Federal Deposit Insurance Corporation • Center for Financial Research

Working Paper Series

Why Do Models that Predict Failure Fail?

Hua Kiefer
Federal Deposit Insurance Corporation

Tom Mayock
UNC Charlotte

August 2020

FDIC CFR WP 2020-05

fdic.gov/cfr

Why Do Models That Predict Failure Fail?*

Hua Kiefer† Tom Mayock‡

August 6, 2020

Abstract

In the first portion of this paper, we utilize millions of loan-level servicing records for mortgages originated between 2004 and 2016 to study the performance of predictive models of mortgage default. We find that the logistic regression model – the traditional workhorse for consumer credit modeling – as well as machine learning methods can be very inaccurate when used to predict loan performance in out-of-time samples. Importantly, we find that this model failure was not unique to the early-2000s housing boom.

We use the Panel Study of Income Dynamics in the second part of our paper to provide evidence that this model failure can be attributed to intertemporal heterogeneity in the relationship between variables that are frequently used to predict mortgage performance and the realized post-origination path of variables that have been shown to trigger mortgage default. Our findings imply that model instability is a significant source of risk for lenders, such as financial technology firms, that rely heavily on predictive statistical models and machine learning algorithms for underwriting and account management.

Key Words: Mortgages, Predictive Modeling, Machine Learning, Fintech, Lending, Lucas Critique

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Deposit Insurance Corporation.

†Federal Deposit Insurance Corporation. Email: hkiefer@fdic.gov.

‡Department of Economics, UNC Charlotte. Email: tmayock@uncc.edu.

1 Introduction

Recent changes in the regulatory environment have opened up the possibility of nearly completely automated mortgage underwriting. The long-run financial viability of these automated underwriting platforms would hinge critically on the out-of-sample predictive accuracy of statistical models and machine learning algorithms that are used in the application decisioning process. The literature on the predictive performance of such models is limited to two studies [Rajan, Seru and Vig, 2010; An et al., 2012] that argued that predictive models of mortgage performance will be poor because of factors related to the "Lucas critique" [Lucas, 1983]. Both of these studies suffer from important limitations. First, both of these analyses were limited to loans that were securitized into private-label mortgage-backed-securities (PLMBS); most of these loans were high-risk and classified as "subprime" or "Alt-A." The PLMBS market largely disappeared during the financial crisis and has yet to recover. That said, it is not clear whether the results from studies that rely solely on PLMBS data are generalizable to the post-crisis mortgage market. A second limitation of existing work is that neither of the aforementioned papers studied the predictive accuracy of machine learning methods which are becoming increasingly popular with financial technology firms ("Fintechs") [Frame, Wall and White, 2018; Jagtiani and Lemieux, 2018] and more generally in consumer credit modeling [Khandani, Kim and Lo, 2010]. Lastly, and perhaps most importantly, previous work has provided little evidence regarding *why* models fail.[1,2]

We address all of these limitations in this study. First, we use data on more than a decade of mortgage originations from across the credit spectrum to provide the first evidence on the predictive stability of mortgage default models throughout the credit cycle. We find that performance of both logistic regression models – the traditional workhorse in consumer credit risk applications – and machine learning methods deteriorates rapidly when used to predict out-of-sample loan performance. To our knowledge, our analysis is the first test of whether the Lucas critique in the context of mortgage modeling also applies to machine learning models. Importantly, our results provide evidence that predictive models are unstable even outside of regimes where private-label securitization was common in the mortgage market. PLMBS issuance collapsed around 2008 and has yet to return to any significant level. Focusing on the post-2008 mortgage origination cohorts thus provides the first evidence on whether the poor performance of predictive default models was a phenomenon unique to the 2000s credit cycle.

After establishing the extent to which traditional statistical models and machine learning methods fail at predicting out-of-time mortgage performance, we posit that predictive models of mortgage performance will be inherently unstable because of intertemporal changes in the relationship between variables used in predictive credit risk models, borrower expectations for the future path of variables known to trigger mortgage delinquency, and realized changes in the economic environment.[3] We provide support for this hypothesis using data on household finances derived from the Panel Study of Income Dynamics (PSID).[4] In the last part of our analysis, we study whether using training data drawn from a mix of economic conditions could potentially alleviate model stability issues; we find that using through-the-cycle data generally reduces instability in the Logit models but has little effect on the performance of the random forest classifier.

The rest of the paper is structured as follows. Section 2 establishes some context for our analysis by discussing the role of predictive models, automation, and model risk in the U.S. mortgage market. In Section

[1] Rajan, Seru and Vig [2010] argue that model failure in the early 2000s was driven by an expansion of private-label securitization. In the context of the present study, this explanation is unsatisfying as we provide evidence of model failure in an era where private-label securitization was virtually non-existent.

[2] An et al. [2012] use a sample of fixed-rate subprime mortgages that were originated in 2003 and securitized into PLMBS to estimate models of mortgage default. The authors used the estimated parameters from this model to predict mortgage defaults for a cohort of subprime loans that were originated in 2006 and found that their model underpredicted defaults by roughly 40 percent. Rajan, Seru and Vig [2015] conduct a similar exercise also using PLMBS data and find that a statistical model estimated using loans issued between 1997 and 2000 systematically underpredicts defaults for PLMBS loans from 2001 onward. The authors of this study interpret their findings as evidence that the increase in securitization in the early 2000s resulted lenders collecting less soft information on borrower quality that in turn changed the relationship between borrower observables and default risk.

[3] The most recent boom-bust cycle in the housing market provides a nice case in point of how the relationship between expected and realized house prices moves can change over time. Using survey data, Case, Shiller and Thompson [2012] document that between 2003 and 2006, household price expectations generally aligned with realized post-survey price appreciation. Starting in 2007, however, price expectations and actual price movements diverged significantly.

[4] In Appendix A we also study intertemporal changes in the relationship between borrower leverage and household expectations for home price appreciation using data from the National Survey of Mortgage Originations.

3 we analyze the predictive performance of mortgage default models estimated using a rich database of loan-level servicing records. Section 4 discusses factors likely driving the instability of model performance and investigates these mechanisms using data from the PSID. In Section 5 we study whether model instability can be alleviated by using training data drawn from a mix of economic conditions. Section 6 concludes.

2 Automation, Mortgage Lending, and Model Risk

The composition of lenders in the U.S. residential mortgage market has undergone a tremendous amount of change in recent years. Non-depository financial institutions ("non-banks") increased their market share of first-lien mortgage originations from 30 percent in 2007 to 50 percent in 2015, with much of this growth coming in the market for high-risk loans guaranteed by the Federal Housing Administration (FHA) [Buchak et al., 2018].[5] In addition to differing from banks in how they are funded [Kim et al., 2018], several of these non-banks are financial technology firms or "Fintechs."[6] There is no agreed-upon definition of a Fintech in the financial economics literature. Previous work on Fintechs, however, has differentiated these firms from other financial institutions based on their use of technology for processing mortgage applications.[7] While automated mortgage underwriting is certainly not new [Frame, Wall and White, 2018], the extent of automation among mortgage lenders in recent years has yielded significant improvements to the efficiency of the loan production process. Fuster, Plosser, Schnabl and Vickery [2018], for instance, find that Fintechs process mortgage applications about 20 percent faster than non-Fintech lenders.

A key driver of this increase in efficiency has been an expansion in the use of predictive models to assess the creditworthiness of a loan applicant. In a typical case for an unsecured loan such as a credit card, historical data would be used to estimate a statistical "scoring" model of loan performance. When a financial institution receives a loan application, the parameters from the scoring model are applied to variables that are derived from the loan applicant's credit bureau files and, now more frequently, other non-bureau data sources [Berg et al., 2018; Jagtiani and Lemieux, 2018]. The resulting score then serves as an input into a loan decisioning engine that can automatically approve or reject the application and, if approved, price the loan. When collateral valuations are not required for underwriting, this entire process can happen almost instantaneously.

