Determinants of Mortgage Loan Delinquency: Application of Interpretable Machine Learning

Taghi Farzad *

University of California, Riverside

Abstract

This paper uses Machine Learning to determine factors influencing mortgage loan delinquencies. Our focus is on 90-day delinquency within the first 12 months after loan origination. We apply a modified Gradient Boosting approach (XGBoost) to data from 2013Q1 to 2017Q4. First, we show XGBoost's prediction accuracy is higher than Logistic regression. Second, we use Permutation Feature Importance method to find the most important factors impacting delinquency. Finally, we use Interpretable Machine Learning to quantify the effect of each factor on delinquency probability. Borrower's credit score, Federal Funds Rate, and original interest rate are found as the most important features in the model. These three features all affect the interest payments; however, credit score may affect other aspects of the contract as well. To interpret the results, we derive Partial Dependence Plots, Individual Conditional Expectations, and Accumulated Local Effects for the important features in the model. Interpretable Machine Learning helps us to find the unbiased effect of each factor. In particular, we can conclude the effect of credit score on delinquency is not merely through the impact on interest rate. This paper uses the loan data over an expansion period and points out the importance of interest rate in determining delinquency. Our results differ from the factors identified in the literature. Notably, in our model Combined Loan-to-Value and unemployment, although important, but are both dominated by the factors mentioned above. This contrast in results draws attention to the role of business cycle in determining delinquency triggers.

Keywords: Mortgage Loan Delinquency, Gradient Boosting, Interpretable Machine Learning.

*taghi.farzad@email.ucr.edu.

Introduction

It is a long-lived question concerning the factors influencing mortgage default. Literature took off by introducing negative equity as the main reason behind mortgage default,[1] however later on the impact of other factors, such as combined loan to value and unemployment status of the mortgagor, were studied and found to be significant. In some studies, these factors are found to be influential in combination to negative equity,[2] while in other cases they are the dominant factor.[3] Providing insight regarding the importance of these features is crucial, as proposing the importance of different factors entails different policy implications.[4] This problem especially drew more attention in the wake of 2008 Great Recession. Availability of Big Data with the prevalence of Machine Learning approaches in the recent years helped the researchers to improve default prediction.[5] Nevertheless, limited number of studies provided interference and interpreted the Machine Learning results.

This paper uses a modified Gradient Boosting algorithm, called Extreme Gradient Boosting (henceforth XGBoost) as a Machine Learning technique to predict +90-day delinquency within 12 months from origination for loans issued during 2013Q1-2017Q4. First, it is shown that applying XGBoost provides a higher prediction accuracy compared to Logistic Regression. Next, XGBoost model is used to find the important factors determining delinquency. Finally, we use Interpretable Machine Learning approaches to quantify and visualize the impact of each factor on delinquency. This paper finds, during 2013Q1-2017Q4, FICO credit score (henceforth credit score), Federal Funds Rate, and original interest rate are the features most important to predict loan delinquency. FFR and original interest rate directly affect the interest payments, while credit score impacts other aspects of the loan contract (such as Loan-to-Value, mortgage insurance, and leniency) as well. This paper suggests any policy recommendation regarding mortgage loan delinquencies should be contingent on the Business-Cycle state of the economy.

This paper is, to the best of our knowledge, the first study that uses Interpretable Machine Learning approaches to find the impact of factors in mortgage default. Sirignano, Sadhwani, and Giesecke (2016) used neural network techniques for this purpose. However, this paper improves their work in several dimensions. First, by implementing XGBoost approach we can

[1] Asay (1979), Foster and Van Order (1984)

[2] For instance Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010) found the combined loan to value (consist of mortgage loan and credit card loans) to be a significant factor determining the mortgage default.

[3] Herkenhoff and Ohanian (2012a) studied the unemployment factor and concluded negative equity is not a necessary condition for mortgage defaults.

[4] As explained by Elul et al. (2010) accepting negative equity theory results in assuming mortgagor will use default as an option when they are in large negative position, while accepting the role of unemployment asserts the mortgagors are unable to pay the mortgage.

[5] For instance see Khandani, Kim, and Lo (2010), Butaru, Chen, Clark, Das, Lo and Siddique (2016), and Bagherpour (2018).

achieve a competitive level of accuracy (compared to other Machine Learning and Artificial Intelligence techniques) at a lower computational cost. Second, Gradient Boosting is based on decision trees which is convenient to understand and interpret. The final product of a Neural Network algorithm is a Black Box that is hard to comprehend and interpret. In addition, Deep Learning lacks a strong theoretical foundation. Gradient Boosting allows us to easily implement interpretability theories. Partial Dependence Plots, Individual Conditional Expectations, and Accumulated Local Effects (all to be explained) can be implemented on an interpretable model such as decision tree. To apply these methods to Black Boxes, we usually need a surrogate model: an interpretable model that behaves closely to the original model. However Gradient Boosting is intrinsically interpretable, hence we can directly apply theories to our model. Finally, we focus on Serious Delinquencies (+90-day delinquency) within the first 12 months after loan origination. 90-day delinquency within first 12 months is correlated with default while itself studies the short-term risk associated with a mortgage loan. This allows us to study a simpler problem while the results can explain determinants of other type of arrears.

We follow Friedman (2001) and Chen and Guestrin (2016) to apply a modified Gradient Boosting algorithm, XGBoost, to construct a classifier tree. XGBoost algorithm penalizes the complexity of the tree (in the same way as in Lasso). First, we use the training data to construct the tree: At each step we need to find the splits and the weights in order to minimize an associated loss function while penalizing model complexity. Then we use the test data to apprise the performance of the model. Since it has a high accuracy in predicting mortgage defaults, we can, with a high level of confidence, use the model to explain the important factors in mortgage default. On the next step, we can provide Partial Dependence Plots (PDPs) to find the marginal effect of the features. PDPs can be used to study the general behavior of a factor (linear, monotone, etc.) in relation to the outcome. We can also construct Individual Conditional Effect (ICE) to study the heterogeneity of the factor behavior for different instances. Finally, Accumulated Local Effects (ALE) provides us a picture of an impact of a factor taking into account the correlation of it with other features. ALEs provide an unbiased effect of a feature on delinquency probability.

Dataset used in this paper includes loan data in addition to local and national level data. Loan data is obtained from Fannie Mae "Single-Family Mortgage Loan Data" for the loans issued between 2013Q1 and 2017Q4. We merge the dataset with quarterly state-level data, including unemployment rate,[6] per capita personal income (seasonally adjusted, annul rate),[7] and Housing Price Index (HPI),[8] and quarterly national-level data, such as Federal Funds

[6]Bureau of Labor Statistics

[7]FRED

[8]Federal Housing Finance Agency

Rate and Slope of Yield Curve.[9] After data cleaning, we end up with $8.8M$ observations that we split into train (80%) and test (20%) data. The out-of-sample study for XGBoost model results in a higher prediction accuracy, $AUC_{XGBoost} = 0.87$, compared to Logistic model, $AUC_{XGBoost} = 0.71$.

Literature Review

Negative Equity proposes a model for mortgage default[10] in which default is considered as a put option on the house. This idea predicts mortgage defaults solely based on negative equity of the house. This model, however, overestimates the frequency of mortgage defaults. Researchers attribute this discrepancy to the transaction costs of exercising the option (such as moving expenses, reputation cost as pointed out by Elul, Souleles, Chomsisengphet, Glennon, and Hunt, 2010). Kau, Keenan, and Kim (1994) pointed out the importance of including these costs in option-based theoretical models. Although the strategic or ruthless default[11] is confirmed in some other works, such as Deng, Quigley, and Van Order (2000) and Bhutta, Dokko, and Shan (2010), they also assert the important role of a "double trigger"[12] event. In other words, negative housing equity is a necessary condition for mortgage default, but not a sufficient one.[13] For instance Foote, Gerardi, and Willen (2008) argue other factors such as expected price appreciation, flow of utility due to house ownership and the mortgage payment size are important factors. Herkenhoff and Ohanian (2012a) studied the mortgage delinquencies and foreclosure as an informal unemployment insurance; hence negated the necessity of negative equity for mortgage defaults. All these opposing conclusions in theoretical works would urge an empirical study that points out the important factors influencing mortgage default.

Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010) emphasis on the effect of illiquidity on mortgage default. They have shown that the combined Loan to Value (henceforth OCLTV), consists of mortgage loans plus credit card loans, is an important factor, in addition to negative equity, impacting the mortgage default decisions. This is in line with the conclusion of Li and White (2009). They have shown, for most of the homeowners, the relation between foreclosure and bankruptcy is complementary; despite the fact that theoretically the two can be substitute as well.

[9] Both obtained from FRED.

[10] See Asay (1979) and Foster and Van Order (1984).

[11] Attributing the mortgage default solely to the decline in the value of property compared to remaining mortgage balance (Gerardi, Herkenhoff, Ohanian, and Willen, 2013).

[12] Double trigger is the combination of negative equity and some adverse life event, such as job loss or health problems (Gerardi et al., 2013).

[13] Foote, Gerardi, and Willen (2008)

Using the individual level data,[14] Gerardi, Herkenhoff, Ohanian, and Willen (2013) found the individual unemployment to be the strongest default predictor. These studies recommend policies that stimulate the employment. This contrasts policy implications of negative equity; while negative equity focuses on reducing the principal and iliquidity, while their model recommends temporary reduction in payments (Elul et al., 2010 and Foote, et al., 2008).
Sirignano, Sadhwani, and Giesecke (2016) studied the factors affecting mortgage risk using Neural Networks. They argue that the traditional Logistic models cannot capture the non-linear behavior of the features. Furthermore, it is impractical to include all the interactions terms in the traditional econometric models. These two motivates them to Deep Learning approaches. Their study presents the state unemployment rate as the most important factor affecting delinquencies.
Predictability performance of Machine Learning techniques has been investigated thoroughly in the literature. Khandani, Kim, and Lo (2010) used Classification and Regression Trees (CART) algorithm to forecast the consumer credit default. Using data of a major commercial bank, they have shown their proposed approach will result in lower Type I and II errors which saves 6% to 23% of losses. Butaru, Chen, Clark, Das, Lo and Siddique (2016), also show Machine Learning techniques (decision trees and random forest) result in a better predictability performance for credit card defaults. Bagherpour (2018) tested the performance of different machine learning approaches in predicting mortgage defaults; all machine learning techniques dominate logistic regression. He has shown among different machine learning approaches, Support-Vector machine performs the best. Fitzpatrick and Mues (2016) have shown boosted trees approach outperform the other Machine Learning techniques and provides predictions that are significantly better than Logistic regression. Kvamme, Sellereite, Aas, and Sjursen (2018) have shown combination of Neural Network and random forest can outdo Logistic regression and each of these approaches individually.
The rest of the paper is organized as follows: First we provide theoretical background for the XGBoost model and Interpretable Machine Learning. Section 3 describes the data used in the paper. Section 4 provides the empirical results. Section 5 concludes the paper.

The Model

Logit Model

We use the Logistic Regression as the baseline model. Let Y be the binary outcome showing the 90-day mortgage delinquency, $Y \in \{0, 1\}$ and let p be delinquency probability, $p = P(Y =$

[14] 2009 PSID Supplement on Housing, Mortgage Distress, and Wealth Data.

1). Logistic Regression can be written as

$$logit(p) = \log\left(\frac{p}{1-p}\right) = \beta X$$

where X is the data points (including the constant term) $\frac{p}{1-p}$ is the odds of delinquency. Equivalently,

$$P(Y = 1|X) = \frac{\exp(\beta X)}{1 + \exp(\beta X)}.$$

Maximum likelihood estimation method is used to estimate, βs, the real coefficients of the above logit model.

Gradient Boosting Model

The goal of Gradient Boosting is to find an estimation of the function $F^*(X)$, that maps characteristic matrix, X, to the outcome vector, Y, while the expected loss function, $L(Y, F(X))$, is minimized. [15] Commonly used loss functions in literature are squared error and absolute error for the real-valued regressions, and negative binomial log-likelihood for classification problems (Friedman, 2001).

