Federal Reserve Bank of Philadelphia

Working Papers
Research Department

WP 21-12
March 2021
https://doi.org/10.21799/frbp.wp.2021.12

The Behavioral Relationship Between Mortgage Prepayment and Default

Arden Hall
Federal Reserve Bank of Philadelphia Supervision, Regulation, and Credit Department

Raman Quinn Maingi
New York University

Consumer Finance Institute

ISSN: 1962-5361

The Behavioral Relationship Between Mortgage Prepayment and Default

Arden Hall*
Federal Reserve Bank of Philadelphia

Raman Quinn Maingi
New York University

March 2021

Abstract

An implication of the dual trigger theory of default is that mortgage borrowers who experience an unexpected financial reverse will prepay their mortgage rather than default if their equity in the house is positive. We test this idea with a new data set created by matching mortgage servicing records and credit bureau records to classify prepayments by what happens subsequently. In particular, we can identify a subset of prepayments that seems consistent with the dual trigger theory. If the theory is correct, these prepayments should exhibit similarities to defaults in the data set rather than other prepayments. We test this idea and find that these prepayments are in fact more closely related to defaults than to other prepayments. However, our data also support a role for strategic default. Understanding these relationships may be critical in predicting mortgage default when house prices decline after a long period of increases. While our work is only a first step in this direction, we believe that a better understanding how prepayments may be driven by financial reverses would be valuable for participants in and regulators of mortgage markets.

Keywords: mortgage finance, prepayment, default, nested logit model

JEL Classification Codes: D12, G51, R21

Contact: Arden Hall, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106; arden.r.hall@gmail.com; Raman Quinn Maingi, rmaingi@stern.nyu.edu.

1. Introduction

The two widely cited theories of mortgage default rely on either option exercise or response to unpredictable events to describe the behavior (Vandell, 1995). *Ruthless default* (or *strategic default*) theory asserts that borrowers default whenever the gain from extinguishing the mortgage by putting the property back to the mortgagee exceeds the negative equity in the house plus the various costs of default to the borrower, such of loss of access to credit. Since these costs may be substantial, the theory predicts that the property value would be significantly below the balance on the mortgage (negative equity) before the borrower would default.

The alternative, the *dual trigger* theory, asserts that two factors must be present to induce default. The borrower must be in a negative equity position as in the ruthless default theory, but in addition, the borrower must have experienced an unexpected negative financial event (liquidity shock), which will prevent him from continuing to make mortgage payments. Because it is the liquidity shock that dictates behavior, the dual trigger theory results in two differences in predicted behavior relative to ruthless default theory: Borrowers will not default even when an option calculation would indicate it is optimal to do so, and borrowers will default when an option exercise is not optimal (i.e., when negative equity is relatively small and the gain from defaulting does not exceed the costs). In observed borrower behavior, these differences will be obscured if actual behavior is a blend of the two theories. In this alternative, if negative equity were small, borrowers would only default if they had experienced a liquidity shock but, as the gain to default increased, they would be increasingly likely to default in the absence of a liquidity shock.

There is another difference between the two models' predictions, although it has not received as much attention in the literature. The dual trigger theory asserts that the borrower receiving a liquidity shock must end her tenure in the house, since she is no longer able to make the required mortgage payments. But if her equity in the property were large enough, it would be preferable to sell the house and prepay the mortgage rather than default. So, the dual trigger theory makes a prediction about prepayment as well as default. Prepayment for this reason would not be driven by the motivations usually cited in the literature: mortgage call exercise or moving to optimize consumption of housing services. Ruthless default theory makes no predictions about prepayment.

If the dual trigger theory operates when home equity is positive, then understanding liquidity shock-driven prepayments would be important not only for predicting prepayment but also for predicting default as house prices vary. With stable home prices, an increase in the frequency of liquidity shocks among homeowners (caused by a recession-induced increase in job losses, for example) can be expected to increase the frequency of default. But if home prices simultaneously fall, the frequency of default will be multiplied as home equity declines. Similarly, the impact of an increase in the frequency of liquidity shocks on mortgage default would be attenuated if it were accompanied by an increase in home prices. An evolution like this may have occurred in the period before the financial crisis of 2008–2010. Here the driver of more frequent liquidity shocks was not changes in the economy, but changes in vulnerability to shocks within the population of mortgagors induced by the continuous loosening of credit standards in that period.[1] Increasingly loose underwriting standards that did not induce large increases in default frequency could be explained by an increase in liquidity shock-driven prepayments.

Identifying liquidity-shock induced prepayments is also important for understanding ruthless default. Suppose that an econometrician could observe mortgagee-level liquidity shock-induced prepayment. Suppose as well that ruthless default does not occur in the population. Then, if house prices are conditionally independent of liquidity shocks, loan-to-value (LTV) ratios should have zero predictive power in distinguishing between inaction and the combination of defaults and liquidity-shock induced prepayments. In other words, ruthless default is an important phenomenon to the extent that house prices are economically important in the first stage of a nested logit model. Our empirical results find an important role for LTV ratios in the first stage of the nested logit model, which seems to show that ruthless default is an important phenomenon in the data.

The dynamic process described previously cannot be accurately modeled by conventional competing risk default and prepayment models that do not distinguish among motives for prepayment. However, if borrowers who prepay because of a liquidity shock (who we will call *disguised defaulters*) could be identified, it should be possible to identify the factors that dispose both them and those who default to suffer liquidity shocks. This knowledge could improve lenders' risk management by measuring the extent to which prepayments in their portfolios will transform into defaults if house prices fall. This

[1] The financial crisis is a large subject that will not be covered here. For a history, see the Financial Crisis Inquiry Commission, 2011. Dokko, Keys, and Relihan (2019) presents a model of the interaction between financial innovation and house prices in the period leading to the financial crisis and reviews the academic literature on the subject.

paper takes a first step in this direction by identifying a superset of disguised defaulters and demonstrating that this group has more in common with defaulters than it has with those who prepay for other reasons.

This paper builds on our earlier paper on reasons for prepayment.[2] In that paper, we constructed a data set by matching mortgage-servicing records to credit bureau records. This permitted us to construct a disjoint set of prepayment outcomes: regular refinance, cash-out refinance, payoff, and move. We estimated a competing hazard model for these outcomes plus default on a sample of prime mortgages and demonstrated that the prepayment outcomes were driven by different factors and that this disaggregated model of prepayment behavior was significantly more accurate than the standard model based on a single overall prepayment outcome. The current paper makes an additional distinction among prepayment outcomes: between those who move, and subsequently take out a new mortgage, and those who move but do not take out a new mortgage. This second group is the superset of the dual trigger prepayers.

The following section describes the construction of the data set that permits the creation of several distinct prepayment outcomes. Section 3 then presents models of prepayment and default, some of with aggregated prepayment outcomes. Comparisons of these models demonstrate the relationship of the outcome including disguised defaults to the default outcome.

2. Data Set Construction and Outcome Definition

Empirical prepayment and default models are typically built using origination and servicing data from a mortgage lender. These data do not support the identification of the reason for mortgage prepayments, since the lender ceases collecting data once the mortgage is prepaid. To obtain the additional information needed, a mortgage data set was matched and merged with credit bureau data covering the period when the mortgage was active. Specifically, the Black Knight McDash Data (McDash) was the mortgage origination and servicing data used, and the Federal Reserve Bank of New York/Equifax Consumer Credit Panel (CCP) was the source for anonymized credit bureau data. Data were taken for mortgages originated between January 2005 and December 2011, and performance was observed until May 2012.

---

[2] See Hall and Maingi, 2019.

Mortgage origination and servicing and credit bureau data contain several variables that should indicate whether records are matched. For the data set we created, matches were based on five data elements: origination date, balance at origination, property zip code, payment amount, and status code for the latest record available from both sources. The matching of data sets is sometimes based on models that score matches based on their likelihood of being correct. We did not take this approach and instead insisted on exact matches. A consequence of our approach is that some matches were rejected because they were not one to one. That is, a single record in one data set could match multiple records in the other data set. In this circumstance, we dropped the observation, trading off sample size for accuracy of matches. A waterfall for the matching process and subsequent adjustments to the sample is given in Figure 1.

