

# Market Analysis and Prepay Prediction to Single-Family Mortgages

Author: Shaun Zheng

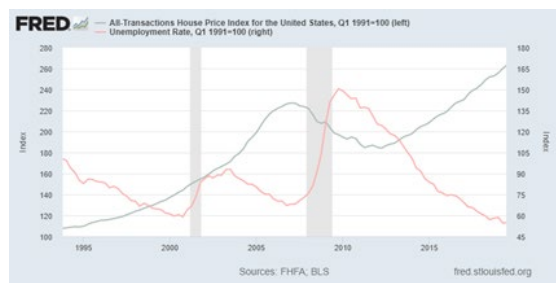
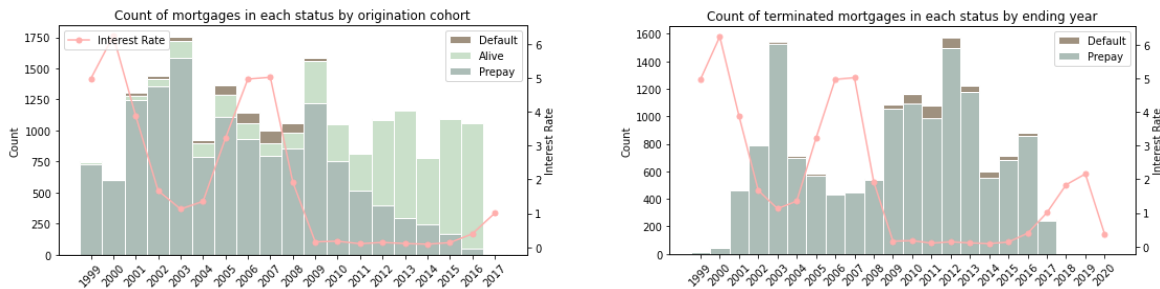
[Google Colab](#)

**Summary:** The project is aimed at analyzing the single-family mortgages from 1999 to 2017, trying to find out the causes for its fluctuations. Through finding the causes and its features, the project further builds a machine learning model to predict the prepay actions of the mortgages during this period. In the end, the project also gives suggestions about building turnover model, avoiding investment risks and increasing investment profits.

## Market Analysis by Mortgage Cohorts

First, I used SQL to query the historical Freddie Mac single-family loan-level data. The range of the dataset is from the Year 1999 to the Year 2017, having 20,000 data points. It covers information about mortgage start date, lasting ages, mortgage status, and other information. The further details are [here](#).

If we use each origination cohorts and each terminal year to analyze the breakdown of mortgage status. We can obtain time series distribution diagrams as follows. I also attach graph from FRED about the House Price Index and Unemployment Rate. Overall, Interest rate moves inversely with the prepay actions, and with making new mortgages. Default actions and unemployment rate move together.



Before 2003, the housing prices steadily increased, and in this period, people were more likely to make minds to make mortgages to buy houses under the concern of steady house prices increasing. Furthermore, the interest rate was decreasing, people are more likely to prepay to refinance. So, we could see in the upper two graphs that the number of new mortgages increased in this period, and many mortgages were prepaid in this period.

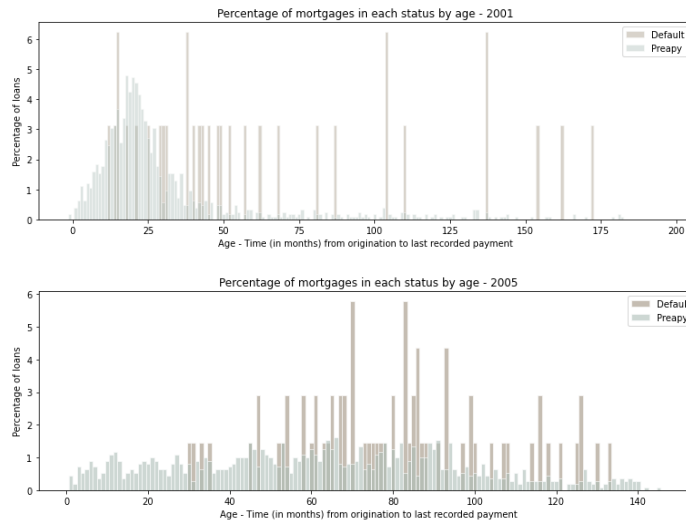
But since 2004, the real estate bubble was building, and the interest rate was increasing. It was too risky to make mortgages and the willingness to refinance decreases. So, the number of new mortgages decreased in the upper left graph, and the prepay decreased in the upper right graph.