Historically the collateral valuation procedure has served as a limit on the extent to which the mortgage process can be automated. Between 2010 and 2018, federal regulations required that mortgages with original outstanding balances in excess of $250,000 be supported by a full appraisal. Residential mortgages with loan amounts below the $250,000 appraisal threshold could be originated based on an "evaluation" in lieu of an appraisal [*Interagency Appraisal and Evaluation Guidlines*, 2010]. The property value used in these evaluations is frequently an estimate of the property's value derived from an automated valuation model (AVM). Like automated decisioning for unsecured lending, AVM-based valuation estimates can be produced almost instantaneously. Because many first-lien mortgages exceed the appraisal limit and full appraisals are required by most secondary mortgage market participants, lenders have not typically used AVMs as the source of valuation for underwriting loans.[8]

Recent changes to federal regulations and underwriting policies at Fannie Mae and Freddie Mac, however, suggest that AVM-based first-lien lending may become much more common in the coming years. In 2016 and 2017, Fannie Mae and Freddie Mac introduced appraisal waiver programs that allowed for the use of values from previous appraisals of the property or AVMs in lieu of full appraisals when certain other conditions are met [Harney, 2017].[9] Even more recently, federal regulators opened the door to the use of AVMs for

[5]FHA borrowers typically have very low down payments and lower credit scores than prime borrowers. The market share of FHA loans increased significantly following the financial crisis. It has been argued that FHA loans are substitutes for many of the high-risk loans colloquially referred to as "subprime" that defaulted at very high rates during the Great Recession [Karikari, Voicu and Fang, 2011].

[6]Fuster, Plosser, Schnabl and Vickery [2018] estimate that the market share for Fintech lenders in the mortgage space quadrupled from 2 percent in 2010 to 8 percent in 2016. They define Fintech lenders as those whose online platforms allow a prospective borrower to obtain a pre-approval entirely online without having to interact with a loan officer or broker. *Rocket Mortgage* from Quicken Loans is arguably the most well-known example of a Fintech mortgage lender.

[7]For a discussion of recent changes in how mortgage applications are processed, see Rodrigues, Gebre and Liu [2016].

[8]In home equity lending, in contrast, because second-lien loan amounts are generally much smaller than first liens and home equity loans are generally held on a lender's portfolio, AVMs have been used more frequently for originations.

[9]In January of 2018, of the loans submitted to Freddie Mac's automated underwriting system (Loan Product Advisor),

origination even further when, in October of 2019, they increased the residential appraisal threshold from $250,000 to $400,000 [*Final Rule: Real Estate Appraisals*, 2019]. An analysis cited in the final rule using 2017 loan originations estimated that this increase in the appraisal threshold would have exempted 72 percent of regulated mortgage transactions from appraisal requirements from 56 percent previously.[10] The appraisal waiver programs and expected increase in the residential appraisal limit for loans, when combined with increasing access to public records data used to estimate and test AVMs, have arguably cleared the way for almost automated first-lien mortgage lending.

The use of statistical models undoubtedly increases the speed at which financial institutions can make decisions, and this increase in efficiency has the potential to be welfare enhancing for both consumers and firms. These potential efficiency gains have been the focus of the burgeoning Fintech literature [Fuster, Plosser, Schnabl and Vickery, 2018; Buchak et al., 2018]. Implicit in much of the discussion of Fintechs, however, is the assumption that relying more heavily on predictive models enables Fintechs to more accurately quantify credit risk.[11] Model stability clearly plays an important role in ensuring the long run viability of mortgage market participants that engage in highly automated lending or buy loans from those that do. Very little attention in the literature, however, has been paid to an important potential downside of the increased reliance on predictive statistical modeling, namely, an increase in model risk, which the regulatory community has defined as "the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports"[*Supervisory Guidance on Model Risk Management*, 2011, p. 3]. Against this backdrop, the following analysis provides important information about the role of model risk in the U.S. mortgage market, the factors that likely drive this risk, and a straightforward way to potentially mitigate such risks.

3 The Failure of Predictive Models of Mortgage Performance

3.1 Data

The primary data source for this portion of our analysis is Black Knight's McDash servicing database. The McDash database contains historical monthly loan-level information for more than 180 million mortgages; since 2004 the McDash data has covered between 37 and 70 percent of the U.S. mortgage market.[12] The McDash data is arguably the premiere source of loan-level mortgage data. Much of the previous literature that is germane to our analysis has been based on data that was limited to loans that were securitized into particular types of mortgage-backed securities, such as PLMBS, loans sold to Fannie Mae or Freddie Mac, or loans packaged into Ginnie Mae securities. Because the mortgage market has changed significantly in recent years, it is not clear to what extent the results from studies based on data for one type of investor generalize to the larger post-crisis mortgage market.[13] In contrast to these investor-specific data sources, the McDash database contains loans that were securitized into PLMBS, Ginnie Mae, Fannie Mae, and Freddie Mac securities as well as loans that lenders chose to retain directly on their portfolios in lieu of securitizing them.

In addition to the breadth of the market coverage, an additional strength of the McDash data is the richness of the set of variables that are reported for each loan. The McDash records include an extensive set of fields capturing the information that was available to market participants at the time that the loan was

approximately 3.7 percent of the loans were eligible for Freddie Mac's appraisal waiver program known as Automated Collateral Evaluation (ACE). "ACE assesses whether a property's estimated value can be used in place of an appraised value to underwrite a mortgage and is based on Freddie Mac's proprietary AVM, called Home Value Explorer"[*An Overview of Enterprise Appraisal Waivers*, 2018, p. 2].

[10] In this context, "regulated transactions" are "transactions originated by FDIC-insured institutions or affiliated institutions, excluding transactions that were sold to the [government sponsered enterprises] GSEs or otherwise insured or guaranteed by a U.S. government agency."

[11] In some cases, the assumption that Fintechs are better able to identify high-quality borrowers is explicit. For example, in a model of the impact of Fintechs on consumer lending in China, Hau et al. [2019] assume that Fintechs possess a superior credit technology that enables them to extract more reliable information from noisy signals about borrower quality.

[12] The first observations in the McDash data are in 1997. In the early years of the data collection, the loans in McDash represented a small fraction of the larger market. The market coverage of the McDash data increased significantly in 2004 as more servicers began contributing to the data collection. For this reason we restrict our analysis to loans that were originated between 2004 and 2016.

[13] The fraction of all mortgages held in Fannie Mae, Freddie Mac, and Ginnie Mae securities has changed significantly over time [*Housing Finance at a Glance: A Monthly Chartbook*, 2018]. The representativity of any analysis based on loans from just one of these investors has thus likely changed significantly on a year-to-year basis.

originated. These static origination fields include but are not limited to: the FICO credit scores ("FICO scores,") that were used to underwrite the loan; the loan balance at origination; the original loan-to-value ratio; the appraised value of the property collateralizing the loan; and the schedule of any rate resets.[14] For each month that the loan is active, the McDash data also reports information on the loan's current balance, the loan's delinquency and foreclosure status, and a variable indicating whether a loan was paid off in a given month.

We use the static variables in the McDash data to estimate statistical models of mortgage default. These models are discussed in more detail in Section 3.2. We use the dynamic variables in the McDash records to define the outcome variable for our default models, which is an indicator variable that is equal to one if a loan ever experiences a delinquency of 90 days-past-due or worse, was charged off, was associated with a bankruptcy, or entered the foreclosure process at some point in the first 24 months following origination.

Regarding data restrictions, we limit our data to first-lien loans collateralized by single-family residences that were originated between 2004 and 2016.[15] We also eliminate records that do not have valid values for our explanatory variables from the data. Lastly, we remove loans from the data that have a "boarding age" in the McDash data that is greater than 2 months.[16]

3.2 Methodology

Our empirical strategy for studying model failure relies upon using historical data to create different hypothetical information sets on which a statistical model of loan repayment behavior could be estimated. We then construct predictions of default behavior based on these hypothetical information sets and compare those predictions against actual loan performance. The deviations between realized and predicted default behavior reflect the extent to which the predictive model "failed." In what follows, we use the term "training data" to refer to the loans that are used to estimate the statistical model that is used to predict borrower repayment behavior. Intuitively, the training data define the information set on which the predictions are based. For example, if the training data is defined to be mortgages originated in 2005, then the information set that is being used to estimate the model reflects that of a booming housing market with rapidly increasing home values and lax underwriting standards. In contrast, if the training data is defined as loans originated in 2010, then the data being used to estimate the predictive model reflect lender and borrower behavior in a market with rapidly declining collateral values and an exceedingly tight credit market. We use the term "validation sample" to refer to the loans that are not included in the training sample. The validation data is the set of the loans that we use to assess how well the model estimated using the training data can predict default.