The credit bureau data provide several data fields useful for distinguishing among types of prepayment. These include:

- a (scrambled) address time series that allows moves to be detected;
- tradeline data with which a new mortgage following prepayment can be detected;
- an indicator for the death of the borrower;
- age of borrower; and
- information on any second mortgage from the same or another lender.

Based on these data elements, five types of prepayment were identified as shown in Table 1. (Table 1 also shows the other outcomes observable in the data: default (defined as becoming 90 days delinquent) and no change in status. Note that the Other outcome has features in common with disguised default — the borrower prepaid and changed addresses but did not buy a new house (as indicated by the lack of a new mortgage). While what happened in the outcome corresponds to what would happen in a disguised default, it does not follow that all Other prepayments are disguised defaults. There are several possible situations besides a liquidity shock that could lead a borrower to prepay and move but not buy a new house. These could include death, moving out of the country (in which case any new mortgage would not show up in American credit reports), moving into a (rental) retirement home, paying all cash for a new house, taking out a private mortgage, taking a job that provides housing, or divorce. There certainly must be other reasons as well. So, the Other outcome should be thought of as a superset of disguised defaulters.

Because the goal of this paper is to analyze the behavior of disguised defaulters, we tried to identify Other prepayments that were not disguised defaults in order to eliminate them from the sample. One such circumstance was the death: We eliminated records in which the prepayment coincided with the death of the borrower. We also expected that older people's moves from owning their own home to rental retirement communities would make up some fraction of Other prepayments, so we deleted mortgages in which the borrower was 65 or older at termination of the mortgage.[3]

Additional filtering and cleaning were applied to the data set. Observations in which there was not a complete time series on addresses were deleted, as were mortgages that were sold or transferred before termination. Observations with missing predictive variables were also deleted. To eliminate any confounding impact of product type on behavior, only 30-year fixed-rate mortgages were retained in the sample. Finally, the sample was divided into three groups based on credit quality: prime, near prime, and subprime, based on the classification introduced by Hancock et al. (2006). The final sample sizes for the three groups are given in Table 2, and outcome frequencies are given in Table 3.

3. Estimating a Model with All Prepayment Outcomes

Our earlier paper presented multinomial logit models for the probability that an outcome occurred in a particular month estimated for prime mortgages, in which the Move and Other outcomes were combined. We explored separating and combining outcomes and were able to show the outcomes we had identified indeed represent different behaviors with different predictors. Here we widen the analysis by distinguishing between the Move and Other outcomes and adding models for near-prime and subprime mortgages. An immediate question is whether the separation of the Move and Other outcomes is appropriate. Table 5 presents hypothesis tests for the Move/Other separation.[4] In each panel of the table, row A represents a model in which all outcomes are separate, and row B represents a model in which the Move and Other outcome are combined. The right-hand side of row B provides the chi-square test statistic and significance level for the hypothesis that the Move and Other outcomes should be combined. For all three credit quality groups, the hypothesis is rejected at a very high level. For completeness, row C for each panel represents a model in which all prepayment outcomes are

[3] Divorce also could cause Other prepayment, so it would be desirable to eliminate prepayments for that reason. However, credit bureaus do not record divorce. We attempted to identify divorces using data on the composition of households over time, which is available, but found that removing the likely divorces had little effect on modeling results, so we chose not to remove them.

[4] The exact specification of each model in terms of outcomes is defined by the formula in the column titled Model. An explanation of the notation is given in Appendix 1.

aggregated and the test statistic and significance level for the hypothesis that they should be combined. Not surprisingly, that hypothesis is also rejected for each credit quality group.

Additional hypothesis tests will be used to investigate the relationship between the Other and Default outcomes, but we should first investigate the models separating all outcomes. Estimates for models that separate all outcomes (including Move and Other) are presented in Tables 6, 7, and 8. Relative to the model presented in our earlier paper (Hall and Maingi, 2019), two things have changed:

- The Move outcome in the earlier paper contained all prepayments that were accompanied by a change in address. In this paper, those prepayments are split between a Move outcome in which a new mortgage is observed following the prepayment, and an Other outcome in which a new mortgage is not observed after the prepayment.
- The earlier paper only reported results for prime mortgages, while this paper also reports results for near-prime and subprime mortgages.

For the separation of Move and Other outcomes, the independence from irrelevant alternatives property of multinomial logit models guarantees that the coefficients for the other outcomes are the same whether the Move and Other outcomes are combined or kept separate. The estimated parameters for the model for the prime group with these two outcomes combined has been discussed in our earlier paper, so that analysis will not be repeated here. Instead, for the prime group, we focus on the differences in estimated parameters between the Move and Other outcomes.

As in the earlier paper, we compare splined variables with the aid of graphs of the part of the model linear form of the model generated by all of the splines associated with a particular variable. These graphs are given in Figures 2, 3, and 4. Comparing the lines representing Move and Other for LTV, Age of Loan, and Rate Incentive shows very little difference between the two relative to differences between these two outcomes versus other outcomes. Comparing coefficients for the nonsplined variable turns up a few differences. The coefficient for Credit Score is positive and significant for the Move outcome, while it is small and not significant for Other. The Other coefficient is very slightly negative, which makes it more similar to the Default coefficient (negative and significant) than to the coefficients for the other outcomes (positive and significant). There are differences by loan type as well. Loan type may be either Jumbo Conforming, Jumbo Nonconforming, or Conforming (the omitted category).

Relative to Conforming, Move is more likely (has a positive coefficient) for Jumbo Conforming and less likely for Jumbo Nonconforming. For the Other outcome, both coefficients are small and not significantly different from zero. Where the current mortgage was related to the purchase of a house, the coefficient for the Move outcome is negative, presumably because these borrowers have more recently adjusted their supply of housing services than a borrower whose mortgage originated as a refinance. The Default coefficient is also negative, but the Other coefficient is positive. The reason for this is unclear, but the magnitude of the impact of all three coefficients is small.

The coefficient for Borrower Age is negative for the Move outcome and positive for the Default outcome, while the Other coefficient is very small and not significantly different from zero. The last interesting relationship between Move and Other coefficients is the unemployment rate. A higher unemployment rate implies that the mortgage borrower is more likely to become unemployed and, if he becomes unemployed, to have a longer spell of unemployment. This should decrease the likelihood that the borrower will prepay to move and buy a new house, and in fact, the coefficient for the Move outcome is negative and significantly different from zero. Since unemployment is a cause of liquidity shocks, higher unemployment rates should increase the likelihood of Other prepayments, to the extent that they are disguised defaults, as well as Default outcomes. However, the coefficient for the Other outcome is negative and about the same size as the Move coefficient, while the Default outcome has the expected positive sign. This is surprising but may be a result of a time series correlation in the data. During the credit crisis, unemployment rates rose as house prices fell. So, more frequent and longer unemployment spells coincided with the decline in house prices which, under the dual trigger theory, shifted those affected away from prepayment and toward default.

We turn now to the near-prime and subprime models. While these models have different coefficients from the model for prime mortgages, the relationships between coefficients by outcome are generally similar. This can be seen in the graphs in Figures 2, 3, and 4 of the splines for LTV, age of loan, and rate incentive. Increasing LTV consistently increases the probability of the Default outcome and reduces the probability of the Payoff outcome across the three credit quality groups. LTV effects are smaller for other outcomes, but they appear consistent across the credit quality groups. For age of loan, and rate incentive as well, relationships across models for the three credit quality groups are consistent for outcome in which effects are large. For example, increasing rate incentive increases the probability of the Regular Refinance outcome across the entire range of the variable and is the most sensitive outcome in all credit quality groups. For the variables that were not splined, relationships of coefficients across outcomes are generally consistent for the three credit quality groups. Magnitudes of coefficients across credit quality groups are more variable, with subprime coefficients having smaller magnitudes than the corresponding prime or near prime coefficients about two-thirds of the time.