Following Friedman (2001), we restrict our attention to the additive expansions of the form

$$F\Big(X; \{\gamma_m, \beta_m\}_1^M\Big) = \sum_{m=1}^{M} \gamma_m h(X; \beta_m). \tag{1}$$

That is, the approximation is the weighted sum of a generic function, $h(X, \beta)$ called the base learner, that depends on the characteristic variables, X, and parameters, β. We are interested in the functions $h(X, \beta)$ that are regression trees generated according to Classification and Regression Tree (CART) algorithm (Breiman, Friedman, Olshen, and Stone, 1983). For the case of finite data, we use the training data, $\{y_i, X_i\}_1^N$, to find the parameters of Equation 1:

$$\{\gamma_m, \beta_m\}_1^M = \arg\min_{\{\gamma'_m, \beta'_m\}_1^M} \sum_{i=1}^{N} L\Big(y_i, \sum_{m=1}^{M} \gamma'_m h(X_i; \beta'_m)\Big)$$

This expression is based on the premise that the solution consists of an initial guess plus a sequence of increments. Each increment is parallel to the direction in which the (expected) loss function has the steepest decent. Finally, the exact amount of increment ("step" or "boost") is determined by a scaling factor that minimizes the (expected) loss function between the estimated function and the outcome.

The above problem may be unsolvable. Hence, as proposed by Friedman (2001), we use the

[15] Friedman (2001)

following "greedy-stagewise" approach. Let

$$F_0(X) = \arg\min_{\rho} \sum_{i=1}^{N} L(y_i, \rho),$$

then for $m = 1,\ 2,\ \ldots,\ M$

$$\{\gamma_m, \beta_m\} = \arg\min_{\{\gamma'_m, \beta'_m\}} \sum_{i=1}^{N} L\Big(y_i, F_{m-1}(X_i) + \gamma' h(X_i; \beta')\Big)$$

and finally

$$F_m(X) = F_{m-1}(X) + \gamma_m h(X; \beta_m).$$

Even solving this simplified problem might be computationally intensive. Thus, following Friedman (2001), we can solve the constrained problem in which the increments are also in the same class of the generic functions $h(X, \beta)$. Therefore, in the m-th iteration, at each data point, X_i, the steepest decent direction of the loss function can be obtained using the gradient:

$$-g_m(X_i) = -\Big[\frac{\partial L(y_i, F(X_i))}{\partial F(X_i)}\Big]_{F(X)=F_{m-1}(X)}$$

Notice this is the gradient at each data point $\{X_i\}_1^N$. That means each $g_m(X_i)$ is unlikely to provide the steepest decent boost for the whole data. One possible candidate for the boost in the m-th iteration is to select a member of the generic functional form that is most parallel to the gradient in the N dimension data: we can find the parameters of γ_m and β_m that provides a boost $h(X, \beta_m)$ that is most parallel to $g_m(X_i)$ at each data point:

$$\{\gamma_m, \beta_m\} = \arg\min_{\{\gamma'_m, \beta'_m\}} \sum_{i=1}^{N} \Big(g_m(X_i) - \gamma'_m h(X_i, \beta'_m)\Big)^2$$

One can think of $h(X, \beta_m)$ to be the constrained gradient explained above; the direction of same generic functional class h that provides the steepest ascent. Next, we can find the scale ("line search") by minimizing the associated loss function:

$$\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} L\Big(y_i, F_{m-1}(X_i) + \rho h(X_i, \beta_m)\Big)$$

Finally, the estimation would be updated according to

$$F_m(X) = F_{m-1} + \rho_m h(X, \beta_m)$$

.

Regression Trees

We focus on Regression Trees *à la* Breiman et al (1983). Let $\{R_j\}_1^J$ be a partition of the data space. A regression tree model with J terminal nodes can be expressed by

$$h(X, \{b_j, R_j\}_1^J) = \sum_{j=1}^{J} b_j I(X \in R_j)$$

where $I(.)$ is the indicator function. Here the base learner is a tree with J terminal nodes that are represented by $\{R_j\}_1^J$. Parameters of the base learner are $\{b_j\}_1^J$. Since the nodes are a partition of the data set, $h(x) = b_j$ if $x \in R_j$. By applying the updating rule 2 to a regression tree described above, in the m-th iteration we have

$$F_m(X) = F_{m-1}(X) + \rho_m \sum_{j=1}^{J} b_{jm} I(X \in R_j m) \tag{2}$$

where$\{R_j m\}_1^J$ construct data partition in the m-th iteration. $\{b_j m\}_1^J$ are simply the square error minimizers for corresponding regions. This expands the tree to include regions that provide more accurate predictions. Finally, all remains is finding the scaling factor, i.e. the line search. Hence, at each iteration the estimation function would be updated according to:

$$F_m(X) = F_{m-1}(X) + \rho_m \sum_{j=1}^{J} b_{jm} I(X \in R_{jm})$$

Finding the optimal partitions, $\{R_j\}_1^J$, and weights will be explained in the next session for the XGBoost algorithm.

XGBoost

In this paper, we have used a modified version of Gradient Boosting, called XGBoost. Since the main goal of this paper is to provide first a model with high prediction accuracy and second interpretable results, it is important that we avoid over-fitting the model. That is, to explain the data with a model (we have especial interest in trees) that is as simple as possible. This can be achieved by penalizing the complexity of trees. We can adopt the previous definition for the regression tree model. let $F(X) = \sum_{m=1}^{M} h_m(X)$ be the model while the base learners can be written as $h(x) = b_j$ if $x \in R_j$, provided that $\{R_j\}_1^J$ is a partition for X. This provides trees with J number of final nodes (leaves) where each leaf is weighted by b_j. Following Chen and Guestrin (2016), define the "regularized Learning Object" as:

$$\mathcal{L}(F) = \sum_{i=1}^{N} L(y_i, F(X_i)) + \sum_{m=1}^{M} \Omega(h_m)$$

where

$$\Omega(h) = \theta J + \frac{1}{2}\lambda \|b\|^2$$

The first term is the loss function introduced above. The second term penalizes complexity of the trees by preferring lower number of leaves and lower norm for the weights. Using the regularized learning object above we can find the optimal tree that needs to be added to the currently constructed one. Chen and Guestrin (2016) provide us an algorithm that generates the optimal tree at the m-th iteration, h_m such that it minimizes

$$\mathcal{L}^m = \sum_{i=1}^{N} L(y_i, F_{m-1}(X_i) + h_m(X_i)) + \Omega(h_m)$$

let $\mu_i = \partial_F L(y_i, F(X_i))$ and $\mu_i = \partial_F L(y_i, F(X_i))$ be the first and second derivatives of the loss function in the current iteration. The second-order approximation around the loss function results in

$$\mathcal{L}^m \simeq \sum_{i=1}^{N} \left[L(y_i, F_m(X_i)) + \mu_i h_m + \frac{1}{2}\nu_i h_m^2 \right] + \Omega(h_m)$$

Ignoring the constant term and plugging the penalty function we have

$$\tilde{\mathcal{L}}^m = \sum_{i=1}^{N} \left[\mu_i h_m + \frac{1}{2}\nu_i h_m^2 \right] + \theta J + \frac{1}{2}\lambda \sum_{j=1}^{J} b_j^2 = \sum_{j=1}^{J} \left[(\sum_{i \in I_j} \mu_i) b_j + \frac{1}{2}(\sum_{i \in I_j} \nu_i + \lambda) b_j^2 \right] + \theta J$$

where $I_j = \{i | X_i \in R_j\}$ is the instance set of leaf j; that is the set indices of all data point in the $j-th$ partition. For a known partition structure $\{R_j\}_1^J$ the optimal weights can be derived as

$$b_j^* = -\frac{\sum_{i \in I_j} \mu_i}{\sum_{i \in I_j} \nu_i + \lambda},$$

which corresponds to the following pseudo-loss function

$$\tilde{\mathcal{L}}^m(\{R_j\}_1^J) = -\frac{1}{2} \sum_{j=1}^{J} \frac{\left(\sum_{i \in I_j} \mu_i\right)^2}{\sum_{i \in I_j} \nu_i + \lambda} + \theta J$$

This can be used to rank different partitions for h_m. However, since considering all the partition candidates is infeasible, a greedy algorithm can be employed that starts from a single leaf and branches will be added iteratively. For each split candidate, let I_L and I_R be the instance sets of the left and right nodes after the split, where $I = I_L \cup I_R$. One can define loss function corresponding to a split as

$$\tilde{\mathcal{L}}^{split} = \frac{1}{2}\left[\sum_{j=1}^{J} \frac{\left(\sum_{i\in I_L}\mu_i\right)^2}{\sum_{i\in I_L}\nu_i + \lambda} + \frac{\left(\sum_{i\in I_R}\mu_i\right)^2}{\sum_{i\in I_R}\nu_i + \lambda} + \frac{\left(\sum_{i\in I}\mu_i\right)^2}{\sum_{i\in I}\nu_i + \lambda}\right] + \theta$$

This “Exact Greedy Algorithm for Split Finding” is a very powerful approach for finding the optimal trees. However, it is computationally expensive. There are approximate algorithms that take into account the feature distribution that are able to efficiently solve for split points (Chen and Guestrin, 2016). Moreover, the algorithm suits unbalanced data with missing values.

Interpretation of Results

Partial Dependence plots (PDPs) show the marginal effect of a subset of features, X_S, on the predicted outcome of the machine learning. These plots provide a picture regarding the behavior of the features in X_S. let $F(X)$ be the approximation in (1) provided by a Machine Learning approach. Partial dependence function is defined as

$$F_{X_S}(X_S) = \int F(X_S, X_C)d\mathbb{P}(X_C) \tag{3}$$

where X_S is the set of our variables of interest and X_C represents the rest of the variables. Hence PDPs show the effect of variables in S on prediction function while the other features are drawn according to their distribution. Although (3) does not restrict the number of variables in S, visualization is only possible when it includes only one or two features. Partial dependence function can be estimated by

$$F_{X_S}(X_S) = \frac{1}{n}\sum_{i=1}^{n} F(X_S, X_C^i),$$

where X_C^i are the actual feature values in the data.
PDPs are intuitive, easy to implement, and provide a clear interpretation (Molnar, 2019), that potentially can provide a causal interpretation (Zhao and Hastie, 2017). However, correlation and/or interdependence of the feature are ignored which urges us to provide Accumulated Local Effects below. Partial dependence plots provide an average effect of features in S on the prediction. This may hide the heterogeneous behavior of individual observations. For instance, assume a feature has a U-shaped effect for males and an inverted U-shape effect on females precisely mirroring the effects on males. If we have the same number of males and females, PDP shows no marginal effect for that feature. To shed further light on the behavior of sub-samples in the data, we can plot the marginal effect for each observation. These

graphs are called Individual Conditional Expectation (ICE) and introduced by Goldstein et al. (2017). The graphs can be generated by plotting $F(X_S, X_C^i)$ for different values of X_S while X_C^i remains unchanged. The net effect of X_S is more clearly shown if we center the curves, i.e. to force all of the curves to start from the same point, e.g. zero. As mentioned above, partial dependence function does not explain the true effect of a variable should it highly correlated to other variables in the data. Finding PD functions may force the prediction function to extrapolate beyond the realm of training data (see for instance the examples in Apley and Zhu, 2016). Assume a variable in X_S is highly correlated to a variable in X_C, for example height and weight. Using the joint distribution for all variables in X_C, PD function simply ignores the possibility of such correlations (Equation 3). Hence Apley and Zhu (2016) proposed an alternative approach to find the effect of variables on predictions of machine learning that is called Accumulated Local Effect (ALE). Assume X_S and X_C both consist of just one variable, X_1 and X_2 respectively. Apley and Zhu (2016) define the ALE main effect function of X_1 as

$$f_{X_1,ALE}(X_1) = \int_{X_{min,1}}^{X_1} \int_{X_2} \mathbb{P}(X_2|Z_1) F^1(Z_1, X_2) dX_2 dZ_1 - Const.$$

where $\mathbb{P}(X_2|Z_1)$ is the conditional distribution of X_2 at $X_1 = Z_1$, $F^1(X_1, X_2) = \frac{\partial F(X_1,X_2)}{\partial X_1}$ is the local effect of X_1 on $F(.)$ and $X_{min,1}$ is just below the lowest observed X_1. The constant term is used to center the plot. The title is self-explanatory: ALE functions find the "local" effects of $X_1 = Z_1$ on $F(.)$, taking into account the correlation of $X_1 = Z_1$ and X_2, and adds them up ("accumulates" the local effects) to the desired X_1. The expression can be generalized to the case that X_C contains more than one variable. However plotting ALEs is feasible only when X_S contains one or two variables. When $|X_S|= 1$, the Uncentered ALE Main Effect Function[16] is defined as

$$g_{S,ALE}(X_S) = \int_{X_{min,S}}^{X_S} \mathbb{E}\Big[F^S(X_S, X_C)|X_S = Z_S\Big] dZ_S$$

The centered ALE main effect is derived by de-meaning the above expression:

$$f_{S,ALE}(X_S) = g_{S,ALE}(X_S) - \mathbb{E}\Big[g_{S,ALE}(X_S)\Big] = g_{S,ALE}(X_S) - \int g_{S,ALE}(X_S) d\mathbb{P}(X_S)$$

[16] Apley and Zhu (2016) define

$$f_{X_S,M}(X_S) = \mathbb{E}\Big[F^S(X_S, X_C)|X_S = Z_S\Big]$$

as Marginal Plot (M plot). Although the correlation between the variables are already taken into account, but due to marginalizing over X_C, M plot interpretations are contaminated by nuisance omitted variable bias.

