The Contagion Effect and Mortgage Delinquencies

David English

George Washington University

Data Science Capstone

April 23rd, 2022

https://github.com/denglish83/Data-Science-Final

Contents

[Glossary of Terms 3](#_Toc101724339)

[Introduction 4](#_Toc101724340)

[Introduction/Background 4](#_Toc101724341)

[Problem Statement 4](#_Toc101724342)

[Literature Review 5](#_Toc101724343)

[Methodology 8](#_Toc101724344)

[Dataset Description 8](#_Toc101724345)

[Data Collection 8](#_Toc101724346)

[Data Preprocessing and/or Feature Engineering 8](#_Toc101724347)

[Data Modeling & Visualizations 11](#_Toc101724348)

[Model 1: Logistic Regression 11](#_Toc101724349)

[Model 2: Two Stage Time Series Logistic Regression 12](#_Toc101724350)

[Model 3: XGBoosted Regression 14](#_Toc101724351)

[Results & Analysis 15](#_Toc101724352)

[Conclusion 17](#_Toc101724353)

[Conclusion 17](#_Toc101724354)

[Project Limitation 17](#_Toc101724355)

[Future Research 18](#_Toc101724356)

[References 19](#_Toc101724357)

[Appendix 20](#_Toc101724358)

[Appendix 1: Full Variable List 20](#_Toc101724359)

[Appendix 2: MSA Level Summary Statistics 25](#_Toc101724360)

[Appendix 3: Fico/LTV Specification 27](#_Toc101724361)

[Appendix 4: Full 60 in 24 Regression Results 29](#_Toc101724362)

# Glossary of Terms

**Mortgage Delinquency**: A mortgage is delinquent if one more payments have been missed and not made up

**Mortgage Default**: A mortgage is in default if the borrower has missed sufficient payments that the lender no longer anticipates that loan will be repaid (usually 4 payments)

**Foreclosure**: A foreclosure is when the lender repossesses the property from the borrower after a loan was defaulted upon

**Conforming (Prime) Loan**: Any loan that meets the criteria to deliver to Fannie Mae or Freddie Mac (LTV<= 97%, DTI <= 45%, FICO > 680, UPB < $646,200)

**LTV**: Loan To Value Ratio, the ratio of the amount of the loan in the first lien position to the value of the house

**CLTV**: Combined Loan To Value Ratio, the ratio of all debts secured by the property (first lien, simultaneous second and any financed mortgage insurance) to the value of the house

**DTI**: Debt to Income Ratio, the ratio of the borrower’s total reported income and their total monthly debt obligation

**UPB**: Unpaid Principal Balance, the size of the loan to be repaid

**MSA**: Metropolitan Statistical Area a city or cities and surrounding communities linked by social and economic ties as determined by OMB

**Origination Date**: The date when a mortgage loan is funded and the property is purchased or refinanced

# Introduction

## Introduction/Background

Buying a house is usually the largest financial decision an American family will make. Being forced to leave that house is an equally large financial and emotional event. During times of economic stress, the government often steps in to try and keep people in their houses under the assumption that preventing families from being forced their homes now also prevents future families from losing their homes. This belief, that people being foreclosed upon now increases the likelihood that other residents near them will also end up delinquent on their mortgages in the future, is called the ‘mortgage contagion effect.’ Originally popularized during the 2008 crisis it has reentered the public consciousness because of the Covid-19 pandemic.

## Problem Statement

Following the 2008 crisis there was concern and anecdotal evidence of a foreclosure 'contagion-effect'. As more people in an area defaulted on their mortgages, the prices of nearby houses declined, and the risks of other borrowers in the neighborhood defaulting increased. This project aims to evaluate that claim in a non-crisis scenario by leveraging the Fannie/Freddie 30 year FRM mortgage set.

# Literature Review

When people discuss mortgage foreclosure contagion effects the usual mechanism is as follows: one or more foreclosures in a neighborhood cause a decline in local home prices; either because of increasing available supply, known as a competitive effect, or decreasing the perceived desirability of the neighborhood, known as a disamenity effect. This decrease in local home values causes an increase in the size of mortgage loan amounts relative to the value of the home. If this new loan amount exceeds the value of the home the borrower will be incentivized to default, a so called ‘strategic default’. Broadly speaking the existing research has been focused on two areas:

1. The impact of foreclosures on local home prices. The evidence here has been robust and a clear relationship has been found, although the exact magnitude is unclear.
2. Borrowers’ tendency to default strategically. The evidence here is much more mixed with some work showing significant effects and others not, depending on the specifications and assumptions of each analysis.

Anenberg (2014) leveraged a unique set of local listings in four large cities: San Francisco, Washington DC, Phoenix, and Chicago as well as default data to observe the immediate impact on local listing prices when a home goes into foreclosure. Testing which of the two effects outlined above (competitive or disamenity) drive the home price declines associated with local foreclosures. They find that “Sellers are generally no more likely to change their list prices in the four weeks before and the four weeks after a new REO listing. However, during the exact week of an REO entry, the probability that a seller adjusts their list price jumps significantly.” (Anenberg, 2014) They further find that a seller is between 6% and 8% more like to adjust their listing price in the week following a new REO listing than in any other week. This effect decreases by about 25% for each standard deviation increase in distance between the two properties. A similar magnitude effect for other new listings in the area suggests that the main driver of their result is the competitive spillover effect.

Similarly, Campbell et al. (2011) leveraged a combined dataset of Massachusetts transaction data and bankruptcy and death records to identify distressed or bankruptcy sales and measure their effect on other sales in the area. They find that “each foreclosure that takes place 0.05 miles away lowers the price of a house by about 1 percent" (Campbell et al. (2011)). Lin et al. used the internal Fannie Mae data to construct a home price model leveraging Fannie Mae’s default data augmented with additional subprime data from other vendors, as Fannie Mae cannot own subprime mortgages and they represented a large share of the contemporaneous default population. They find that “within a 0.9-km radius and a 5-year period from liquidation, the occurrence of foreclosure can effectively depress the neighborhood property values by as much as 8.7% per event.” (Lin et al. (2009)) In a subsequent paper they find a slightly smaller effect of “having a neighboring property in the process of foreclosure can result in a discount to market value of up to 1% per nearby distressed property” (Harding, 2013)

Papers that examine ‘strategic defaults’ often have an identification problem: only the borrower knows if they are unable to continue making payments or if they were simply unwilling to do so. Survey data can help mitigate this identification problem by clarifying the thinking of borrowers at the time of default. Guiso (2013) conducted a survey in which they asked respondents to state the number of mortgage defaults they had observed in the last 6 months (or whatever), and how many of those defaults they believed could have continued to be paid by the borrower. “By taking a ratio of the two, we obtain an estimate of the percentage of actual defaults that are considered ‘strategic’ by the defaulters’ acquaintances.” (Guiso, 2013) They find an estimate that “the relationship between default and shortfall appears to be nonlinear, with a peak in the sensitivity of default to shortfall when the value of the shortfall is 50% of the value of the house.” (Guiso, 2013) Both Campbell and find about 1% price declines per nearby default suggesting that there would need to be a large general home price decline and a large number of local defaults before the price impact would be large enough for a strategic default to be likely. Unsurprisingly they “do not find any evidence for the clustering effect: ceteris paribus, knowing somebody who defaulted does not affect the moral attitude toward defaulting.” (Guiso, 2013) They do, however, find evidence of “the learning hypothesis: knowing somebody who strategically defaulted reduces the perceived probability that a bank would go after a borrower who defaults.” (Guiso, 2013). Local foreclosures should not impact a borrowers delinquency probability unless they are strategic; a fact that is unobservable in the Fannie Mae data. Strategic defaults are also not a large share of the population as “this study estimates that, in 2008, 17% of all U.S. defaults were strategic” (Guiso, 2013) They also find that “the willingness to default has remained fairly stable over time.” (Guiso, 2013) Importantly, they find a racial component stating that "Blacks and Hispanics appear much more likely to walk away from an underwater mortgage. Blacks are 87% more likely than the sample mean to default strategically than whites, Hispanics 82%.” (Guiso, 2013) Seiler et al. (2017) build upon these results to try to explain why some people are more willing to default than others. I will return to the correlation between race and strategic defaults later in the paper.

Towe and Lawley (2010) and Goodstein et al. (2017) both model the foreclosure contagion effect more directly. Towe and Lawley (2010) utilize a parcel level dataset for Maryland combined with loan data, census data and unemployment data to build a social interaction survivor model. They find that “a one unit increase in neighboring foreclosures increases the hazard of foreclosure by as much as 3%. The mean time to foreclosure among those neighbors that do foreclose, increases the hazard rate of foreclosure by less than 1%.” (Towe, 2010). Goodstein et al. (2017) utilize data from a national mortgage payment processing firm LPS. They specifically mention the under-reporting of subprime mortgages as a potential problem with their data, a point I will expand on later. Following Guiso et al. (2013) they attempt to control for strategic defaults by using FICO scores above 720 as a proxy for financially savvy borrowers who are more likely to strategically default. They also control for Government Sponsored Entity (GSE) borrowers like Fannie Mae and Freddie Mac, as GSE borrowers behave differently than non-GSE borrowers. They find that

“[F]or the general population the coefficient on the area delinquency rate does not affect the probability that a loan will enter into default. However, for borrowers that are more likely strategic defaulters we find that a one percent increase in the area delinquency rate results in a 1.1–2.5% increase in the probability of default. Moreover, the coefficient on the area delinquency rate for borrowers in this group is statistically different from borrowers less at risk of strategically defaulting.” (Goodstein, 2017)

So, their division between savvy and non-savvy borrowers seems warranted.