4. Testing for Disguised Default

The implication of the dual trigger theory for this data set is that some of those with the Other outcome are similar to those with the Default outcome in that their behavior was triggered by a liquidity shock. We are unable to observe that event directly. However, a testable consequence is that some of those with Other outcomes will exhibit similarities to those with Default outcomes, except in their reaction to changes in LTV ratio, in which the relationship should be opposite: Higher LTVs should increase the probability of the Default outcome relative to Other. A natural way to test this idea is to compare a model that embodies this relationship between Default and Other with one that does not.[5] If there were no relationship, then the appropriate model would be the single-stage model treating all outcomes as distinct, which was analyzed previously. The presence of a relationship between Default and Other can be represented by a nested model in which the two outcomes are combined and treated as distinct from other outcomes in the first stage, but are then separated in a second-stage model. Diagrams of these two models are shown in Figure 5. The diagram at the bottom of Figure 5 shows a third model. Here, the Move and Other outcomes are combined in the first stage and then separated in a second-stage model. This structure was considered because these two outcomes are related in that they are the only prepayment outcomes that involve a change in address. Further, there have been a series of papers analyzing models treating prepayment with a change of address as a distinct outcome.[6] This suggests an alternate hypothesis that is inconsistent with the dual trigger theory, and so it should be tested. Each of the models shown in Figure 5 have been estimated for each of the credit quality groups.

The nested models represent alternatives to, rather than restricted versions of, the single-stage model covering all outcomes. As a result, the hypothesis that the relationship embodied in the nested model is a better representation of the data than the single-stage model cannot be tested with a likelihood ratio test. However, because each model predicts for the same outcomes, they can be compared using the Akaike Information Criterion (AIC). These comparisons are provided in Table 9. The three panels cover

[5] This test is not precise because an unknown proportion of Other outcomes were driven by different reasons, and not a liquidity shock. However, these reasons are not likely to be related to those leading to Default. So, the impact of these Other outcomes not caused by a liquidity shock would be to dilute the relationship between Other and Default outcomes and bias the test against find a relationship.

[6] See Clapp et al. (2001); Clapp, Deng, and An (2006); and An, Clapp, and Deng (2010).

prime, near-prime, and subprime mortgages and provide the AICs for the single-stage model for all outcomes, the nested model combining Other and Default in the first stage, and the nested model combining Other and Move in the first stage. For each credit quality group, the single-stage model outperforms the nested model combining Other and Move in the first stage, but the nested model combining Other and Default in the first-stage outperforms both models. The differences in the AIC of the other models relative to the model combining Other and Default in the first stage are so large that it is extremely unlikely that either of the other models is a better representation of the data than the nested Other/Default model.[7]

The AIC results provide substantial support for the hypothesis that the disguised default is related to true default as hypothesized in the dual trigger theory. Since the Other category includes outcomes that are not disguised defaults and have no plausible relationship to default, these empirical tests are biased *against* finding a relationship. Further, the dual trigger theory hypothesizes, and results for the single-stage model reviewed previously, finds that LTV has the opposite impact on default and disguised default outcomes. But in the first stage of the nested Other/Default model, those two outcomes are combined and the combined outcome can only have single relationship to LTV. This is a model misspecification that is not found in the single-stage model. This limitation should also bias the AIC comparisons against finding a relationship between the Default and Other outcomes. Considering these biases, the AIC comparisons in Table 9 understate the strength of the true disguised default–default relationship.

5. Ruthless Default

That the dual trigger theory appears to explain some default does not imply that ruthless default never occurs. Some borrowers not facing a liquidity shock and capable of continuing to make mortgage payments may choose to default if the benefit is sufficiently high. We can test for the existence of this behavior using our nested model. The implication of the dual trigger theory is that the triggering event for either default or disguised default is a liquidity shock. The borrower's equity is thus only relevant to the decision whether to default or prepay. The implication for our model is that, were there no ruthless default, borrower equity would not affect the first-stage model and would only become predictive in the nested model distinguishing between default and disguised default. To test this hypothesis, we reran our first-stage model eliminating the LTV spline variables for the combined Default and Other outcome.

[7] See Burnham et al. (2001). They provide (p. 25) a standard: Models with differences greater than 20 have "essentially no empirical support" vis-à-vis the model with the lowest AIC.

These versions were then tested with likelihood ratio tests. The results are shown in Table 10. Tests for all three credit quality groups are highly significant, indicating that the current LTV affected borrows' decisions to choose the combined Other and Default group in the first stage. This is consistent with the ruthless default theory but not the dual trigger theory. These results support the idea that both theories have a role in explaining borrower behavior.

6. Implications

Our earlier paper demonstrated the substantial differences in the motivations for different types of mortgage prepayment. For example, borrowers who prepay simply to lower the interest rate on their mortgage show much more sensitivity to market interest rates than those who prepay to pay off their mortgages, while members of that group are much more sensitive to their current LTV. Behavioral differences like these might be of interest to investors in mortgages since they offer the potential for better prepayment prediction and therefore more accurate valuation of mortgages.

This paper provides empirical support for an aspect of the dual trigger theory of mortgage default: Some borrowers who suffer a liquidity shock prepay their mortgage rather than defaulting. The primary driver of the choice between prepayment and default is the amount of equity the borrower has in the house as indexed by the LTV. This relationship makes it difficult to sort out the pure impact of model variables related to the likelihood of a liquidity shock by modeling only default, or even modeling default jointly with a single overall prepayment outcome. For example, rising unemployment might reasonably be expected to increase the frequency of liquidity shocks, but if it is accompanied by rising house prices, the total effect of the unemployment rate on default will be muted. The effect will show up largely as increased prepayment, but it is not likely to be large, relative to the frequency of prepayment for other reasons, and so it may easily be missed. Similarly, relaxation of lending standards should introduce more borrowers at a higher risk of a liquidity shock into mortgage portfolios, but again, higher house prices will mute the impact on default and potentially mislead lenders.

Our results also show a role for ruthless default in the mortgage crisis. For all three credit grades, we find a strong, significant, and positive relationship between LTV ratio and the Other/Default nest in the first stage of the nested logit model. Assuming our coefficients are well identified,[8] this means having a higher LTV ratio causally increases the likelihood of entering the troubled nest. This cannot be explained

[8] This requires that house price shocks are, conditional on our other covariates, independent of liquidity shocks. Since we control for unemployment shocks, we think this is a reasonable benchmark.

by the dual trigger theory; only a liquidity shock should predict entry into the troubled nest. This evidence is best explained by ruthless default, where agents who do not receive a liquidity shock still elect to default due to the negative NPV of retaining their house. While this evidence does not directly quantify the frequency of ruthless default, the importance of ruthless default in the first stage of the nested logit suggests it was important during the crisis.[9]

The obvious follow-up question: "What are the relative strengths of these two explanations of default?" cannot be answered with our data. That would require a data set that accurately identified disguised defaults, not a superset of them. Unfortunately, such mortgage data sets do not currently exist. The proportion of disguised defaults among our Other outcomes is sufficiently high, and the impact on default is sufficiently strong that we are able to demonstrate the relationship, but demonstrating its existence is not the same as accurately measuring it. So, refinement in the definition of the Other outcome, either through clever use of credit bureau data or through matching to other data sources that provide information about events like job loss, extended illness, or serious financial loss could facilitate modeling that would accurately measure the factors behind the dual trigger theory.

Constructing this effective data set seems to us to be a difficult task, but the return from success would be high. First, mortgage lenders and risk managers could better price and manage mortgage credit risk. Given the size of the U.S. mortgage market, this would be an important improvement. Beyond that, accurately measuring the underlying frequency of liquidity shocks and its relationship to predictive variables and then the sliding relationship between prepayment and default mediated by the mortgagors home equity could provide a view of the dynamics of mortgage credit risk that is not available any other way. Lending practices in the prelude to the credit crisis seem to demonstrate that mortgage lenders and investors were not even aware of the mechanism described here. A better understanding by market participants and regulators ought to lead to better decisions and less risk in mortgage lending.