Apley and Zhu (2016) also provide estimations of Uncentered and Centered ALEs. First choose a partition on X_j; for $k = 1, 2, \ldots, K$ let $N_j(k) = (Z_{k-1,j}, Z_{k,j}]$ be a partition on X_j. Define $n_j(k)$ as the number of observations in the training data that fall into $N_j(k)$. Finally, let $k_j(x)$ to be the index of the interval in which x lays; that is $x \in (Z_{k_j(x)-1,j}, Z_{k_j(x),j}]$. According to Apley and Zhu (2016), the Uncentered ALE can be estimated by

$$\hat{g}_{j,ALE}(X) = \sum_{k=1}^{k_j(X)} \frac{1}{n_j(k)} \sum_{\{i: x_{i,j} \in N_j(k)\}} \Big[F(z_{k,j}, X_C) - F(z_{k-1,j}, X_C)\Big]$$

and likewise, the Centered ALE is derived:

$$\hat{f}_{j,ALE}(X) = \hat{g}_{j,ALE}(X) - \frac{1}{n} \sum_{k=1}^{K} n_j(k) \hat{g}_{j,ALE}(z_{k,j})$$

or equivalently

$$\hat{f}_{j,ALE}(X) = \hat{g}_{j,ALE}(X) - \frac{1}{n} \sum_{i=1}^{N} \hat{g}_{j,ALE}(x_{i,j})$$

Comparing the expressions for the estimated Partial Dependence and Accumulated Local Effect functions reveal that not only ALE provides a more accurate interpretation for the effect of variables (it provides an unbiased estimate of the feature effect), it is also computationally less expensive.

Data

The Dataset includes quarterly mortgage loans data, as well as some local and national economic variables. The loan data contains acquisition and performance information obtained from Fannie Mae single family dataset.

Mortgage Loan Data

Loan data used in this study is Single Family Loan Acquisition and Performance Dataset, which is publicly available by Fannie Mae. This paper abstracts away from the complexities introduced by structural breaks impacts; hence we focus on the loans originated after 2013. Data includes information about 8.8M mortgage loans from 2013Q1 to 2017Q4. Data contains information for "fully amortizing, full-documentation, single family, conventional fixed-rate mortgages".[17]

[17] Glossary of dataset asserts "[it] does not include data on adjustable-rate mortgage loans, balloon mortgage loans, interest-only mortgage loans, mortgage loans with prepayment penalties, government-insured mortgage loans, Home Affordable Refinance Program® (HARP®) mortgage loans, Refi Plus™mortgage loans, or

Acquisition data includes information at the time of loan origination date. It includes three type of information; information about i) borrower's characteristics, ii) property features, and iii) loan's properties. Borrower characteristics contain information such as borrower's (and if applicable co-borrower's) credit score at origination, number of borrowers, OCLTV and original Debt-to-Income ratio (DTI). Property features include information such as ZIP Code info (only first three digits), and the type of property. Original loan-to-value, interest rate, and loan purpose are examples of information regarding the loan properties.

Acquisition data is accompanied by monthly performance data. Performance data includes information such as the balance on the loan and delinquency or foreclosure status. Performance data continues until the loan either liquidates (matures, gets repurchased, etc.) or goes under either delinquency or foreclosure. Our data prior to cleaning includes 9,833,900 observations for mortgage loans originated between 2013 and 2017. Our data cleaning process includes removing variables that are missing in 70% or more of the instances. After matching the loan data with the local and national data (explained below), we end up with 8,838,161 instances for this study.[18] Delinquency rate for our sample is 0.21%. Majority of the mortgage loans are granted to Single-Family homes (60%), while Project Unit Developments (PUD) are the second rank with 29%, and condos are ranked third (10%). Manufactured Houses[19] and Co-Opstogether compose 1% of mortgage loans. Data comprise single-unit (97.8%), two-unit (1.5%), three-unit (0.3%), and four-unit properties (0.3%). Majority of the loans (51%) are acquired for refinancing of an existing property (29% without Cash-Out and 22% with Cash-Out), while the rest of the loans are for purchasing a new property. We discard the instances, including only three observations, with undefined loan purpose. 69% of the loans are channeled through Correspondents and the rest through Retail.

We use two categories for the number of buyers; 49.5% of the loans are granted to the properties with one buyer and the rest are granted to properties with two or more buyers. First-time home buyers comprise 81% of the mortgagors and less than 1% of the loans are relocation mortgage loans.[20] The data also shows 0.021% of total loans are modified during the first year after origination. Debt-to-Income, Borrower's Credit Score, and Combined-Loan-to-Value are of great importance. Hence, we impute the missing values of these features. If either Debt-to-income and Credit Score is missing, we impute the value with the sample average of that variable. Missing Combined-Loan-to-Value data is imputed by Original-Loan-to-Value; that is, other loans rather than the mortgages are ignored.

non-standard mortgage loans."

[18] Difference between summary statistics before and after the data cleaning is not significant.

[19] Homes that are constructed off-site.

[20] As defined by Fannie Mae Single Family Loan Glossary, "Made to borrowers whose employers relocate their employees".

Local and National Economic Factors

To improve the explanatory power of the model, we follow the literature and add some local data to our dataset that potentially can affect the loan performance. This includes unemployment data, house price index, and per capita personal income. Bureau of Labor Statistics (BLS) provides local area unemployment statistics at different levels. We use the monthly unemployment rate at the state level.We also include Housing Price Index (HPI), obtained from Federal Housing Finance Agency, at the state level to control for the price of properties. Per capita household income data for each state is also added to the dataset. This data is obtained from Bureau of Census. Slope of the Yield Curve (difference between the yield rate of 10-year and 3-month Treasury Bonds) and Federal Funds Rate are also added to the dataset to control for the national level fluctuations in the interest rate. Table 1 provides data summary statistics.

	Mean	SD
Per Capita State Income	49108.00$	7388.45$
State HPI	155.70$	32.94$
State Unemployment Rate	5.61%	1.49%
Federal Funds Rate	0.36%	0.36%
Slope of the Yield Curve	1.90%	0.48%
Credit Score	755.78	45.69
Original Rate	3.96%	0.57
OCLTV	74.11%	17.13
OLTV /100	73.44%	17.15%
DTI	33.279%	9.159%
Original House Value	336006.00$	220576.00$
Original Loan Amount	230585.00$	118762.00$

Table (1) Summary Statistics

Empirical Results

We split the data described in the previous section; we train the XGBoost model using 80% of observations and then will use the rest of the observations to test it. We train the model using mean squared errors as the loss function, learning rate ρ_m equal to 0.1, maximum tree depth of four,[21] and penalize the complexity with $\lambda = 1$. Accuracy of delinquency prediction reflects the goodness-of-fit for our model. We compare prediction accuracy of XGBoost and Logistic models. Since XGBoost outperforms Logistic, we use it to find the important features

[21] In each step, the maximum number of edges between a node and the root is four.

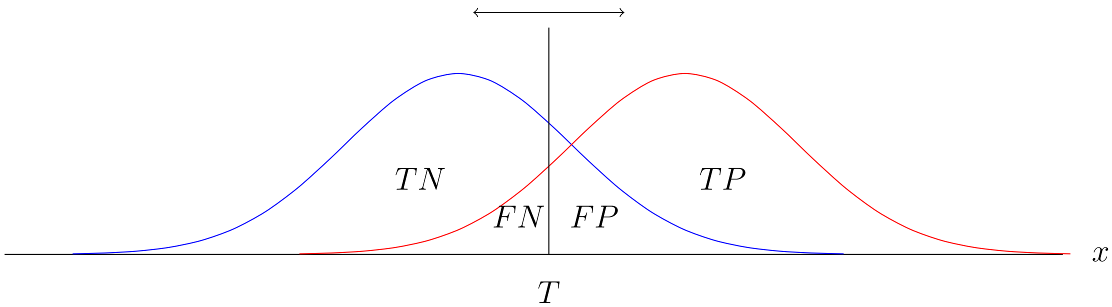


Figure (1) Schematic Graph Explaining ROC Construction.

describing the delinquency. The next step is to interpret the results of the Machine Learning model. In particular, we are interested in the impact of each feature on prediction power, the behavior of features (i.e. linear vs. non-linear, monotone vs. non-monotone, etc.), and the interaction between features. Table 2 shows the Logistic regression results. Although we discard this model for our main purposes (feature importance and interpretability), it is worth mentioning that Logistic results are generally in agreement with XGBoost results.

Out of Sample Analysis

In Section 2 we introduced two classifier models; Logistic Regression as the baseline model and Gradient Boosting as an alternative Machine Learning approach. We need to examine the prediction accuracy of each classifier models using the test data. That is, we compare the predictions of a model with the actual data. This helps us to construct Receiver Operating Characteristic (ROC) curve for each model. For each observation in the test data, either the model correctly predict the default (TP), or incorrectly predicts the default (FP), or incorrectly predicts no default (FN), or finally, correctly predicts no default (TN). Notice FP and FN cases refers to the Type I and Type II errors, respectively.

Predicted Positive and Negative distributions provide insight about prediction accuracy. Assume we plot the two distributions against one variable, x.[22] Figure 1 shows the mentioned distributions over the range of x. The classifying problem usually reduces to choosing a threshold, T, such that the instance is classified positive[23] for $x > T$. A reduction in the overlap between two distributions results in a more accurate prediction, i.e. lower Type I and II errors. Define True Positive Rate as

$$TPR = \frac{TP}{TP + FN}$$

[22]The idea can be generalized to the case that X contains several features.
[23]Without loss of generality

and False Negative Rate as

$$FPR = \frac{FP}{FP + TN}$$

Changing the threshold, T will change the False Negative and False Positives and consequently TPR and FPR. ROC curve plots the true positive rate against the false positive rate for different classification thresholds. ROC Curves for Logit and Machine Learning models are depicted in Figure 2. As indicated in the figure, for each value of False Positive, XGBoost provides a higher True Positive. This means XGBoost has a higher prediction accuracy.