All of our analyses require that we first generate predictions of default behavior for the validation data. In Section 3.2.1 and Section 3.2.2, we describe the models that we use to generate these predictions.[17] Section 3.3 discusses the performance metrics that we use to discuss the performance of the predictive models, and Section 3.4 describes how we change the nature of the training data to study the predictive accuracy of models of loan repayment behavior.

3.2.1 Logistic Regression Models

The first model that we use to study loan repayment behavior is the traditional workhorse in consumer credit modeling: the logistic regression (or "Logit") model. Specifically, if i indexes loans, we define the default

[14] The McDash data does not contain personally identifiable information.

[15] The property type restriction eliminates loans used to purchase condominiums from our sample.

[16] Boarding age is defined as the loan's age at the time that the loan's repayment activity is first reported in the McDash records. Because of lags between loan origination and when the loan is added to McDash's database, boarding ages of 1 or 2 months are not uncommon. A boarding age greater than 2 months, however, suggests that a non-trivial number of monthly performance observations may be missing. In the context of our analysis, the loans with greater boarding ages are more likely to be misclassified as non-defaults.

[17] Our choice of predictive models of mortgage performance is similar to that of Fuster, Goldsmith-Pinkham, Ramadorai and Walther [2018] who use the logistic regression model and the random forest classifier to study the potential impact of lenders' adoption of machine learning methods on minority households' access to credit.

indicator variables as follows

$$D_i = \begin{cases} 1 \text{ if the loan is ever 90 days-past-due in the 24 months following origination} \\ 0 \text{ otherwise} \end{cases}$$

Under the Logit specification, the probability that a default is observed is defined as

$$\Pr\{D_i = 1\} = \frac{\exp[X_i\beta]}{1+\exp[X_i\beta]} \tag{1}$$

where X_i is a vector of borrower and loan variables that are known at the time that the loan is originated and β is a parameter vector. The vector X_i includes variables that are known to be used in underwriting and have been shown to be predictive of mortgage default in previous work such as the loan's LTV, the FICO score used to underwrite the loan, the borrower's debt-to-income (DTI) ratio, a flag for owner-occupancy, and an indicator variable that identifies whether a loan had a fixed or variable interest rate. X_i also includes a set of binary variables that indicate whether a loan was conventional without private mortgage insurance (PMI), conventional with PMI, guaranteed by the Department of Veterans Affairs, or guaranteed by the Federal Housing Administration.[18]

In a second set of models, in addition to the borrower and loan variables that were known at the time of origination, we include the rolling window growth rate in the following three variables in the 24 months following origination: the county-level unemployment rate reported by the Bureau of Labor Statistics, the 30-year mortgage rate reported by Freddie Mac, and the unadjusted county-level Case-Shiller house price index from Moody's Analytics. While the post-origination path of these factors was clearly not known to the lender at the time of origination, lenders frequently use forecasts of such variables in the approval and pricing process as well as for capital and loss reserve planning. That said, the results from models that include the realized paths of these variables in the set of regressors represent the potential accuracy of models that utilize forecasts of post-origination risk factors when these forecasts are perfect. In what follows we refer to these as "perfect foresight" models.

For each of the various definitions of the training data discussed below, we estimate Equation 1 via maximum likelihood. We then use the estimated parameters from each of these models to predict the probability of default for each loan in the validation sample. As is common in the literature, we estimate the model separately for purchase and refinance mortgages.

3.2.2 The Random Forest Classifier

While the Logit model has traditionally been the most commonly used framework for modeling the performance of retail lending products, lenders – and particularly Fintechs – are increasingly using machine learning methods to model consumer behavior. To see if the results from our analysis using Logit models generalize to machine learning methods, for each of the different training-validation sample combinations we also use a common machine learning algorithm – the random forest classifier [Breiman, 2001] – to predict the probability of default for every loan in the validation sample. In the context of our analysis, the random forest (RF) algorithm, in essence, takes a weighted average of the votes from many simple default models – typically referred to as "trees" in the machine learning literature – to produce an overall default probability that is assigned to the loan [Hastie, Tibshirani and Friedman, 2009].[19]

The strength of the RF classifier is that it is a nonparametric model that allows the predictors to affect the outcome variable, which in our case is the occurrence of mortgage default, in a highly non-linear and potentially interactive fashion. In comparison to the Logit models that rely on strong distributional and functional form assumptions, the RF classifier can be viewed as a much more flexible way of modeling the relationship between variables used in loan underwriting and loan repayment behavior. A comparison of the forecasting accuracy of the RF classifier and the Logit models can thus give us a sense of whether the poor

[18] Previous work has shown that the relationships between mortgage default and several of the variables in our model are highly non-linear. For example, default risk increases rapidly as the loan's LTV approaches unity. To allow for these non-linearities, these variables enter our model specification in flexible functional forms.

[19] In our implementation of the RF algorithm, we do not place any limits on the depth of the trees or the number of observations that must be contained in the final nodes.

predictive performance of retail mortgage models discussed in the previous literature were at least driven in part by overly restrictive assumptions about the data generating process.

Though single decision trees are easy to visualize and understand, they are not robust in the sense that their accuracy is highly sensitive to noise in the data. Furthermore, single decision trees are prone to overfitting. The advantage of the RF classifier over the use of a single decision tree is that the RF classifier has a lower forecast variance. The "forest" component of the RF moniker derives from the fact that the reduction in variation is achieved by combining the predictions from many individual decision trees. More specifically, RF combines the output from individual decision trees using a bootstrap aggregation ("bagging") procedure [Breiman, 2001]. The tree bagging procedure first draws different bootstrap subsamples from the data and trains a separate decision tree to each of the subsamples. Because trees are trained on these subsamples and not the entire sample, the bagging procedure serves to lessen correlation between the trees. To further reduce correlation between the trees, the RF algorithm randomly selects a subset of predictors at each split in the decision tree. This feature randomization procedure serves to prevent strong predictors from appearing in most of the end nodes, thereby reducing the correlation between tree-specific predictions.[20]

There are several complications associated with using the RF algorithm to predict defaults. First, our data is "imbalanced" in the sense that non-defaults are far more common than defaults; this imbalance issue is endemic in credit risk modeling. Classification algorithms like RF are designed to minimize the rate of misclassification in the full sample. In our application, because defaults are quite rare, the RF algorithm will have a tendency to focus on predicting the outcome of non-defaults as non-defaults have an outsized impact on the calculation of the overall error rate. Because of this issue, when machine learning classifiers like RF are applied to extremely unbalanced data, they will tend to correctly classify very few of the positive outcomes.

The second issue associated with implementing the RF procedure is computational in nature. The RF algorithm is very computationally intensive, and implementing the algorithm on the full sample of loans is computationally infeasible.[21] The third and final complication associated with using the RF classifier is deciding how the tree-specific predictions are aggregated to create a final classification of a loan as a default or non-default. The default method for aggregating across trees is to simply use "majority rules" in the sense that the most common classification vote of the individual trees is taken to be the overall classification. In practice, however, the manner in which tree-level predictions are aggregated will be "tuned" to optimize measures of classification accuracy.

In our implementation of the RF classifier, we address these three complications as follows. Regarding the imbalance problem, instead of training the RF algorithm on the raw mortgage data, we utilize the Synthetic Minority Over-Sampling Technique (SMOTE) to improve balance in our data using "synthetic" default observations [Chawla et al., 2002].[22,23] Specifically, if N_j^D and N_j^{ND} denote the default and non-default observations in origination cohort j, we use SMOTE to create $\widetilde{N_j^D}$ synthetic defaults such that the sum of the true and augmented defaults comprise 20 percent of the "SMOTEd" data for origination cohort j. That is

$$\frac{N_j^D + \widetilde{N_j^D}}{N_j^D + \widetilde{N_j^D} + N_j^{ND}} = 0.2 \tag{2}$$

[20]We implement the RF procedure using the "randomForest" package in R. We chose to use 2000 trees in our forest based on a tuning exercise that demonstrated little improvement in model performance with larger forests. Regarding the other parameters used to implement the algorithm, in our tuning exercise we conducted a grid search over: the number of variables randomly sampled as candidates at each split ("MTRY"), the minimum number of terminal nodes, and the size of the sample used to grow the tree. Based on this grid search, we determined that the optimal value of MTRY was the number of features in our data divided by 3, a value that coincided with the default value of MTRY in the package. As model performance was largely insensitive to variation in the minimum node size and sample size, we also set those parameters at their default values.