7. Conclusion

The hypothesis that some mortgage prepayments are driven by some of same factors driving mortgage default is implicit in the dual trigger theory of mortgage default, in which the primary driver of behavior is an unexpected liquidity shock. This idea is not testable with standard mortgage data sets. We present

[9] To use this evidence to quantify ruthless default, we would need to estimate a structural model of prepayment and default with the LTV coefficients as targeted moments. While this could be an interesting exercise, it goes beyond the scope of this paper, and we leave it for future work.

a new data set that is based on matched mortgage servicing and credit bureau records, which enables us to identify a fraction of prepayments that are could be driven by liquidity shocks. Unfortunately, there are other possible reasons that a prepayment might fall within the group we define, so we are left with a superset of the mortgages we would like to study. Despite this limitation, we are able to show that a nested model embodying the idea that the prepayment outcomes we have identified as having something in common with the default outcomes in our data set outperforms a model based on the assumption that there is no relationship. The result is consistent across the credit spectrum for mortgages. The model is also consistent with ruthless default, implying both types of behavior occur.

The implication of this work is that there is a subtle process at work that can, at times, affect the frequency of default in mortgage portfolios in an unexpected way. A period of increasing house prices will lead many mortgagors receiving an unexpected liquidity shock to prepay their mortgage rather than default. Thus, lending to borrowers who appear at greater risk of a liquidity shock may not cause large increases in default. But the opposite can happen when house prices decline — a liquidity shock that might have driven a prepayment now causes a default. These effects cannot be incorporated in standard competing hazard prepayment and default models. Developing mortgage data sets that enable the study of this relationship between prepayment and default would provide a better understanding of mortgage default of benefit to banks, bank regulators, and other investors in mortgages or mortgage-backed securities.

References

An, X., J. Clapp, and Y. Deng. (2010). "Omitted Mobility Characteristics and Property Market Dynamics: Application to Mortgage Termination." *Journal of Real Estate and Financial Economics* 41: 245–271.

Burnham, K., D. Anderson, and K. Huyvaert. (2011). "AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons." *Behavioral Ecology and Sociobiology* (65), 23–35.

Clapp, J., G. Goldberg, J. Harding, and M. LaCour-Little. (2001). "Movers and Shuckers: Interdependent Prepayment Decisions." *Real Estate Economics* 29(3), 411–450.

Clapp, J., Y. Deng, and X. An. (2006). "Unobserved Heterogeneity in Models of Competing Mortgage Termination Risks." *Real Estate Economics* 34(2), 243–273.

Dokko, J., B. Keys, and L. Relihan. (2019). "Affordability, Financial Innovation, and the Start of the Housing Boom," Federal Reserve Bank of Chicago Working Paper 2019-01, 2019.

Financial Crisis Inquiry Commission. (2011). "Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States," 2011. https://www.govinfo.gov/content/pkg/GPO-FCIC/pdf/GPO-FCIC.pdf.

Hall, A., and R. Maingi.(2019). "The Mortgage Prepayment Decision: Are There Other Motivations Beyond Refinance and Move?" Federal Reserve Bank of Philadelphia Working Paper 19-39, 2019.

Hancock, D., A. Lehnert, W. Passmore, and S. Sherlund. (2006). "The Competitive Effects of Risk-Based Bank Capital Regulation: An Example from U.S. Mortgage Markets." *Federal Reserve Board, Finance and Economics Discussion Series* 2006-46.

Vandell, K. D. (1995). "How Ruthless Is Mortgage Default? A Review and Synthesis of the Evidence." *Journal of Housing Research* 6(2), 245–264.

Appendix 1

To present the results, of our hypothesis tests, it will be helpful to use a standardized notation for describing a model. Following An, Clapp, and Deng (2010), outcomes will be represented by letters, as indicated in Table 1. An estimated model will be represented by a set of letters in circumflexes, so:

(1) {R,C,P,M,O,D,xN},

represents a model estimated with each of the five termination types treated as a distinct outcome and the omitted category indicated by an "x" preceding it. (The code N indicates the outcome that the loan was not terminated.) Aggregation of outcomes is indicated by parentheses, so:

(2) {(R,C,P,M,O),D,xN},

represents the standard two terminal state competing hazard model in which all types of prepayment are aggregated. Finally, nested models are indicated by the addition of notation describing the second-stage model, so:

(3) {(R,C,P,M,O),D,xN} + {R,C,P,M,xO},

represents a model in which all types of prepayment are aggregated in the first stage, then disaggregated in the second-stage nested model, with M (move) as the omitted category in the second-stage model.

It is important to note that the model in (2) differs from those in (1) and (3) in a fundamental way: It does not provide an estimated probability for each outcome. We will call the model in (2) *incomplete* to distinguish it from the *complete* models in (1) and (3). Likelihood functions for incomplete models are not comparable to those for complete models because they do not cover the same set of outcomes. Thus, model (2) is not a restricted version of model (1) or (3), and the hypothesis that outcomes aggregated in (2) are indistinguishable cannot be tested using their estimated likelihood functions. The solution to this dilemma is to complete model (2) in a way that is consistent with the hypothesis. If the outcomes aggregated in model (2) are indistinguishable under the hypothesis, then they must be randomly assigned and equally likely. Our notation for that uses square brackets, so:

(4) [R,C,P,M,O],

represents a second-stage model that randomly assigns the four outcomes. So a complete version of model (2) would be:

(5) {(R,C,P,M,O),D,xN} + [R,C,P,M,O].

