Bad Banking Behavior:

Analyzing Bank Mortgages during the 2008 Housing Bubble

Michael Siebel, M.S.

Nima Zahadat, Ph.D.

The George Washington University

Abstract

In 2008, structural issues in the U.S. housing market led to mass foreclosures and ultimately to a global economic crisis. Prior analysis of mortgage data often observed which individuals were at greatest risk of housing foreclosure. This research observes the lending behavior of banks and how reforms in behavior could have lessen the risk of defaults during the housing bubble. Housing foreclosures are modeled from the nine largest bank lenders using a model ensembling approach. These models are then used to explore alternative scenarios to understand what lending practices could have possibly led to fewer foreclosures.

Keywords: Foreclosure, Mortgage Default, Credit Score, Debt-to-Income, Loan-to-Value, Model Ensemble, Vote Classifier, Machine Learning

Introduction

It has been over a decade since a mass sell-off of single-family mortgage loans and housing foreclosures popped the housing bubble and plunged the global economy into a financial crisis. To create greater transparency, Fannie Mae began releasing single-family loan performance data in 2015. This data has usually been analyzed by the data's level of analysis – individuals. In other words, prior analysis often observed which individuals were at greatest risk of housing foreclosure.

I investigated this data by changing the level of analysis to that of the bank. This research observes banking behavior during the housing bubble and how reforms in behavior could have lessen the risk of defaults. My objective is to investigate the behavior of the nine largest banks during the housing bubble and alternative scenarios in which better behavior could have led to fewer foreclosures. In other words, I predict under a hypothetical scenario of how the banks could have shaped their lending practices. Under these scenarios, I generate predictions of mortgage foreclosures to compare how these nine banks would have fared during the housing bubble.

Literature Review

In 2008, the U.S. and then the world fell into a deep economic recession as financial markets collapsed from a mass sell-off of real estate loans. Banks were exchanging high volumes of subprime mortgages from secondary markets without much consideration paid to their risk. Banks would provide single-family mortgage loans, which were bought on a secondary market, grouped with similar mortgage loans, and then sold to other financial institutions as aggregated, mortgage-backed securities (MBS). The secondary market was primary run by Fannie Mae. Many financial institutions formed a highly speculative market, selling and reselling MBSs until the original value of the MBSs became opaque. Once a sizable number of homeowners defaulted on their loans, these financial institutions grasped the uncertainty of many of these MBSs and—not knowing which were originally valued as safe and which were valued as risky— triggered a mass sell-off.

Literature on Theory

Literature regarding the causes of mortgage defaults and housing foreclosures broadly fall into three categories: macro-economic theory, personal responsibility theory, and the predatory market theory. The period preceding the financial crisis was one of low interest rates, unemployment, and house prices, bringing in many new homeowners into the market (Mayer, Pence, & Sherlund, 2008). First-time homeowners are often earlier in their career stage and potential have less stable income and shorter credit histories than those with previously paid mortgages. High entry into the market led to surging housing prices, transitioning the market to expect and rely on constant increasing prices. MBSs further fueled surging housing prices, encouraged to overleverage in order to benefit from future high housing prices. Meanwhile, homeowners assumed constant increasing prices and bought houses under the belief that any future personal financial struggles could be solved with refinancing on higher housing values (FDIC, 2017). Feldstein (2008) argued that because recourse is limited in mortgages (i.e., repossession is limited to the house itself and not smaller items of property), once one is struggling to make mortgage payments, often the only solution is foreclosure. With little middle-ground between making payments and housing foreclosures, a domino-effect occurred with housing prices: foreclosures increased the supply of houses, lowering the value of nearby houses, which reduces equity of homeowners who then have less options when struggling to meet mortgage payments. Once credit markets froze in response to foreclosures, banks sold at fire-sale prices, increasing the supply of houses, and further reducing the value of nearby houses (FDIC, 2017). However, Hardarson and Vuono (2019) noted that their bank default prediction models using macro-economic variables performed worse than models based purely on bank-specific variables, indicating that macro-economics is quite limited in explaining mass foreclosures during the financial crisis.

Personal responsibility theory focuses on the financial mistakes and/or bad behavior of mortgage recipients. There is vast literature that discusses broadly the importance of financial literacy, particularly given the greater personal financial responsibility Americans face in a modern economy and increasingly complicated financial products (see Allgood & Walstad, 2016; Hastings, Madrian, & Skimmyhorn, 2013). Gerardi, Goette, and Meier (2013) show that inability to perform certain, rather-basic mathematical calculations was associated with the

likelihood to face housing foreclosures. While much of this literature does not lay blame at individuals for their difficulties in meeting complex financial responsibilities, during and immediately following the financial crisis, several newspaper opinion-editorials faulted the intentions of those defaulting on mortgages. In the Washington Post, Harney (2009) wrote about individuals choosing to default on their mortgages, quoting Fannie Mae spokesman Brian Faith, "there's a moral dimension to this as homeowners who simply abandon their homes contribute to the destabilization of their neighborhood and community." Luigi Zingales (2010) wrote that individuals were participating in "strategic default" which was "morally reprehensible."

Most literature attempts to explain the role of lenders in explaining what made 2006-2008 different from other periods of increased housing sales. The predatory market theory places blame at banks and mortgage lenders for the customers they targeted and the quality of the loans they passed into MBSs. As far back as 2000, the U.S. Department of Housing and Urban Development documented the rapid growth of subprime mortgages and its role in "widespread predatory practices" (FDIC, 2017). In 2007, Brooks and Simon noted news accounts steering mortgage recipients into subprime adjustable-rate mortgages when they could have qualified for fixed-rate or prime mortgages. During this period, subprime mortgages originated rose from 1.1 million in 2003 to 1.9 million in 2005 (Mayer, Pence, & Sherlund, 2008). Foreclosures for subprime loans was 15.6% in 2008, compared to 3.3% for prime conventional loans (U.S. Census, 2018).

Keys, Mukherjee, Seru, and Vig, (2008) noted that mortgage brokers changed their lending process to an "originate-to-distribute" model, where mortgages where made with the intent to resell them quickly to secondary financial institutions that then securitized them. This process lowered incentives to steer applicants into safer, more manageable mortgages. A Federal Deposit Insurance Corporation (FDIC) report (2017) noted that this divorced participants in the securitization process from risk inherent in risky mortgages. Mayer, Pence, and Sherlund (2008; p. 24) argued that lenders pushed mortgages that "had interest rates that adjusted in potentially confusing ways; did not require full documentation of income and assets; allowed borrowers to postpone paying off mortgage principal; or imposed fees if borrowers prepaid their mortgages within a certain period of time." Corbae and Quintin (2015) claimed that high loan-to-value

(LTV) mortgages (i.e., mortgages with low down payments) increased the sensitivity of default rates to rising home prices and income shocks, because loan recipients possessed lower equity early in the life of the loan. Instead, these recipients acquired debt through repeated refinancing that took equity from homes.

Literature on Modeling

Most literature on modeling mortgage defaults and housing foreclosures focus on low credit scores as a predictor of foreclosure (Chan, Gedal, Been, & Haughwout, 2011; Haughwout, Peach, & Tracy, 2008). Arguably, credit score is the predictor most directly associated with the performance of a mortgage as it is a measure of one's past ability to pay off debt. High debt-to-income (DTI) ratio at the time of loan origination is associated with higher foreclosure rates (Been et al., 2011; Foote, Gerardi, Goette, & Willen, 2009). Corbae and Quintin (2015) found that 40% to 65% of the initial spike in foreclosure rates can be attributed to the higher availability of high LTV mortgages during the housing boom which proceeded the financial crisis – a finding similar to Campbell and Cocco (2012) who investigated LTV and loan-to-income. Chan et al., (2013) researched mortgage defaults in New York City neighborhoods and found that low median income neighborhoods were associated with higher mortgage defaults.

In research for the FDIC, Kiefer and Mayock (2020) studied the challenges related to modeling mortgage performance. They note that model instability is common due to intertemporal changes in macro-economics and loan standards. Because foreclosures have few highly correlated variables and the housing market's relationship to those variables is changing year-after-year, their models were unable to adequately predict out-of-time samples. Deng (2016) was able to predict three mortgage payment outcomes (Paying, Default, and Prepay) with an F1 score of 0.49, while Hamid and Ahmed (2016) achieved similar results (an F1 score of 0.43) on a different dataset of general bank loans (i.e., not necessarily mortgages). Models can perform better if not limited to information available at loan origination such as current actual unpaid principle balance (UPB) and remaining month to maturity as found by Sealand (2018) who achieved F1 scores above 0.85.