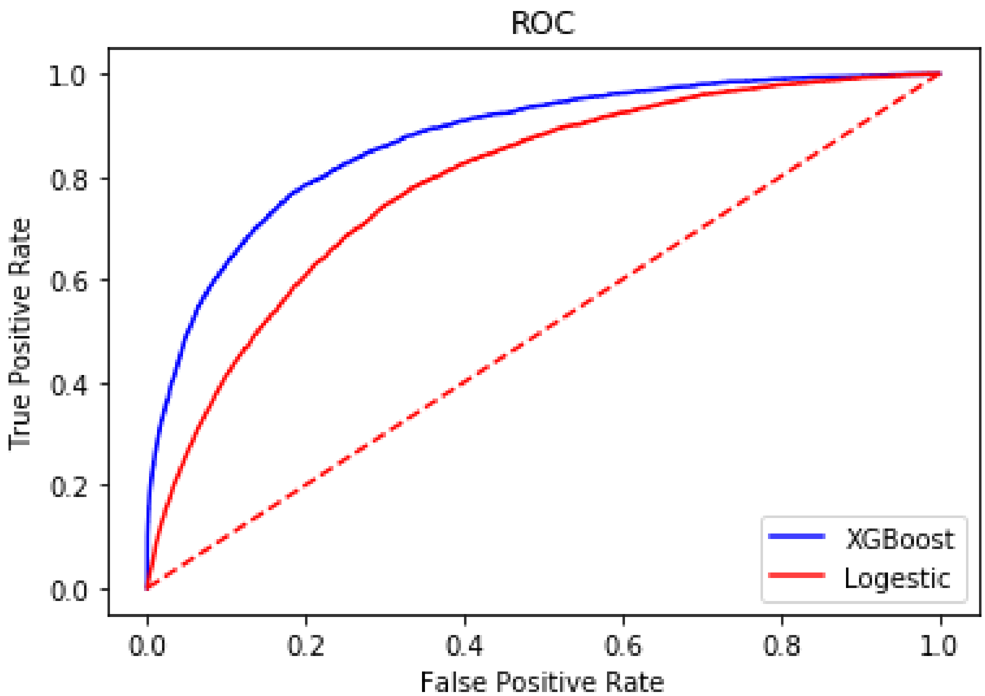


Figure (2) ROC for Gradient Boosting and Logistic Models: XGBoost clearly outperforms Logistic model.

Consider a randomly chosen defaulted observation and a randomly chosen non-defaulted observation. An ideal classifier always ranks the default probability of the former higher than the latter. It turns out that the Area Under the ROC Curve (AUC) reflects the probability that a randomly chosen positive instance is ranked higher than the randomly chosen negative instance. AUC is an aggregate measure of performance for all of the thresholds shown in Fig 1. For XGBoost and Logistic model AUC is 0.87 and 0.71, respectively. Deep Learning method in Sirignano et al. (2016) results in an AUC equal to .070 for +90-day delinquencies, which is close to Logistic model's AUC (0.67).[24]

[24] The purpose of this juxtaposition is not to compare the performance of these models, but to provide a reference from the literature. In Sirignano et al. (2016) study a multi-class model containing foreclosure, in

Feature Importance

Since we demonstrated the superiority of Gradient Boosting over Logistic model and have shown it is comparable to Deep Learning methods, it can be used to provide insight over the important factors influencing mortgage delinquency. There are three measures that can be used to find the relative importance of a feature in our XGBoost tree model. We can either find the weight (or frequency), cover, or gain associated with each feature. The frequency measures the number of times a feature is used to split the data across all the derived trees. Cover calculates the average decrease in the loss function when a feature is used for spiting the data. Finally, gain calculates the weighted frequency; that is, the number of times a feature is used to split the data across all derived trees, weighted by the number of training data that pass through those splits.[25] Unfortunately, these approaches may render different results. Hence, in this paper we use Permutation Feature Importance based on Breiman (2001). We use the model-agnostic version introduced by Fisher, Rudin, and Dominici (2018).

Based on Permutation approach, the higher the increase in the error resulted from shuffling a feature's values, the more important the impact of that feature in the model.[26] First, we calculate the model error, e^{orig}. Then for each feature j, we can permute the observations in feature j (we shuffle the values of feature j among the observations) and calculate the associated error of the model, e^{Perm}. Feature importance for factor j can be calculated either by e^{Perm}/e^{orig} or $e^{Perm} - e^{orig}$. Figure 3 shows that Borrower's Credit Score,[27] Federal Funds Rates, the original interest rate, the number of borrowers, Debt-to-Income,[28] state Personal Income, the regional (south) dummy, House Price Index in the State, Original Combined Loan-to-Value, Loan Modification dummy, Original Loan-to-Value, and unemployment in the state are the most important features in the model.

The first three important features are the all determinants of the mortgage loan contract. While FFR and original rate directly determine interest payments, credit score affect other aspects of the loan contract as well. For instance, the loan approval, Loan-to-Value, and mortgage insurance all, to a great extent, depend on the credit score. This unique feature of our model shows during the expansion period of 2013-2018, when there is no drastic reduction in the housing values by external factors, delinquency is mainly determined by interest payments. Also notice, during this period, since the economy is generally growing, income and unemployment cannot attribute to the delinquency. The contrast between our results and the results in the literature, especially the results Gerardi et al. (2013) and Sirignano

addition to different types of delinquency. The reported AUC's above are the highest values in the +90-day delinquencies.

[25] Scott Lundberg: Interpretable Machine Learning with XGBoost.

[26] Molnar (2019).

[27] In presence of a co-borrower, the average of borrower and co-borrower's FICO scores.

[28] Total monthly obligation divided by the total monthly income at origination.

et al. (2016), suggests the policy recommendations from these studies should be contingent on business cycle. This is in line with Bagherpour (2018) where the importance of factors explaining the mortgage default changes when we split the data into before, during, and after Great Recession periods.

Credit score and the relation with mortgage default is investigated in many studies. Elul et al. (2010) found the effect of FICO score statistically significant, a higher credit score decreases probability of default. Li and White (2009) also studied the relation between bankruptcy (credit card loans) and foreclosure (mortgage loan) and found the two are generally complementary. Here we found credit score is an important factor in predicting the mortgage default. In the next section we describe the nature of this relation and investigate whether the data confirms the results of these two studies.

Literature has also studied the loan modification and the redefault risk. Calem, Jagtiani, and Maingi, and Abell (2018) find the positive effect of modification on reducing the redefault risk for during 2008-2011. However, Hwang, Jang, and Watcher (2016) and Chen, Xiang, and Yang (2018) find that the modified loans are more likely to default. This paper first emphasizes on the effect of loan modification on the redefault risk. Second it confirms the latter effect; modified loans are more exposed to redefault risk.[29]

The importance of employment, at different levels, have been recognized in Gerardi et al. (2013) and Sirignano et al. (2016), to name a few. In Gerardi et al. (2013), the unemployment data is at the individual level, while in Sirignano et al. (2016) it is at the state level. Our dataset contains unemployment rate and per capita personal income at the state levels. Our model confirms the importance of unemployment but introduces the income as even a more important factor. Hence, this study suggests the labor status is not only important at the extensive margin, but also at the intensive margin.

Interpreting the Results: Effect of Factors on Delinquency

In this section, we study the effect of each factor on the mortgage loan delinquency. First, we start by Partial Dependence Plots; the marginal effect of factors on the outcome. Panel (a) in Figure 4 shows the effect of credit score. This graph shows credit score and delinquency probability are, as expected, inversely related. Lower credit score is associated with a riskier mortgagor; hence lenders will impose a stricter contract which in turn results in a higher delinquency risk. Furthermore, the plot suggests a nonlinear relation between credit score and delinquency. Hence imposing a linear relation between would misrepresent the nature of the effect. As shown in the graph, the same amount of increases in the lower credit score levels causes a higher drop in delinquency probability. We still need to examine whether

[29]The adverse effect of loan modification cannot be deducted by Feature Importance; this conclusion is based on the results of the following section.

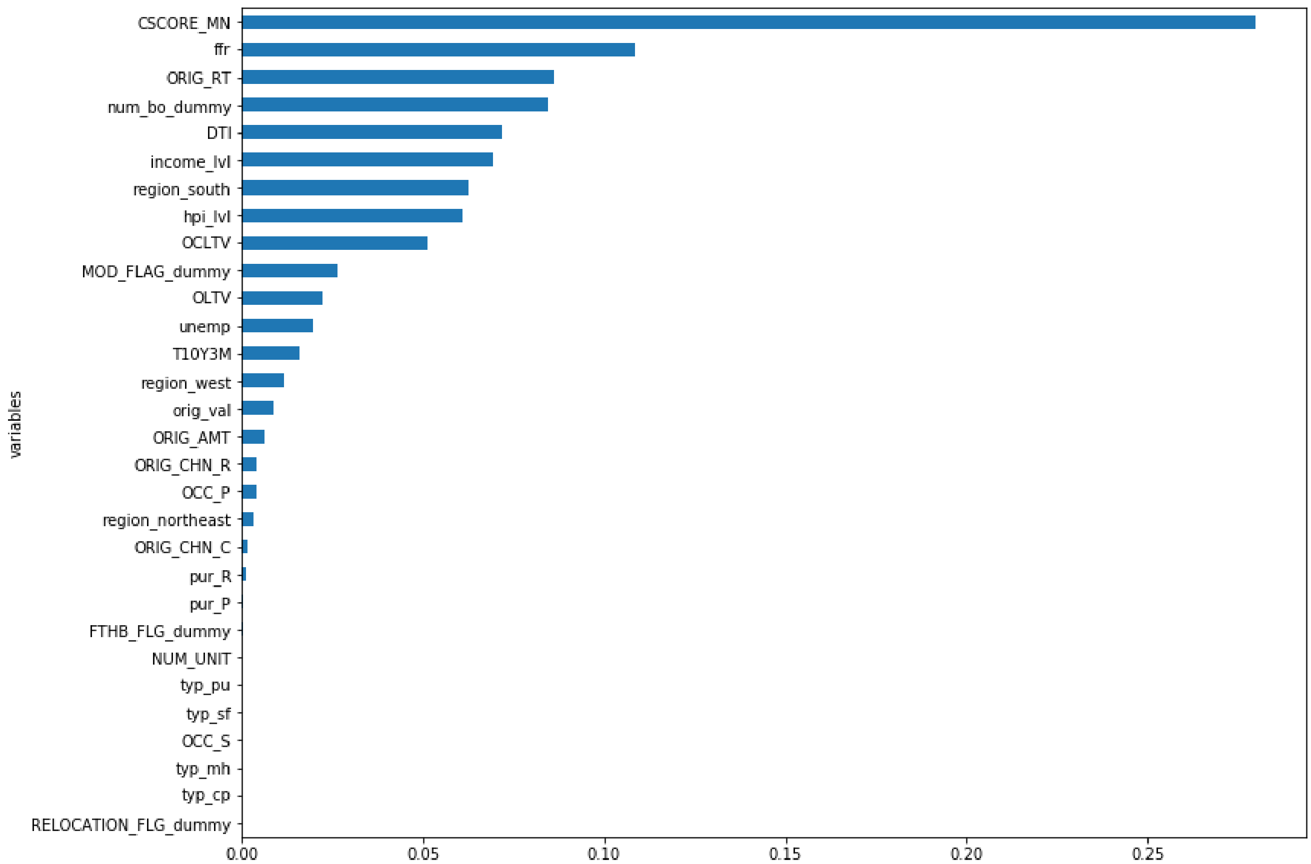


Figure (3) Permutation Feature Importance of the XGBoost Model

the effect of credit score is merely through interest rate or it does affect other aspects of the contract as well.

The effect of interest rate is further examined in panels (b) and (c) of Figure 4. The higher the interest rate, the higher is the delinquency probability. Since the loan interest rate depends on nominal interest rate (Federal Funds Rate), it is highly expected that delinquency also depends on FFR. Panel (c) depicts this relation.

