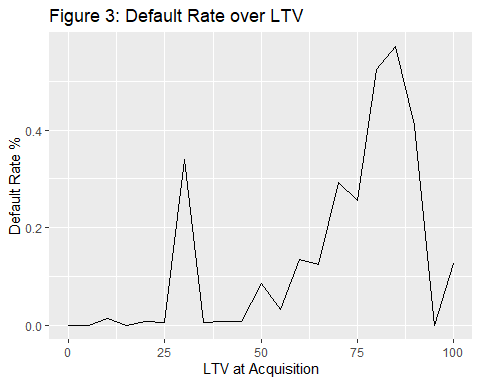
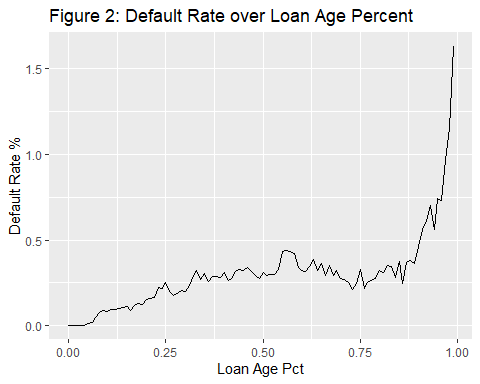
Where is the variable at the reporting date, and $variable\_{t-12$ is the variable 12 months prior. This is a common transformation for economic variables, as it considers the trend in the variable as opposed to short-term movements. Since this transformation uses the year-ago value, seasonality is also removed in this transformation.

**Change since origination:**

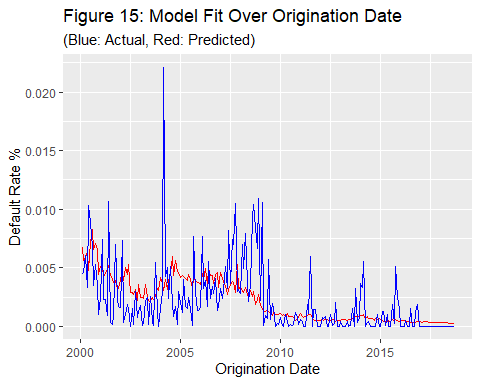
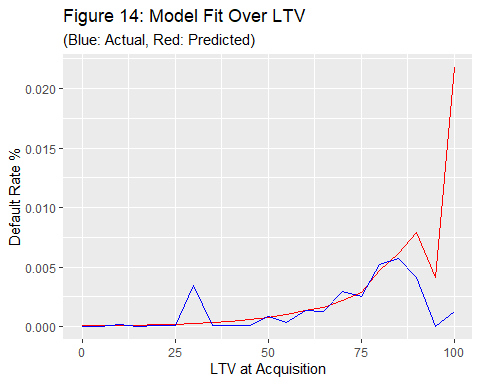
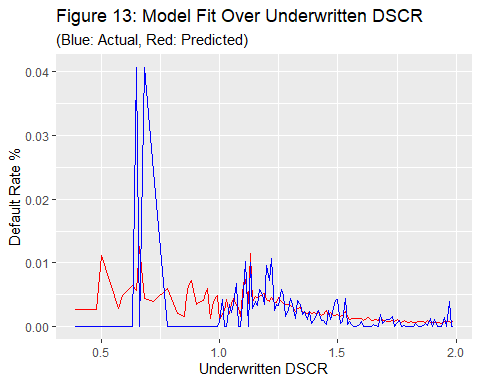
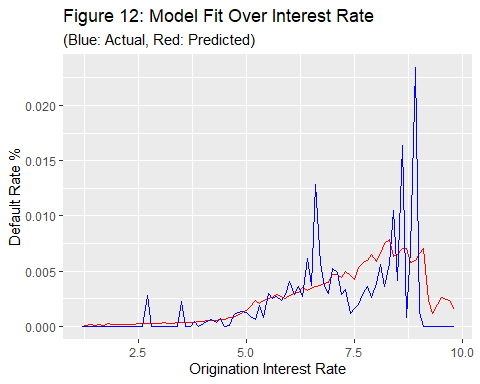
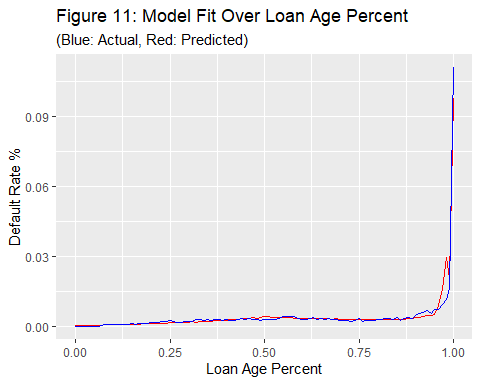
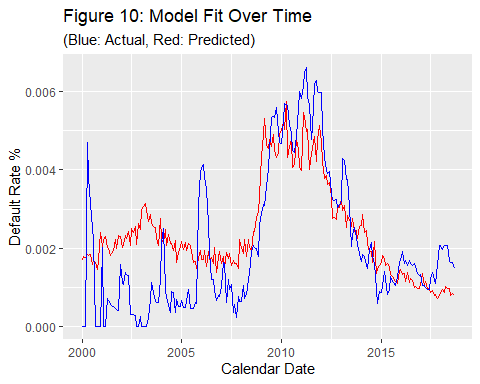
Where is the variable at the reporting date, and is the variable at the origination date of the loan. This is a transformation used in loan-level credit modeling, as it better captures changes in the economic environment than contemporaneous economic variables. This also has the advantage of incorporating the economic environment at origination of the loan, which tends to impact the quality of underwriting. Finally, a shock to the economic environment as compared to the original environment is what makes payments unmanagable and leads to default, so from a theoretical perspective this transformation is attractive.