[21]The computational environment that we used to conduct this analysis was a high-performance computing cluster with 512 gigabytes of RAM.

[22]Resampling approaches to addressing class imbalance are often referred to as "data-level" solutions in contrast to "algorithm-level" solutions in the machine learning literature. See Chawla, Japkowicz and Kotcz [2004] for a discussion of SMOTE and other resampling methods in applied machine learning.

[23]In unreported robustness checks, we also trained the RF classifier on the raw mortgage data as well as on an "undersampled" dataset in which the non-defaults were down-sampled to match the number of defaults in the sample in a given year. The RF classifier's predictive accuracy when trained on these alternative training samples was generally worse than the accuracy of the RF classifier trained on the SMOTE sample. These results are available upon request.

The SMOTEd data serves as the sampling base for the bootstrap samples of our RF algorithm. When defining the training sample for the implementation of the RF algorithm, we first partition the SMOTEd data into loans from an origination cohort j that default and those that do not. Let $S_j^D = N_j^D + \widetilde{N_j^D}$ and $S_j^{ND} = N_j^{ND}$ denote the set of defaulted and non-defaulted loans, respectively. When growing tree b, we randomly sample 50,000 observations from S_j^D with replacement to create a sample of defaults denoted $S_j^{D,b}$. Similarly, we create a sample of non-defaults, denoted $S_j^{ND,b}$, by randomly selecting 50,000 observations from S_j^{ND} with replacement. Lastly, we use a sample comprised of the observations in $S_j^{D,b}$ and $S_j^{ND,b}$ to grow tree b. The resulting dataset is perfectly balanced and small enough to render the algorithm computationally feasible. This process is repeated $B = 2,000$ times, where B denotes the total number of trees in the random forest. Regarding the aggregation of the tree-specific predictions, we used data from the 2004 origination cohort to tune a voting rule based on several common metrics of classification accuracy. This tuning procedure led us to classify a loan as a default if at least 70 percent of the tree-specific votes were default. We discuss the sampling procedure and parameter tuning in more detail in Appendix D.

3.3 Assessing Model Performance

One of the primary goals of our analysis is assessing the out-of-sample predictive performance of statistical models of loan default. Conducting this analysis thus requires that we adopt an operational definition of "predictive performance." One approach to assessing the accuracy of binary dependent variable models is to report the fraction of observations for which the dependent variable was classified correctly; in such an exercise, an observation is typically classified as a 1 if its predicted probability from the model exceeds 50 percent, 0 otherwise. Such measures are ill-suited for measuring the performance of consumer credit models. Even in the most lax of lending environments, defaults are still relatively rare, and the 50-percent decision boundary will result in very few observations being classified as defaults. A measure of the fraction of loans correctly classified under this standard would thus tell us little about how well the model actually predicts failure.[24] Lenders thus generally eschew using classification rates under a fixed decision boundary and instead assess the quality of credit risk models using alternative statistics that summarize a model's predictive accuracy.[25,26]

One of the most widely used methods for quantifying a model's predictive accuracy is to compare the expected number of defaults in some sample of loans with the actual number of defaults that were realized in that subpopulation; such exercises are frequently referred to as "expected-versus-actual" (EvA) analyses and form the basis for formal goodness-of-fit tests for classification models [Hosmer, Lemeshow and Sturdivant, 2013]. For the Logit model, the standard approach to conducting EvA analyses is to first use the estimated Logit model to predict the probability that a loan defaults. Let j index the training datasets, k index the validation datasets of size N_k, and let $\widehat{PD}_{ji}$ for $i = 1, ..., N_k$ denote the predicted probability that loan i defaults based on the Logit model estimated using training data j. The expected number of defaults in the validation sample based on a Logit model trained on sample j, which we denote ED_{jk}, can be written as

[24] Wooldridge [2010, p. 574] notes that "while the percent correctly predicted is useful as a goodness-of-fit measure, it can be misleading. In particular, it is possible to get rather high percentages correctly even when the least likely outcome is very poorly predicted." In the context of lending, for example, if the default rate in the population is 10 percent, even if a model did not predict that a single loan in the sample would default based on the 50 percent decision boundary, the percent correctly classified would be equal to 90 percent.

[25] In addition to measuring a model's predictive accuracy, lenders often also assess a model's ability to discriminate between defaults and non-defaults. Two of the most commonly used measures of a model's discriminatory power are the Komolgorov-Smirnov (KS) statistic and the Area Under the Receiver Operating Curve (AUROC). For book-length discussions of the use, development, and testing of predictive models in retail lending, see Anderson [2007] and Thomas [2009]. Classification models can exhibit strong discriminatory power even when they do a poor job of predicting the overall level of default. For this reason and because the mechanisms that we discuss in Section 4 primarily concern the ability of models to predict outcomes – not rank order risk – in our analysis we focus on measures that summarize a model's ability to predict the prevalence of default.

[26] It is worth noting that one component of the Basel Accords requires lenders to compare actual defaults in their portfolios against the number of defaults predicted by their credit risk models as a part of the regulatory capital quantification process [Thomas, 2009, p. 137]. In this context, the predictive accuracy of the risk models is sometimes referred to as "calibration." Other applications of credit risk models that rely on the predictive accuracy of the model output include loan pricing, loss forecasting, and loan-loss reserve estimation.

$$ED_{jk} = \sum_{i=1}^{N_k} \widehat{PD}_{ji} \tag{3}$$

where N_k denotes the number of loans in the validation sample.

The RF procedure does not directly produce a predicted probability of default but rather a classification of whether a loan will or will not default. This classification is achieved by aggregating the votes from all the individual trees in the forest, where each tree is constructed using a randomly drawn subset of the training data. Let $\widehat{Y}_{jbi}$ denote the vote for loan i from tree b based on training set j, and let $\widetilde{Y}_{ji}$ denote the final prediction from the RF procedure based on an aggregation of all B of the tree-specific votes.[27] The RF algorithm determines $\widetilde{Y}_{ji}$ as follows

$$\widetilde{Y}_{ji} = \begin{cases} 1 \text{ (default) if } \frac{1}{B}\sum_{b=1}^{B} \widehat{Y}_{jbi} > \tau \\ 0 \text{ (non-default) if } \frac{1}{B}\sum_{b=1}^{B} \widehat{Y}_{jbi} \leq \tau \end{cases} \tag{4}$$

where $\frac{1}{B}\sum_{b=1}^{B} \widehat{Y}_{jbi}$ measures the fraction of trees that classify a loan as being in default and τ is a cutoff threshold that is used in the aggregation.[28] As discussed in Section 3.2, based on the results of a parameter tuning procedure we selected $\tau = 0.7$ as the threshold that determines whether a loan is classified as a default in our accuracy analysis.[29] After using the procedure above to classify the loans in the validation samples as defaults and non-defaults, we construct the expected number of defaults in validation set k based on the RF classifier as

$$ED_{jk} = \sum_{i=1}^{N_k} \widetilde{Y}_{ji} \tag{5}$$

In what follows, we summarize the predictive accuracy of both the Logit and RF models using the following measure

$$EvA_{jk} = 100 * \left(\frac{ED_{jk} - \sum_{i=1}^{N_k} D_i}{\sum_{i=1}^{N_k} D_i} \right) \tag{6}$$

where $\sum_{i=1}^{N_k} D_i$ is the number of *actual* defaults in validation set k. Large positive values of EvA_{jk} indicate that when training set j is used to predict the performance of loans in set k, the model is overpredicting defaults by a significant margin. Likewise negative values of EvA_{jk} indicate that the model is underpredicting defaults. Because of the large number of different training and validation sets that we use to test model performance, we use heat maps to summarize EvA_{jk} across various $j - k$ combinations in Section 3.6.

3.4 Changing the Information Set

Our strategy for studying the failure of predictive models relies upon systematically altering the nature of the training data – and thus the information set that is used to estimate the predictive model – and studying how these changes to the training data affect the out-of-sample predictive accuracy of the models.[30] For our analysis, we first define the training data as loans that were originated in 2004. The performance of these loans over the course of the next 24 months is used to estimate the Logit and RF models. The output from these models is then used to predict mortgage defaults for every loan origination cohort year between

[27] The number of trees (B) is set to be 2,000 for all training data sets. The size of the forest is thus not indexed by the training data.