Much of the literature agrees that there is a contagion effect for mortgage delinquencies but the size of the effect, the population affected and the mechanism by which mortgage delinquencies spread are up for debate.

# Methodology

## Dataset Description

The Federal Housing Finance Authority (FHFA) has required both Fannie Mae and Freddie Mac to publish anonymized, loan-level data on all of their prime acquisitions. The data is at the loan/payment month level. Each loan has a record for every payment month containing loan characteristics as well as the current number of payments that the borrowers have made and the number expected (note that even if the borrower prepays a loan can never be more than current). As part of the anonymization process, Fannie Mae only publishes the first 3 digits of the property’s zip code as opposed to all 5 digits. This is an unavoidable feature of my data set and the implications will be discussed more below.

## Data Collection

Quarterly acquisition data sets are available publicly at the Fannie Mae website: <https://datadynamics.fanniemae.com/data-dynamics/#/reportMenu;category=HP> and are updated in the second month of each quarter. I downloaded all data back to 2011 in Q1 of 2022. I generated a base population using the 2017/2018 origination population and then the MSA/Zip level default data using acquisitions back to 2011 originations. This truncation may lead to a slight undercounting of total local delinquencies as loans originated prior to 2011 that go into default will be missed.

## Data Preprocessing and/or Feature Engineering

Due to the large (in excess of 200 Gigabytes and 1.2 billion rows) size of the data set I dropped many observations. Previous work has taken stratified random cuts of the population, but random downsampling is potentially problematic given the geographic nature of the data. Additionally, the anonymized nature of the data makes it impossible to calculate the exact distance between any two properties. To address both of these issues I instead decided to focus on Metropolitan Statistical Areas (MSAs) and then use the zip codes in each MSA as a proxy for the local neighborhood. This approach has some support for the literature “the zip code level provides a far more reasonable delineation [than MSA], which is correlated to factors that impact house value.” (Lin, 2009) To limit the size of my data set I chose to focus on only a subset of MSA with enough observations for analysis and enough zip codes to delineate different neighborhoods. Finally, to avoid fitting to different state legal systems as opposed local contagion effects, all MSAs that covered multiple states were removed. Table 1 below shows the results of this preprocessing.

Table



I ranked the top 15 MSAs the were entirely in the same state by number of distinct transactions and number of distinct Zip-3 codes. I repeated this for both 2017 Q1 and 2018Q1 to ensure stability. Any MSA that placed in the top 15 in any measure appears in Table 1. Any MSA that placed in the top 15 in all 4 measures was included in the final analysis set. The final list of MSAs used in my analysis are the 10 MSAs highlighted in green above.

Table

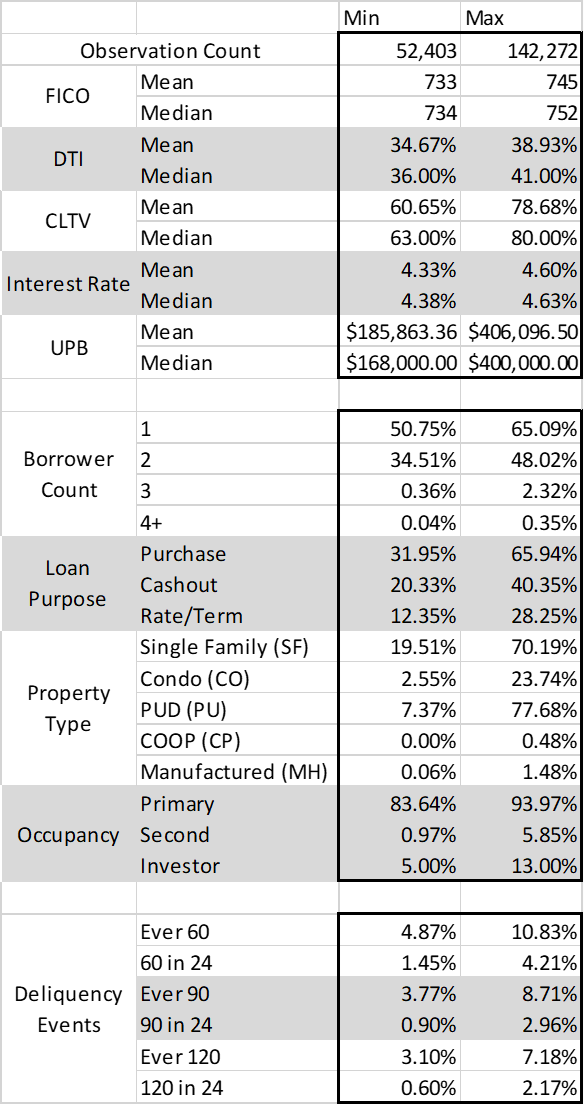


Table 2 shows the minimum and maximum MSA level values for the subset of columns that I used in my analysis[[1]](#footnote-1). Most variables are comparable across MSAs, however, a few require some additional discussion. The largest MSA (by observation count) is Los Angeles and the smallest is San Francisco. The remainder are clustered fairly closely between 80,000 and 100,00 observations.

Every MSA had a mean and median CLTV between 70 and 80 except San Francisco and Los Angeles, both in California. These two had substantially (almost 10 percentage points) lower CLTVs. Their mean and median UPBS are also about $100,000 larger than in any other city. Combined these two factors indicate that the underlying home prices in California are, unsurprisingly, much higher than in the rest of the MSAs in my analysis set. Given the much larger loan amounts, Los Angeles and San Francisco also had the highest shares of 2 and 3 borrower loans. All three MSAs in California (the third is Riverside) were majority refinances whereas all other MSAs were majority purchases

MSAs in the sunbelt, like Dallas, Atlanta and Phoenix, had much higher rates of Planed Urban Developments and much lower rates of Single-Family dwellings compared to cities outside of the sunbelt. Huston had higher delinquency rates than any other MSA in my set by every measure.

The contagion effect usually is described as increasing a borrower’s odds of going into foreclosure based on the number of borrowers in the same area who have gone into foreclosure. However, going into foreclosure takes several months of missed payments plus a sometimes quite lengthy judicial process for the bank to repossess and then sell the house. For this reason, it is standard practice to study delinquencies instead of foreclosures. They are more timely, and a necessary intermediate step before a loan goes into foreclosure. A borrower is expected to make a payment every month. If a borrower misses a payment, they are said to be 30 days delinquent. This is true regardless of if a payment was late or never arrived at all. For this reason 30 day delinquencies are rarely studied in the literature. The number of missed payments is too small relative to the number of accidentally late payments to be of much analytic value. A loan is 60 days delinquent if two payments have been missed and not made up (i.e. the borrower did not make two payments in one month to catch back up). For empirical analysis, the more missed payments researchers require before declaring a delinquency event the more likely that delinquency is to result in default and, ultimately, foreclosure. But, the higher the standard for delinquency events, the fewer observations you will have. The convention in the literature is to examine 60-day delinquencies (Gerardi, 2013). As a robustness check I will also check 90-day and 120-day delinquencies (when a loan is considered in default). Finally, a loan is ‘ever 60’ if at some point during its life the borrower misses two payments without catching up. This means that, the longer a borrower has been making payments, the more likely they are to be ‘ever 60,’ making intertemporal comparisons challenging. Instead, I examine if the borrower misses 2 (3, 4) payments in their first 24 months of payments. This measure is referred to as 60- (90-, 120-) in 24 measure of delinquency.

The Fannie Mae loan-level prime mortgage historical performance data set contains 108 columns. Many of these columns had to do with either the details of the loan’s rate adjustments or disposition status[[2]](#footnote-2). These were all removed as Adjustable Rate Mortgages (ARMs) were not considered conforming loans during the time periods I analyzed and so were not included in the final set. Disposition status variables would be potentially colinear with foreclosure in a way that would detract from the model’s performance and so these were also removed. Categorical variables were then 1-hot encoded. I also calculated the loan’s representative FICO score from the borrower and coborrower FICO scores as outlined in the Fannie Mae selling guide[[3]](#footnote-3).

## Data Modeling & Visualizations

### Model 1: Logistic Regression

The first model that I constructed was a logit regression at time of origination. The delinquency variable was sufficiently imbalanced that it noticeably affected model performance. To address this I implemented a SMOTE technique to artificially increase the number of delinquency observations. I then tested various FICO and LTV specifications, such as continuous CLTV, introducing a discontinuity at 80 CLTV (to reflect the addition of Mortgage Insurance), and splinning and binning following the Fannie Mae Pricing Buckets.[[4]](#footnote-4) FICO bins and continuous LTV were used as they provided the best performance. This also controls for savvy and non-savvy borrowers as outlined by Goodstein et al. Appendix 3 has the performance data for all of the specifications we tested. The model was estimated for each MSA to control for local economic factors (as the MSA should be economically integrated). Below is the output from a single model (Los Angeles) as an example. All other model outputs are available in Appendix 4.