Piecewise Linear Spline: This transformation allows a variable to have a piecewise linear effect. This transformation is applied for the loan age percent variable, where the marginal effect at first slopes upward slowly, then picks up in the middle values, and spikes sharply upward at the end. Polynomial transformations have difficulty capturing such a nuanced shape (without an exorbitant number of parameters) and there is a nice interpretation of a linear effect, even if that linear effect changes at specific points. In order to apply this transformation, knot points need to be picked where the marginal effect will be allowed to change (hinge points in the spline). These knots are chosen by examening the univariate plot of the default rate across all values of loan age percent.

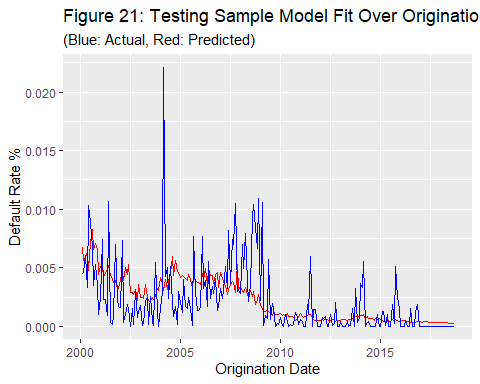
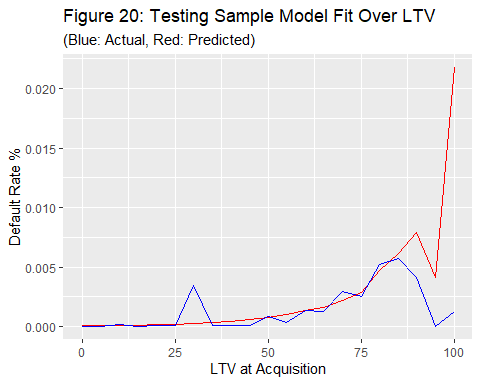
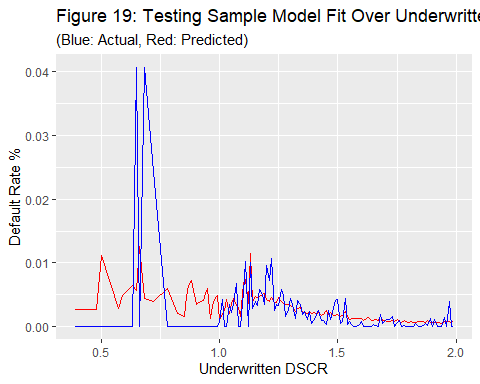
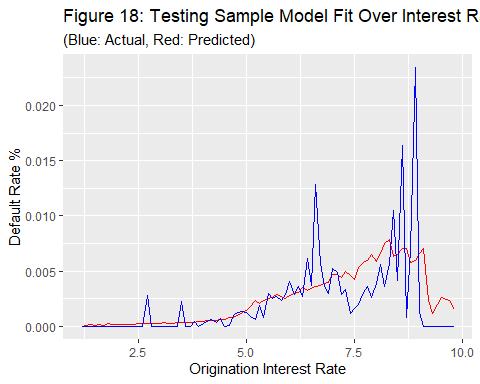
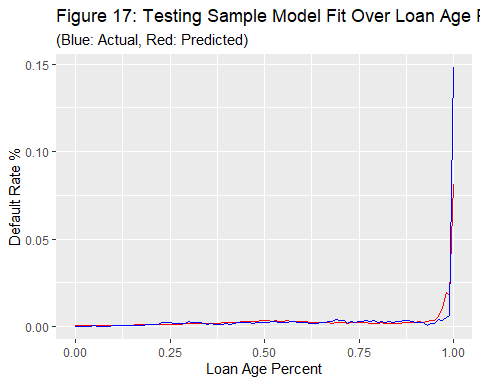
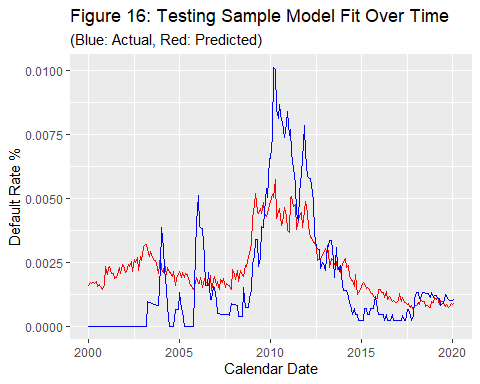
## Univariate Plots of Default Rate vs Predictor Variables



## In-Sample Univariate Model Fit



## Out-of-Time/Out-of-Sample Univariate Model Fit



## Citations

This report was produced using RMarkdown and all analysis was conducted using the R statistical software, version 4.2.0.