[28] In the standard presentation of the RF model, the aggregation is conducted via "majority rule" in the sense that the modal prediction in the set of T predictions is the final classification that is applied to the loan [Hastie, Tibshirani and Friedman, 2009]. Because of the binary nature of the prediction in our application, this majority rule aggregation is equivalent to setting $\tau = 0.5$ in Equation D.2.

[29] For details on how this tuning was performed, see Appendix D.

[30] This strategy is similar to the methodology of testing for model failure advanced by Rajan, Seru and Vig [2015]. In this study, the authors estimated a logistic regression model using PLMBS data for loans originated between 1997 and 2000. The parameters from this model were then used to predict defaults on PLMBS loans that were originated between 2001 and 2006. Based on this methodology, the authors found that their statistical model systematically underpredicted defaults, a finding that they attributed to a reduction in the amount of soft information that was collected during underwriting in the early 2000s.

2004 and 2016, and we construct $EvA_{2004,k}$ for $k = 2004$ through $k = 2016$ as in Equation 6. We then repeat this process by iteratively taking loans originated in each of the years between 2005 and 2016 to be the training sample. The result of this process is 169 EvA_{jk} measures, which include both in-sample and out-of-sample predictions. Given the large amount of variation in underwriting standards, house price dynamics, and adverse shocks to household finances during our sample period, comparing the actual and predicted default behavior in the validation sample across models estimated using these different training data samples provides evidence on the net impact of broad intertemporal changes in credit markets on the predictive accuracy of default models.[31]

3.5 Descriptive Statistics

In Table 1 and Table 2 we report the average values of the key variables from the McDash data for purchase and refinance mortgages, respectively, that we use to estimate and test the Logit and RF models. Given the significant changes that occurred in the mortgage market and the larger economy during our sample period, we report these averages on a year-by-year basis. Default risk varied enormously over the course of our sample. For both purchase and refinance mortgages, default rates rose steadily in the early years of sample, peaking at roughly 12 percent in the 2007 origination cohorts. With tightening mortgage credit and an improving economy, default rates then declined after 2008, reaching a low of 1.4 percent for purchase mortgages and 0.5 percent for refinance mortgages in 2015. Interestingly, the 2016 origination cohorts experienced the first year-on-year increase in delinquencies since 2007.

As expected, the post-origination path of house prices, interest rates, and unemployment rates also exhibited sizable variation across our loan cohorts. The 2004 and 2005 origination cohorts were characterized by rising home values, rising interest rates, and declining unemployment rates. In stark contrast, the loans originated between 2006 and 2008 experienced sharply declining real estate values, falling mortgage rates, and dramatic increases in the county-level unemployment rate in the two years following origination. The final cohorts in our sample were typically exposed to increasing home values and declining unemployment rates.

Regarding loan underwriting variables, the back-end DTI ratio and LTV ratio both remained relatively stable between 2004 and 2016, as did the fraction of loans that were collateralized by owner-occupied units. The average FICO score for new originations increased moderately between 2007 and 2008 and remained elevated for the duration of our sample. Lastly, while loans with variable interest rates were relatively common in the early years in our sample, starting in 2008 virtually all new originations had fixed rates.

3.6 Results

3.6.1 EvA Analysis

In Figure 1 and Figure 2 we summarize the EvA_{jk} measures defined in Equation 6 for the various training-validation year combinations for purchase and refinance mortgages, respectively.[32,33] The leftmost panel in each of these figures is a heatmap depicting the EvA_{jk} measures for the Logit model estimated using the standard mortgage risk factors that are observable at the time a loan is underwritten. The center and rightmost panels include EvA heatmaps for the Logit and RF models, respectively, that were estimated using the variables aggregate variation in post-origination dynamics in housing prices, mortgage rates, and unemployment rates (the "perfect foresight" variables) in addition to the standard mortgage risk factors that are known at the time of origination. In all of these heatmaps, cells that are close to white in color represent training-validation combinations where the EvA_{jk} measure was close to zero. Said differently, the cells that are close to white in color are the training-validation combinations where there was little difference between the number of defaults predicted by the model and the number of loans that actually default in that validation cohort. Cells with blue shading in these heat maps represent cases where the model is underpredicting default behavior, with increasingly darker shades of blue representing worsening

[31] For example, predicting default behavior for loans originated in 2013 using a model estimated with 2008 development data provides evidence on the accuracy of default predictions for loans originated during an expansion based on information from a period of extreme duress in the credit market.

[32] The full set of EvA_{jk} measures used to construct these figures is available upon request.

[33] We report the full regression results for the Logit models in Appendix B.

Table 1: Summary Statistics for Purchase Mortgages by Year

		Average Values												
									Owner	Fixed	Conventional			
Year	Observations	D_i[1]	ΔHPI_i[2]	$\Delta Rate_i$[3]	$\Delta Unemp_i$[4]	DTI[5]	FICO	LTV[6]	Occupied?[7]	Rate?[8]	Without PMI	With PMI	FHA	VA
2004	306651	3.52%	21.14%	9.26%	-15.99%	27.23%	708.39	79.06%	91.22%	58.25%	70.66%	20.09%	7.36%	1.89%
2005	743497	5.09%	7.48%	9.03%	-7.92%	32.38%	707.56	79.38%	88.82%	65.84%	71.10%	19.02%	7.24%	2.64%
2006	734438	8.58%	-5.37%	-5.73%	27.42%	33.41%	704.76	81.02%	90.65%	77.92%	65.64%	20.36%	9.66%	4.34%
2007	906161	11.79%	-12.00%	-20.65%	106.25%	35.76%	704.70	83.35%	90.84%	89.71%	55.27%	29.21%	10.62%	4.89%
2008	1157604	8.42%	-9.73%	-22.31%	71.71%	35.70%	710.02	85.01%	86.83%	96.34%	37.13%	21.59%	35.07%	6.22%
2009	1189008	3.77%	-5.89%	-12.39%	-2.66%	36.74%	717.49	87.16%	91.31%	98.74%	34.33%	9.48%	48.23%	7.97%
2010	1095221	3.29%	-2.47%	-22.07%	-15.95%	36.81%	722.12	87.76%	91.83%	96.88%	37.64%	5.92%	47.67%	8.77%
2011	1046435	2.70%	10.25%	-8.91%	-18.18%	33.85%	726.56	87.21%	91.13%	95.58%	39.90%	9.45%	41.47%	9.18%
2012	945750	2.10%	14.98%	14.19%	-23.39%	31.97%	729.68	86.97%	92.47%	96.71%	40.33%	13.16%	37.39%	9.13%
2013	944053	1.77%	10.78%	-2.41%	-27.50%	34.05%	732.46	86.33%	92.11%	96.06%	42.90%	17.01%	30.18%	9.91%
2014	685955	1.55%	9.77%	-12.88%	-18.83%	33.90%	732.46	85.78%	92.46%	94.62%	41.11%	21.26%	26.71%	10.91%
2015	678931	1.42%	11.01%	3.10%	-16.17%	33.91%	736.88	85.11%	92.63%	95.85%	41.08%	23.17%	26.94%	8.81%
2016	690784	1.81%	11.88%	26.26%	-19.26%	34.86%	738.34	84.97%	92.42%	96.98%	40.78%	28.56%	22.64%	8.02%

[1] D_i is equal to one if mortgage i is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income.
[6] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
[7] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[8] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.