There was a large economic recession since 2009, we can see in the FRED graph that the housing prices was dropping, and unemployment rate was increasing so much in this period. People had less money for the monthly installments, so the default increased at this period in the upper right graph. We

can also trace those mortgages in the upper left one, knowing these mortgages were made in the bubble building period since 2005. Because housing prices dropped so much, the mortgage made before the recession was risky. Many lenders at this recession were experiencing bankruptcy. But under this scenario, how could prepay increase in this period? That is because in 2009, asked by the regulator (the congress), Fannie Freddie entered the market to make mortgages to stabilize the market and allow people to prepay. At the same time, the money supply was increased so that the interest rate in 2009 was deep low. People were willing to refinance to make their mortgages newly affordable.

### Market Analysis by Mortgage Ages

If we focus on mortgages respectively lent in 2001 and 2005, reorganizing the data by age to validate the explanation above. We can obtain two main graphs as follows. Since we organize the data by age, the "Alive" couldn't tell us much information. Therefore, in the following graphs, alive mortgage instances are not taking into graphs. For the mortgages that originated in 2001, we can see most people had prepaid their mortgages within the first 25 months which are located at around 2003. For the mortgages that originated in 2005, we can see that there is not a peak for prepayment and most defaults happened within the 60 months of their first pmt. That time begins in 2010, which corresponds to the default increase since 2010.

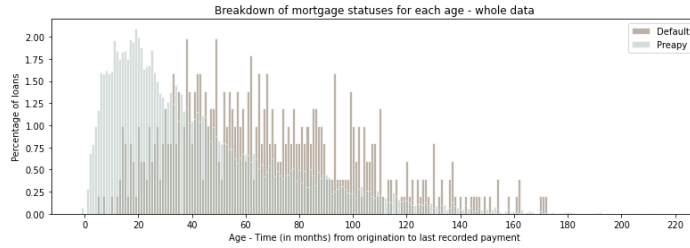


If we aggregate the whole data by age, we can obtain the below graph. From this graph, we can see that the prepay and default actions both peak within first 60 months. To determine the probabilities of mortgage behavior in the first five years. I firstly filter the original data to only get the mortgages that are terminated (as the status of "Default" & "Prepay") within less than 60 months. And then I filter the original data to obtain the mortgage who are still in the "Alive" status when their ages are more than 60 months. The sum of these three numbers will be the denominator. And each of these three numbers will be the nominator. Therefore, the results are as follows:

The probability of a mortgage has defaulted in the first 5 years is: 1.45 %

The probability of a mortgage has prepaid in the first 5 years is: 64.1 %

The probability of a mortgage remains alive after the first 5 years is: 34.48 %

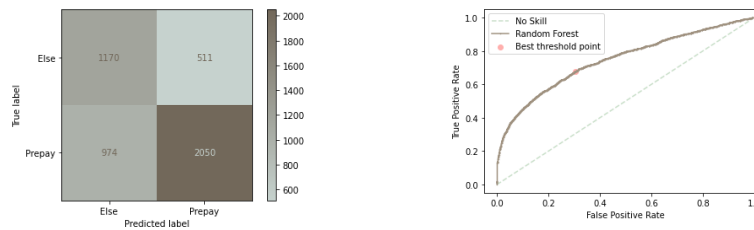


Based on these discussions, the first five year is a good window to predict the prepay and default actions of mortgages. To build the “prepay” prediction model, I will take the “Current and Expectation” factors of the interest rate, the unemployment rate, and the housing price index into the building process. Of course, other factors about the fundamental conditions of borrowers like the debt-to-income ratio and local real estate markets are also important to predict the prepayment and default and should be considered.