Data

Most of the data comes from Fannie Mae's Loan Acquisition and Performance Data. The data comes in two forms: 1) acquisition data and 2) performance data. The acquisition data includes one observation for each loan with each feature representing knowledge Fannie Mae has when acquiring the loan (e.g., balance, primary lender, credit score, etc.). Each quarter contains around a half million loans that represent a population of over 35 million mortgage loans held by Fannie Mae. The performance data includes observations for each month each loan is held and information on the payment of the loan. Each quarter contains around 6.5 million observations. Non-standard or refinanced mortgage loans are not included in this data.

While the acquisition data contains many possible features, the performance data contains the target variable: mortgage loan foreclosures. Foreclosures are defined as mortgage loans with zero balance remaining due to one of the following five reasons: 1) Third Party Sale, 2) Short Sale, 3) Repurchased, 4) Deed-in-Lieu (REO), and 5) Note. Zero balance mortgages with prepaid or matured and reperforming loan sale will serve as a non-foreclosure classification. Each observation represents a loan originating or refinanced between 2006-2008. The financial crisis is often noted as occurring between 2007-2009; most loans that lend to housing foreclosures originated or were refinanced only a year to three years before foreclosure, leading to a peak in loans destined for foreclosure occurring in 2007. Refinanced loans are limited to loans originating in fiscal year 2000 or later. The data is further limited only to loans made by the nine largest banks at the peak of risking lending (2007). Features modeled to predict foreclosures are limited to information known at the time of loan origination except for the date of last reporting (i.e., the window in which a foreclosure could have occurred and be in the dataset).

With an extraction, transformation, and loading (ETL) process, I wrangled and merged external data sources used to model loan foreclosures. Federal Reserve Economic Data (FRED) includes macro-economic data that is related to the housing market. This data is merged on the date variable (mm/yyyy); it includes monthly data and quarterly data, the latter used carryforward hard coding (each quarter represented the beginning of the quarter) to cover each month. Some FRED sets included four Census region subsets; these were merged on the date variable

(mm/yyyy) and the property state variable, the latter was mapped to the four Census regions of Northeast, Midwest, South, and West. Values were converted to quarterly and yearly deltas (e.g., the change in housing vacancies from 2006 Q4 to 2007 Q1 or 2006 Q1 to 2007 Q1).

Federal Deposit Insurance Corporation (FDIC) data includes information on FDIC-backed banks, such as their number of employees, assets, debts, etc. I used regular expressions to map FDIC data to the Bank variable; this included summing various instances of the same bank (from a different branch or functional area). This data is merged on the bank variable and the date variable (mm/yyyy); it includes quarterly data, which used carrybackward hard coding (each quarter represented the end of the quarter) to cover each month. Values were converted to quarterly and yearly deltas (e.g., the change in Bank of America liabilities from 2006 Q4 to 2007 Q1 or 2006 Q1 to 2007 Q1).

The ETL process can be summarized in five steps:

1) Fannie Mae Loan Acquisition and Performance Data [Individual Mortgage Loans],
2) Past Fannie Mae Loan Acquisition Data [Changes in Loan Amounts],
3) U.S. Census Bureau, Small Area Estimates Branch [Median Household Income by County],
4) Federal Reserve Economic Data (FRED) [Macroeconomic Data related to the Housing Market], and
5) Federal Deposit Insurance Corporation (FDIC) Data [Information on FDIC-backed Banks]

Software

In Python 3.6, I used the *sklearn* package for implementing the machine learning components along with standard packages such as *pandas* and *numpy* for common data transformations. This involved splitting the data into training and testing frames, standardizing the variables on their standard deviations, building a model, making predictions, and comparing the F1 score of those predictions. All data visualizations were created using *matplotlib*. Visualizations and code can be found here:

https://siebelm.gitub.io/Bad_Banking_Behavior

Methodology

The main challenge of modelling mortgage foreclosures is that they represent a rare event –
defined as a dichotomous variable in which an occurrence of the event is low relative to cases in
which that event could occur. Consequently, counts in the numerator of the rate are expected to
be small by comparison to the denominator. Such small numerators in a target variable are
known to create higher bias in estimation and, when forming a proportion, higher variance (King
& Zeng, 2001). In laymen terms, classification models are likely to become better at predicting
non-occurrences than occurrences simply because there are more examples of non-occurrences
to learn from. This results in models over-predicting non-occurrences. In very imbalanced
datasets, this can result in zero predictions of an occurrence.

Further, the features in the Fannie Mae do not correlate well with foreclosures. To this end, I
linked many features from other data. Still, it remains is difficult to predict foreclosures without
overfitting the training data. To solve these two problems, I use two primary ensembling
techniques. The first, algorithm-based ensembling, rebalances the data to augment foreclosure
predictions. The second uses vote ensembling to increase generalization of each model's results.

Balancing Predictions

Figure 1 displays a hypothetical dataset in which the target variable occurs at a rate of 20%. In
order to ensure a predictive model remains balanced in its learning from occurrences and non-
occurrences, an analyst can use an artificial intervention to balance the classes. One approach is
to draw small samples with balanced classes. These samples produce relatively weak and simple
models, which are then ensembled into stronger predictions.

Moving from Step 1 to Step 2 displays how this example data could be sampled using an
ensembling technique. The analyst converts the training data into four balanced datasets in
which the target variable occurs at a rate of 50%. These samples can encompass the full training
data and yet each model learns from the same number of occurrences as it does non-occurrences.
Ensembling these simpler models involves averaging their predicted probabilities and then
converting those predicted probabilities into a dichotomous result using a classification threshold
(i.e., the hardlimit). Finally, that prediction model is used to predict on the validation data for

model evaluation. The validation data remains in its raw form (i.e., without balanced classes) to evaluate these predictions as with natural data (i.e., no analyst interventions).

While this approach overcomes the issue of over-predicting non-occurrence, it is prone to over-predicting occurrences. By over-predicting the rare events, the ensembled model is shown to know what situations lead to preventing or causing a mortgage foreclosure, although it is biased towards predicting mortgage foreclosures at higher rates. Fortunately, this bias can be minimized by optimizing the classification threshold.

As shown in Step 3, the testing data remains imbalanced and the predicted probability predictions are equally likely to be above or below 0.5 – the standard classification threshold. Because the analyst intervened and forced balanced classes in the training data, the analyst should compensate by increasing the threshold to an appropriate level. In this example, the analyst could choose a 0.8 threshold, which would provide perfect predictions on the test set. Note, this decision is made on validation data, separate from the testing data which will make the final model evaluations.

Figure 1
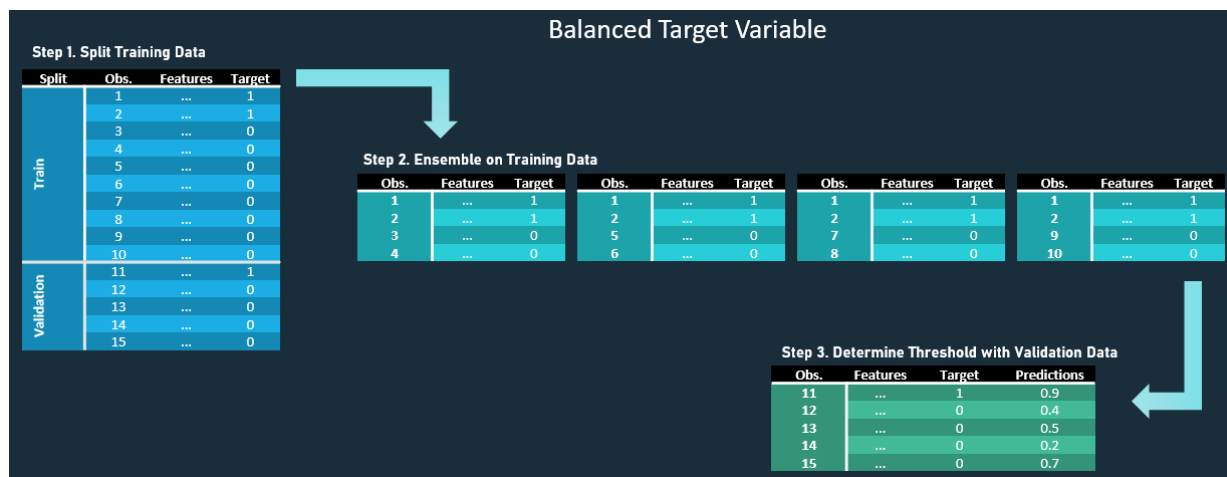
In modeling mortgage foreclosures, the classification threshold that maximizes F1 score is selected as this is the best at balancing recall and precision. Many models were run on different cohorts and using different algorithms; more on this in the next section. Table 1 shows that the

results were quite poor: one of the 3 first models on all nine banks resulted in only 0.24 precision and, despite the stricter threshold, it predicted 19% foreclosures when the testing data contained only 10% foreclosures. (Note, while the full data contained 9.7% foreclosures, the testing data contained 10%.)