Table 2: Summary Statistics for Refinance Mortgages by Year

		Average Values												
									Owner	Fixed	Conventional			
Year	Observations	D_i[1]	ΔHPI_i[2]	$\Delta Rate_i$[3]	$\Delta Unemp_i$[4]	DTI[5]	FICO	LTV[6]	Occupied?[7]	Rate?[8]	Without PMI	With PMI	FHA	VA
2004	483952	1.69%	21.87%	9.46%	-16.54%	28.41%	707.28	64.37%	94.72%	61.82%	92.51%	6.12%	1.15%	0.22%
2005	1022559	2.53%	5.23%	8.74%	-5.21%	32.81%	702.99	66.54%	93.66%	68.02%	91.25%	7.50%	1.04%	0.21%
2006	883143	6.16%	-11.29%	-5.88%	37.56%	34.40%	697.16	67.61%	93.18%	71.61%	89.36%	7.12%	3.24%	0.28%
2007	1043423	12.09%	-17.22%	-20.50%	111.65%	35.92%	695.24	70.00%	92.33%	84.63%	81.62%	10.50%	7.46%	0.42%
2008	1199856	9.08%	-11.40%	-19.60%	82.89%	35.17%	710.74	70.79%	90.05%	94.37%	65.52%	11.14%	22.62%	0.72%
2009	1680750	2.21%	-6.03%	-9.87%	-0.89%	33.34%	745.35	67.33%	94.78%	98.15%	72.29%	8.29%	17.41%	2.00%
2010	1439271	1.26%	-1.92%	-22.05%	-16.27%	33.47%	754.14	68.51%	94.42%	95.60%	82.50%	2.88%	12.51%	2.11%
2011	1250577	1.05%	9.26%	-9.02%	-17.37%	31.24%	755.72	66.68%	92.85%	93.81%	85.73%	3.56%	8.21%	2.51%
2012	1222559	0.76%	15.30%	14.57%	-23.77%	29.12%	756.10	69.62%	92.39%	94.92%	84.21%	4.61%	7.25%	3.93%
2013	1159307	0.88%	11.97%	0.47%	-28.25%	30.97%	743.44	68.48%	88.99%	95.58%	83.76%	6.13%	5.72%	4.39%
2014	367232	0.84%	10.39%	-11.89%	-21.17%	32.16%	736.42	67.64%	89.78%	93.08%	85.31%	5.99%	4.41%	4.30%
2015	455509	0.51%	11.53%	4.47%	-18.08%	31.42%	746.63	65.78%	92.26%	93.39%	89.34%	4.01%	3.32%	3.33%
2016	537799	0.73%	12.41%	25.87%	-19.55%	31.68%	746.76	65.38%	93.46%	94.23%	87.11%	5.13%	3.11%	4.65%

[1] D_i is equal to one if mortgage i is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income.
[6] For refinance mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the property's appraised value.
[7] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[8] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.

underpredictions. Similarly, red cells represent cases where the model under consideration is overpredicting delinquencies, with dark shades of red representing the most significant overpredictions. Diagonal cells in all of these plots represent the in-sample model fit in terms of the EvA_{jk} measure; for the Logit models, the prediction error is mechanically zero in these cells [Hosmer, Lemeshow and Sturdivant, 2013, p. 9].

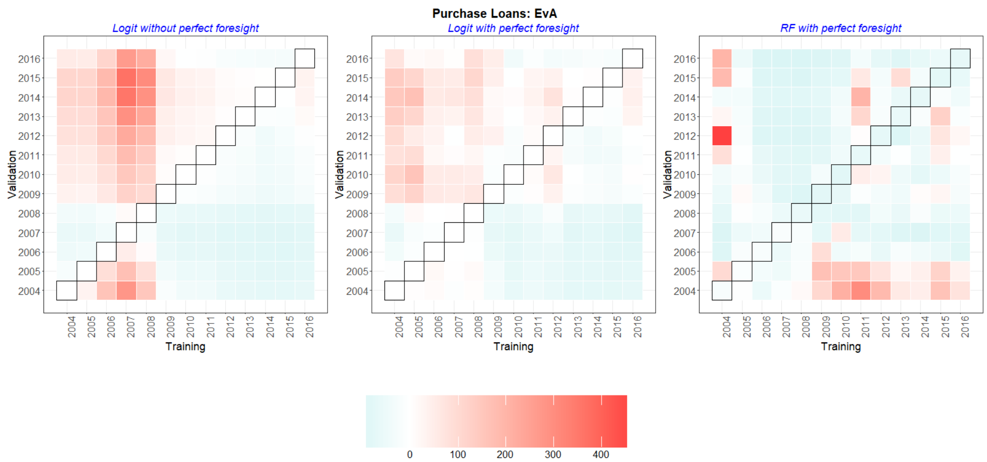

Figure 1: EvA Analysis of Year-by-Year Estimation: Purchase Loans

Turning first to the Logit models that do not include the perfect foresight variables, model performance deteriorates significantly as the temporal distance between the training and validation sample increases. Consider the 2007 purchase loan origination cohort, which experienced the highest default rate in our sample. The Logit model trained using the 2007 origination cohort overpredicted defaults for every out-of-time validation sample, in many cases by a very wide margin, with $EvA_{2007,k}$ ranging from a low of 12 percent in 2008 to a high of 363 percent in 2015. We see a similar deterioration in predictive performance when data from a low-default regime is used to train a predictive model that is used to forecast performance in a high-default regime. For example, the default rate in the 2007 origination cohort was more than 8 times higher than the default rate in the 2015 cohort for purchase mortgages, and the model trained using 2015 originations underestimated defaults in the 2007 validation sample by more than 84 percent.

Even small intertemporal deviations between the training and the validation samples can result in large prediction errors. For example, the models estimated using data on 2006 originations – the cohort with the second highest default rate – underpredicted defaults in the 2007 cohort by 37 percent for purchase mortgages and by more than 52 percent for refinance mortgages.

The model fits no better in terms of predictive performance when 2007 originations are used to predict the performance of 2006 originations, in this case *overestimating* purchase mortgage defaults by 47 percent and *overestimating* refinance mortgage defaults by 113 percent. The wild swings in the out-of-sample performance when the training data changes by only one year speak to how poorly predictive models of mortgage default can perform when macroeconomic conditions are changing rapidly.

As expected, increasing the similarity between the macroeconomic conditions in the training data and those in the validation data increases the predictive accuracy of the forecasts. For example, consider the case of predictions from the Logit model trained on 2012 purchase mortgage origination data. By 2012, most of the post-crisis structural changes to the mortgage market that characterize the remaining years in our sample had taken place, and as summarized in Table 1, the default rates for loans originated between 2012 and 2016 varied little across cohorts. In the post-2013 validation samples, $EvA_{2012,k}$ was quite stable, ranging from -20 percent in 2016 to 5 percent in 2013.

Comparing the leftmost panels with the corresponding center panels in Figure 1 and Figure 2 summarizes the extent to which the instability of the Logit models of default could potentially be remedied by including

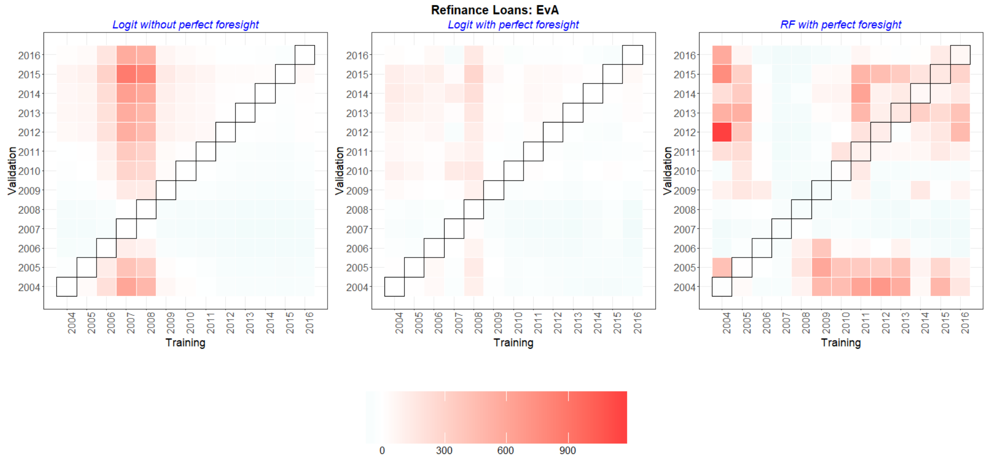

Figure 2: EvA Analysis of Year-by-Year Estimation: Refinance Loans

highly accurate forecasts of key risk drivers – mortgage rates, house prices, and unemployment rates – that previous work has shown to be predictive of mortgage performance. As above, we focus our discussion on the 2006 and 2007 origination cohorts as this time period was characterized by extreme economic turmoil and the 2007 training sample had the highest default rate in our sample.

As evidenced by the reduction in dark red cells when comparing the left and center panels, including the perfect foresight variables does reduce the tendency of the model to overestimate defaults when trained on data from years with elevated delinquencies. Even with perfect knowledge of the post-origination path of interest rates, house price indices, and unemployment rates, however, models estimated using the 2006 training data underestimate defaults in the 2007 cohort by 12 percent for purchase loans and by 32 percent for refinance loans.