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 223218

Model: Logit Df Residuals: 223200

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6732

Time: 16:39:50 Log-Likelihood: -50565.

converged: True LL-Null: -1.5472e+05

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.3425 0.023 -14.652 0.000 -0.388 -0.297

Brwr Cnt -1.3828 0.018 -77.461 0.000 -1.418 -1.348

DTI 0.0357 0.001 29.978 0.000 0.033 0.038

Orig CLTV 0.0457 0.001 71.099 0.000 0.044 0.047

cnt\_120\_msa 0.0088 0.000 82.034 0.000 0.009 0.009

cnt\_120\_zip 0.0037 0.000 17.690 0.000 0.003 0.004

constant -4.1215 0.100 -41.313 0.000 -4.317 -3.926

FTMB\_Y -0.5559 0.029 -18.913 0.000 -0.613 -0.498

Loan Purp\_P -0.9699 0.025 -39.493 0.000 -1.018 -0.922

Loan Purp\_R -1.6996 0.030 -56.334 0.000 -1.759 -1.640

Occ Stat\_P -1.3504 0.025 -54.963 0.000 -1.399 -1.302

Occ Stat\_S -4.0435 0.256 -15.794 0.000 -4.545 -3.542

fico\_bin\_1 -4.0565 0.040 -102.481 0.000 -4.134 -3.979

fico\_bin\_2 -4.3742 0.036 -122.270 0.000 -4.444 -4.304

fico\_bin\_3 -4.6273 0.033 -141.856 0.000 -4.691 -4.563

fico\_bin\_4 -4.9715 0.033 -152.893 0.000 -5.035 -4.908

fico\_bin\_5 -5.5764 0.037 -151.305 0.000 -5.649 -5.504

fico\_bin\_6 -5.4207 0.028 -196.220 0.000 -5.475 -5.367

The first variable, Orig Int Rate, is the interest rate at the time of origination of the loan. Brwr Cnt is the count of borrowers on the loan. DTI is the debt to income ratio. Orig CLTV is the CLTV at origination of the loan. Cnt\_120\_msa is the count of loans in the MSA, excluding those in the same zip as the current loan, that went 120+ days delinquent (i.e. into default) in the 2 years leading up to the origination date of the current loan. Cnt\_120\_zip is the count of loans in the same zip code as the current loan that went 120+ days delinquent in the 2 years leading up to the origination of the current loan and is my variable of interest. Constant is the constant term. FTMB\_Y is a dummy variable for if one or more of the borrowers is a first-time homebuyer. Loan Purp P and R control for Purchase and Rate/Term Refinances with Cashout Refinance as the reference value. Occ Stat P and S control for the Primary Residence and Second Home occupancy with Investor as the reference value. The final 6 variables control for which FICO pricing bucket the loan falls into with ‘0’ (FICO < 640) as the reference value.

A potential problem with the first model is that it predicts the odds of going delinquent using loans from the 2 years before origination. This is outside the time period that most of the literature suggests the contagion effect should be significant over. My second model addresses this concern.

### Model 2: Two Stage Time Series Logistic Regression

The second model is a two-stage logistic regression. The first step predicts the odds at origination of a loan going delinquent in its first 24 payments. It uses the same construction as first model with the exception that neither of the two delinquency variables appear in it.

The second step is at the payment month level as opposed to the loan level. I take the output from the first regression and interact it with the current payment month then add the number of 120+ day delinquencies in the prior 1, 3, 6, and 12 months at both the MSA and Zip code level (again, delinquencies in the zip are excluded from the MSA)

Below is the output from each stage for Los Angeles.

Step 1 - Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 223218

Model: Logit Df Residuals: 223202

Method: MLE Df Model: 15

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.4864

Time: 17:22:05 Log-Likelihood: -79473.

converged: True LL-Null: -1.5472e+05

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate 1.3115 0.016 83.888 0.000 1.281 1.342

Brwr Cnt -1.2023 0.013 -91.458 0.000 -1.228 -1.177

DTI 0.0625 0.001 65.875 0.000 0.061 0.064

Orig CLTV 0.0325 0.000 66.644 0.000 0.032 0.033

constant -5.3015 0.082 -64.836 0.000 -5.462 -5.141

FTMB\_Y -0.2552 0.020 -12.602 0.000 -0.295 -0.215

Loan Purp\_P -0.4951 0.019 -26.286 0.000 -0.532 -0.458

Loan Purp\_R -1.3099 0.023 -57.752 0.000 -1.354 -1.265

Occ Stat\_P -0.3083 0.020 -15.744 0.000 -0.347 -0.270

Occ Stat\_S -3.4136 0.197 -17.320 0.000 -3.800 -3.027

fico\_bin\_1 -3.7362 0.033 -111.758 0.000 -3.802 -3.671

fico\_bin\_2 -3.9240 0.031 -126.067 0.000 -3.985 -3.863

fico\_bin\_3 -3.8120 0.028 -135.881 0.000 -3.867 -3.757

fico\_bin\_4 -4.1226 0.028 -148.146 0.000 -4.177 -4.068

fico\_bin\_5 -4.4172 0.029 -151.985 0.000 -4.474 -4.360

fico\_bin\_6 -4.4792 0.024 -183.940 0.000 -4.527 -4.432

=================================================================================

Step 2 - Logit Regression Results

==============================================================================

Dep. Variable: del No. Observations: 5202274

Model: Logit Df Residuals: 5202243

Method: MLE Df Model: 30

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.2376

Time: 17:22:47 Log-Likelihood: -2.7490e+06

converged: True LL-Null: -3.6059e+06

Covariance Type: nonrobust LLR p-value: 0.000

===============================================================================

coef std err z P>|z| [0.025 0.975]

-------------------------------------------------------------------------------

cnt\_1\_msa 0.0094 0.000 69.410 0.000 0.009 0.010

cnt\_3\_msa 0.0003 7.27e-05 3.591 0.000 0.000 0.000

cnt\_6\_msa -0.0050 5.33e-05 -93.654 0.000 -0.005 -0.005

cnt\_12\_msa -0.0030 3.59e-05 -84.096 0.000 -0.003 -0.003

cnt\_1\_zip -0.0092 0.001 -11.532 0.000 -0.011 -0.008

cnt\_3\_zip -0.0028 0.001 -5.085 0.000 -0.004 -0.002

cnt\_6\_zip 0.0136 0.000 32.143 0.000 0.013 0.014

cnt\_12\_zip -0.0116 0.000 -39.925 0.000 -0.012 -0.011

constant -0.5814 0.002 -301.052 0.000 -0.585 -0.578

payments\_10 1.4045 0.014 97.748 0.000 1.376 1.433

payments\_11 1.8348 0.014 129.023 0.000 1.807 1.863

payments\_12 1.4738 0.015 98.392 0.000 1.444 1.503

payments\_13 1.8937 0.013 145.403 0.000 1.868 1.919

payments\_14 0.7268 0.016 46.104 0.000 0.696 0.758

payments\_15 1.3684 0.013 103.048 0.000 1.342 1.394

payments\_16 1.5730 0.013 118.410 0.000 1.547 1.599

payments\_17 2.5618 0.012 210.507 0.000 2.538 2.586

payments\_18 2.9721 0.012 250.266 0.000 2.949 2.995

payments\_19 4.0161 0.012 326.588 0.000 3.992 4.040

payments\_20 3.8643 0.012 330.212 0.000 3.841 3.887

payments\_21 4.7150 0.012 383.311 0.000 4.691 4.739

payments\_22 4.6476 0.012 373.796 0.000 4.623 4.672

payments\_23 4.6316 0.013 368.011 0.000 4.607 4.656

payments\_24 5.1661 0.013 409.134 0.000 5.141 5.191

payments\_3 1.9568 0.016 121.753 0.000 1.925 1.988

payments\_4 2.4398 0.014 168.730 0.000 2.411 2.468

payments\_5 1.4578 0.017 86.804 0.000 1.425 1.491

payments\_6 0.7903 0.018 43.985 0.000 0.755 0.825

payments\_7 1.2327 0.017 70.877 0.000 1.199 1.267

payments\_8 1.7110 0.014 118.111 0.000 1.683 1.739

payments\_9 1.1961 0.016 75.781 0.000 1.165 1.227

===============================================================================

Payment months 1 and 2 are dropped because it is impossible to go 2 months delinquent in your first payment month and the count of people who missed their first two payments was 0. Similarly, when modeling 90-day delinquencies periods 1, 2, and 3 were dropped[[5]](#footnote-5) and when modelling 120-day delinquency periods 1, 2, 3, and 4 were dropped.

### Model 3: XGBoosted Regression

The final set of models I estimated were XGBoosted models using the same independent variables as the first model. XGBoosted models have greater accuracy than logistic regression, but lack the global coefficients that Logistic regression models have because of their tree structure. Instead, I derived their estimation of the impact of additional local defaults by manually changing the data and observing the changes in predicted delinquency percentages. I will discuss these results in more depth in the Results section.

# Results & Analysis

Table 3 shows in- and out-sample accuracy measures for model 1 (for compactness only 60 in 24 values are presented) as well as the beta associated with the number of 120+ day delinquencies in the same zip code as the loan being modeled. Betas highlighted in yellow are not statistically different from 0 and betas highlighted in red have the wrong sign (an increase in local defaults decreases the probability of another loan going delinquent).

Table



As the severity of the delinquency being modeled increases the coefficient is significant more often with all but 2 being significant when the modeled severity is 120 days. Table 3 also shows the magnitude of a one unit increase in local delinquencies on the predicted 60 in 24 probability for the median loan.