Figure 1

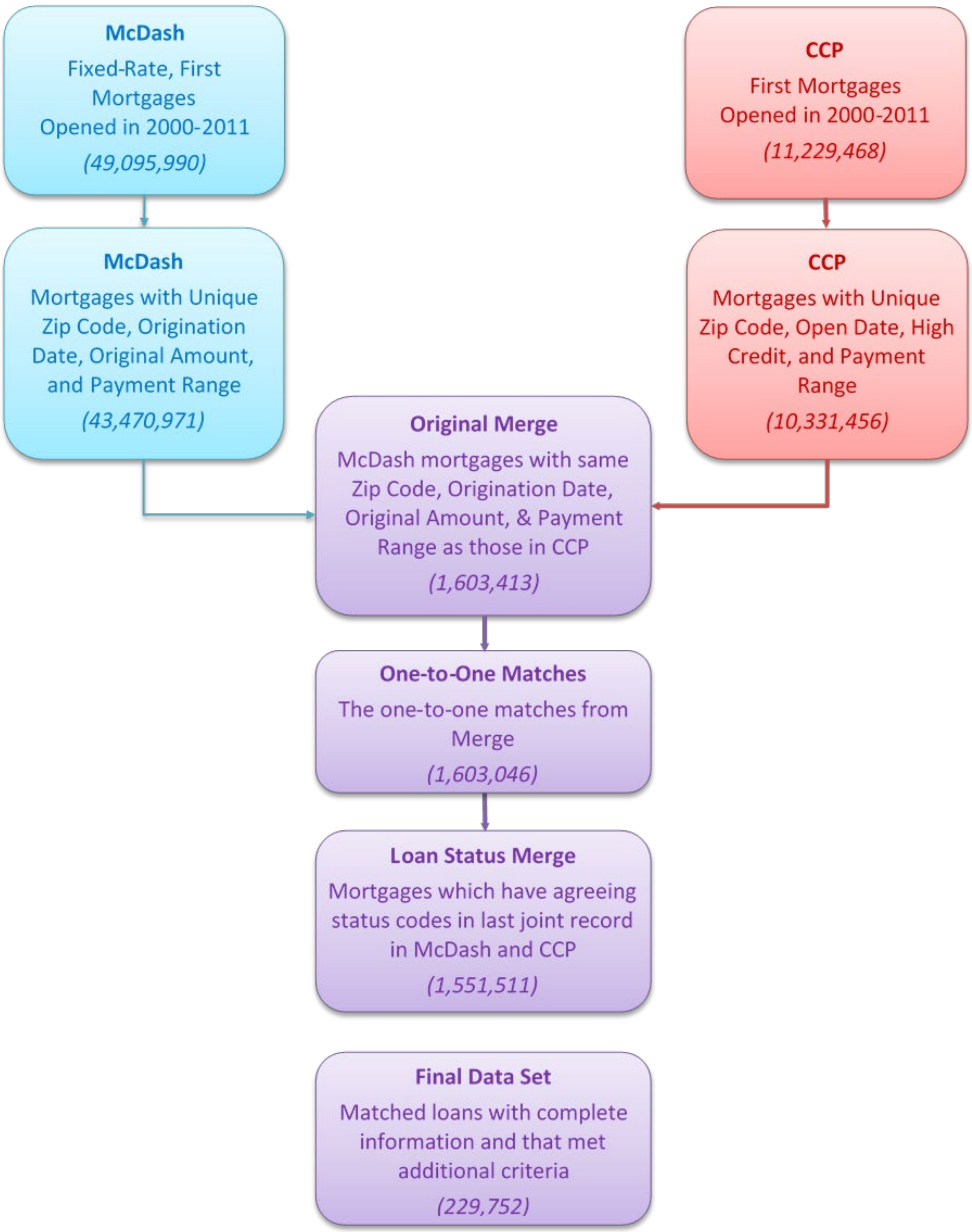

McDash
Fixed-Rate, First Mortgages Opened in 2000-2011
(49,095,990)
McDash
Mortgages with Unique Zip Code, Origination Date, Original Amount, and Payment Range
(43,470,971)
CCP
First Mortgages Opened in 2000-2011
(11,229,468)
CCP
Mortgages with Unique Zip Code, Open Date, High Credit, and Payment Range
(10,331,456)
Original Merge
McDash mortgages with same Zip Code, Origination Date, Original Amount, & Payment Range as those in CCP
(1,603,413)
One-to-One Matches
The one-to-one matches from Merge
(1,603,046)
Loan Status Merge
Mortgages which have agreeing status codes in last joint record in McDash and CCP
(1,551,511)
Final Data Set
Matched loans with complete information and that met additional criteria
(229,752)

Figure 2

LTV Splines

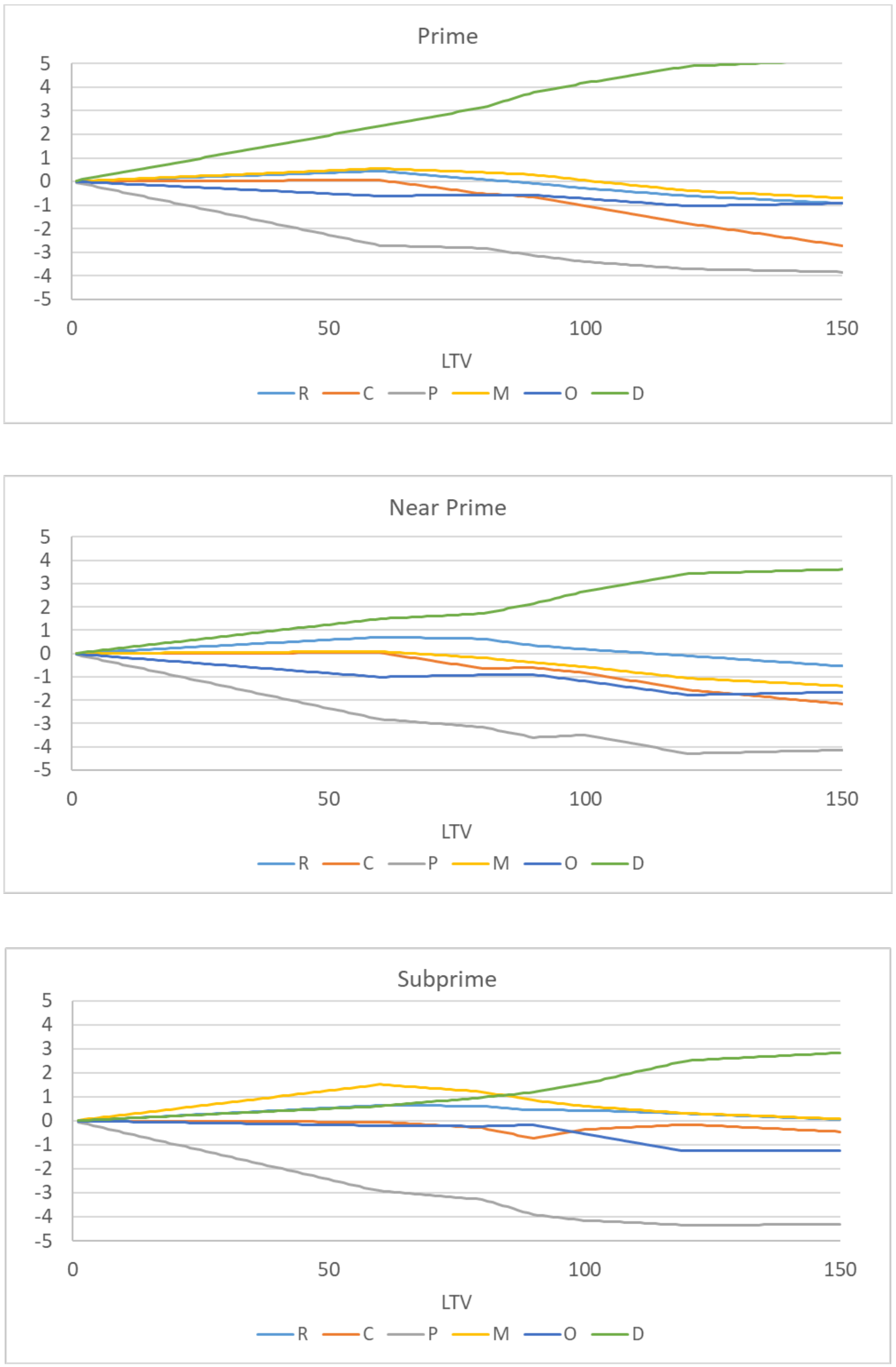
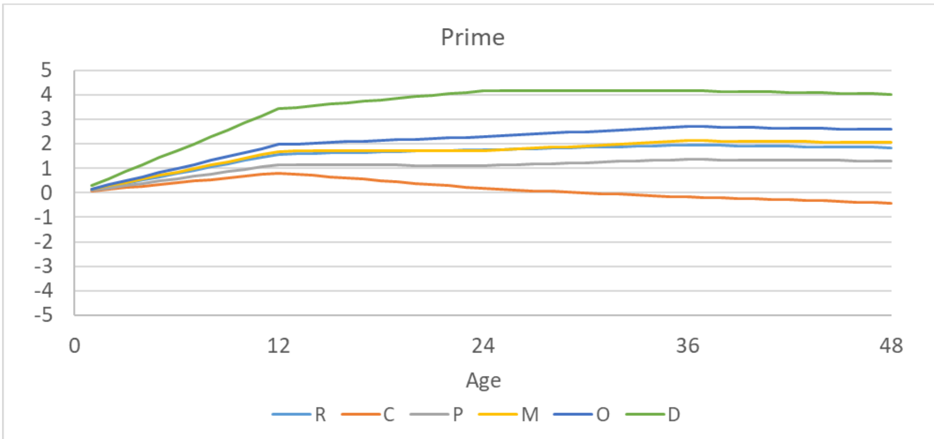
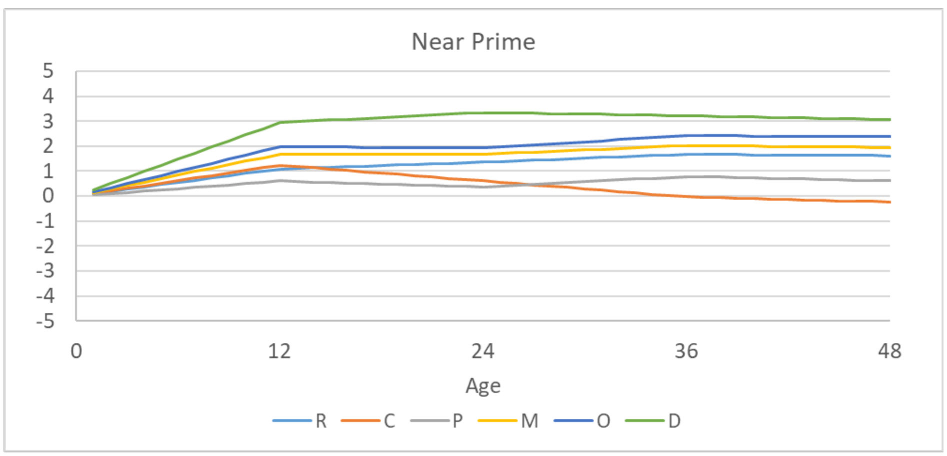