Panels (d), (e), and (f) in Figure 4 investigate the effect of borrowing on delinquency from different angles. As seen in these plots, higher Debt-to-income, OCLTV, and OLTV increase the delinquency rate. However, low levels of these features have no effect on delinquency. Delinquency increases only when DTI is between 25% to 45%. OCLTV and OLTV both start to have impact around 60%. The first three panels in Figure 5 show the effect of three state-level variables on delinquency: unemployment rate, per capita income, and HPI. Since the data is highly concentrated for these values, we focus our attention on $[0.2, 0.8]$ quantiles. Panels (a) and (b) show the effect of unemployment and income on delinquency. As seen in these graphs the relation between these factors and delinquency is not monotonic. One explanation might be the urban-rural difference in the distribution of these factors. Urban areas have a higher income levels and higher unemployment rates. Hence a state with higher urban population is expected to have higher income and unemployment as well. Accessing more detailed data regarding the location for which the loan is originated can shed more light on the nature of this relation. Finally, as panel (c) presents, higher state HPI is associated with higher delinquency rate. This result is against the negative equity. If according to negative equity low values of a house (compared to the mortgage loan) stimulates the default, we should have derived the opposite result.

Figure 5 also shows the effect of dummy variables which recognized as important features determining delinquency. Panel (d) confirms the negative impact of loan modification on redefault risk.[30] Panel (e) shows higher number of borrowers will decrease default possibility; it is less likely that two, or more, borrowers miss a payment. Panel (f) shows higher possibility of delinquency in southern states.

To further investigate the extent of heterogeneity in the effect of features on delinquency we show the Individual Conditional Expectation (ICE) plots in Figure 6 and 7. This is important, in particular, for DTI, OCLTV, and OLTV where PDP shows there is no effect in some intervals. We have derived the centered ICE plots; hence the negative numbers show a decrease in delinquency probability. The thick black line represents the PDP. The y-axis in these graphs show the log-odds[31] and x-axis is re-scaled to represent the quantiles of the variable of interest. As seen in the graphs, individual observations have homogeneous behavior. Therefore, the average (PDP) truly represents the factor's marginal effect.

We can also use the 2D-PDPs to visualize the joint effect of variables. It is worth noticing whether behavior of a feature changes when we draw its joint PDP against another feature. Panel (a) in Figure 8 shows the joint effect of credit score and OCLTV. As seen in the graph, OCLTV increases the delinquency rate when $OCLTV > 60\%$. The general trend confirms the individual PDPs for each feature; the higher (lower) the credit score (OCLTV), the lower the delinquency probability. Joint PDPs help us to discover anomalies in the behavior of features; for instance, for credit scores above 800, higher OCLTV may reduce delinquency probability. Since it covers less than 10% of our observations, this solo anomaly does not disapprove our model. We can also study the effect of state income and unemployment rate. Panel (b) in Figure 8 shows the bimodal behavior we have seen already in 1D-PDPs. Generally, higher (lower) unemployment rate (income level), increases delinquency probability. The block at the bottom-left corner of the graph represents an area with low income and low unemployment rate that has an unusual high delinquency rate. As mentioned before, urban-rural difference can explain this bimodal behavior. Hence, we can assume this area represents a state that is highly rural.

[30] Hwang, Jang, and Watcher (2016) and Chen, Xiang, and Yang (2018).

[31] This helps us to see the behavior of PDP in the presence of the extreme individual cases that have a response an order of magnitude larger than PDPs.

Finally, we study the effect of each feature, taking into account the correlation with other ones. If a feature is highly correlated with other features, PDPs provide a biased estimate on its marginal effect. In these cases, Accumulated Local Effect plots provide a more realistic estimate of the effect. Figure 9 and Figure 10 show the ALE plots for the important variables in our model.[32] The results are generally comparable to PDPs. However, one instance is worth noticing. The high correlation between OCLTV and OLTV makes the PDPs biased. Comparing PDPs and ALEs for these two features shows PDP for OCLTV (OLTV) is biased upward (downward); that is, PDP overestimates the effect of OCLTV and underestimates the effect of OLTV. Focusing on PDPs and ALEs of state unemployment rate and per capita income shows the unbiased effect of these factors are larger than PDP estimation. ALE graphs show that higher unemployment rate (income) results in a higher increase (decrease) in delinquency rate. This, once again, suggests a correlation between unemployment rate and per capita income at the state levels. A dataset with more detailed local information can further shed light on this issue.

Figure 11 shows the 2D-ALE plot for the effect of credit score and FFR, when we consider the correlation of these factors with the other ones. As seen in the graph, credit score's effect is analogous to PDP and 1D-ALE plot. Therefore, we can conclude the effect of credit score is not solely through impacting the interest rate. That is, the credit score's importance is due to its impact on all aspects of mortgage contract.

conclusion

This paper studies the factors affecting mortgage loan delinquency. First, we applied a Machine Learning technique, XGBoost, to mortgage loan data from 2013Q1 to 2017Q4 to predict +90-day delinquency and showed its superiority to traditional Logistic Model. Then we used Permutation Feature Importance to find the most important factors in the model. Finally, we use Partial Dependence Plots (PDPs), Individual Conditional Expectation (ICE) plots, and Accumulated Local Effect plots to study the impact of important factors on delinquency probability. We find for the period of 2013Q1-2017Q4, the most important factors are the ones affecting contract terms, interest rate in particular. Credit scores, Federal Funds Rates, and the original interest rate are the most important factors predicting a 90-day delinquency. These factors dominate other factors mentioned in the literature such as OCLTV (based on negative equity) and unemployment.

XGBoost approach is a Machine Learning approach that is easy to implement and inter-

[32] Dummies and classes are not proper candidates for ALE plots as we cannot properly derive the "accumulated" effects for them.

pret. In that sense, it is very suitable to be used in economic empirical studies. Applying this approach to a dataset containing more information about the local variables can further illustrate the effect of such variables on delinquency.

References

Apley D. W. and J. Zhu (2016),“Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models”. Forthcoming.

Asay M. R. (1979), “Rational mortgage pricing.” Technical report, Ph.D. dissertation. University of Southern California, Los Angeles.

Bagherpour, A. (2018), “Predicting Mortgage Loan Default with Machine Learning Methods”, Ph.D. dissertation. University of California, Riverside.

Bhutta, N., J. Dokko, and H. Shan, (2010), “Consumer Ruthlessness and Mortgage Default During the 2007 to 2009 Housing Bust”, Journal of Finance, Forthcoming.

Breiman, L. (2001), “Random Forests.” Machine Learning 45 (1). Springer: 5-32.

Breiman L., J. H. Friedman, R. Olshen, and C. Stone (1983), “Classification and Regression Trees”. Wadsworth.

Butaru, F., Q. Chen, B. Clark, S. Das, A. Lo and A. Siddique (2016), “Risk and risk management in the credit card industry”, Journal of Banking and Finance 72, 218–239.

Calem, P., J. Jagtiani, R. Maingi, and D. Abell, (2018), “Redefault Risk in the Aftermath of the Mortgage Crisis: Why Did Modifications Improve More than Self-Cures?” FRB of Philadelphia Working Paper No. 18-2.

Chen, T. and C. Guestrin. (2016), “XGBoost: A Scalable Tree Boosting System.” 785-794.

Chen, J., J. Xiang, and T. T. Yang, (2018),“Re-Default Risk of Modified Mortgages”, International Real Estate Review, 21(1), 1-44.

Deng, Y., J. M. Quigley, and R. Van Order (2000), “Mortgage terminations, heterogeneity and the exercise of mortgage options”. Econometrica, 68(2):275–307.

Elul, R., N. S. Souleles, S. Chomsisengphet, D. Glennon, and R. Hunt (2010), “What “Triggers” Mortgage Default?” American Economic Review, 100 (2): 490-94.

Fisher, A., C. Rudin, and F. Dominici (2018), “Model Class Reliance: Variable importance measures for any machine learning model class, from the ‘Rashomon’ perspective.” http://arxiv.org/abs/1801.01489.

Fitzpatrick, T. and C. Mues (2016), “An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market”, European Journal of Operational Research, Volume 249 (2), 427-439.

Foote, C. L., K. Gerardi, and P S. Willen (2008), "Negative equity and foreclosure: Theory and evidence", Journal of Urban Economics, Volume 64, Issue 2, 234-245.

Foster, C. and R. Van Order (1984), "An Option-based Model of Mortgage Default." housing finance Review, 3(4): 351–72.

Friedman, J. H. (2001), "Greedy function approximation: A gradient boosting machine." Ann. Statist. 29, no. 5, 1189-1232.

Gerardi, K., K. F. Herkenhoff, L. E. Ohanian, and P. S. Willen (2013), "Unemployment, Negative Equity, and Strategic Default", Unpublished.

Herkenhoff, K. F. and L. Ohanian (2012a), "Foreclosure delay and us unemployment." Federal Reserve Bank of St. Louis Working Paper Series.

Hwang M., I. Jang, and S. Watcher (2016), "Evaluating Loan Modifications: 2008-2014". Unpublished.

Kau, J. B., D. C. Keenan, and T. Kim (1994). "Default Probabilities for Mortgages." Journal of urban Economics, 35(3): 278–96.

Khandani, A., A. Kim, and A. Lo (2010), "Consumer credit-risk models via machine-learning algorithms", Journal of Banking and Finance 34(11), 2767–2787.

Kvamme, H., N. Sellereite, K. Aas, and S. Sjursen (2018) "Predicting mortgage default using convolutional neural networks", Expert Systems with Applications, Volume 102, 207-217,

Li, W. and M. White (2009), "Mortgage Default, Foreclosure, and Bankruptcy". National Bureau of Economic Research Working Paper 15472.

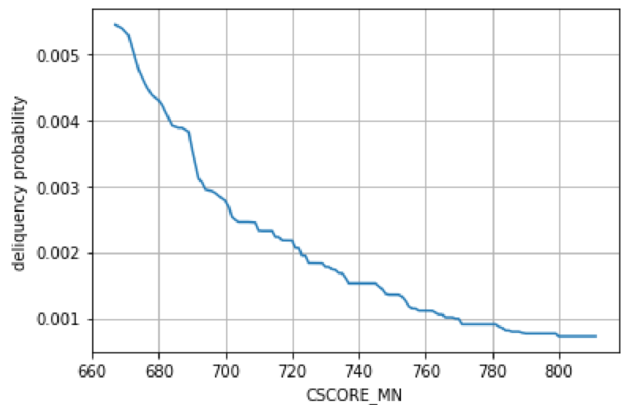
Molnar, C. (2019), "Interpretable Machine Learning A Guide for Making Black Box Models Explainable."

Sirignano, J., A. Sadhwani, and K. Giesecke (2016), "Deep Learning for Mortgage Risk". arXiv:1607.02470v2

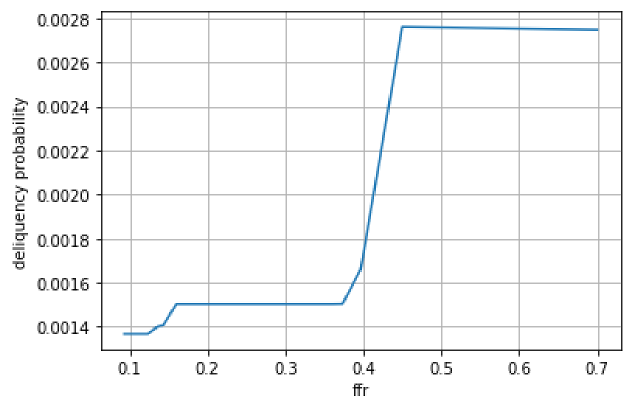
Zhao, Q. and T. Hastie (2019). "Causal Interpretations of Black-Box Models". Journal of Business & Economic Statistics, 1-19.