Table 1

First Model	
Predicted Foreclosures	19%
Actual Foreclosures	10%
F1 Score before Voting	0.32
Recall before Voting	0.48
Precision before Voting	0.24
Accuracy before Voting	0.80

Voting Architecture

Three models on each of the nine banks plus on an all banks dataset were run (for a total of 30 models).[1] As described above, the three models resampled the data to better balance foreclosures to non-foreclosures. However, while one substitutes an even balance (1:1 between the two outcomes), one substitutes a 3:1 balance, and the other substitutes a 5:1 balance. This is done to ensure greater diversity between models. The two that favor non-foreclosures still oversample foreclosures as a true dataset (such as the validation and testing data) favor non-foreclosures at 10:1.

The 1:1 model is a random forest that uses principal component analysis (PCA) to reduce the features to 10. It tunes two hyperparameters: minimum samples split and the number of estimators. The 3:1 model is an AdaBoost decision tree algorithm called SAMME.R, which uses all 42 variables. It tunes one hyperparameter: the learning rate. The 5:1 model is another random forest that selects a different square root amount of the 42 variables per tree. It also tunes minimum samples split and the number of estimators. Each of the three models receives one vote per bank.

---

[1] See Appendix A for a full list of features used.

Figure 2 illustrates the architecture in which these models were ensembled—simplified to only show three banks. The bottom layer displays the models, and the arrows represent a vote of foreclosure or not being sent to another layer. The all bank dataset is a special case, which receives a middle layer for the first round of voting. The all banks dataset attempts to capture cases of foreclosure that are likely to occur regardless of bank-specifics. As a result, I imposed a strict criterion of unanimous vote, meaning the three models must all agree that a prediction is a foreclosure for it to be marked as a foreclosure. The top layer receives three votes from the three models per bank plus a vote from the model layer's all bank dataset. The top layer requires a majority vote—three out of the four votes must agree that a prediction is a foreclosure for it to be marked as a foreclosure.

This voting architecture is designed to use the algorithm-based ensembling to ensure model attention on what cases a foreclosure, and vote ensembling to keep foreclosures rare.

Figure 2
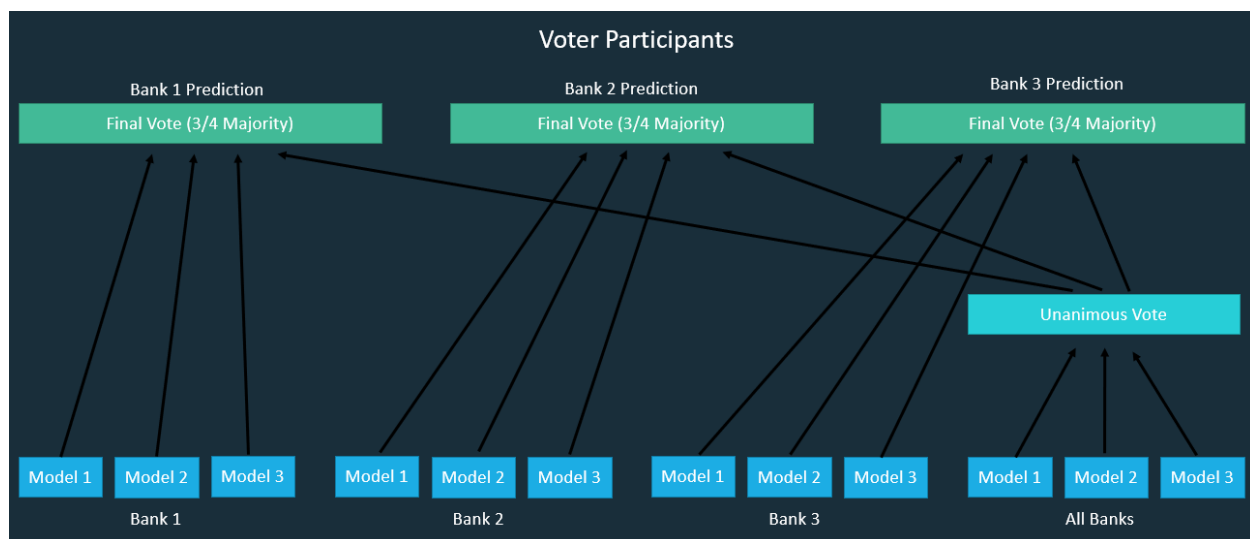
Modeling Performance

After all votes were cast, the final model, which stacks each of the final bank results from the top layer on one another, correctly predicted a foreclosure rate of 10%. The overall accuracy rate appears high at 0.88, but it contains issues which can be understood by the low F1 score (0.38).

The F1 score is a better measure of accuracy as it accounts for the low prevalence of foreclosures.

Certainly, the voting architectures improved the results of the individual models. It was low on bias, but high in variation; each prediction was prone to error (i.e., high variation), but the errors were about as likely to be a false positive (0.37 precision) as a false negative (0.39 recall).

A similar pattern held across banks, with all but one bank containing predicted foreclosures only a percentage point off actual foreclosures (Table 2). SunTrust Mortage performed the worst with two percentage points off actual foreclosures, predicting 8% foreclosures instead of 10. It contained a 0.09 point gap between precision (0.41) and recall (0.32), causing more false negatives (i.e., bias towards non-foreclosures). All other banks contained between 0 – 0.05 point gap between precision and recall. F1 scores ranged from 0.31 (AmTrust Bank) to 0.43 (Bank of America).[2]

Table 2

Final Model	
Predicted Foreclosures	10%
Actual Foreclosures	10%
F1 Score after Voting	0.38
Recall after Voting	0.39
Precision after Voting	0.37
Accuracy after Voting	0.88

Therefore, these predictions appear valuable in the aggregate (e.g., bank predictions), but not as individual predictions (e.g., mortgage recipient predictions). In other words, this model seems reasonably equipped to predict foreclosure rates for bank lenders, but not equipped to predict whether a loanee will eventually have their home foreclosed upon.

---

[2] See Appendix B for model result tables of each bank dataset.

Analysis

Foreclosures

Mortgages made between 2006–2008 foreclosed at a rate of 9.7% among the top nine banks. These foreclosures peaked in 2007 at 10.2% (Figure 3). Note, the research paper and Data Mining scripts show that these nine banks issued loans more likely to be foreclosed upon than smaller banks, increasing this rate from the total U.S. rate.

Of importance, the average mortgage amount not only grew dramatically since the year 2001, but mortgage amounts on loans that were eventually foreclosed upon increased at a higher rate, matching (and briefly exceeding) the mortgage amounts on loans not foreclosed upon. This implies that large banks may have been willing to make riskier loans for inflated housing prices.

Figure 3

Banks

This research investigates the differences between the nine largest banks. Bank of America is the largest as defined by the sheer quantity of loans it made, While PNC is the smallest of the nine (note, this is not factoring in the amount of those loans).

Loans destined to be foreclosed upon begin to notably increase after 2004. As these loans
increase in foreclosures, the difference between the banks increase.

Table 3 shows that Flagstar Bank had the highest foreclosures (11.7%) between 2006 and 2008
with Bank of America immediately following it (11.6%). Wells Fargo Bank and JPMorgan
Chase had the lowest number of foreclosures (7.6%) with Wells Fargo Bank containing the
lowest before rounding to the first decimal place. CitiMortgage immediately followed (7.7%).

Table 3

	Worst Actors		Best Actors	
	Bank	Foreclosed (%)	Bank	Foreclosed (%)
2006	Flagstar Bank	11.6	Wells Fargo Bank	7
2007	Bank of America	12	JPMorgan Chase	7.7
2008	SunTrust Mortgage	13.3	CitiMortgage	7.6
Overall	Flagstar Bank	11.7	Wells Fargo Bank	7.6

Analysis was conducted on the full dataset, combining the training, validation, and testing data.
Below contains information on five features: credit score, debt-to-income ratio, loan-to-value
ratio, median household income at the 3-digit zip code-level, and the dollar amount change in
mortgage loans made 1 year ago and 5 years ago. The latter feature was created by taking total
loan amount during a fiscal year quarter for each bank within a 3-digit zip code.