Figure 3

Age Splines





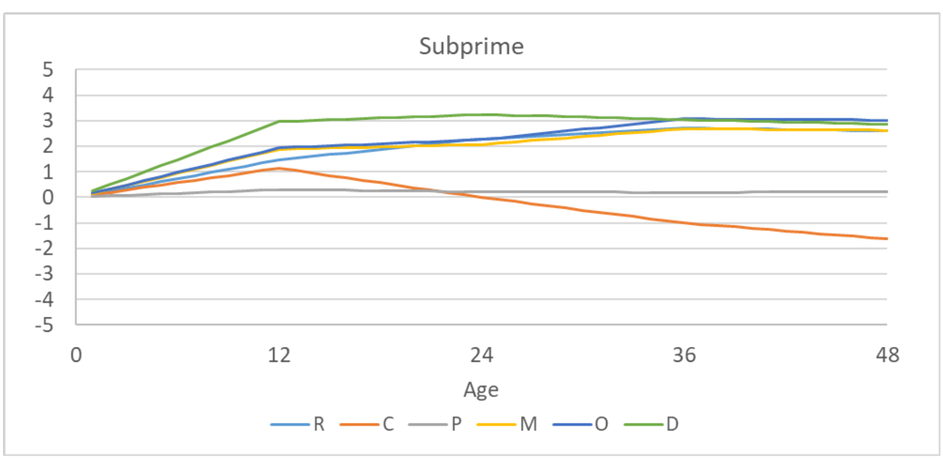
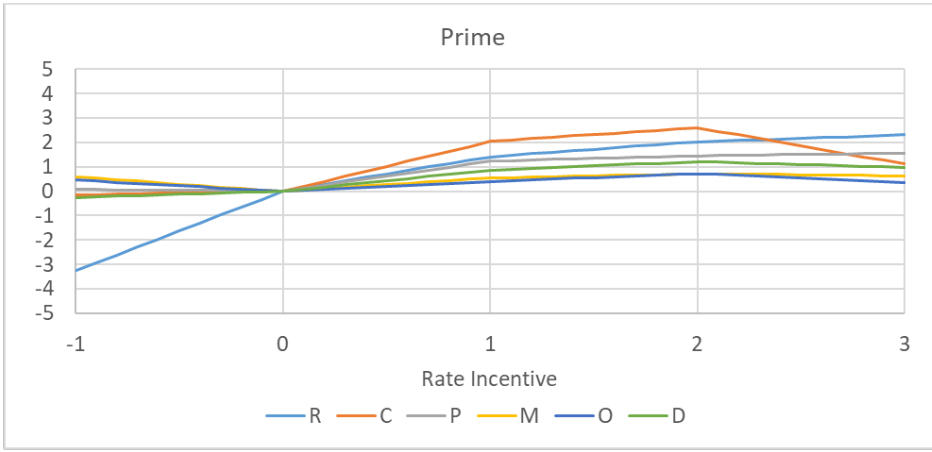
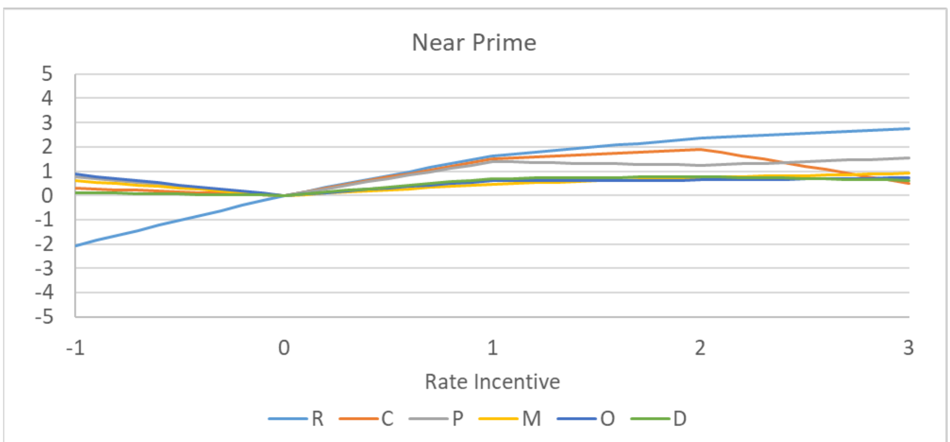


Figure 4

Rate Incentive Splines





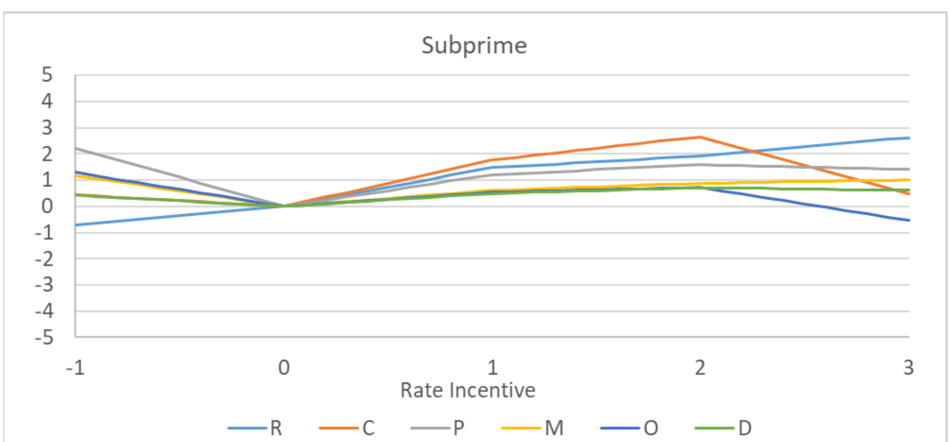
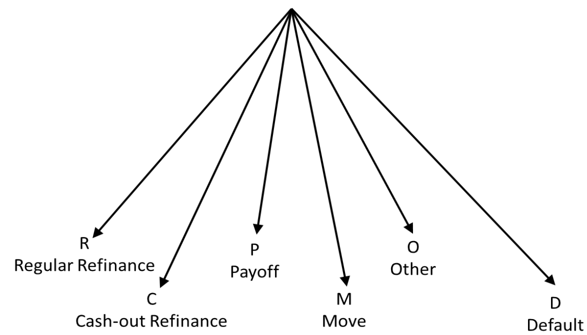
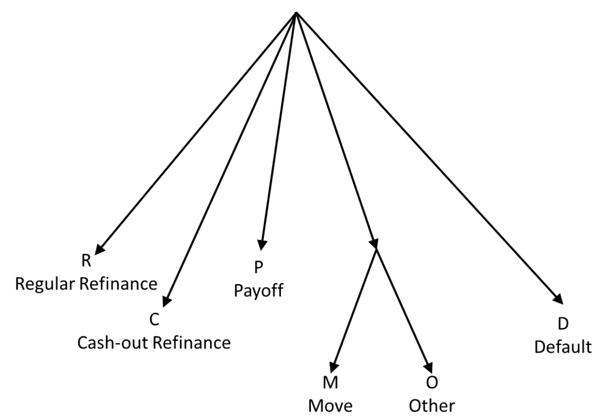