	Marginal	(SE)
Per Capita State Income /1000	-0.022^{***}	(0.001)
State HPI /1000	13.53^{***}	(0.413)
State Unemployment Rate $\times 100$	0.064	(1.01)
Federal Funds Rate	0.446^{***}	(0.040)
Slope of the Yield Curve	0.074^{*}	(0.034)
Credit Score /100	-1.722^{***}	(0.019)
Original Rate	0.430^{***}	(0.026)
OCLTV /100	1.874^{***}	(0.245)
OLTV /100	-0.102	(0.269)
DTI /100	2.912^{***}	(0.119)
Original House Value /$1M	-0.435^{*}	(0.218)
Original Loan Amount /$1M	$-.024$	(0.303)
Loan Modification (Dummy)	7.003^{***}	(0.084)
Number of Units (Dummy)	.066	(0.051)
Relocation (Dummy)	0.765^{***}	(0.163)
First Time Home Buyer (Dummy)	0.079^{**}	(0.025)
Number of Borrowers (Dummy)	-0.660^{***}	(0.019)
Region (West)	-0.773^{***}	(0.031)
Region (South)	0.060	(0.032)
Region (North East)	0.037	(0.040)
Channel (Retail)	-0.209^{***}	(0.029)
Channel (Correspondent)	-0.170^{***}	(0.030)
Property Type (Single-Family)	0.241^{***}	(0.032)
Property Type (Co-Op)	-0.725^{**}	(0.022)
Property Type (Manufactured Home)	0.488^{***}	(0.109)
Property Type (PUD)	0.272^{***}	(0.035)
Occupancy Type (Principal)	0.312^{***}	(0.042)
Occupancy Type (Secondary)	0.196^{**}	(0.070)
Loan Purpose (Purchase)	0.175^{***}	(0.027)
Loan Purpose (Refinance)	0.143^{***}	(0.028)

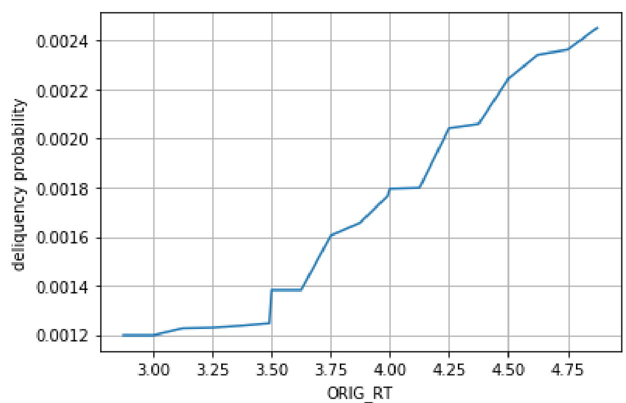
Table (2) Logistic Regression Results



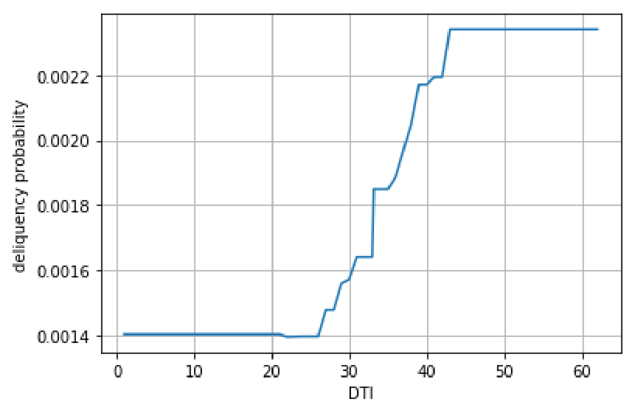
(a) PDP: Effect of Credit Score on Delinquency probability



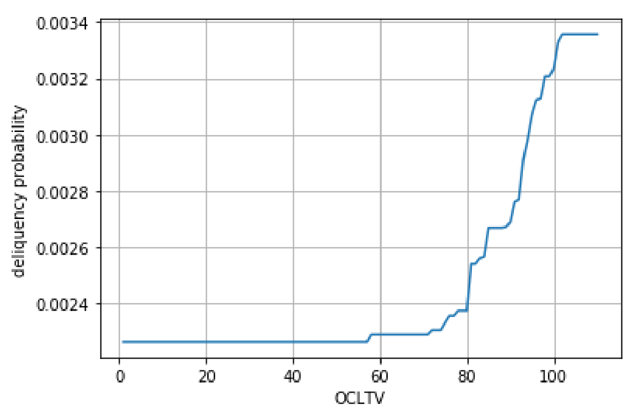
(b) PDP: Effect of FFR on Delinquency probability



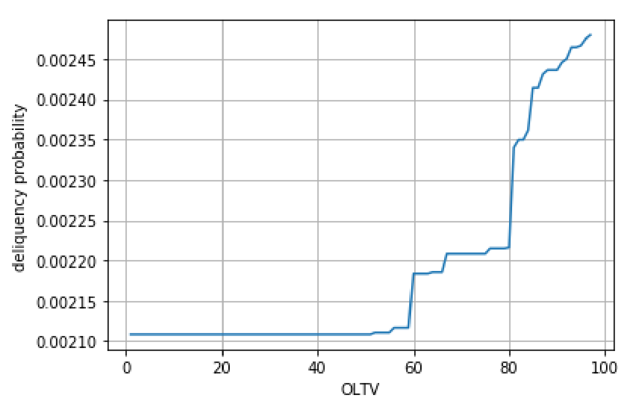
(c) PDP: Effect of Original Interest Rate on Delinquency probability



(d) PDP: Effect of Debt-to-income on Delinquency probability

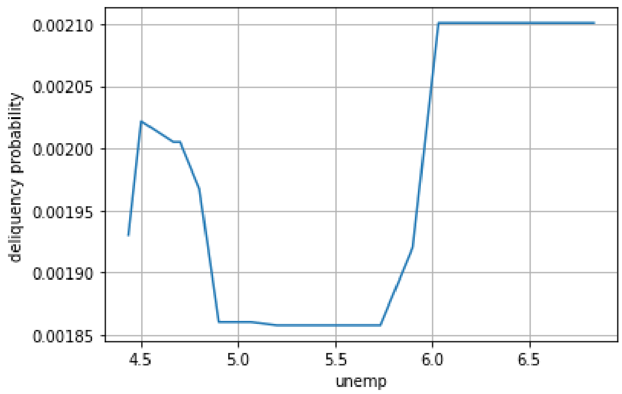


(e) PDP: Effect of OCLTV on Delinquency probability

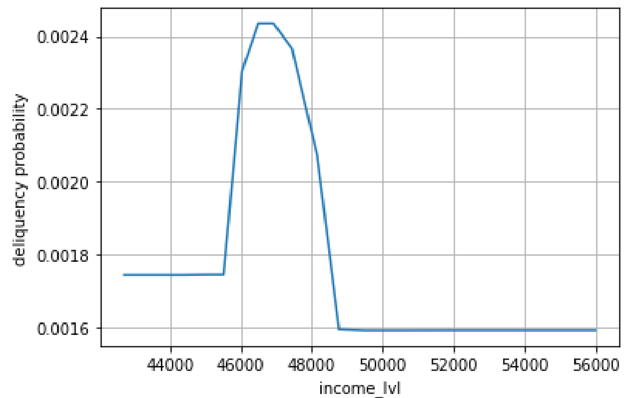


(f) PDP: Effect of OLTV on Delinquency probability

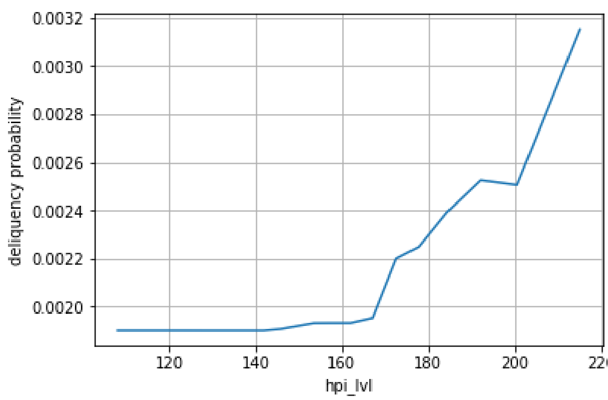
Figure (4) PDP Plots



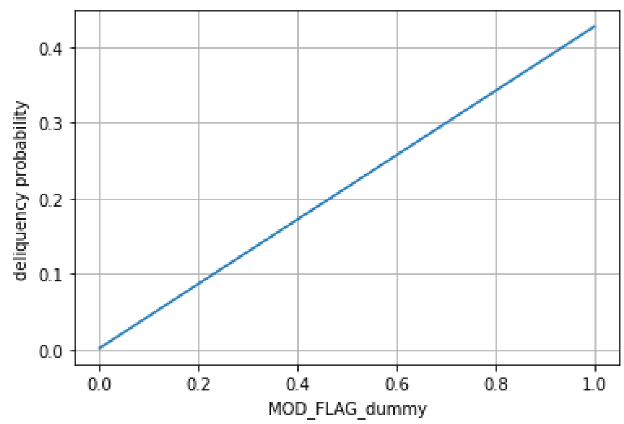
(a) PDP: Effect of State Unemployment Rate on Delinquency probability



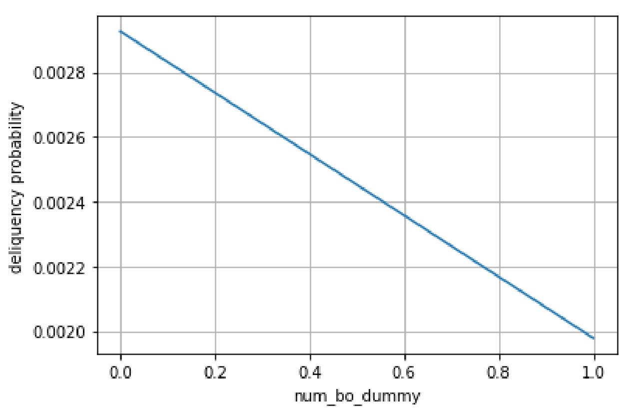
(b) PDP: Effect of State Per Capita Income on Delinquency probability



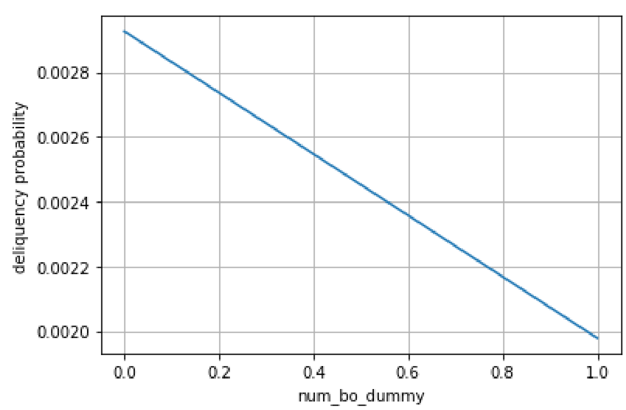
(c) PDP: Effect of State HPI on Delinquency probability