Using data mining techniques, each feature was examined at each bank. Then, the feature was
replaced by an improved and a weakened assumption based on the inter-quartile range (25-75
percentiles) of the feature across all banks and predicted probabilities were generated to see what
the expected foreclosure rate would be if each banks' behavior were different. For example, a
high credit score is associated with fewer foreclosures. Among all banks, the average credit
score was 719 (on a scale of 300 to 850). I modified the credit score at each bank to the 75th
percentile—an improved assumption of a 770 credit score—and to the 25th percentile—a
weakened assumption of 675 credit score. I left all other feature values unchanged. I ran these
values through the saved model detailed in the section above and analyzed the change in

foreclosure rates. One can interpret the findings as: "If GMAC Mortgage only lent to those with a credit score of 770, with all other considerations staying the same, its foreclosure rate is predicted to fall from 9.7% to 1%."

Credit Score

Credit score is perhaps the feature most directly associated with the performance of a mortgage loan as it is a measure of one's past ability to pay off debt. It contained large difference between loans foreclosed (with an average credit score of 668) and loans not destined to be foreclosed upon (with an average credit score of 722). However, there is little difference between banks, ranging from the best actor PNC Bank (728) to worst actor GMAC Mortgage (710). Notably, Wells Fargo Bank, boasting the lowest foreclosure rate, only possesses the third highest credit score, and Flagstar Bank, holding the highest foreclosure rate, only possesses the third lowest credit score.[3]

This seems to indicate that a change in targeting individuals with certain credit scores should highly influence foreclosure rates, but that banks appeared unwilling to change their credit score standards if other banks did not follow suit.

As shown in Table 5, foreclosure rates were predicted to improve from 9.7% to 1.6% if all mortgage holders had a credit score of 770 and weaken to 14.1% if all mortgage holder had a credit score of 675. Bank of America had the largest change, improving from 11.6% to 1.9%. Meanwhile, the bank with the lowest credit score, PNC Bank, is predicted to have the least improvement from 8.8% to 3.1%. However, Bank of America is expected to have the largest increase in foreclosures, if credit scores were weakened to 675, from 11.6%–17.3%, compared to GMAC Mortgage's 9.7%–12.2%.[4]

---

[3] See Appendix C for descriptive statistics for each feature.
[4] See Appendix D for best and worst bank predictions for each feature.

Table 5

	Foreclosures (2006-2008)	Credit Score Improved	Weakened
Bank of America	11.6	1.9	17.3
Wells Fargo Bank	7.6	1.2	11.9
CitiMortgage	7.7	1	11.3
JPMorgan Chase	7.6	1.6	10.9
GMAC Mortgage	9.7	1	12.2
SunTrust Mortgage	10.3	1.6	13
AmTrust Bank	9.3	2.8	12.3
PNC Bank	8.8	3.1	13.8
Flagstar Bank	11.7	2.1	15.8
All Banks	9.7	1.6	14.1

Credit score displayed a substantial relationship with foreclosures, but average credit scores differ little between banks. Bank of America had the most impact by changes in credit score. Most notably, the highest predicted foreclosure rate among all features analyzed in this section was when Bank of America's credit score assumptions weakened.

Debt-to-Income (DTI) Ratio

The average DTI ratio for loans destined for foreclosure is 41.4% and 37.8% for those not destined for foreclosure. Similar to credit score, banks did not range greatly with Flagstar Bank the highest (40.9%) and CitiMortgage the lowest (35.3%). The best actor, CitiMortgage, had roughly as similar proportions of loan recipients with DTI ratios at 60% as it did at 20%. A general rule of thumb in personal finance is to keep a DTI ratio of 35% or less, implying that loan recipients between 2006-2008 were exceeding conventional standards for DTI.

As shown in Table 6, foreclosure rates were predicted to improve from 9.7% to 6.5% if all mortgage holders had a DTI ratio of and weaken to 29% and weaken to 11.8% if all mortgage holders had a DTI ratio of 47%. Most banks were predicted to reduce foreclosures to 5%–5.5%, with Wells Fargo Bank improving to 3.5% and Bank of America improving to only 8.9% as outliers. Under improved assumptions, Flagstar Banks was predicted to have the largest decrease in foreclosures from 11.7% to 5.3%. Under weakened assumptions, PNC Bank was predicted to have the largest increase in foreclosures from 8.8% to 12.5%. Again, Bank of

America was predicted to have the highest foreclosure rate of 14.8% under a DTI ratio of 47%.
However, one bank, SunTrust Mortage, decreased its foreclosure rate with weakened
assumptions, showing that not all potential recipients with DTI's were risky.

Table 6

	Foreclosures (2006-2008)	Debt-to-Income	
		Improved	Weakened
Bank of America	11.6	8.9	14.8
Wells Fargo Bank	7.6	3.5	9.4
CitiMortgage	7.7	5.4	9.7
JPMorgan Chase	7.6	5.5	8.9
GMAC Mortgage	9.7	5	11
SunTrust Mortgage	10.3	5.4	9.3
AmTrust Bank	9.3	5.5	9.9
PNC Bank	8.8	5.2	12.5
Flagstar Bank	11.7	5.3	11.9
All Banks	9.7	6.5	11.8

All banks were approving loans to recipients with high DTI ratios. In general, foreclosures
decreased substantially with improved DTI assumptions, but did not increase as substantially
when assumptions were weakened.

Loan-to-Value (LTV) Ratio

LTV data displays three spikes around 95, 90, and 80. There is a large tail in which presumably
wealthy buyers were putting down 40% (an LTV of 60) or more for their down payment. Loans
destined to foreclose averaged a down payment of 20.5% (an LTV of 79.5), and loans not
eventually foreclosed upon averaged a down payment of 28.7% (an LTV of 71.3). Loans with
an LTV's of 90 and 95 displayed particularly high levels of foreclosure. Bank LTV's ranged
from only 69.8 (GMAC Mortgage) to 74.6 (AmTrust Bank). AmTrust Bank contained few
outliers: people who made large down payments. Interestingly, Wells Fargo Bank, the bank
with the lowest foreclosure rate, had the second highest LTV.

As shown in Table 7, foreclosure rates were predicted to improve from 9.7% to 5.3% if all mortgage holders had a LTV of 63 and weaken to 13.2% if all mortgage holder had a LTV of 84. Under improved assumptions, SunTrust Mortage changed the most from 10.3% to 3.4%. Under weakened assumptions, Bank of America changed the most from 11.6% to 16.4% – again, the worst foreclosure rate. SunTrust Bank was the only one to improve (slightly) as LTV increased.

Table 7

	Foreclosures (2006-2008)	Loan-to-Value (LTV)	
		Improved	Weakened
Bank of America	11.6	7.3	16.4
Wells Fargo Bank	7.6	3.2	9.9
CitiMortgage	7.7	3.8	10.8
JPMorgan Chase	7.6	3.4	9.3
GMAC Mortgage	9.7	5.4	14.4
SunTrust Mortgage	10.3	3.4	9.8
AmTrust Bank	9.3	4.3	9.8
PNC Bank	8.8	5	12.6
Flagstar Bank	11.7	5.9	13.1
All Banks	9.7	5.3	13.2

LTV was the second most impactful feature analyzed in this document after credit score. Bank of America increased in foreclosures more than any other bank to an extremely high rate under weakened assumptions.

Median Household Income
Median household income at the 3-digit zip code does not necessarily reflect the income of the loan recipient but does indicate whether banks are focusing their loan efforts in poorer or wealthier areas. The difference between the average income of loans destined for foreclosure ($48,204) was very close to those that did not foreclose ($48,797). Between banks, income only ranged from $47,264 (SunTrust Mortage) to $49,491 (CitiMortgage).

Due to low data variation, median household income was adjusted only slightly from an improved assumption of $53,615 to a weakened assumption of $43,298 (Table 8). Perhaps,

because a $10,000 income difference is too small, banks overall did not reduce predicted foreclosures under improved assumptions or increase foreclosure under weakened assumptions.