Figure 5

Non-nested Model



Nested Model – Move and Other



Nested Model – Other and Default

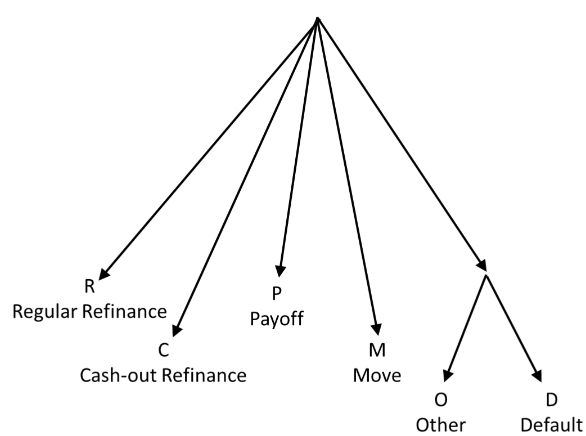


Table 1

Outcome Definitions

Outcome	Code	Definition
Regular Refinance	R	Prepay, take out a new first mortgage for the balance of the prepaid mortgage and do not change address
Cash-out Refinance	C	Prepay, take out a new first mortgage for a larger amount than the prepaid mortgage and do not change address
Payoff	P	Prepay, do not take out a new first mortgage and do not change address
Move	M	Prepay, take out a new first mortgage and change address
Other	O	Prepay, do not take out a new first mortgage, and change address
Default	D	Fall 90 days behind in mortgage payments
No Change	N	Do not change status

Table 2

Loan Count by Credit Grade

Credit Grade	Number	Frequency
Prime	222,349	45.55%
Near Prime	110,034	22.54%
Subprime	155,754	31.91%
Total	488,137	

Table 3

Observations and Outcomes by Loan-Month

		Prime		Near Prime		Subprime	
Outcome	**Code**	**Number**	**Frequency**	**Number**	**Frequency**	**Number**	**Frequency**
Regular Refinance	R	50,430	0.74%	15,155	0.43%	9,810	0.23%
Cashout Refinance	C	26,132	0.38%	10,899	0.31%	12,284	0.29%
Payoff	P	6,720	0.10%	2,537	0.07%	1,710	0.04%
Move	M	8,915	0.13%	3,729	0.11%	2,892	0.07%
Other	O	3,269	0.05%	1,628	0.05%	1,489	0.03%
Default	D	6,596	0.10%	10,364	0.29%	15,264	0.36%
No Change	N	6,750,633	98.51%	3,506,185	98.75%	4,241,217	98.99%
Total		6,852,695		3,550,497		4,284,666	

Table 4: Predictive Variables

Name	Definition
LTV, LTV > 60, LTV > 80, LTV > 90, LTV > 100, LTV > 120	Current combined loan-to-value ratio lagged 5 months (splined with knot points at 60%, 80%, 90%,100%, and 120%)
Credit Score	Credit score at origination
Loan Age, Loan Age > 12, Loan Age > 24, Loan Age > 36	Age of loan (splined with knot points at 12, 24, and 36 months)
Rate Incentive, Rate Incentive > 0, Rate Incentive > 1, Rate Incentive > 2	Interest rate incentive to prepay (rate at origination minus current rate) lagged 2 months
Spread at Origination	Mortgage coupon spread at origination
Low Doc	Low-doc Indicator
No Doc	No-doc Indicator
Jumbo Conforming	Jumbo conforming indicator
Jumbo Nonconforming	Jumbo nonconforming indicator
Wholesale	Wholesale channel indicator
Correspondent	Correspondent channel indicator
Purchase	Purchase mortgage indicator
Borrower Age	Borrower age at termination
Second Mortgage	Second mortgage indicator
Unemployment Rate	County unemployment rate lagged 2 months[10]

[10] Source: Bureau of Labor Statistics, Local Area Unemployment Statistics

Table 5

Likelihood Ratio Tests for Combining Move and Other Outcomes and for Combining All Prepayment Outcomes

	Model	Description	-2log(likelihood)	Degrees of Freedom	H_0	Additional likelihood term to represent H_0	-2log(likelihood) under H_0	Likelihood ratio test statistic	Likelihood ratio test degrees of freedom	Likelihood ratio test significance level
	Prime									
A	**{**R,C,P,M,O,D,xN**}**	Completely disaggregated model	1,236,369	156						
B	{R,C,P,(M,O),D,xN}+[M,O]	Model with Move and Other	1,222,656	130	The Move and Other outcomes can be combined	16,891	1,239,546	3,178	26	0.000
C	{(R,C,P,M,O),D,xN}+[R,C,O,P,M,O]	hazard prepayment/default model	1,026,598	52	All prepayment outcomes can be combined	307,293	1,333,891	97,522	104	0.000
	Near Prime									
A	**{**R,C,P,M,O,D,xN**}**	Completely disaggregated model	564,467	156						
B	{R,C,P,(M,O),D,xN}+[M,O]	Model with Move and Other	557,988	130	The Move and Other outcomes can be combined	14,368	572,355	7,888	26	0.000
C	{(R,C,P,M,O),D,xN}+[R,C,O,P,M,O]	Standard competing hazard prepayment/default model	488,160	52	All prepayment outcomes can be combined	109,274	597,435	32,967	104	0.000
	Subprime									
A	**{**R,C,P,M,O,D,xN**}**	Completely disaggregated model	560,413	156						
B	{R,C,P,(M,O),D,xN}+[M,O]	Model with Move and Other	554,921	130	The Move and Other outcomes can be combined	21,160	576,082	15,669	26	0.000
C	{(R,C,P,M,O),D,xN}+[R,C,O,P,M,O]	Standard competing hazard prepayment/default model	501,665	52	All prepayment outcomes can be combined	90,724	592,389	31,977	104	0.000

Table 6

Seven Outcome Model for Prime Mortgages

	Outcomes					
	R	C	P	M	O	D
Intercept	-13.525***	-6.897***	-6.810***	-9.650***	-9.269***	-6.430***
LTV	0.007***	0.001	-0.045***	0.009***	-0.010***	0.039***
LTV > 60	-0.026***	-0.030***	0.039***	-0.017***	0.013***	0.000
LTV > 80	0.003	0.015***	-0.025***	-0.001	-0.004	0.026***
LTV > 90	-0.006	-0.024***	0.004	-0.015	-0.013	-0.026**
LTV > 100	0.006	0.001	0.012	0.002	-0.002	-0.004
LTV > 120	0.005	0.006	0.011	0.012	0.020*	-0.027***
Loan Age	0.132***	0.068***	0.096***	0.141***	0.165***	0.286***
Loan Age > 12	-0.117***	-0.121***	-0.101***	-0.140***	-0.138***	-0.225***
Loan Age > 24	0.001	0.024***	0.026***	0.034***	0.008	-0.060***
Loan Age > 36	-0.025***	0.009***	-0.027***	-0.042***	-0.045***	-0.012***
Rate Incentive	3.265***	0.149***	-0.075	-0.585***	-0.457***	0.257
Rate Incentive > 0	-1.851***	1.895***	1.311***	1.127***	0.864***	0.610***
Rate Incentive > 1	-0.804***	-1.486***	-1.025***	-0.371***	-0.097	-0.522***
Rate Incentive > 2	-0.328***	-2.045***	-0.100	-0.254	-0.658**	-0.579***
Credit Score	0.009***	0.002***	0.001***	0.004***	0.000	-0.013***
Spread at Origination	-0.587***	-1.011***	-0.558***	-0.199***	-0.394***	-0.727***
Low Doc	-0.406***	-0.301***	-0.204***	-0.076	0.034	0.421***
No Doc	0.155***	0.223***	0.080*	0.220***	0.318***	0.102**
Jumbo Conforming	1.125***	-0.208***	0.012	0.324***	-0.046	-1.013***
Jumbo Non-conforming	0.218***	-0.763***	0.034	-0.247***	-0.078	0.004
Wholesale	-0.021*	0.280***	0.270***	-0.142***	-0.098**	0.296***
Correspondent	-0.307***	-0.108***	-0.222***	-0.442***	-0.419***	-0.112***
Purchase	-0.168***	0.047***	0.383***	-0.073***	0.200***	-0.163***
Borrower Age	-0.014***	-0.009***	0.001	-0.032***	0.001	0.019***
Second Mortgage	0.480***	0.903***	0.270***	0.364***	0.249***	-0.288***
Unemployment Rate	0.016***	-0.094***	-0.037***	-0.063***	-0.067***	0.033***