Like the "without perfect foresight models," the predictive performance of the "with perfect foresight" models worsens considerably as we move further off the diagonal in Figure 1 and Figure 2. For example, the perfect foresight models estimated using the 2005 training data *overestimate* default in the 2015 validation data by 109 percent for purchase mortgages and 86 percent for refinance mortgages, while the "perfect foresight" models estimated using 2015 training data *underestimate* default in the 2005 purchase and refinance samples by 74 percent and 55 percent, respectively. Thus while it appears that including post-origination forecasts of aggregate economic variables known to drive mortgage performance has the potential to increase the accuracy of statistical models of default, even in the case where the forecasts of these variable are perfect, the predictive accuracy of statistical default models deteriorates rapidly with changes in the macroeconomic environment.

Previously published work on the predictive accuracy of models of mortgage performance was based on PLMBS loans that were originated in the late 1990s and early 2000s, and the structure of the mortgage market has undergone extraordinary change since then. In 2006, for example, roughly 42 percent of first-lien originations were placed in PLMBS, 32 percent of such loans were sold to Fannie Mae or Freddie Mac, 3 percent of loans were packaged into Ginnie Mae securities, and 23 percent of loans were held on portfolio. Through the end of the third quarter, in contrast, 37 percent of first-lien originations in 2019 were held on portfolio, 42 percent were sold to Fannie or Freddie, 19 percent were securitized through Ginnie Mae, and only 2 percent of loans were in PLMBS [*Housing Finance at a Glance: A Monthly Chartbook*, 2019]. Given the significant differences in the structure of the early-2000s PLMBS deals and the GSE, Ginnie Mae, and portfolio loan programs, it is not clear that the results from the existing literature would generalize to the post-crisis lending environment where the originate-to-distribute lending model requires repeatedly interacting with a very small number of well-informed financial institutions. An important contribution of our analysis is thus providing the first evidence that predictive models are unstable in the post-crisis mortgage market.

Visually, the predictive performance of our models when estimated and validated only on post-crisis data is summarized in the cells that are above the 2008 line on the "Validation" axis and to the right of the 2008 line on the "Training" axis. As in the pre-2008 years, the predictive performance of our models typically worsens as the difference in economic conditions between the training and validation samples increases. Using the 2009 training data, for instance, even when using the perfect foresight model we overestimate defaults in the 2015 validation data by 42 percent for purchase mortgages and by 52 percent for refinance loans. Our findings suggest that the poor performance of mortgage default models discussed in previous work was not a statistical anomaly driven by the quirks of the mid-2000s mortgage market; rather, predictive models of mortgage default models appear to perform poorly in general when the economic environment characterizing the training sample diverges from the macroeconomic conditions experienced in the validation sample.

As discussed in Section 1, machine learning techniques have experienced a surge in popularity among lenders using predictive models to automate decisioning for consumer lending. If the poor performance of our Logit models is being driven not by factors related to the Lucas critique but rather by statistical model misspecification, then machine learning methods – which allow for highly complex and flexible relationships between covariates and the outcome variable – should exhibit superior out-of-sample predictive performance when compared with the Logit models. The right panels in Figure 1 and Figure 2 that report the predictive performance of the perfect foresight RF model do not support this hypothesis.

Before discussing the predictive performance of the RF classifier, it should be emphasized that because of the sampling method we used to balance defaults and non-defaults algorithm and the bootstrapping procedure that is used when training the RF classifier, the validation sample on the diagonal elements of the rightmost panels of the heatmaps in Figure 1 and Figure 2 actually differs from the training sample.[34] The diagonal results for the Logit models, on the other hand, are mechanically equal to zero and represent true in-sample predictions in the sense that the set of observations used to construct the EvA_{jj} measures is identical to the set of observations used to estimate the Logit model. The diagonal cells for the RF and Logit models are thus not comparable, while the off-diagonal cells are comparable.

In spite of the fact that the RF model allows for far more flexibility in the relationship between the covariates and mortgage default, the out-of-sample predictive performance of the perfect foresight RF model is typically worse than that of the perfect foresight Logit models, in some cases significantly so.[35] Our finding that the Logit model outperforms the RF classifier may be due in part to the nature of our prediction problem. The RF classifier tends to perform well when the feature set contains many variables that are unrelated to the outcome variable, a situation that is referred to as "sparsity" in the literature [Athey and Imbens, 2019]. Like all mortgage servicing records, our data contains a relatively limited number of features, so our prediction problem is not characterized by sparsity. Second, the RF classifier is adept at discovering non-monotonic relationships between the features and the outcome variables but can struggle with recovering linear and quadratic relationships. In our application, it is natural to believe that the main drivers of mortgage default such as LTV and FICO scores have a monotonic impact on default risk, potentially limiting the usefulness of the RF classifier.

The difference in the predictive performance between the perfect foresight Logit model and perfect foresight RF model is evidenced by a greater share of darker blue and darker red cells in the rightmost panels in Figure 1 and Figure 2 as compared to the center panels. Interestingly, relative to the Logit model, the predictive performance of the RF classifier appears to deteriorate much more quickly with slight temporal deviations between the training and validation sample, especially in time periods where defaults were quite rare. For example, when the perfect foresight Logit model was trained using 2015 data, it underestimated defaults in the 2016 validation sample by 53 percent for purchase loans and overestimated defaults by 141 percent for refinance loans. The RF model that was trained using 2016 data, in contrast, underestimated

[34] The sampling procedures used to implement the RF classifier are discussed in Section 3.2.

[35] In our preliminary analyses, we trained the RF classifier on the exact same training data that we used to estimate the Logit model without using the sampling scheme to rebalance the prevalence of defaults and non-defaults. When the RF classifier was trained in this way, the out-of-sample predictive accuracy of the RF model was even worse than the results presented here as measured by EvA_{jk}. The superior predictive performance of the Logit model trained on the raw data thus does not appear to be solely an artifact of the balanced sampling scheme we use when training the RF model. It is worth noting that our finding that the Logit model outperformed the RF model based on certain measures of prediction accuracy is not unique in the literature [Kirasich, Smith and Sadler, 2018].

defaults in the 2015 validation sample by 63 percent for purchase loans and overestimated defaults by 278 percent for refinance loans.

Like the Logit results, our analysis of the RF classifier indicates that the predictive instability of models of mortgage default was not a phenomenon that was unique to the early-2000s boom-era mortgage market. In fact, our results suggest that if anything, the RF classifier is an even less accurate predictor of default in the post-crisis lending environment.

3.6.2 Marginal Effects Analysis

In the context of our analysis, the Lucas critique implies that relationship between the mortgage default indicator and the regressors in our Logit model will change over time. That is, under the Lucas critique the estimated parameters on the regressors should exhibit sizable changes as the training data is changed. To get a sense of the magnitude of intertemporal changes in the relationship between common mortgage risk factors and loan performance, we constructed the average marginal effects (AME) for the FICO score variables and LTV ratio variables that were used to generate the out-of-sample predictions discussed in Section 3.6.1.[36,37] The results of this analysis are presented in Tables 3 and 4.

Consistent with the Lucas critique, the estimated AMEs change considerably as we change the training data cohort. Consider first the estimated impact of LTV on default risk. When using the 2007 origination cohort to estimate the Logit model, we find that relative to loans with an LTV below 60, purchase and refinance loans with zero down payment ($LTV > 99$) were roughly 18 percentage points more likely to default. When estimating the same model using the 2016 origination cohort, the AME on the $LTV > 99$ term was a mere 0.9 percentage points for purchase loans and 0.8 percentage points for refinance loans. The relationship between default risk and FICO scores is similarly unstable over time. The estimated AMEs based on the 2007 cohorts imply that relative to loans with a FICO score of less than 550, purchase and refinance loans with a FICO score in excess of 800 were 36 and 33 percentage points less likely to default, respectively. The estimated AME on the $FICO > 800$ term based on the 2016 training data, in contrast, is -4.8 percentage points for purchase loans and -2.3 percentage points for refinance loans. In light of the extreme changes in the estimated AMEs associated with, in some cases, only a one year change in the training data, the instability of the predictive default models summarized in Figures 1 and 2 is unsurprising.

[36] The models on which the AMEs are based were estimated using the full set of regressors described in Section 3.2. To conserve space, we do not report the AMEs for the other variables.