Table 4, below, shows similar information as Table 3 but for the two-step model. The Pseudo R-squared and F1 scores decline noticeably relative to the first model. This is because I am now trying to predict not just whether or not a given loan will go delinquent, but exactly which month it will go delinquent in. Again, betas in yellow are not statistically significantly different than 0 and betas in red have the wrong sign.

Table



For 60-day delinquency the two-step regression tells a very different story than the simple logistic regression. For more than half of the MSAs I examined additional defaults in the same zip code actually decrease the odds that the loan being modeled will go delinquent. As the modeled delinquency gets more and more severe the coefficient moves to the expected direction and, when modeling 120-day delinquency, all 10 models have statistically significant positive coefficients.

The final set of models I developed were the XGBoosted models. As mentioned previously, XGBoosted models do not have clean interpretations the way that Logistic regression models do. There are no global coefficients to interpret because of the tree structure of the model. Instead, I developed a two-step process to derive a similar output to a coefficient to test my hypothesis. First, an XGBoosted model is trained on my dataset. Next, I manually changed the number of 120-day delinquencies in the zip code to 0 and re-ran the model on the modified dataset to compare the predicted delinquency percentages and the prior predictions to see what the shift was. As a robustness check I also added the delinquencies from the zip to the MSA. Table 5 presents the percentage of the time the delinquency percentage decreased under both treatments compared to staying the same or increasing. When highlighted yellow, a plurality, but not a majority, of cases had their probability decrease. When highlighted red, a minority of cases had their delinquency percentage decrease.

Table



For all three definitions of delinquency, all 10 models decrease the predicted odds of going delinquent more than 50% of the time when the number of defaults in the zip are set to 0. When the delinquencies were moved to the MSA, instead of simply removed, a few MSAs stopped showing evidence of a contagion effect. Unlike the previous models the number of models showing evidence of a contagion effect decreased as the delinquency measure got more severe.

# Conclusion

## Conclusion

All three models showed strong evidence for a mortgage foreclosure contagion effect, especially for the more severe 90- and 120-day delinquencies. The exact MSAs that showed evidence of a contagion effect changed based on the model specification and delinquency definition used. This suggests that there may noise in the data, especially for less sever delinquencies. That all 3 model specifications found similarly strong evidence of a contagion effect functions as a robustness check. A slightly more troubling result is that the model that showed the weakest correlation (at least for 60-in-24) was the two-step logistic regression which should have been best equipped to detect a contagion effect. Additionally, the magnitude of the change in delinquency odds varies substantially from a few basis points (logistic regression) to approximately 10bps (XGBoost) to more than a percentage point (2-step logistic regression). A possible explanation is that a spike in local defaults makes people already close to going delinquent more likely to, thereby pulling forward delinquencies that would have occurred in the future anyway.

## Project Limitation

This style of analysis has several limitations, as discussed in the literature review already. Harding et al. explain that “A general problem with the hedonic specification is that it is impossible to observe all house, location and local market characteristics and thus the coefficient estimates of the included variables are subject to omitted variable bias. The omitted variable problem arises very frequently in the study of urban externalities, including the impact on home values of environmental problems, schools, and commercial development.” (Harding, 2009). Campbell et al. cite an additional limitation saying “A challenge in interpreting this result is that local economic shocks, such as plant closings, may drive both house prices and foreclosures. Furthermore, foreclosures are endogenous to house prices because homeowners are more likely to default if they have negative equity, which is more likely as house prices fall. Ideally, I would like an instrument that influences foreclosures but that does not influence house prices except through foreclosures; however, I have not been able to find such an instrument.” (Campbell, 2011) As MSAs should be economically integrated fitting my model at an MSA level should help minimize this effect but it cannot be entirely discounted. A third potential issue, outlined in Towe, is that “The composition of neighborhoods is the outcome of a sorting process, wherein households with similar socioeconomic characteristics sort into neighborhoods according to variation in the suite of amenities offered by different neighborhoods (Ioannidies and Zabel 2002). In this case, spatial clustering of foreclosures occurs because households with higher inherent propensity to foreclose have sorted into neighborhoods.” (Towe, 2010). Given the racial discrepancy in strategic default rates found by Guiso et al. and the correlation between race and wealth I may be finding a proxy for the racial composition of a neighborhood instead of a true contagion effect. Most perniciously I cannot clearly show the direction of causality:  do price declines cause foreclosure or do foreclosures cause price decline? Regression is unable to show causality, simply correlation. The contagion effect is a claim about causality.

More specific to this analysis and dataset are the limitations presented by the anonymized nature of the data. Campbell et al. find a foreclosure spillover of only .05 miles while Lin et al. find a more generous .9km, but either measure is likely smaller than a zip code. I just have a partial zip code. This means that any measure of contagion effect will be muted by including properties that are too far to have an impact on the property being modeled for us to expect them to have any impact. The Fannie Mae data also only includes prime mortgages and when default rates are much higher among subprime mortgages. Subprimes as a share of the mortgage market have declined since their peaks immediately prior to the 2008 crisis, but I am still likely systematically undercounting total defaults in a way that is highly correlated with neighborhood.

## Future Research

An easy first step for additional research would be to model additional MSAs. Baltimore, Columbus and Pittsburg were all excluded from this analysis because their total observation count were too low, but their counts are still likely to be sufficient for model training. A simple way to expand the counts would be to model additional time periods. I was computationally constrained and so limited my analysis to 2017/2018 originations, but Fannie Mae’s data goes all the way back to 2000 so there is an abundance of available data.

A more complex avenue for additional research would be to add in additional time variant data like the Case-Shiller home price index or unemployment data. Because the data does not exist at the zip-3 level and the model is fit at an MSA level it should already be included in the intercept term. Even so there is likely a way to augment the time series model with those, or other variables, for additional accuracy.

The final, and most complex, potential avenue for future research would be to get access to non-anonymized data as Harding et al. and Lin et al. had. This would allow for precise measurements of distance between properties and the incorporation of more time series data. It would also allow for a national, or state level estimations as opposed to many MSA level models. Ultimately this would be necessary to truly demonstrate the existence of a mortgage contagion effect.

# References

1. Anenberg, E., & Kung, E. (2014). Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures. The American Economic Review, 104(8), 2527–2551. <http://www.jstor.org/stable/42920898>
2. Campbell, J. Y., Giglio, S., & Pathak, P. (2011). Forced Sales and House Prices. The American Economic Review, 101(5), 2108–2131. <http://www.jstor.org/stable/23045632>
3. Gerardi, Kristopher S. and Herkenhoff, Kyle and Ohanian, Lee E. and Willen, Paul S., Unemployment, Negative Equity, and Strategic Default (July 12, 2013). Available at SSRN: <https://ssrn.com/abstract=2293152> or <http://dx.doi.org/10.2139/ssrn.2293152>
4. Goodstein, R., Hanouna, P., Ramirez, C. D., &amp; Stahel, C. W. (2017). Contagion effects in strategic mortgage defaults. ScienceDirect. Retrieved April 20, 2022, from <https://doi.org/10.1016/j.jfi.2016.10.001>
5. Guiso, L., Sapienza, P., &amp; Zingales, L. (2013). The Determinants of Attitudes toward Strategic Default on Mortgages. Wiley Online Library. Retrieved April 21, 2022, from <https://doi.org/10.1111/jofi.12044>
6. Harding, J. P., Rosenblatt, E., &amp; Yao, V. W. (2009). Journal of Urban Economics. The contagion effect of foreclosed properties. Retrieved April 20, 2022, from  <https://www.sciencedirect.com/journal/journal-of-urban-economics/vol/71/issue/1>
7. Lin, Z., Rosenblatt, E. & Yao, V.W. Spillover Effects of Foreclosures on Neighborhood Property Values. J Real Estate Finan Econ 38, 387–407 (2009). <https://doi.org/10.1007/s11146-007-9093-z>
8. Seiler, M. J., Seiler, V. L., Lane, M. A., &amp; Harrison, D. M. (2012). Fear, Shame and Guilt: Economic and Behavioral Motivations for Strategic Default. Wiley Online Library. Retrieved April 21, 2022, from <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6229.2012.00344.x>
9. Towe, Charles, and Chad Lawley. "The Contagion Effect of Neighboring Foreclosures." American Economic Journal: Economic Policy 5, no. 2 (2010): 313-35. <http://www.jstor.org/stable/43189336>. Copyright statement: <https://www.aeaweb.org/journals/policies/copyright>