* Significant at the 10% level against H_0: coefficient = 0

** Significant at the 5% level against H_0: coefficient = 0

*** Significant at the 1% level against H_0: coefficient = 0

Table 7

Seven Outcome Model for Near-Prime Mortgages

	Outcomes					
	R	C	P	M	O	D
Intercept	-13.820***	-5.555***	-4.844***	-8.905***	-8.264***	-5.414***
LTV	0.012***	0.001	-0.047***	0.002	-0.017**	0.025***
LTV > 60	-0.016***	-0.034***	0.030***	-0.015*	0.021*	-0.014*
LTV > 80	-0.025***	0.036***	-0.028***	-0.007	-0.004	0.033***
LTV > 90	0.013**	-0.024***	0.057***	0.002	-0.029	0.009
LTV > 100	0.002	-0.014	-0.053***	-0.007	-0.002	-0.015***
LTV > 120	-0.001	0.016	0.045***	0.014	0.034***	-0.032***
Loan Age	0.091***	0.103***	0.050***	0.140***	0.164***	0.246***
Loan Age > 12	-0.068***	-0.156***	-0.070***	-0.140***	-0.168***	-0.213***
Loan Age > 24	0.003	-0.001	0.054***	0.030***	0.044***	-0.043***
Loan Age > 36	-0.030***	0.037***	-0.049***	-0.035***	-0.045***	-0.002
Rate Incentive	2.061***	-0.295***	-0.766***	-0.607***	-0.870***	-0.107
Rate Incentive > 0	-0.429**	1.790***	2.152***	1.067***	1.480***	0.805***
Rate Incentive > 1	-0.893***	-1.076***	-1.529***	-0.184	-0.589***	-0.620***
Rate Incentive > 2	-0.358***	-1.831***	0.439	-0.100	0.072	-0.235*
Credit Score	0.008***	0.000*	0.000	0.004***	0.001***	-0.010***
Spread at Origination	-0.226***	-0.516***	-0.192***	-0.025	-0.030	-0.337***
Low Doc	-0.416***	-0.494***	-0.646***	-0.143	-0.053	0.621***
No Doc	0.086***	0.291***	0.066	0.186***	0.286***	0.189***
Jumbo Conforming	1.241***	0.281***	-0.397*	0.397***	-0.486*	-0.177
Jumbo Non-conforming	0.001	-0.664***	0.183*	-0.176**	-0.177	0.081
Wholesale	-0.135***	0.237***	0.493***	-0.039	-0.062	0.328***
Correspondent	-0.289***	-0.050**	-0.248***	-0.358***	-0.401***	0.041*
Purchase	-0.116***	0.004	0.221***	-0.182***	0.011	-0.178***
Borrower Age	-0.010***	-0.005***	-0.003	-0.034***	-0.018***	0.014***
Second Mortgage	0.725***	0.971***	0.531***	0.599***	0.382***	-0.081**
Unemployment Rate	-0.005	-0.115***	-0.055***	-0.105***	-0.105***	0.017***

* Significant at the 10% level against H_0: coefficient = 0

** Significant at the 5% level against H_0: coefficient = 0

*** Significant at the 1% level against H_0: coefficient = 0

Table 8

Seven Outcome Model for Subprime Mortgages

	Outcomes					
	R	C	P	M	O	D
Intercept	-13.896***	-5.507***	-4.321***	-10.822***	-9.128***	-4.784***
LTV	0.011	-0.001	-0.049***	0.025	-0.003	0.010
LTV > 60	-0.012	-0.011	0.031**	-0.040	0.002	0.008
LTV > 80	-0.016	-0.032***	-0.046**	-0.021	0.008	0.003
LTV > 90	0.017*	0.081***	0.039*	0.011	-0.045**	0.015
LTV > 100	-0.007	-0.027***	0.014	0.010	0.003	0.012***
LTV > 120	-0.001	-0.020***	0.011	0.007	0.036***	-0.037***
Loan Age	0.122***	0.095***	0.025*	0.157***	0.162***	0.246***
Loan Age > 12	-0.053***	-0.191***	-0.033*	-0.141***	-0.135***	-0.222***
Loan Age > 24	-0.034***	0.011*	0.005	0.035***	0.041***	-0.042***
Loan Age > 36	-0.045***	0.033***	0.005	-0.056***	-0.072***	0.003
Rate Incentive	0.697***	-0.433***	-2.219***	-1.181***	-1.297***	-0.428***
Rate Incentive > 0	0.784***	2.207***	3.428***	1.789***	1.830***	0.924***
Rate Incentive > 1	-1.042***	-0.893***	-0.829***	-0.337**	-0.347*	-0.280***
Rate Incentive > 2	0.265**	-3.068***	-0.552	-0.145	-1.439***	-0.322**
Credit Score	0.007***	-0.002***	0.000	0.004***	0.002***	-0.010***
Spread at Origination	-0.222***	-0.430***	-0.145***	0.018	-0.084*	-0.285***
Low Doc	-0.447***	-1.160***	-0.672*	0.007	0.408	0.337***
No Doc	0.449***	1.019***	0.460***	0.659***	0.857***	0.588***
Jumbo Conforming	1.404***	0.767***	-0.598*	0.402**	0.106	0.013
Jumbo Non-conforming	0.570***	0.103	-0.004	0.103	0.440	0.201*
Wholesale	0.113***	0.344***	0.461***	0.057	0.159**	0.222***
Correspondent	-0.471***	-0.278***	-0.220***	-0.611***	-0.454***	-0.015
Purchase	-0.174***	-0.013	0.401***	-0.097**	0.333***	-0.002
Borrower Age	-0.003***	0.005***	-0.002	-0.033***	-0.038***	0.012***
Second Mortgage	0.683***	0.950***	0.724***	0.866***	0.458***	0.057
Unemployment Rate	-0.029***	-0.082***	-0.106***	-0.124***	-0.139***	-0.002

* Significant at the 10% level against H_0: coefficient = 0

** Significant at the 5% level against H_0: coefficient = 0

*** Significant at the 1% level against H_0: coefficient = 0

Table 9

Model Comparisons

Model Description	First Stage	First Stage -2log(likelihood)	First Stage Number of Parameters	Second Stage	Second Stage -2log(likelihood)	Second Stage Number of Parameters	Total -2log(likelihood)	Total Number of Parameters	Akaike Information Criterion (AIC)	Difference from Mininum AIC
Prime										
Single Stage Model	{R,C,P,M,O,D,xN}	1,236,369	156	None	-	-	1,236,369	156	1,236,681	586
Nested Model Combining Move and Other	{R,C,P,(M,O),D,xN}	1,222,656	130	{M,xO}	13,721	26	1,236,377	156	1,236,689	595
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN}	1,228,346	130	{xO,D}	7,436	26	1,235,782	156	1,236,094	-
Near Prime										
Single Stage Model	{R,C,P,M,O,D,xN}	564,467	156	None	-	-	564,467	156	564,779	657
Nested Model Combining Move and Other	{R,C,P,(M,O),D,xN}	557,988	130	{M,xO}	6,471	26	564,458	156	564,770	648
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN}	557,444	130	{xO,D}	6,366	26	563,810	156	564,122	-
Subprime										
Single Stage Model	{R,C,P,M,O,D,xN}	560,413	156	None	-	-	560,413	156	560,725	307
Nested Model Combining Move and Other	{R,C,P,(M,O),D,xN}	554,921	130	{M,xO}	5,493	26	560,415	156	560,727	309
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN}	552,974	130	{xO,D}	7,131	26	560,105	156	560,417	-

Table 10

LTV Hypothesis Tests

Prime										
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN} includes LTV	1,228,346	130	{xO,D}	7,436	26	1,235,782	156	2,957	0
Nested Model Combining Other and Default	{(R,C,P,M,O) ,D,xN} excludes LTV	1,231,304	124	{xO,D}	7,436	26	1,238,740	150	-	6
Near Prime										
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN} includes LTV	557,444	130	{xO,D}	6,366	26	563,810	156	3,012	0
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN} excludes LTV	560,456	124	{xO,D}	6,366	26	566,822	150	-	6
Subprime										
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN} includes LTV	552,974	130	{xO,D}	7,131	26	560,105	156	2,989	0
Nested Model Combining Other and Default	{R,C,P,M,(O,D),xN} excludes LTV	555,963	124	{M,xO}	7,131	26	563,094	150	-	6