Table 8

	Foreclosures (2006-2008)	Median Household Income Improved	Weakened
Bank of America	11.6	14.4	13.3
Wells Fargo Bank	7.6	8.4	8.7
CitiMortgage	7.7	8.9	7.7
JPMorgan Chase	7.6	8.2	7.4
GMAC Mortgage	9.7	10.3	9.2
SunTrust Mortgage	10.3	8.9	8.1
AmTrust Bank	9.3	8.7	10.2
PNC Bank	8.8	11.3	9.9
Flagstar Bank	11.7	11.5	10.4
All Banks	9.7	11.2	10.4

Foreclosure predictions seem random: only slightly change and often in the wrong direction. This could be related to two issues: 1) there is little correlation between median household income at the 3-digit zip code, and/or 2) there is little data variation in this feature. Given the latter potential explanation, I adjusted the assumptions to the maximum (an income of $101,651) and the minimum ($28,832).

After adjusting assumptions to the richest region ($101,651) and the poorest region ($28,832), the predictions begin looking more like one would expect (Table 9). Emphasizing the highest region is predicted to decrease foreclosures from 9.7% to 4.9%, although emphasizing the poorest region is not predicted to change foreclosures greatly (from 9.7% to 9.9%).

Table 9

	Foreclosures (2006-2008)	Median Household Income	
		Improved	Weakened
Bank of America	11.6	6.5	12.9
Wells Fargo Bank	7.6	4.7	7.4
CitiMortgage	7.7	3.5	6.7
JPMorgan Chase	7.6	3.4	7.3
GMAC Mortgage	9.7	3.2	10.5
SunTrust Mortgage	10.3	4.3	8.6
AmTrust Bank	9.3	3.2	4.4
PNC Bank	8.8	3.3	10.2
Flagstar Bank	11.7	5.1	10.7
All Banks	9.7	3.6	7.7

Loan Change

Banks increasing the total value of their loans over 1 year did not correspond with lower foreclosures. However, there appears to be a minor effect among banks increasing their loans over 5 years; loans destined for foreclosure on average increased $63,530 over 5 years compared to a $61,1932 increase in non-foreclosed loans. There are dramatic bank differences in loan changes with AmTrust increasing only $34,664 over 5 years and Well Fargo increasing $96,423.

Under improved assumptions, 1 year loan changes increased $2,292 and 5 year loan changes increased $37,980, while overall foreclosure rates were predicted to improve from 9.7% to 7.9% (Table 10). Under weakened assumptions, 1 year loan changes increased $27,661 and 5 year loan changes increased $86,358, while overall foreclosure rates were predicted to decline from 9.7% to 12.9%. Wells Fargo Bank decreased in foreclosures the most from 7.6% to 3.6%, while Bank of America increased in foreclosures the most from 11.6% to 17%.

Table 10

	Foreclosures (2006-2008)	Loan Change Improved	Loan Change Weakened
Bank of America	11.6	11.1	17
Wells Fargo Bank	7.6	3.6	6.4
CitiMortgage	7.7	5.4	10.7
JPMorgan Chase	7.6	4.2	11
GMAC Mortgage	9.7	7	11.4
SunTrust Mortgage	10.3	6.7	10.1
AmTrust Bank	9.3	9.9	13.6
PNC Bank	8.8	8.7	13.6
Flagstar Bank	11.7	13.6	13.8
All Banks	9.7	7.9	12.9

Changes in loans vary across banks more than any other feature examined in this document. Likewise, predicted changes in foreclosure varied greatly between banks. Wells Fargo Bank substantially decreased foreclosures under improved assumptions but also slightly decreased foreclosures under weakened assumptions. Overall, changes in loans over time impacted foreclosures more than median household income at 3-digit zip codes but less than the other features examined in this document.

Conclusion

After 2005, mortgage loans issued augmented in likelihood to foreclose as well as differences in bank foreclosure rates. Wells Fargo Bank and JPMorgan Chase held the lowest foreclosure rates between 2006-2008, while Flagstar Bank and Bank of America held the highest foreclosure rates.

Of the five features focused on in this analysis, credit score had the highest impact on foreclosures but littlest discrepancy between banks. With improved assumptions, a high credit score of 770 is predicted to reduce foreclosures to between 1%–3.1%. With weakened assumptions, a low credit score of 675 is predicted to increase foreclosures to 10.9%–17.3%. Of the five features focused on in this analysis, loan change had the lowest impact on foreclosures but highest discrepancy between banks. With improved assumptions, a low 1 year ($2,292) and 5 year ($37,980) increase in mortgage loans made is predicted to reduce foreclosures to between

3.6%–13.6%. With weakened assumptions, a high 1 year ($27,661) and 5 year ($86,358) increase in mortgage loans is predicted to increase foreclosures to 6.4%–17%.

Bank of America always displayed the highest foreclosure rate when assumptions were weakened, even though it possesses a slightly lower rate to Flagstar Bank without any assumptions made. Similarly, Wells Fargo usually (but not always) had the lowest foreclosure rate when assumptions were improved, even though it possesses the same rate as JPMorgan Chase without any assumptions made.

Biography

Michael Siebel joined Fors Marsh Group in 2015 as a researcher for the Department of Defense's (DoD) Office of People Analytics (OPA) Program. His primary responsibilities include designing, managing, and conducting data analysis on military personnel. This research requires deployment of techniques for machine learning, natural language processing, HTML-based visualization, and data warehousing on SQL Servers. He is currently completing a M.S. in Data Science from George Washington University. He holds a M.A. in Political Science (focusing on survey methodology) from the University of Missouri-St. Louis and a graduate certificate in Survey Design and Data Analysis from George Washington University.

Dr. Nima Zahadat is a professor of data science, information systems security, and digital forensics. His research focus is on studying the Internet of Things, data mining, information visualization, mobile security, security policy management, and memory forensics. He has been teaching since 2001 and has developed and taught over 100 topics. Dr. Zahadat has also been consultant with the federal government agencies, the US Air Force, Navy, Marines, and the Coast Guard. He enjoys teaching, biking, reading, and writing.

Appendix A
Full List of Modeling Features

Features Modeled	
Reported Period	First Time Home Buyer_N
Original Interest Rate	First Time Home Buyer_Y
Original Mortgage Amount	Property Type_CO
Original Loan Term	Property Type_CP
Original Date	Property Type_MH
Loan-to-Value (LTV)	Property Type_PU
Single Borrower	Property Type_SF
Debt-to-Income	Occupancy Type_I
Loan Purpose	Occupancy Type_P
Number of Units	Occupancy Type_S
Mortgage Insurance %	Relocation Mortgage Indicator_N
Credit Score	Relocation Mortgage Indicator_Y
Loan Change (1 Yr)	File Quarter_Q1
Loan Change (5 Yr)	File Quarter_Q2
Median Household Income	File Quarter_Q3
Number of Employees	File Quarter_Q4
Loan Liabilities (5 Yr)	Macroeconomy FRED PCA 1
Loan Liabilities (1 Yr)	Macroeconomy FRED PCA 2
Origination Channel_B	Macroeconomy FRED PCA 3
Origination Channel_C	Macroeconomy FRED PCA 4
Origination Channel_R	Macroeconomy FRED PCA 5
Bank_AmTrust Bank	Bank_JPMorgan Chase
Bank_Bank of America	Bank_PNC Bank
Bank_CitiMortgage	Bank_SunTrust Mortgage
Bank_Flagstar Bank	Bank_Wells Fargo Bank
Bank_GMAC Mortgage	

Appendix B
Final Bank Model Results

Bank of America Model	
Predicted Foreclosures	13%
Actual Foreclosures	12%
F1 Score after Voting	0.43
Recall after Voting	0.45
Precision after Voting	0.40

Wells Fargo Model	
Predicted Foreclosures	8%
Actual Foreclosures	8%
F1 Score after Voting	0.32
Recall after Voting	0.32
Precision after Voting	0.31

CitiMortgage Model	
Predicted Foreclosures	8%
Actual Foreclosures	8%
F1 Score after Voting	0.33
Recall after Voting	0.34
Precision after Voting	0.33

JPMorgan Chase Model	
Predicted Foreclosures	8%
Actual Foreclosures	8%
F1 Score after Voting	0.36
Recall after Voting	0.36
Precision after Voting	0.36

GMAC Mortgage Model	
Predicted Foreclosures	9%
Actual Foreclosures	10%
F1 Score after Voting	0.33
Recall after Voting	0.32
Precision after Voting	0.34

SunTrust Mortgage Model	
Predicted Foreclosures	8%
Actual Foreclosures	10%
F1 Score after Voting	0.36
Recall after Voting	0.32
Precision after Voting	0.41