# Appendix

## Appendix 1: Full Variable List

Taken from <https://datadynamics.fanniemae.com/data-dynamics/#/resources/HP>

1. **Reference Pool ID** A unique identifier for the reference pool.
2. **Loan Identifier** A unique identifier for the mortgage loan.
3. **Monthly Reporting Period** The month and year that pertains to the servicer’s cut-off period for mortgage loan information.
4. **Channel** The origination channel used by the party that delivered the loan to the issuer.
5. **Seller Name** The name of the entity that delivered the mortgage loan to Fannie Mae.
6. **Servicer Name** The name of the entity that serves as the primary servicer of the mortgage loan.
7. **Master Servicer** Fannie Mae
8. **Original Interest Rate** The original interest rate on a mortgage loan as identified in the original mortgage note.
9. **Current Interest Rate** The rate of interest in effect for the periodic installment due.
10. **Original UPB** The dollar amount of the loan as stated on the note at the time the loan was originated.
11. **UPB at Issuance** The unpaid principal balance of the loan as of the cut-off date of the reference pool.
12. **Current Actual UPB** The current actual outstanding unpaid principal balance of a mortgage loan, reflective of payments actually received from the related borrower.
13. **Original Loan Term** For fixed-rate, adjustable-rate and Interest-only mortgages, the number of months in which regularly scheduled borrower payments are due at the time the loan was originated.
14. **Origination Date** The date of each individual note.
15. **First Payment Date** The date of the first scheduled mortgage loan payment to be made by the borrower under the terms of the mortgage loan documents.
16. **Loan Age** The number of calendar months since the mortgage loan's origination date. For purposes of calculating this data element, origination means the date on which the first full month of interest begins to accrue.
17. **Remaining Months to Legal Maturity** The number of calendar months remaining until the mortgage loan is due to be paid in full based on the maturity date as defined in the mortgage documents.
18. **Remaining Months To Maturity** The number of calendar months remaining until the outstanding unpaid principal balance of the mortgage loan amortizes to a zero balance, taking into account any additional prepayments, which could lead to the loan paying off earlier than its maturity date.
19. **Maturity Date** The month and year in which a mortgage loan is scheduled to be paid in full as defined in the mortgage loan documents.
20. **Original Loan to Value Ratio (LTV)** The ratio, expressed as a percentage, obtained by dividing the amount of the loan at origination by the value of the property.
21. **Original Combined Loan to Value Ratio (CLTV**) The ratio, expressed as a percentage, obtained by dividing the amount of all known outstanding loans at origination by the value of the property.
22. **Number of Borrowers** The number of individuals obligated to repay the mortgage loan.
23. **Debt-To-Income (DTI)** The ratio obtained by dividing the total monthly debt expense by the total monthly income of the borrower at the time the loan was originated.
24. **Borrower Credit Score at Origination** A numerical value used by the financial services industry to evaluate the quality of borrower’s credit. Credit scores are typically based on a proprietary statistical model that is developed for use by credit data repositories. These credit repositories apply the model to borrower credit information to arrive at a credit score. When this term is used by Fannie Mae, it is referring to the "Classic" FICO score developed by Fair Isaac Corporation.
25. **Co-Borrower Credit Score at Origination** A numerical value used by the financial services industry to evaluate the quality of borrower’s credit. Credit scores are typically based on a proprietary statistical model that is developed for use by credit data repositories. These credit repositories apply the model to borrower credit information to arrive at a credit score. When this term is used by Fannie Mae, it is referring to the "Classic" FICO score developed by Fair Isaac Corporation.
26. **First Time Home Buyer Indicator** An indicator that denotes if the borrower or co-borrower qualifies as a first-time homebuyer.
27. **Loan Purpose** An indicator that denotes whether the mortgage loan is either a refinance mortgage or a purchase money mortgage. Purpose may be the purchase of a new property or refinance of an existing lien (with cash out or with no cash out).
28. **Property Type** An indicator that denotes whether the property type secured by the mortgage loan is a condominium, co-operative, planned urban development (PUD), manufactured home, or single-family home.
29. **Number of Units** The number of units comprising the related mortgaged property (one, two, three, or four).
30. **Occupancy Status** The classification describing the property occupancy status at the time the loan was originated.
31. **Property State** A two-letter abbreviation indicating the state or territory within which the property securing the mortgage loan is located.
32. **Metropolitan Statistical Area (MSA)** The numeric Metropolitan Statistical Area Code for the property securing the mortgage loan. MSAs are established by the US Office of Management and Budget. An area usually qualifies as an MSA if it is defined by the Bureau of the Census as an urbanized area and has a population of 50,000 or more in a total metropolitan area of at least 100,000. An MSA may consist of one or more counties.
33. **Zip Code Short** Limited to the first three digits of the code designated by the U.S. Postal Service where the subject property is located.
34. **Mortgage Insurance Percentage** The original percentage of mortgage insurance coverage obtained for an insured conventional mortgage loan and used following the occurrence of an event of default to calculate the insurance benefit, as defined by the underlying master primary insurance policy.
35. **Amortization Type** The classification of the loan as having either a fixed- or an adjustable-interest rate at the time the loan was originated.
36. **Prepayment Penalty Indicator** An indicator that denotes whether the borrower is subject to a penalty for early payment of principal.
37. **Interest Only Loan Indicator** An indicator that denotes whether the loan only requires interest payments for a specified period of time beginning with the first payment date.
38. **Interest Only First Principal And Interest Payment Date** For interest-only loans, the month and year that the first monthly scheduled fully amortizing principal and interest payment is due.
39. **Months to Amortization** For interest-only loans, the number of months from the current month to the first scheduled principal and interest payment date.
40. **Current Loan Delinquency Status** The number of months the obligor is delinquent as determined by the governing mortgage documents.
41. **Loan Payment History** The coded string of values that describes the payment performance of the loan over the most recent 24 months. The most recent month is located to the right.
42. **Modification Flag** An indicator that denotes if the mortgage loan has been modified.
43. **Mortgage Insurance Cancellation Indicator** An indicator that denotes if the mortgage insurance (MI) has been cancelled since origination.
44. **Zero Balance Code** A code indicating the reason the loan's balance was reduced to zero or experienced a credit event, if applicable.
45. **Zero Balance Effective Date** Date on which the mortgage loan balance was reduced to zero.
46. **UPB at the Time of Removal** The unpaid principal balance of the loan at the time of removal.
47. **Repurchase Date** The date on which a Reversed Credit Event Reference Obligation occurs with respect to a loan.
48. **Scheduled Principal Current** The minimum principal payment the borrower is obligated to pay for the corresponding reporting period, based on the terms provided in the related mortgage loan documents, provided that the payment is collected from the borrower by the servicer and reported to Fannie Mae for the corresponding reporting period.
49. **Total Principal Current** The change between the prior reporting period’s disclosed Current Actual UPB and the current reporting period’s disclosed Current Actual UPB.
50. **Unscheduled Principal Current** The principal amount received in excess of the scheduled principal payment collected from the borrower by the servicer and reported to Fannie Mae for the corresponding reporting period.
51. **Last Paid Installment Date** The due date of the last paid installment that was collected for the mortgage loan.
52. **Foreclosure Date** The date on which the completion of the legal action of foreclosure occurred.
53. **Disposition Date** The date on which Fannie Mae’s interest in a property ends through either the transfer of the property to a third party or the satisfaction of the mortgage obligation.
54. **Foreclosure Costs** Expenses associated with obtaining title to property from the mortgagor, valuing the property, and maintaining utility services to the property. Such costs include costs and fees associated with bankruptcy and foreclosure.
55. **Property Preservation and Repair Costs** The expenses associated with securing and preserving the property including two major categories: maintenance and repairs. Maintenance costs are associated with preserving the property through normal upkeep, while repairs are associated with either avoiding deterioration of the asset or a marketing decision to help maximize sales proceeds upon final disposition.
56. **Asset Recovery Costs** Expenses associated with removing occupants and personal property from an occupied property post foreclosure. Such expenses include relocation assistance, deed-in-lieu fee, and fees and costs associated with eviction actions.
57. **Miscellaneous Holding Expenses and Credits** Expenses and credits associated with preserving the property, including Homeowners Association and other dues; flood, hazard, and MI premiums and refunds; rental income; and title insurance costs.
58. **Associated Taxes for Holding Property** Payment of taxes associated with holding the property.
59. **Net Sales Proceeds** Total cash received from the sale of the property net of any applicable selling expenses, such as fees and commissions, allowable for inclusion under the terms of the property sale, as currently reported on the HUD-1 or other settlement statement.
60. **Credit Enhancement Proceeds** Proceeds from claims on primary and certain other limited mortgage insurance policies, and recourse and indemnification payments from lenders under arrangements designed to limit credit exposure to Fannie Mae.
61. **Repurchase Make Whole Proceeds** Amounts received by Fannie Mae under the terms of our representation and warranty arrangements for the repurchase of the mortgage loan or the subject property or loss reimbursement subsequent to property disposition.