(d) PDP: Effect of Loan Modification Flag on Delinquency probability

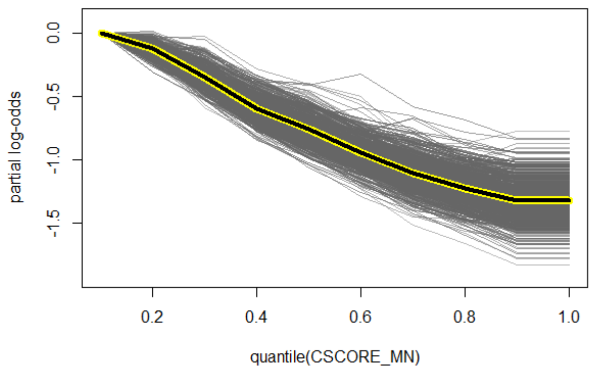


(e) PDP: Effect of Number of Borrowers Dummy on Delinquency probability

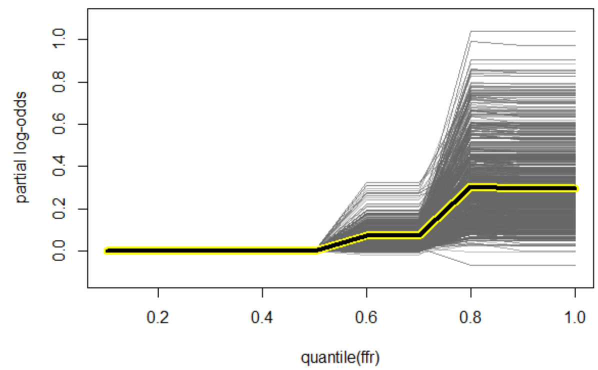


(f) PDP: Effect of South Region on Delinquency probability

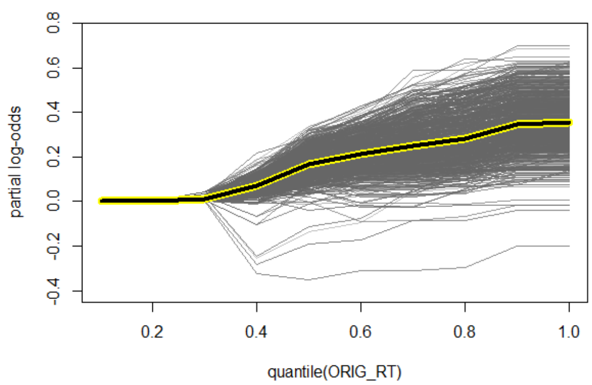
Figure (5) PDP Plots



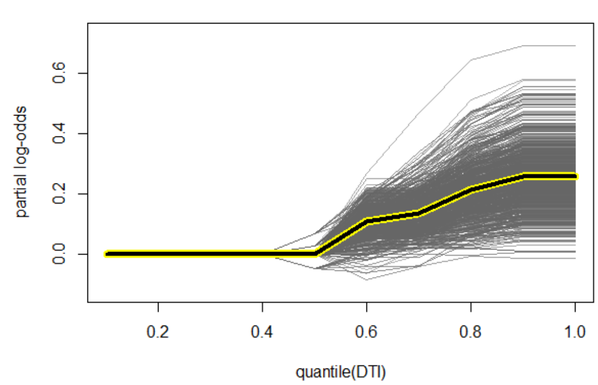
(a) ICE Plot for Credit Score



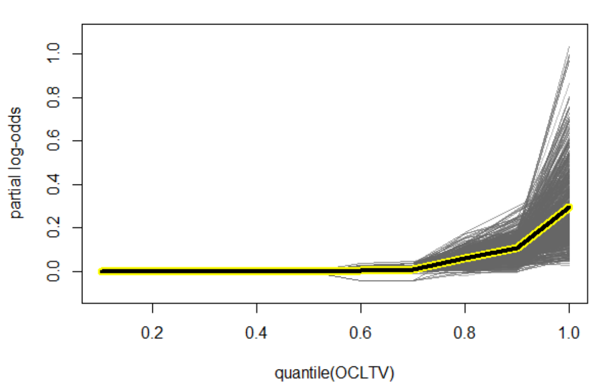
(b) ICE Plot for Federal Funds Rate



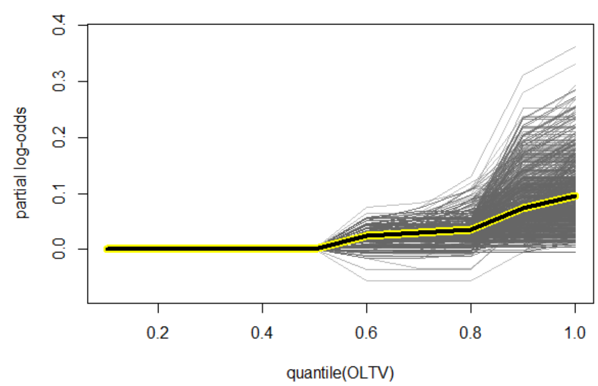
(c) ICE Plot for Original Interest Rate



(d) ICE Plot for Debt-to-Income

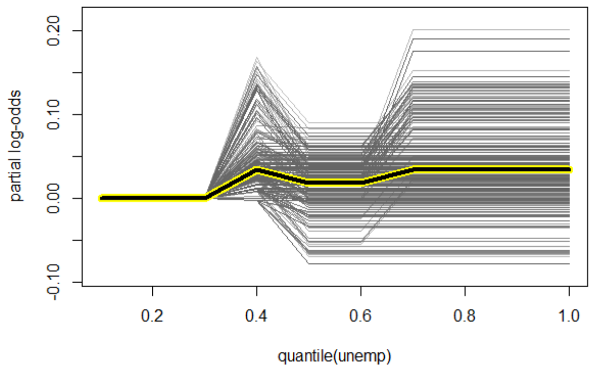


(e) ICE Plot for OCLTV

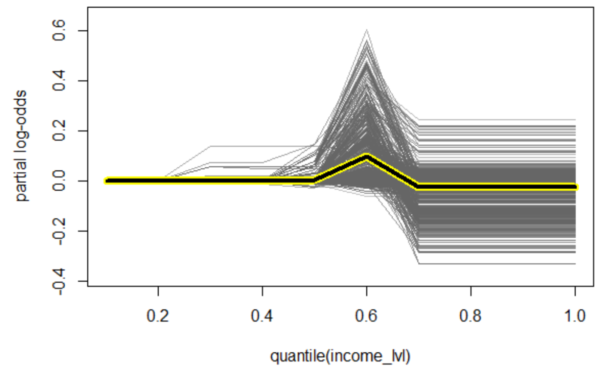


(f) ICE Plot for OLTV

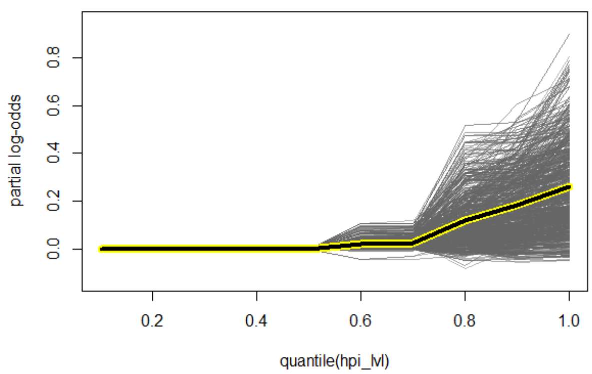
Figure (6) ICE Plots



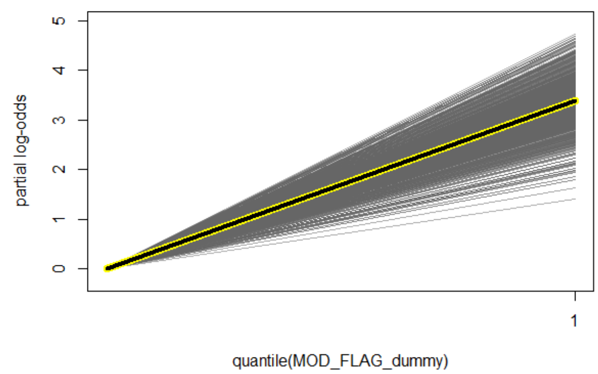
(a) ICE Plot for State Unemployment Rate



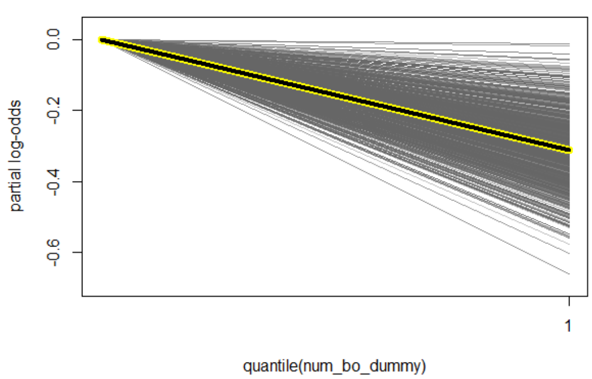
(b) ICE Plot for State Per Capita Income