AmTrust Bank Model	
Predicted Foreclosures	9%
Actual Foreclosures	10%
F1 Score after Voting	0.31
Recall after Voting	0.3
Precision after Voting	0.33

PNC Bank Model	
Predicted Foreclosures	10%
Actual Foreclosures	9%
F1 Score after Voting	0.32
Recall after Voting	0.34
Precision after Voting	0.31

Flagstar Bank Model	
Predicted Foreclosures	11%
Actual Foreclosures	12%
F1 Score after Voting	0.37
Recall after Voting	0.35
Precision after Voting	0.39

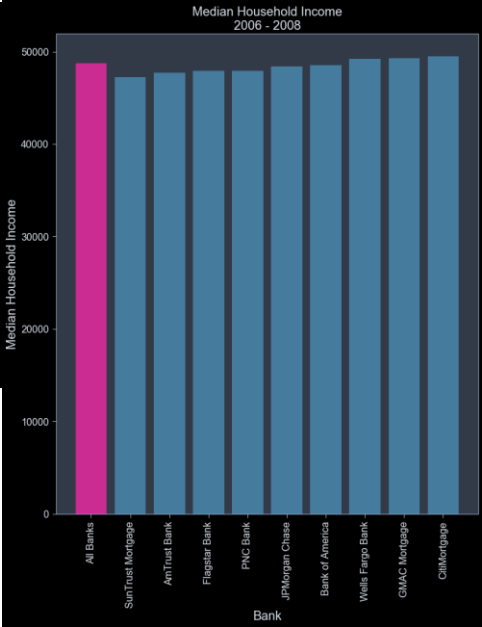**Appendix C**
Feature Descriptive Statistics
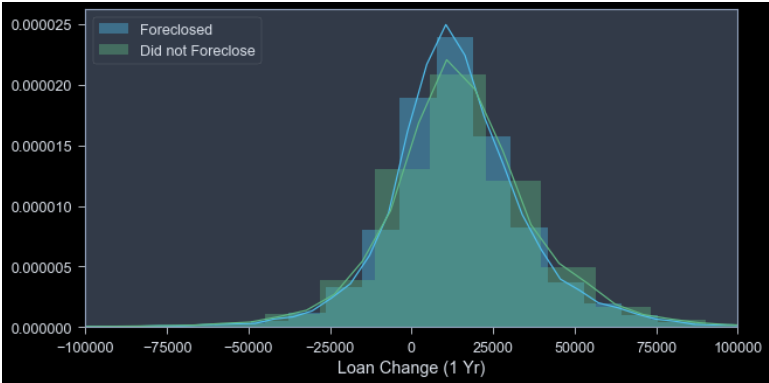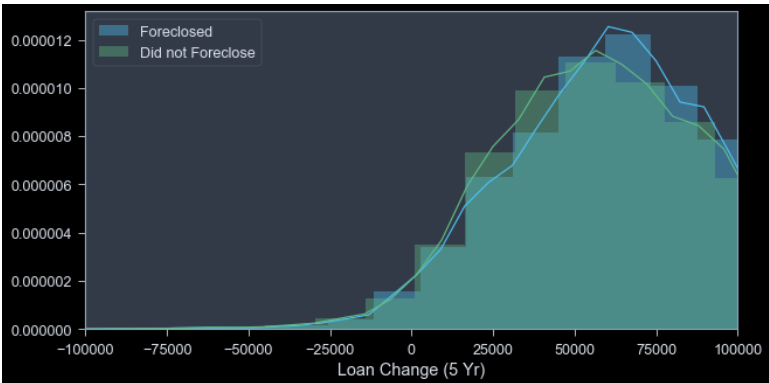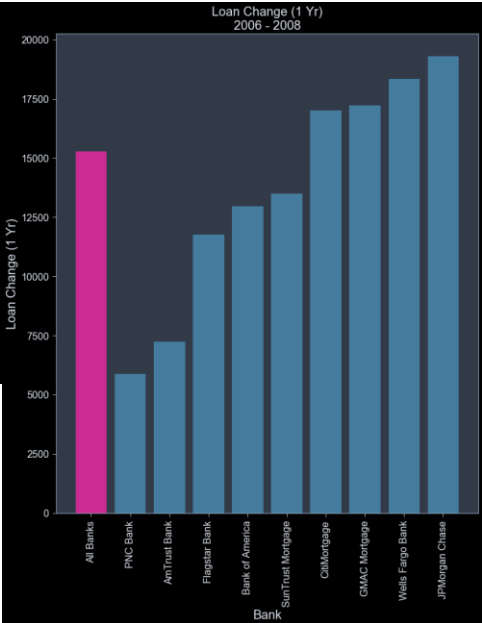