62. **Other Foreclosure Proceeds** Amounts, other than sale proceeds, received by Fannie Mae following foreclosure of a property, including redemption proceeds received from the mortgagor.
63. **Non-Interest Bearing UPB** A portion of the UPB, as a result of an eligible loan modification, that will not accrue interest.
64. **Principal Forgiveness Amount** A reduction of the UPB owed on a mortgage by a borrower that is formally agreed to by the lender and the borrower, usually in conjunction with a loan modification.
65. **Original List Start Date** The agreed upon date, between a property seller and a broker, authorizing the broker to begin the process to procure a buyer or tenant for the property seller’s real property.
66. **Original List Price** The initial price at which a real property is offered for sale by the property seller.
67. **Current List Start Date** The agreed upon date, between a property seller and a broker, authorizing the broker to begin the process to procure a buyer or tenant for the property seller’s real property.
68. **Current List Price** The price at which a real property is offered for sale.
69. **Borrower Credit Score At Issuance** A numerical value used by the financial services industry to evaluate the quality of borrower credit. Credit scores are typically based on a proprietary statistical model that is developed for use by credit data repositories. These credit repositories apply the model to borrower credit information to arrive at a credit score. When this term is used by Fannie Mae, it is referring to FICO Score 5(1) developed by Fair Isaac Corporation and provided by Equifax Inc. and is distinct from the FICO Score referenced in Fannie Mae's Selling Guide, which may be provided by any of the three major credit repositories
70. **Co-Borrower Credit Score At Issuance** A numerical value used by the financial services industry to evaluate the quality of borrower credit. Credit scores are typically based on a proprietary statistical model that is developed for use by credit data repositories. These credit repositories apply the model to borrower credit information to arrive at a credit score. When this term is used by Fannie Mae, it is referring to FICO Score 5(1) developed by Fair Isaac Corporation and provided by Equifax Inc and is distinct from the FICO Score referenced in Fannie Mae's Selling Guide.
71. **Borrower Credit Score Current** A numerical value used by the financial services industry to evaluate the quality of borrower credit. Credit scores are typically based on a proprietary statistical model that is developed for use by credit data repositories. These credit repositories apply the model to borrower credit information to arrive at a credit score. When this term is used by Fannie Mae, it is referring to FICO Score 5(1) developed by Fair Isaac Corporation and provided by Equifax Inc and is distinct from the FICO Score referenced in Fannie Mae's Selling Guide, which may be provided by any of the three major credit repositories.
72. **Co-Borrower Credit Score Current** A numerical value used by the financial services industry to evaluate the quality of borrower credit. Credit scores are typically based on a proprietary statistical model that is developed for use by credit data repositories. These credit repositories apply the model to borrower credit information to arrive at a credit score. When this term is used by Fannie Mae, it is referring to FICO Score 5(1) developed by Fair Isaac Corporation and provided by Equifax Inc and is distinct from the FICO Score referenced in Fannie Mae's Selling Guide.
73. **Mortgage Insurance Type** The entity that is responsible for the Mortgage Insurance premium payment.
74. **Servicing Activity Indicator** An indicator that denotes a change in servicing activity during the corresponding reporting period.
75. **Current Period Modification Loss Amount** The loss amount calculated for a mortgage loan resulting from a modification event for the corresponding reporting period.
76. **Cumulative Modification Loss Amount** The cumulative loss amount calculated for a mortgage loan resulting from a modification event.
77. **Current Period Credit Event Net Gain or Loss** The net realized gain or loss amount calculated for a mortgage loan resulting from a credit event for the corresponding reporting period.
78. **Cumulative Credit Event Net Gain or Loss** The cumulative net realized gain or loss amounts for a mortgage loan resulting from a credit event.
79. **HomeReady® Program Indicator** "An indicator that denotes if the borrower participated in Fannie Mae’s HomeReady program. HomeReady is our affordable, low down payment mortgage product designed to expand the availability of mortgage financing to creditworthy low-to-moderate-income borrowers."
80. **Foreclosure Principal Write-off Amount** Amounts that Fannie Mae or its loan servicers have determined to be uncollectable under applicable state laws, due to foreclosure statute of limitations.
81. **Relocation Mortgage Indicator** An indicator that denotes whether or not the type of mortgage loan is a relocation mortgage loan, made to borrowers whose employers relocate their employees.
82. **Zero Balance Code Change Date** The most recent date in which a loan status change was identified, resulting from corresponding change to the Zero Balance Code.
83. **Loan Holdback Indicator** An indicator that denotes if a loan has been moved temporarily into a ‘hold’ status to allow Fannie Mae to further evaluate unique situations that may otherwise result in a credit event or loan removal. Such situations may include loans with reported data anomalies, loans currently in forbearance due to a natural disaster or loans refinanced under the High LTV program that will continue to be included in the reference pool.
84. **Loan Holdback Effective Date** The date of the latest Loan Holdback indicator change.
85. **Delinquent Accrued Interest** The lost accrued interest amount calculated for a mortgage loan that becomes subject to a credit event for the corresponding reporting period.
86. **Property Valuation Method** An indicator that denotes the method by which the value of the subject property was obtained.
87. **High Balance Loan Indicator** An indicator that denotes if the original principal balance of a mortgage loan is greater than the general conforming loan limit and up to the high-cost area loan limit.
88. **ARM Initial Fixed-Rate Period ≤ 5 YR Indicator** For an adjustable-rate mortgage loan, an indicator that denotes if the Initial Fixed-Rate Period is less than or equal to five years.
89. **ARM Product Type** For an adjustable-rate mortgage loan, a string that denotes the Initial Fixed-Rate Period, the subsequent Interest Rate Adjustment Frequency, and the Original Loan Term.
90. **Initial Fixed-Rate Period** For an adjustable-rate mortgage loan, the number of months between the first full month the mortgage loan accrues interest and the initial interest rate change date.
91. **Interest Rate Adjustment Frequency** For an adjustable-rate mortgage loan, the number of months between scheduled rate changes. For loans with an Initial Fixed-Rate Period, the number of months between subsequent rate adjustments.
92. **Next Interest Rate Adjustment Date** For adjustable-rate loans, the month and year that the interest rate is next subject to change.
93. **Next Payment Change Date** For an adjustable-rate mortgage loan, the next date on which the payment amount due from the borrower is subject to change.
94. **Index** For adjustable-rate loans, the description of the index on which adjustments to the interest rate are based.
95. **ARM Cap Structure** For an adjustable-rate mortgage loan, a numeric string that explains the interest rate caps on the ARM. The first number is the Initial Interest Rate Cap Up Percent (i.e., the maximum percentage points the interest rate can adjust upward at the initial rate change date). The second number is the Periodic Interest Rate Cap Up Percent (i.e., the maximum percentage points the interest rate can adjust upward at each interest rate change date after the initial interest rate change date). The third number is the Lifetime Interest Rate Cap Up Percent (i.e., the maximum percentage points that the interest rate can adjust upward over the life of the loan relative to the initial interest rate).
96. **Initial Interest Rate Cap Up Percent** For an adjustable-rate mortgage loan, the maximum percentage points the interest rate can adjust upward at the initial interest rate change date.
97. **Periodic Interest Rate Cap Up Percent** For an adjustable-rate mortgage loan, the maximum percentage points the interest rate can adjust upward at each interest rate change date after the initial interest rate change date.
98. **Lifetime Interest Rate Cap Up Percent** For an adjustable-rate mortgage loan, the maximum percentage points that the interest rate can adjust upward over the life of the loan relative to the initial interest rate.
99. **Mortgage Margin** For an adjustable-rate mortgage loan, the rate that is added to the index value to establish the new interest rate (after applying all applicable caps and floors) accruing on the loan at each interest rate change date.
100. **ARM Balloon Indicator** For an adjustable-rate mortgage loan, a code that denotes if the loan has a balloon feature.
101. **ARM Plan Number** For an adjustable-rate mortgage loan, a code identifying the standardized plan under which the mortgage loan was delivered to Fannie Mae. The ARM plan outlines the characteristics of the adjustable-rate mortgage loan, including the ARM Index, the Initial Fixed-Rate Period, the Cap Structure, look-back days, assumability, and the option to convert to a fixed-rate mortgage loan..
102. **Borrower Assistance Plan** An indicator that denotes the type of assistance plan that the borrower is enrolled in that provides temporary mortgage payment relief or an opportunity for the borrower to cure a mortgage delinquency over a defined period.
103. **High Loan to Value (HLTV) Refinance Option Indicator** An indicator that denotes if an eligible original reference loan is refinanced under Fannie Mae’s HLTV refinance option, which results in such mortgage loan remaining in the Reference Pool, as further defined in each individual CRT document, if applicable.
104. **Deal Name** The title of the series issuance.
105. **Repurchase Make Whole Proceeds Flag** Indicates if Fannie Mae received proceeds under the terms of its representation and warranty arrangements for the repurchase of the mortgage loan.
106. **Alternative Delinquency Resolution** An indicator that denotes the loss mitigation solution designed to resolve delinquencies and help homeowners remain in their homes in accordance with the servicer’s contractual obligation, while allowing the loan to remain in the security.
107. **Alternative Delinquency Resolution Count** The total number of Alternative Delinquency Resolutions as reported by the servicer for a specific loan.
108. **Total Deferral Amount** The total non-interest-bearing deferral amount related to one or more Alternative Delinquency Resolutions.