### Machine Learning Model Building

In the “Prepay and default prediction model in the first five year”, I select “prepay” as a response to building classifier. The process is displayed in the Jupyter Notebook. During this process, I firstly wrangled the data (dropping not useful columns, filtering NA values, transforming some columns’ type, and encoding the categorical variables) and added interest rate, housing price index, and the unemployment rate as variables. Because the sizes of data points are in different statuses, I undersampled the training dataset to let different statuses have similar numbers of instances. In the end, I run the parameter grid cv to tune the parameters and use roc to trade off the threshold to further improve the classifier. The results are as follows.

- The predicted prepay rate is 54.4%
- The actual prepay rate is 64.3%
- The accuracy rate is 68.4%



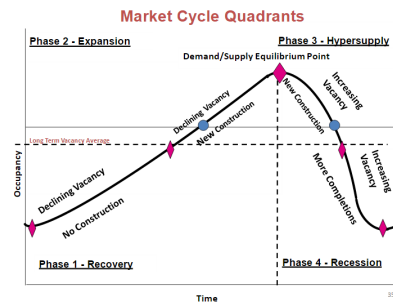
I still believe the model does a good job. Because it is conservative, it is more likely to falsely classify the datapoint into “Else” where might cause defaults and bring harms to real estate finance firms, instead of falsely classifying the datapoints into “Prepay”. In the future, I will further think about construct the variables to describe borrowers expectation in the first five year, and tune the parameters and add more local market variables to make predictions more accurate.

### Effect of Mortgage Prepayment and Default

There is a lasting impact in the real estate finance firms’ equity agreement from the prepayment behaviors of homeowners’ mortgages. The prepayment of a mortgage means homeowners are in good financial condition, making sure they have availabilities to pay the real estate taxes, insurance, etc. on time, which potentially increases the property prices in the market. But that also means they might be likely to buy out real estate finance firms’ options after the first five years using low-interest-rate mortgages in the market (if there is), which may harm real estate finance firms’ aim of investing in the

long-term real estate market to make profits. On the other side, default behavior means the homeowner is in bad financial condition. That increases the risks of leaving the house under maintenance, decreasing housing value, and even foreclosure.

Real estate market is a semi-efficient market, and the construction lag effect exists. If we want to know more about the local market and to build the turnover model. Then data like the building permits, the demographic information, the interest rate, the economic growth, and the information about the family member will be needed. Because we need to have information about market supply, market demand, and market consumption in the foreseen future. Like the graph I show below, real estate has its



own cycle. Information mentioned above can really help us locate which stage the local real estate market is in. Besides these, data about the residents should also be collected. Because real estate is not like other standard and highly transacted commodities, it is also closely related to the residents' lives. A homeowner may not be willing to sell his/her house if their kids are taking schools nearby. To establish the parameter, I will use the "The residual sum of squares" statistical method to make model better learn from the past and predict the future.

Different market conditions change the behaviors of turnover. If the interest rate goes up, the housing market will be relatively inactive and that increase the turnover. If the unemployment rate is increasing, that might also lead to an inactive housing market, increasing the turnover. If the home price appreciation is in the ascending period, then people will be more likely to buy houses. Demands will increase and lead to a decrease in turnover.

### Conclusion

To sum up, the mortgage market is vital to real estate finance firms because the mortgages attached to each house directly decide the current capacity of the equity where real estate finance firms can invest as options and become the co-investor. Besides this, a healthy mortgage means the homeowners can pay the taxes, maintenance fees, etc. on time, bringing the potential of the value increase. More importantly, the mortgage market is the indicator of the real estate market, the better understandings of the single-family mortgage market can make sure real estate finance firms live through cyclical recessions.

Other indicators of investment risks are interest rate, unemployment rate, building permits in local markets (Supply), economic conditions (real GDP growth and risk premium), and demographic information in the metropolitan statistical areas (Demand). If a market has a continuous population input and an economic growth potential. Then the market is desirable to be invested. If the building permits are growing too fast and the unemployment rate is also growing. Then the market will not bring the house value to increase.

The two ways to enhance the investment returns are first, REA keeps the possibility of buying more options during the process if the homeowners need to cash out and the local market still has the potential of growth (even currently in the fluctuation stage). The second way is to package and sell portions of some

options, which have potential growth and are adored by the market, in the option market. In this way, real estate finance firms can obtain more cash to invest in other markets that have greater potential.