Credit Score
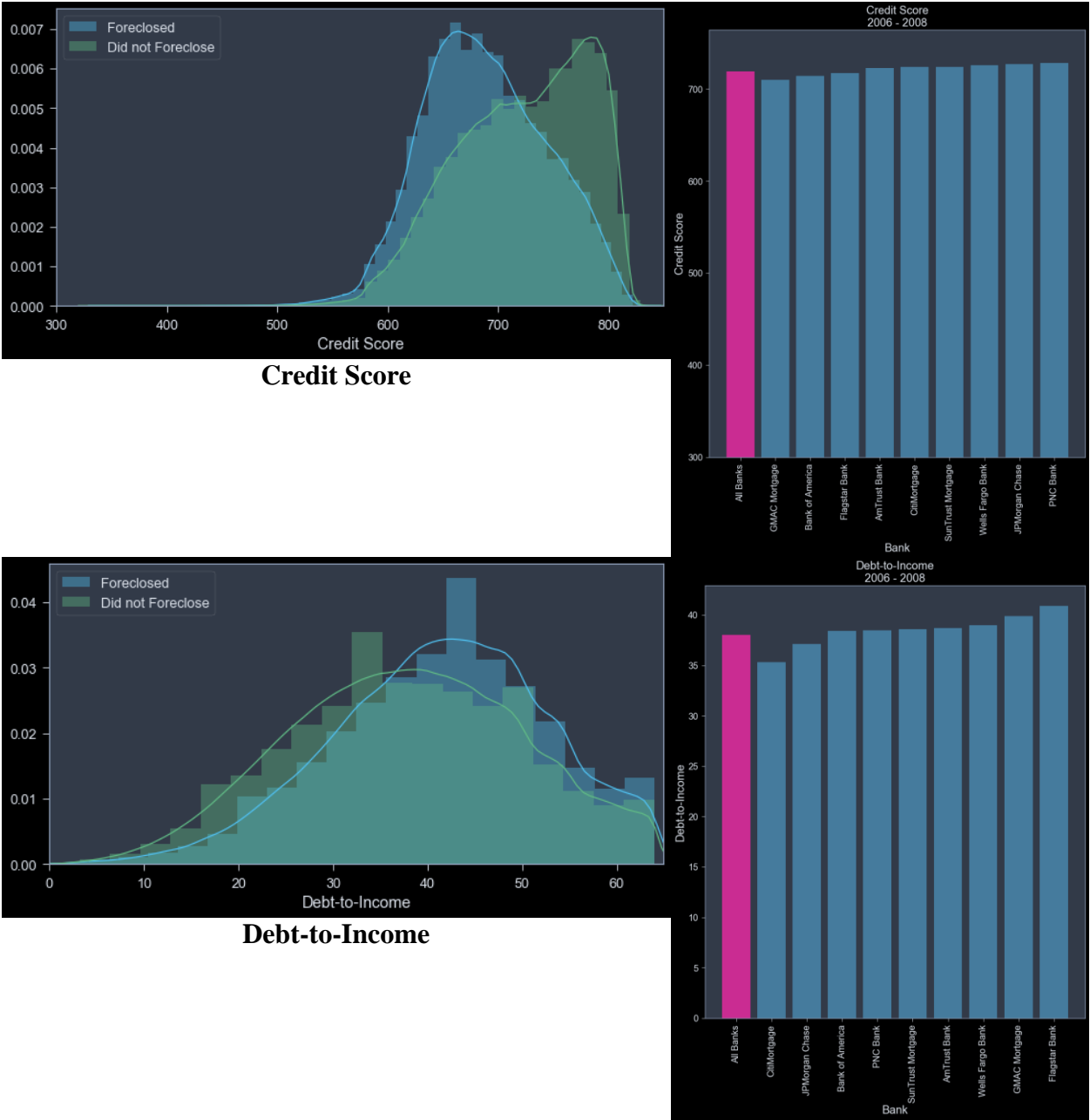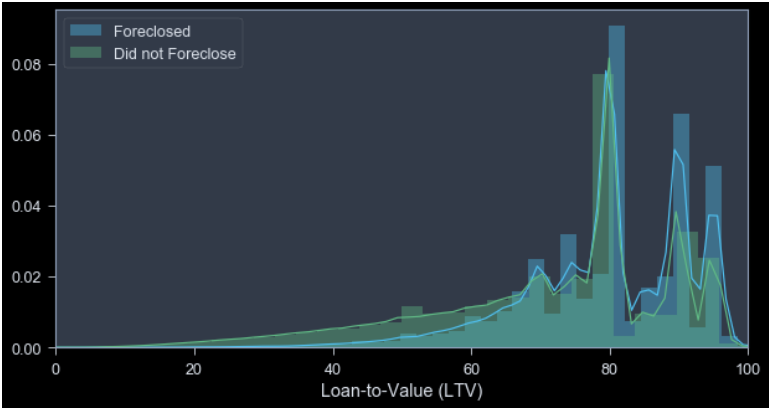
Debt-to-Income
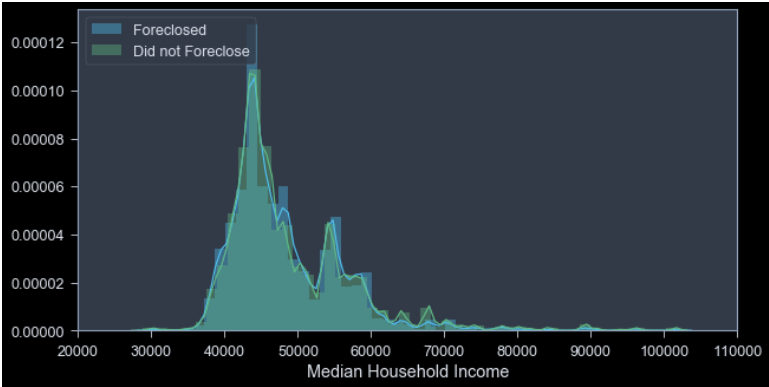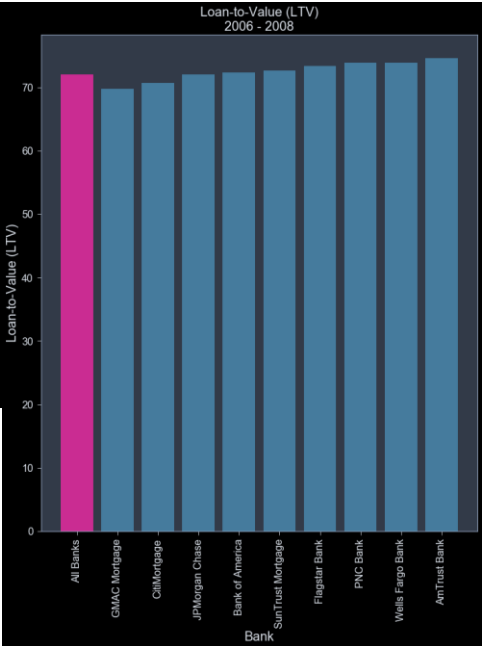
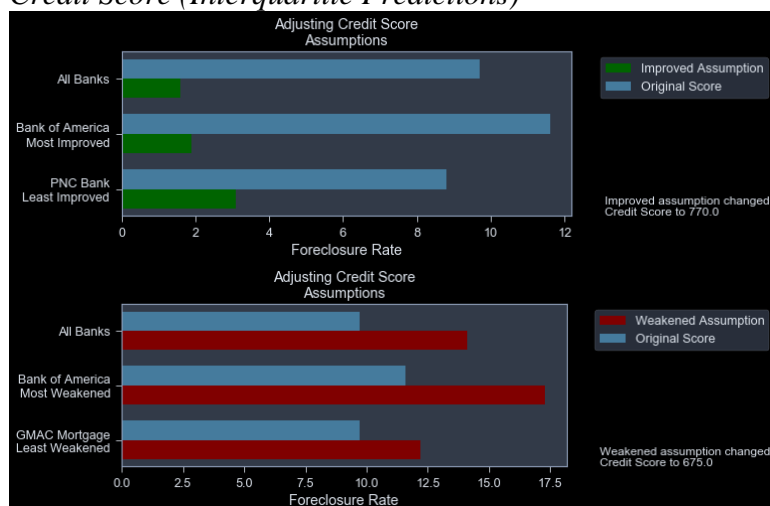**Loan-to-Value**

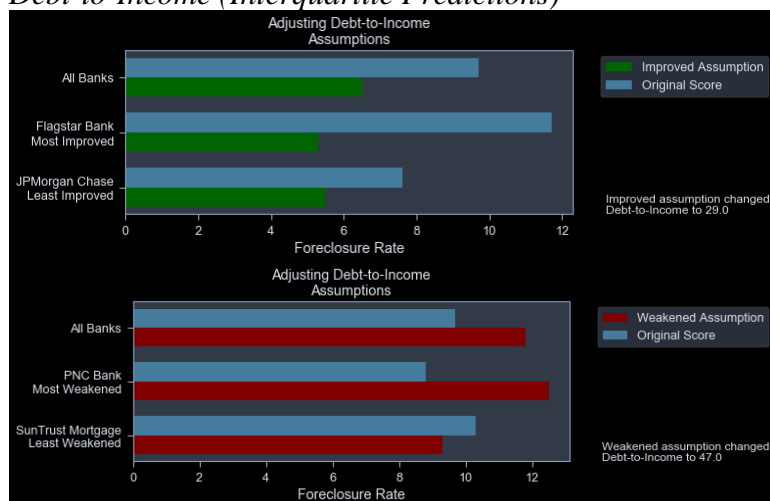**Median Household Income**

Loan Change (1 Yr)
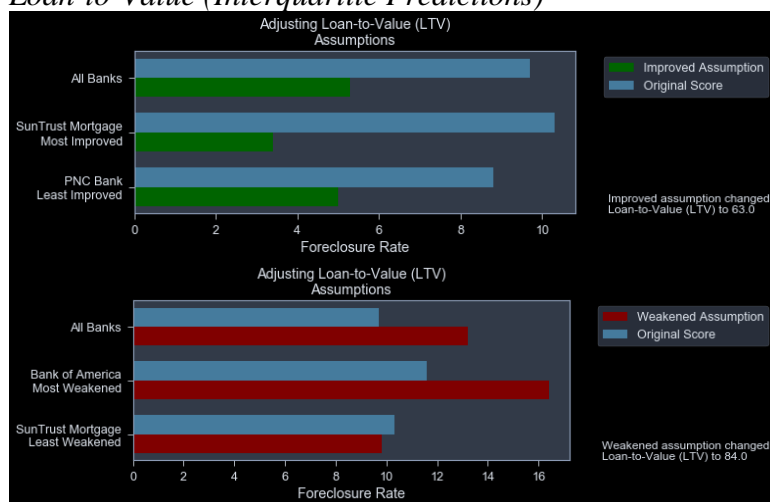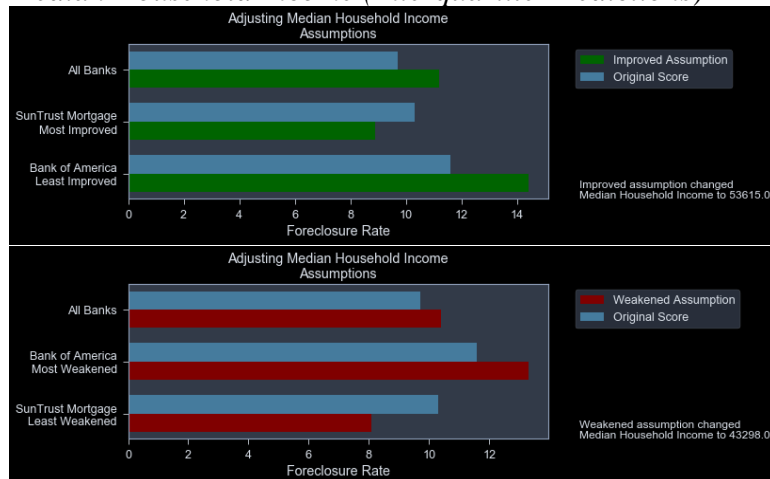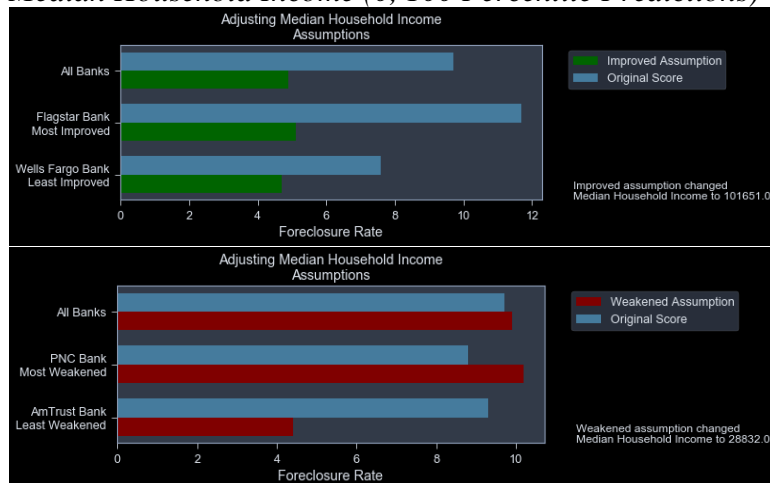