## Appendix 2: MSA Level Summary Statistics





## Appendix 3: Fico/LTV Specification





## Appendix 4: Full 60 in 24 Regression Results

Modeling MSA 31080

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 223218

Model: Logit Df Residuals: 223200

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6732

Time: 16:39:50 Log-Likelihood: -50565.

converged: True LL-Null: -1.5472e+05

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.3425 0.023 -14.652 0.000 -0.388 -0.297

Brwr Cnt -1.3828 0.018 -77.461 0.000 -1.418 -1.348

DTI 0.0357 0.001 29.978 0.000 0.033 0.038

Orig CLTV 0.0457 0.001 71.099 0.000 0.044 0.047

cnt\_120\_msa 0.0088 0.000 82.034 0.000 0.009 0.009

cnt\_120\_zip 0.0037 0.000 17.690 0.000 0.003 0.004

constant -4.1215 0.100 -41.313 0.000 -4.317 -3.926

FTMB\_Y -0.5559 0.029 -18.913 0.000 -0.613 -0.498

Loan Purp\_P -0.9699 0.025 -39.493 0.000 -1.018 -0.922

Loan Purp\_R -1.6996 0.030 -56.334 0.000 -1.759 -1.640

Occ Stat\_P -1.3504 0.025 -54.963 0.000 -1.399 -1.302

Occ Stat\_S -4.0435 0.256 -15.794 0.000 -4.545 -3.542

fico\_bin\_1 -4.0565 0.040 -102.481 0.000 -4.134 -3.979

fico\_bin\_2 -4.3742 0.036 -122.270 0.000 -4.444 -4.304

fico\_bin\_3 -4.6273 0.033 -141.856 0.000 -4.691 -4.563

fico\_bin\_4 -4.9715 0.033 -152.893 0.000 -5.035 -4.908

fico\_bin\_5 -5.5764 0.037 -151.305 0.000 -5.649 -5.504

fico\_bin\_6 -5.4207 0.028 -196.220 0.000 -5.475 -5.367

=================================================================================

Modeling MSA 19100

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 181562

Model: Logit Df Residuals: 181544

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6771

Time: 16:42:49 Log-Likelihood: -40632.

converged: True LL-Null: -1.2585e+05

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.3823 0.022 -16.994 0.000 -0.426 -0.338

Brwr Cnt -1.4697 0.021 -68.559 0.000 -1.512 -1.428

DTI 0.0299 0.001 24.053 0.000 0.027 0.032

Orig CLTV 0.0694 0.001 69.834 0.000 0.067 0.071

cnt\_120\_msa 0.0068 8.27e-05 81.820 0.000 0.007 0.007

cnt\_120\_zip 0.0011 7.61e-05 14.026 0.000 0.001 0.001

constant -3.0677 0.113 -27.079 0.000 -3.290 -2.846

FTMB\_Y -0.8976 0.027 -32.982 0.000 -0.951 -0.844

Loan Purp\_P -0.8752 0.027 -32.869 0.000 -0.927 -0.823

Loan Purp\_R -2.2826 0.055 -41.547 0.000 -2.390 -2.175

Occ Stat\_P -1.3889 0.032 -43.286 0.000 -1.452 -1.326

Occ Stat\_S -3.1052 0.191 -16.294 0.000 -3.479 -2.732

fico\_bin\_1 -3.5628 0.036 -98.066 0.000 -3.634 -3.492

fico\_bin\_2 -4.3120 0.037 -114.992 0.000 -4.385 -4.238

fico\_bin\_3 -4.4187 0.035 -125.307 0.000 -4.488 -4.350

fico\_bin\_4 -5.1540 0.039 -130.967 0.000 -5.231 -5.077

fico\_bin\_5 -5.4326 0.041 -131.109 0.000 -5.514 -5.351

fico\_bin\_6 -5.5179 0.031 -177.010 0.000 -5.579 -5.457

=================================================================================

Modeling MSA 12060

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 136878

Model: Logit Df Residuals: 136860

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.7101

Time: 16:43:23 Log-Likelihood: -27508.

converged: True LL-Null: -94877.

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.4632 0.031 -15.106 0.000 -0.523 -0.403

Brwr Cnt -1.8812 0.033 -56.252 0.000 -1.947 -1.816

DTI 0.0434 0.001 30.980 0.000 0.041 0.046

Orig CLTV 0.0767 0.001 61.046 0.000 0.074 0.079

cnt\_120\_msa 0.0124 0.000 71.574 0.000 0.012 0.013

cnt\_120\_zip 0.0008 0.000 6.047 0.000 0.001 0.001

constant -3.5493 0.151 -23.484 0.000 -3.846 -3.253

FTMB\_Y -0.7284 0.028 -25.641 0.000 -0.784 -0.673

Loan Purp\_P -1.0690 0.033 -32.444 0.000 -1.134 -1.004

Loan Purp\_R -2.6647 0.071 -37.415 0.000 -2.804 -2.525

Occ Stat\_P -1.3301 0.039 -34.215 0.000 -1.406 -1.254

Occ Stat\_S -3.0168 0.131 -22.982 0.000 -3.274 -2.760

fico\_bin\_1 -4.1235 0.057 -72.610 0.000 -4.235 -4.012

fico\_bin\_2 -4.7072 0.052 -90.817 0.000 -4.809 -4.606

fico\_bin\_3 -5.1058 0.048 -105.480 0.000 -5.201 -5.011

fico\_bin\_4 -5.5809 0.049 -114.827 0.000 -5.676 -5.486

fico\_bin\_5 -6.0694 0.051 -118.361 0.000 -6.170 -5.969

fico\_bin\_6 -6.2131 0.043 -143.448 0.000 -6.298 -6.128

=================================================================================

Modeling MSA 38060

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 170532

Model: Logit Df Residuals: 170514

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.7043

Time: 16:45:12 Log-Likelihood: -34952.

converged: True LL-Null: -1.1820e+05

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate 0.0024 0.023 0.106 0.916 -0.042 0.047

Brwr Cnt -1.5265 0.024 -63.725 0.000 -1.573 -1.480

DTI 0.0450 0.001 31.899 0.000 0.042 0.048

Orig CLTV 0.0838 0.001 70.641 0.000 0.081 0.086

cnt\_120\_msa 0.0060 0.000 44.828 0.000 0.006 0.006

cnt\_120\_zip -8.34e-05 0.000 -0.295 0.768 -0.001 0.000

constant -3.5016 0.120 -29.201 0.000 -3.737 -3.267

FTMB\_Y -0.3903 0.026 -15.129 0.000 -0.441 -0.340

Loan Purp\_P -1.2590 0.032 -39.523 0.000 -1.321 -1.197

Loan Purp\_R -2.3481 0.056 -41.969 0.000 -2.458 -2.238

Occ Stat\_P -1.7386 0.038 -45.525 0.000 -1.813 -1.664

Occ Stat\_S -3.5013 0.132 -26.471 0.000 -3.761 -3.242

fico\_bin\_1 -3.8607 0.042 -92.733 0.000 -3.942 -3.779

fico\_bin\_2 -4.5278 0.041 -109.826 0.000 -4.609 -4.447

fico\_bin\_3 -4.8812 0.039 -125.473 0.000 -4.957 -4.805

fico\_bin\_4 -5.4173 0.040 -135.027 0.000 -5.496 -5.339

fico\_bin\_5 -6.7342 0.053 -127.510 0.000 -6.838 -6.631

fico\_bin\_6 -6.2227 0.038 -165.628 0.000 -6.296 -6.149

=================================================================================

Modeling MSA 41860

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 82558

Model: Logit Df Residuals: 82540

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6731

Time: 16:47:11 Log-Likelihood: -18707.

converged: True LL-Null: -57225.

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.2197 0.040 -5.511 0.000 -0.298 -0.142

Brwr Cnt -1.2426 0.029 -43.120 0.000 -1.299 -1.186

DTI 0.0149 0.002 8.391 0.000 0.011 0.018

Orig CLTV 0.0333 0.001 36.015 0.000 0.031 0.035

cnt\_120\_msa 0.0310 0.001 56.198 0.000 0.030 0.032

cnt\_120\_zip -0.0008 0.000 -3.577 0.000 -0.001 -0.000

constant -2.7405 0.162 -16.914 0.000 -3.058 -2.423

FTMB\_Y 0.1099 0.047 2.335 0.020 0.018 0.202

Loan Purp\_P -0.9412 0.043 -21.838 0.000 -1.026 -0.857

Loan Purp\_R -1.6511 0.049 -33.696 0.000 -1.747 -1.555

Occ Stat\_P -1.1184 0.039 -28.595 0.000 -1.195 -1.042

Occ Stat\_S -3.3441 0.457 -7.321 0.000 -4.239 -2.449

fico\_bin\_1 -4.2768 0.074 -57.691 0.000 -4.422 -4.131

fico\_bin\_2 -4.4174 0.061 -72.285 0.000 -4.537 -4.298

fico\_bin\_3 -4.6529 0.057 -81.778 0.000 -4.764 -4.541

fico\_bin\_4 -5.1078 0.057 -90.399 0.000 -5.219 -4.997

fico\_bin\_5 -6.0145 0.067 -89.486 0.000 -6.146 -5.883

fico\_bin\_6 -5.2020 0.046 -114.254 0.000 -5.291 -5.113

=================================================================================

Modeling MSA 19740

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 122776

Model: Logit Df Residuals: 122758

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6920

Time: 16:55:55 Log-Likelihood: -26213.

converged: True LL-Null: -85102.

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.4719 0.032 -14.968 0.000 -0.534 -0.410

Brwr Cnt -1.7541 0.029 -61.102 0.000 -1.810 -1.698

DTI 0.0294 0.002 18.588 0.000 0.026 0.032

Orig CLTV 0.0579 0.001 51.472 0.000 0.056 0.060

cnt\_120\_msa 0.0179 0.000 46.687 0.000 0.017 0.019

cnt\_120\_zip 0.0022 0.000 4.770 0.000 0.001 0.003

constant -0.9623 0.143 -6.716 0.000 -1.243 -0.681

FTMB\_Y -0.6988 0.037 -19.060 0.000 -0.771 -0.627

Loan Purp\_P -1.0946 0.034 -32.595 0.000 -1.160 -1.029

Loan Purp\_R -2.3024 0.059 -38.956 0.000 -2.418 -2.187

Occ Stat\_P -1.5585 0.039 -39.896 0.000 -1.635 -1.482

Occ Stat\_S -3.4558 0.242 -14.263 0.000 -3.931 -2.981

fico\_bin\_1 -4.1302 0.054 -77.031 0.000 -4.235 -4.025

fico\_bin\_2 -4.4342 0.050 -89.130 0.000 -4.532 -4.337

fico\_bin\_3 -4.7524 0.045 -105.568 0.000 -4.841 -4.664

fico\_bin\_4 -5.3609 0.047 -113.576 0.000 -5.453 -5.268

fico\_bin\_5 -6.1097 0.053 -114.285 0.000 -6.215 -6.005

fico\_bin\_6 -6.0722 0.041 -146.780 0.000 -6.153 -5.991

=================================================================================

Modeling MSA 26420

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 112284

Model: Logit Df Residuals: 112266

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6028

Time: 17:03:13 Log-Likelihood: -30917.

converged: True LL-Null: -77829.