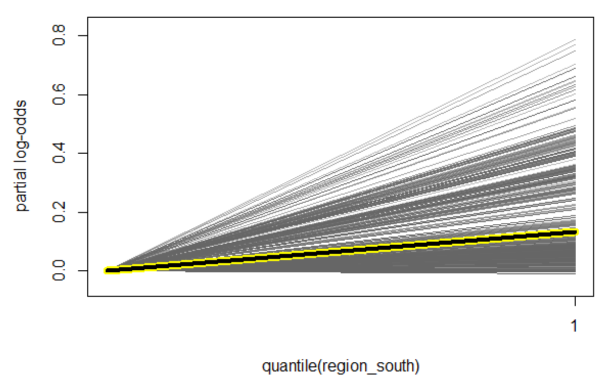
(c) ICE Plot for State HPI



(d) ICE Plot for Loan Modification Dummy

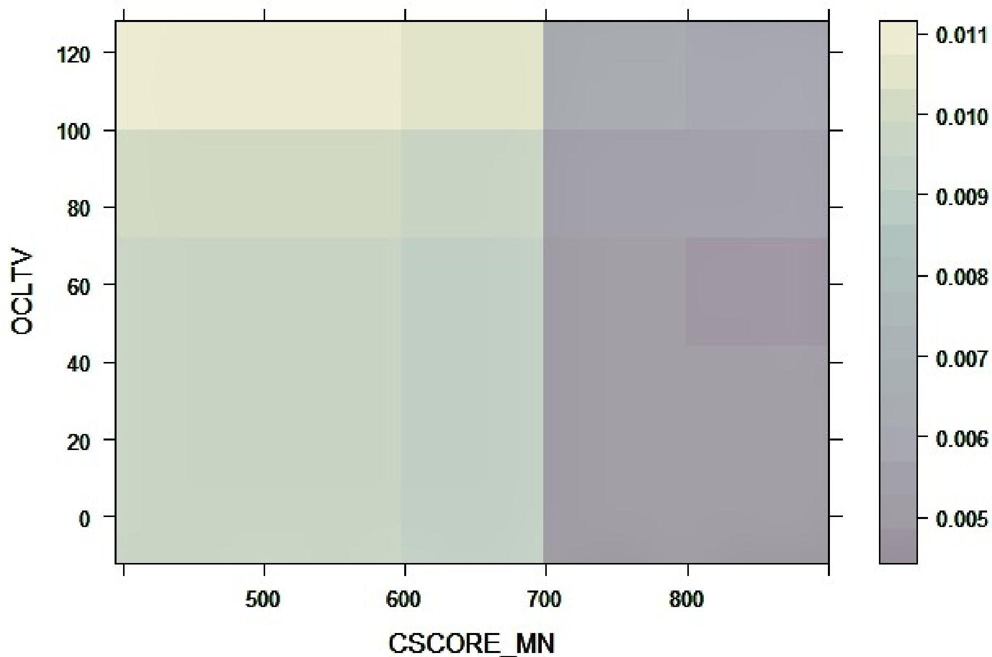


(e) ICE Plot for Number of Borrowers

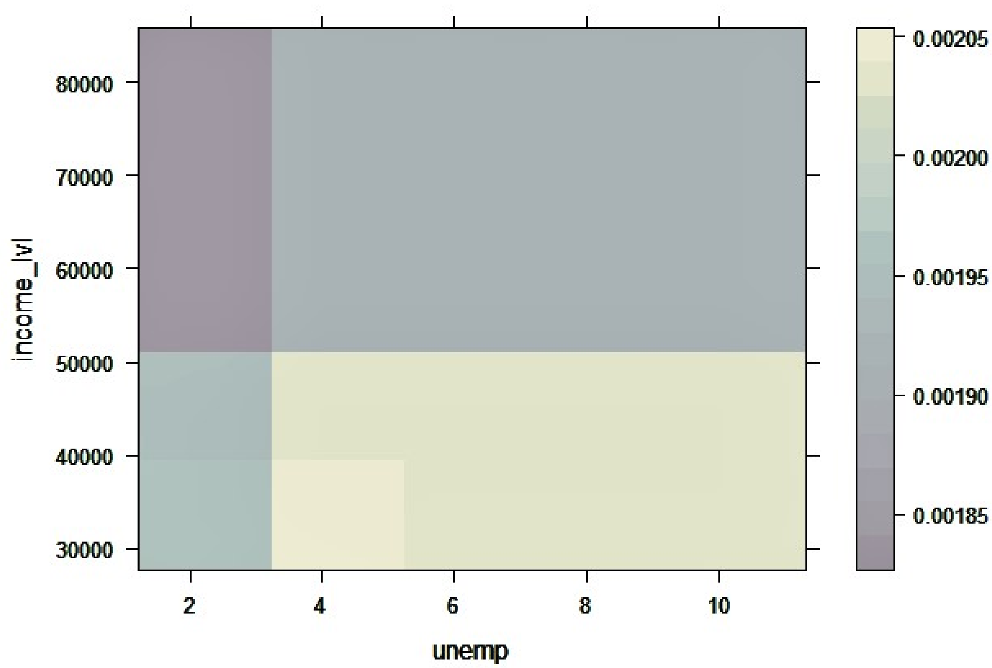


(f) ICE Plot for South Region

Figure (7) ICE Plots

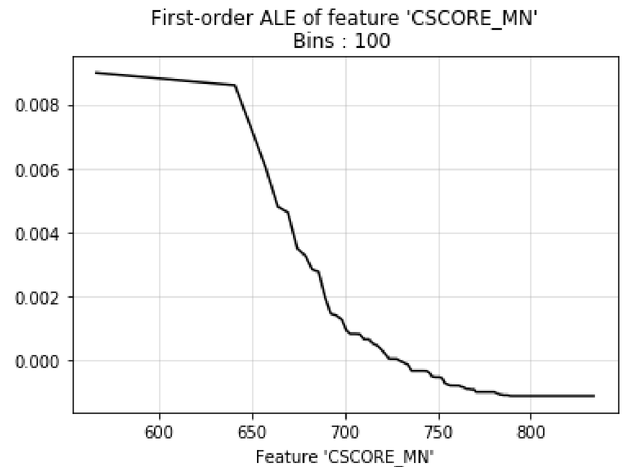


(a) PDP: The Joint Effect of Credit Score and OCLTV on Delinquency probability



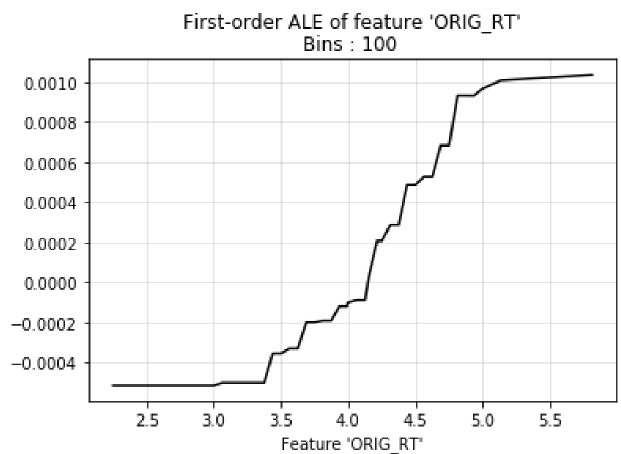
(b) PDP: The Joint Effect of State Unemployment Rate and Income Level on Delinquency probability

Figure (8) 2D Partial Dependence Plots

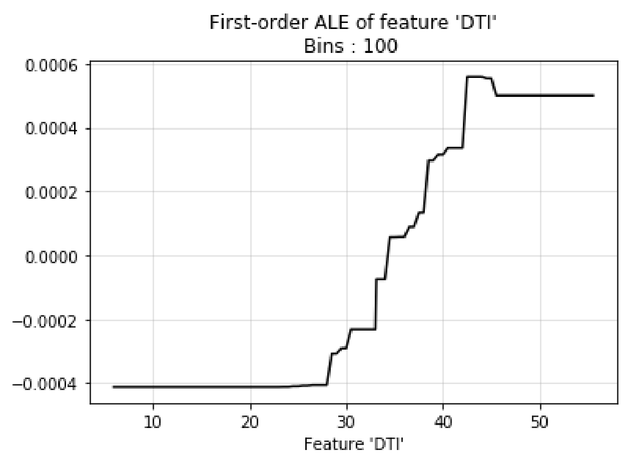


(a) ALE: Unbiased Effect of Credit Score on Delinquency probability

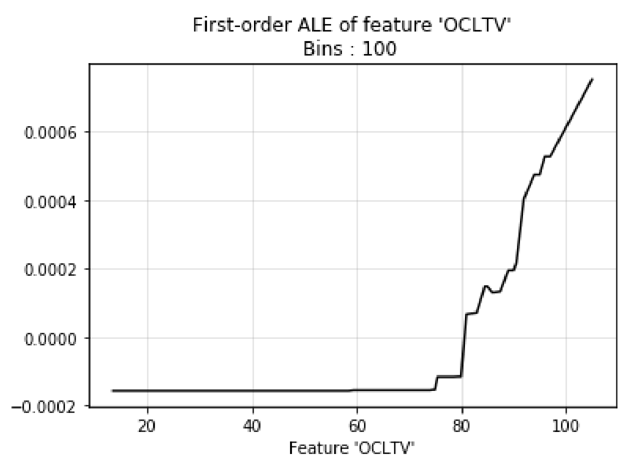
(b) ALE: Unbiased Effect of FFR on Delinquency probability



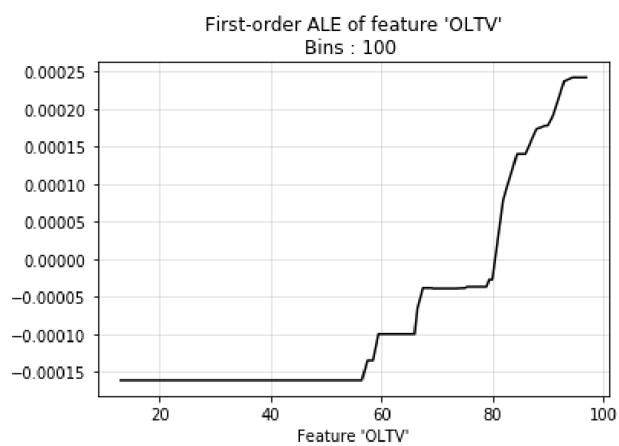
(c) ALE: Unbiased Effect of Original Interest Rate on Delinquency probability



(d) ALE: Unbiased Effect of DTI on Delinquency probability

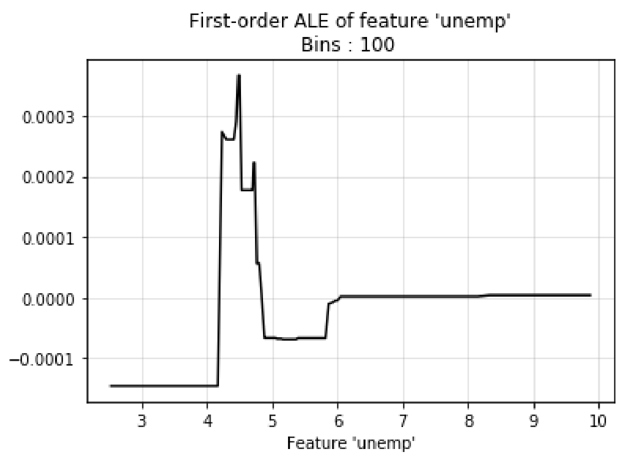


(e) ALE: Unbiased Effect of OCLTV on Delinquency probability

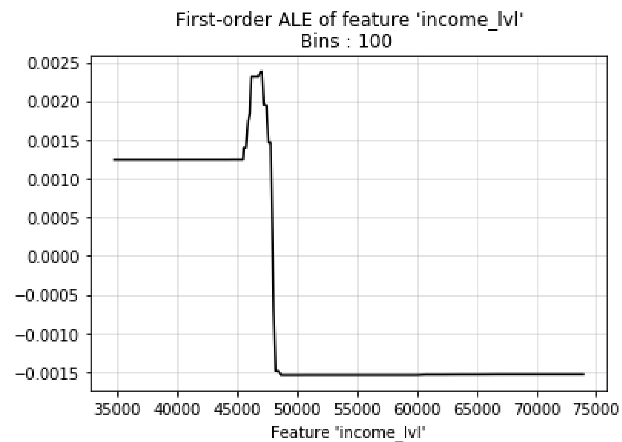


(f) ALE: Unbiased Effect of OLTV on Delinquency probability

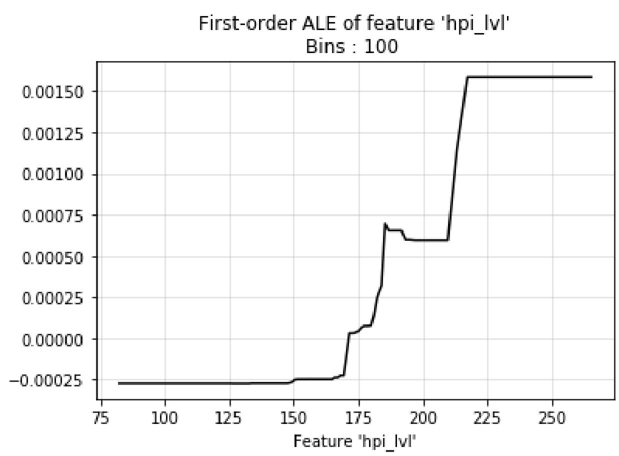
Figure (9) ALE Plots



(a) ALE: Unbiased Effect of State Unemployment Rate on Delinquency probability



(b) ALE: Unbiased Effect of State Per Capita Household Income on Delinquency probability



(c) ALE: Unbiased Effect of State HPI on Delinquency probability

Figure (10) ALE Plots

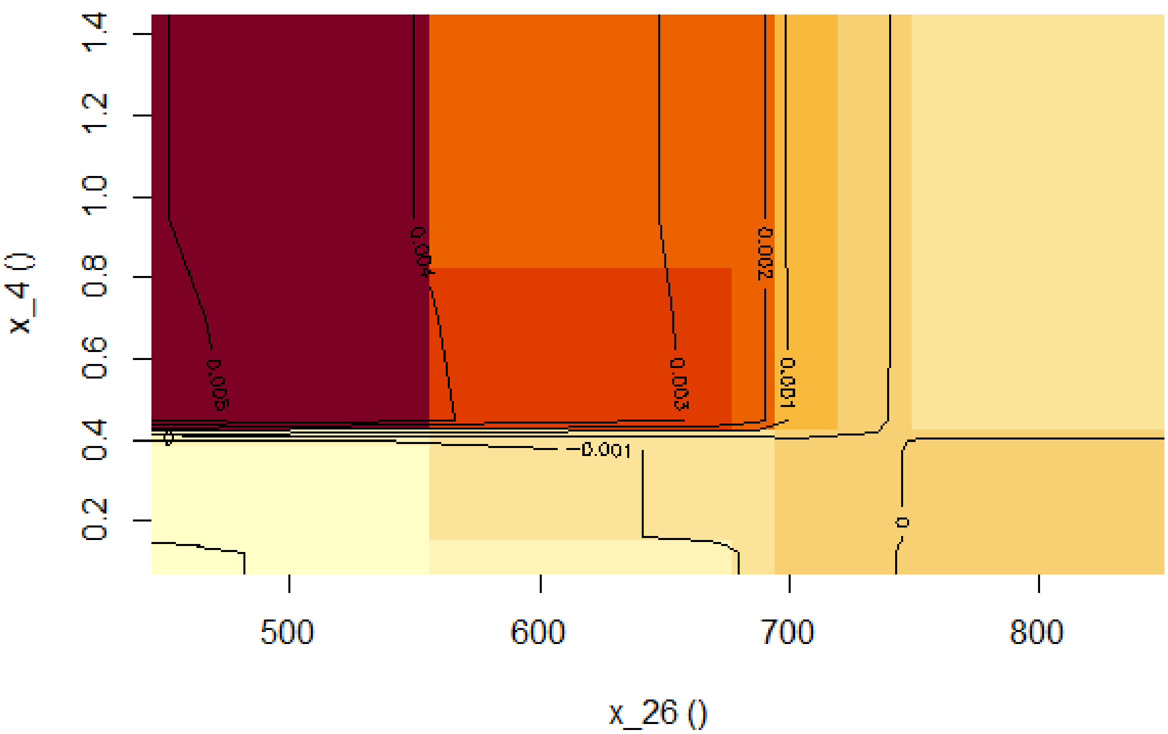


Figure (11) 2D-ALE: The joint effect of credit score (x-axis) and FFR (y-axis) on mortgage delinquency.