Loan Change (5 Yr)
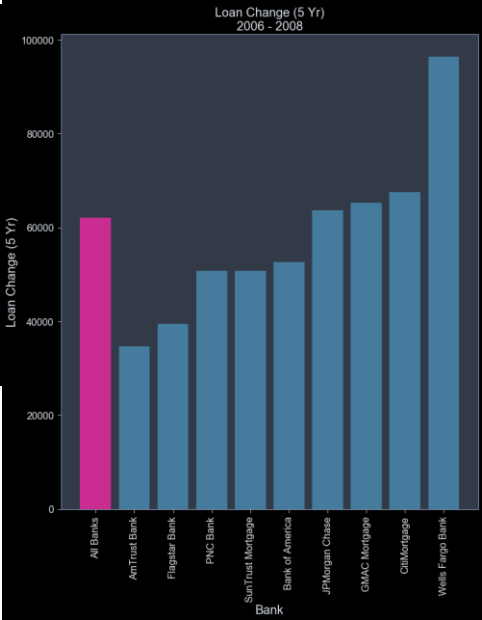
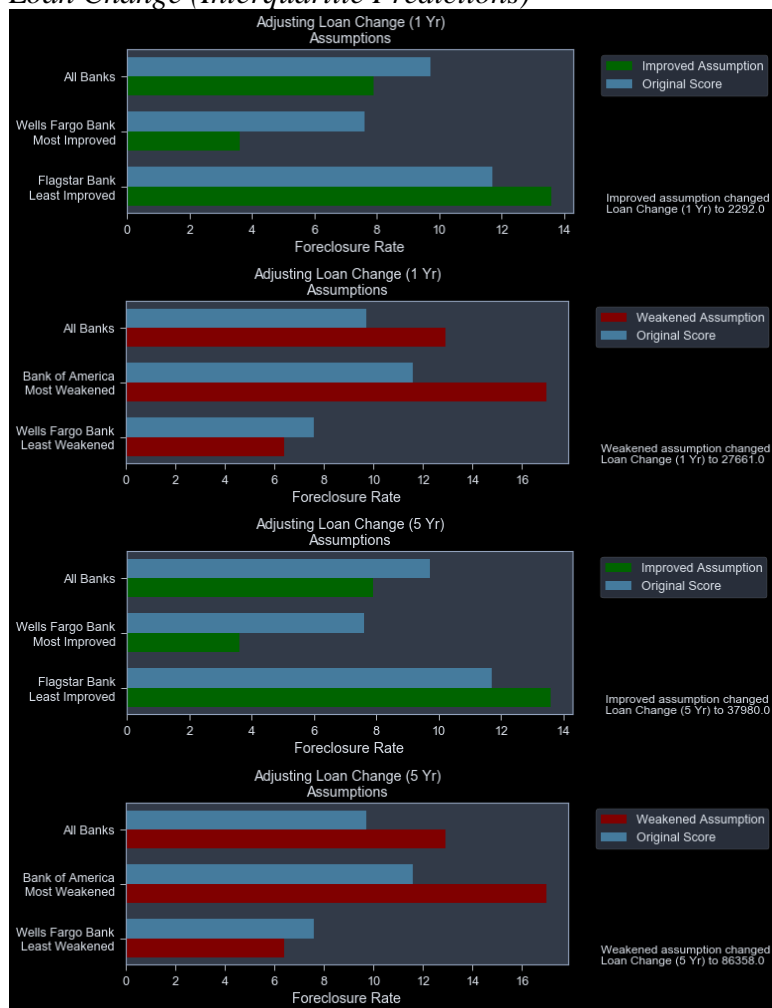## Appendix D
Best and Worst Bank Predictions

Credit Score (Interquartile Predictions)

Debt-to-Income (Interquartile Predictions)

Loan-to-Value (Interquartile Predictions)

Median Household Income (Interquartile Predictions)

Median Household Income (0, 100 Percentile Predictions)

Loan Change (Interquartile Predictions)

Works Cited

Allgood, S., & Walstad, W. B. (2016). The effects of perceived and actual financial literacy on financial behaviors. *Economic inquiry*, *54*(1), 675-697.

Been, V., Chan, S., Ellen, I. G., & Madar, J. R. (2011). Decoding the foreclosure crisis: Causes, responses, and consequences. *Journal of Policy Analysis and Management*, 388-396.

Brooks, R., and Simon, R. (2007). "Subprime Debacle Traps Even Very Credit Worthy." The Wall Street Journal, December 3, p. A1. http://online.wsj.com/article/SB119662974358911035.html.

Campbell, J. Y., & Cocco, J. F. (2012). A Model of Mortgage Default. *unpublished paper, Harvard University and London Business School*.

Chan, S., Gedal, M., Been, V., & Haughwout, A. (2013). The role of neighborhood characteristics in mortgage default risk: Evidence from New York City. *Journal of Housing Economics*, *22*(2), 100-118.

Corbae, D., & Quintin, E. (2015). Leverage and the foreclosure crisis. *Journal of Political Economy 123*: 65. 10.1086/677349.

Deng, G. (2016). "Analyzing the Risk of Mortgage Default." University of California – Berkeley. https://www.stat.berkeley.edu/~aldous/Research/Ugrad/Grace_Deng_thesis.pdf

Federal Deposit Insurance Corporation (FDIC). (2017). *Crisis and Response: An FDIC History, 2008–2013*. Washington, DC: FDIC.

Feldstein, M. (2008, November 18) "How to Help People Whose Home Values Are Underwater," Wall Street Journal editorial.

Foote, C. L., Gerardi, K., Goette, L., & Willen, P. (2009). Reducing foreclosures. *FRB of Boston Public Policy Discussion Paper*, (09-2).

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences*, *110*(28), 11267-11271.

Hamid, A. J. and Ahmed, T. M. (2016). "Developing Prediction Model Of Loan Risk In Banks Using Data Mining." *Machine Learning and Applications: An International Journal (MLAIJ)*3.1: 1-9.

Harney, K. (2009, November 28). "The Nation's Housing: Walking Away from a Mortgage," Washington Post editorial

Harðarson, B. and Vuono, M. (2019). Predicting Bank Insolvency with Random Forest
    Classification. (Master's thesis, Lund University School of Economics and Management).

Haughwout, A., Peach, R., & Tracy, J. (2008). Juvenile delinquent mortgages: Bad credit or bad
    economy? *Journal of Urban Economics*, *64*(2), 246-257.

Keys, B. J., Mukherjee, T., Seru, A., & Vig, V. (2010). Did securitization lead to lax screening?
    Evidence from subprime loans. *The Quarterly journal of economics*, *125*(1), 307-362.

Kiefer, H., & Mayock, T. (2020). Why Do Models that Predict Failure Fail? *Available at SSRN
    3616889*.

Mayer, C., Pence, K., & Sherlund, S. (2008). The rise in mortgage defaults: Facts and
    myths. *Journal of Economic Perspectives*.

Sealand, J. C. (2018). Short-term Prediction of Mortgage Default using Ensembled Machine
    Learning Models (Doctoral dissertation, Slippery Rock University).

U.S. Census. (2018, October 2018). US Census Library Publications. Retrieved from US Census:
    ftp.census.gov/library/publications/2010/compendia/statab/130ed/tables/

Zingales, L. (2010). "The Menace of Strategic Default," City Journal, p. 47-51.