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.3287 0.025 -13.303 0.000 -0.377 -0.280

Brwr Cnt -1.3797 0.025 -54.746 0.000 -1.429 -1.330

DTI 0.0217 0.001 15.660 0.000 0.019 0.024

Orig CLTV 0.0630 0.001 57.212 0.000 0.061 0.065

cnt\_120\_msa 0.0001 1.89e-05 5.699 0.000 7.07e-05 0.000

cnt\_120\_zip 0.0004 6.82e-05 6.103 0.000 0.000 0.001

constant 1.8139 0.141 12.896 0.000 1.538 2.090

FTMB\_Y -0.5582 0.030 -18.549 0.000 -0.617 -0.499

Loan Purp\_P -1.2172 0.030 -40.696 0.000 -1.276 -1.159

Loan Purp\_R -1.6514 0.045 -36.777 0.000 -1.739 -1.563

Occ Stat\_P -1.0451 0.037 -28.409 0.000 -1.117 -0.973

Occ Stat\_S -2.8370 0.184 -15.420 0.000 -3.198 -2.476

fico\_bin\_1 -3.4979 0.044 -79.720 0.000 -3.584 -3.412

fico\_bin\_2 -4.0372 0.043 -93.632 0.000 -4.122 -3.953

fico\_bin\_3 -4.2473 0.041 -104.780 0.000 -4.327 -4.168

fico\_bin\_4 -4.7962 0.044 -108.661 0.000 -4.883 -4.710

fico\_bin\_5 -4.9131 0.044 -111.668 0.000 -4.999 -4.827

fico\_bin\_6 -4.9893 0.033 -150.373 0.000 -5.054 -4.924

=================================================================================

Modeling MSA 40140

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 115022

Model: Logit Df Residuals: 115004

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6926

Time: 17:13:41 Log-Likelihood: -24504.

converged: True LL-Null: -79727.

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate -0.2810 0.033 -8.571 0.000 -0.345 -0.217

Brwr Cnt -1.5306 0.027 -56.941 0.000 -1.583 -1.478

DTI 0.0273 0.002 16.373 0.000 0.024 0.031

Orig CLTV 0.0690 0.001 59.946 0.000 0.067 0.071

cnt\_120\_msa 0.0114 0.000 60.833 0.000 0.011 0.012

cnt\_120\_zip 0.0007 0.000 4.163 0.000 0.000 0.001

constant -3.9367 0.147 -26.834 0.000 -4.224 -3.649

FTMB\_Y -0.5254 0.037 -14.239 0.000 -0.598 -0.453

Loan Purp\_P -0.9622 0.033 -29.513 0.000 -1.026 -0.898

Loan Purp\_R -2.5206 0.063 -39.982 0.000 -2.644 -2.397

Occ Stat\_P -1.6525 0.037 -44.913 0.000 -1.725 -1.580

Occ Stat\_S -3.1486 0.114 -27.715 0.000 -3.371 -2.926

fico\_bin\_1 -4.3786 0.057 -77.272 0.000 -4.490 -4.268

fico\_bin\_2 -4.2339 0.047 -90.376 0.000 -4.326 -4.142

fico\_bin\_3 -4.7139 0.045 -104.811 0.000 -4.802 -4.626

fico\_bin\_4 -5.1798 0.046 -112.571 0.000 -5.270 -5.090

fico\_bin\_5 -6.0718 0.058 -104.621 0.000 -6.186 -5.958

fico\_bin\_6 -5.5471 0.041 -136.818 0.000 -5.627 -5.468

=================================================================================

Modeling MSA 19820

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 94256

Model: Logit Df Residuals: 94238

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.7022

Time: 17:14:14 Log-Likelihood: -19453.

converged: True LL-Null: -65333.

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate 0.1310 0.034 3.896 0.000 0.065 0.197

Brwr Cnt -1.7435 0.038 -46.363 0.000 -1.817 -1.670

DTI 0.0316 0.002 19.636 0.000 0.028 0.035

Orig CLTV 0.0654 0.001 47.178 0.000 0.063 0.068

cnt\_120\_msa 0.0143 0.000 41.152 0.000 0.014 0.015

cnt\_120\_zip -0.0002 0.000 -0.450 0.653 -0.001 0.001

constant -3.8916 0.175 -22.186 0.000 -4.235 -3.548

FTMB\_Y -0.7254 0.037 -19.816 0.000 -0.797 -0.654

Loan Purp\_P -1.0188 0.038 -26.536 0.000 -1.094 -0.944

Loan Purp\_R -2.3809 0.064 -37.396 0.000 -2.506 -2.256

Occ Stat\_P -0.8537 0.054 -15.886 0.000 -0.959 -0.748

Occ Stat\_S -3.5797 0.417 -8.577 0.000 -4.398 -2.762

fico\_bin\_1 -3.8633 0.058 -66.632 0.000 -3.977 -3.750

fico\_bin\_2 -4.6430 0.058 -80.339 0.000 -4.756 -4.530

fico\_bin\_3 -4.9703 0.053 -93.200 0.000 -5.075 -4.866

fico\_bin\_4 -4.7814 0.050 -95.242 0.000 -4.880 -4.683

fico\_bin\_5 -5.7454 0.059 -97.559 0.000 -5.861 -5.630

fico\_bin\_6 -5.7450 0.046 -126.117 0.000 -5.834 -5.656

=================================================================================

Modeling MSA 42660

Logit Regression Results

==============================================================================

Dep. Variable: 60 in 24 No. Observations: 115704

Model: Logit Df Residuals: 115686

Method: MLE Df Model: 17

Date: Sun, 24 Apr 2022 Pseudo R-squ.: 0.6879

Time: 17:14:39 Log-Likelihood: -25031.

converged: True LL-Null: -80200.

Covariance Type: nonrobust LLR p-value: 0.000

=================================================================================

coef std err z P>|z| [0.025 0.975]

---------------------------------------------------------------------------------

Orig Int Rate 0.0083 0.031 0.267 0.789 -0.053 0.069

Brwr Cnt -1.3463 0.025 -52.929 0.000 -1.396 -1.296

DTI 0.0488 0.002 28.600 0.000 0.045 0.052

Orig CLTV 0.0628 0.001 52.371 0.000 0.060 0.065

cnt\_120\_msa 0.0157 0.000 46.873 0.000 0.015 0.016

cnt\_120\_zip 0.0040 0.000 11.189 0.000 0.003 0.005

constant -4.8079 0.148 -32.441 0.000 -5.098 -4.517

FTMB\_Y -0.4905 0.033 -14.868 0.000 -0.555 -0.426

Loan Purp\_P -1.0631 0.036 -29.278 0.000 -1.134 -0.992

Loan Purp\_R -1.6606 0.055 -30.455 0.000 -1.767 -1.554

Occ Stat\_P -1.6102 0.039 -41.059 0.000 -1.687 -1.533

Occ Stat\_S -3.6638 0.311 -11.768 0.000 -4.274 -3.054

fico\_bin\_1 -4.1964 0.059 -71.006 0.000 -4.312 -4.081

fico\_bin\_2 -4.7985 0.056 -85.944 0.000 -4.908 -4.689

fico\_bin\_3 -4.7486 0.048 -99.795 0.000 -4.842 -4.655

fico\_bin\_4 -5.0807 0.047 -108.502 0.000 -5.172 -4.989

fico\_bin\_5 -5.9107 0.053 -112.055 0.000 -6.014 -5.807

fico\_bin\_6 -5.8520 0.042 -138.426 0.000 -5.935 -5.769

=================================================================================

1. For all values by MSA interested readers should consult Appendix 2: MSA Level Summary Statistics [↑](#footnote-ref-1)
2. A full list of all columns with a short description is available in Appendix 1: Full Variable List [↑](#footnote-ref-2)
3. Current logic available at: <https://selling-guide.fanniemae.com/Selling-Guide/Origination-thru-Closing/Subpart-B3-Underwriting-Borrowers/Chapter-B3-5-Credit-Assessment/Section-B3-5-1-Credit-Scores/1032990981/B3-5-1-02-Determining-the-Representative-Credit-Score-for-a-Mortgage-Loan-09-01-2021.htm> [↑](#footnote-ref-3)
4. Available at: <https://singlefamily.fanniemae.com/media/9391/display> [↑](#footnote-ref-4)
5. San Francisco experienced 0 90-day delinquencies in period 6 so it was necessary to drop period 6 in that regression as well [↑](#footnote-ref-5)