[37] The analogue to the marginal effects analysis for the Random Forest model is the partial dependence plot (PDP) [Greenwell, 2017]. The PDPs for CLTV and FICO scores for the year-by-year RF models were also unstable across different training samples. To conserve space, we focus our discussions on the Logit average marginal effects. The PDP plots for the RF model are available upon request.

Table 3: Average Marginal Effects of LTV and FICO Variables for Year-by-Year "Perfect Foresight" Logit Models: Purchase Mortgages

	Dependent Variable: Default within 24 Months of Origination[1] **Average Marginal Effects**[2] *Purchase Loans* Training Data Cohort												
Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
LTV Buckets[3]													
$60 < LTV \leq 70$	0.025	0.026	0.049	0.073	0.039	0.000	0.007	0.010	0.007	0.007	0.006	0.003	-0.001
$70 < LTV \leq 80$	0.038	0.051	0.099	0.110	0.057	0.003	0.015	0.012	0.007	0.008	0.007	0.003	0.001
$80 < LTV \leq 90$	0.048	0.068	0.126	0.135	0.077	0.011	0.025	0.019	0.017	0.020	0.013	0.006	0.002
$90 < LTV \leq 95$	0.060	0.077	0.134	0.156	0.086	0.014	0.032	0.024	0.023	0.026	0.016	0.009	0.004
$95 < LTV \leq 99$	0.064	0.079	0.141	0.167	0.090	0.015	0.033	0.026	0.023	0.028	0.019	0.010	0.005
$LTV > 99$	0.075	0.097	0.156	0.176	0.095	0.027	0.049	0.041	0.032	0.032	0.020	0.010	0.009
FICO Buckets													
$550 < FICO \leq 650$	-0.031	-0.038	-0.066	-0.097	-0.059	-0.025	-0.018	-0.021	-0.014	-0.021	-0.023	-0.020	-0.011
$650 < FICO \leq 700$	-0.064	-0.082	-0.126	-0.172	-0.120	-0.056	-0.044	-0.041	-0.029	-0.032	-0.033	-0.029	-0.022
$700 < FICO \leq 750$	-0.089	-0.116	-0.172	-0.220	-0.164	-0.086	-0.071	-0.065	-0.048	-0.049	-0.047	-0.042	-0.036
$750 < FICO \leq 800$	-0.117	-0.157	-0.238	-0.300	-0.228	-0.120	-0.099	-0.090	-0.068	-0.066	-0.061	-0.054	-0.048
$FICO > 800$	-0.116	-0.163	-0.268	-0.360	-0.279	-0.138	-0.106	-0.100	-0.076	-0.076	-0.063	-0.058	-0.048

[1] A loan is classified as defaulting if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] To estimate the average marginal effect (AME) associated with moving from the reference category of a categorical variable to some category of interest x, we proceed as follows. First, for every observation in the training sample we set the categorical variable equal to the reference category and use the estimated parameters from the Logit model to predict the probability of default while holding the other covariates at their observed values. Second, we change the value of x to one and predict the probability default. Third, we take the difference between these two predicted probabilities. Lastly, to construct the AME we average over all of these observation-specific differences.
[3] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.

Table 4: Average Marginal Effects of LTV and FICO Variables for Year-by-Year "Perfect Foresight" Logit Models: Refinance Mortgages

	Dependent Variable: Default within 24 Months of Origination[1] **Average Marginal Effects**[2] *Refinance Loans* Training Data Cohort												
Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
LTV Buckets[3]													
$60 < LTV \leq 70$	0.008	0.009	0.034	0.073	0.053	0.011	0.007	0.004	0.003	0.002	0.001	0.001	0.002
$70 < LTV \leq 80$	0.012	0.017	0.061	0.119	0.082	0.019	0.010	0.008	0.004	0.004	0.003	0.002	0.003
$80 < LTV \leq 90$	0.020	0.028	0.090	0.150	0.097	0.026	0.017	0.013	0.009	0.007	0.006	0.004	0.004
$90 < LTV \leq 95$	0.025	0.039	0.107	0.167	0.105	0.030	0.020	0.016	0.010	0.008	0.008	0.005	0.006
$95 < LTV \leq 99$	0.030	0.043	0.111	0.178	0.112	0.034	0.021	0.017	0.011	0.009	0.007	0.005	0.006
$LTV > 99$	0.034	0.050	0.117	0.188	0.119	0.035	0.026	0.023	0.015	0.013	0.011	0.007	0.008
FICO Buckets													
$550 < FICO \leq 650$	-0.022	-0.026	-0.051	-0.086	-0.053	-0.021	-0.016	-0.010	-0.008	-0.009	-0.009	-0.006	-0.007
$650 < FICO \leq 700$	-0.036	-0.046	-0.090	-0.146	-0.107	-0.032	-0.022	-0.015	-0.012	-0.014	-0.015	-0.010	-0.012
$700 < FICO \leq 750$	-0.049	-0.066	-0.123	-0.194	-0.150	-0.047	-0.029	-0.021	-0.018	-0.020	-0.020	-0.014	-0.017
$750 < FICO \leq 800$	-0.062	-0.087	-0.169	-0.267	-0.219	-0.068	-0.041	-0.031	-0.026	-0.029	-0.027	-0.019	-0.021
$FICO > 800$	-0.068	-0.092	-0.190	-0.332	-0.277	-0.083	-0.048	-0.037	-0.029	-0.032	-0.031	-0.019	-0.023

[1] A loan is classified as defaulting if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] To estimate the average marginal effect (AME) associated with moving from the reference category of a categorical variable to some category of interest x, we proceed as follows. First, for every observation in the training sample we set the categorical variable equal to the reference category and use the estimated parameters from the Logit model to predict the probability of default while holding the other covariates at their observed values. Second, we change the value of x to one and predict the probability default. Third, we take the difference between these two predicted probabilities. Lastly, to construct the AME we average over all of these observation-specific differences.
[3] For refinance mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the property's appraised value.

4 Why Do Models That Predict Failure Fail?

4.1 Potential Mechanisms

In Section 3 we provided strong evidence that the accuracy of predictive models of mortgage performance deteriorates rapidly when the economic environment of the training data differs significantly from that of the validation data. Additionally, we demonstrated that the estimated relationships between default risk and two of the most widely-used variables in mortgage risk models – LTV and FICO scores – are wildly unstable across different training samples. These findings beg an obvious question: why do models that predict failure fail?

To understand the econometric issues associated with predicting mortgage performance, it is helpful to first discuss the two leading schools of thought regarding the mechanisms that cause mortgage default. Under both the "strategic" and "double trigger" models, a consumer defaults on debt when some trigger event occurs. If $Event_{i,s,s+t}$ is equal to 1 if loan i experiences an event that leads to default between s and $s+t$, 0 otherwise, and $NegEquity_{i,s,s+t}$ and $LiquidShock_{i,s,s+t}$ are indicator variables taking values of 1 if the borrower experiences negative equity or a liquidity shock, respectively, over the performance horizon, then under the strategic default model,

$$Event_{i,s,s+t} = \begin{cases} 1 \text{ if } NegEquity_{i,s,s+t} = 1 \\ 0 \text{ if } NegEquity_{i,s,s+t} = 0 \end{cases} \tag{7}$$

and under the double trigger model

$$Event_{i,s,s+t} = \begin{cases} 1 \text{ if } NegEquity_{i,s,s+t} \times LiquidShock_{i,s,s+t} = 1 \\ 0 \text{ if } NegEquity_{i,s,s+t} \times LiquidShock_{i,s,s+t} = 0 \end{cases} \tag{8}$$

To think about the mechanisms driving the failure of predictive models discussed in Section 3, let us consider an econometric model of the relationship between a variable indicating default between period s and $s+t$, which we denote $D_{i,s,s+t}$, and $Event_{i,s,s+t}$. To keep the exposition as simple as possible, we limit our discussion to the context of a linear probability model. In this setting, the empirical analog to both the "strategic" and "double trigger" models can be written as

$$D_{i,s,s+t} = \beta_0 + \beta_1 Event_{i,s,s+t} + \varepsilon_{i,s,t+s} \tag{9}$$

Note that β_0 and β_1 in Equation 9 should be time-invariant if either of the theories under consideration capture the actual mechanism that causes mortgage default.

When building predictive models of mortgage performance, lenders clearly do not know if the borrower will experience negative equity or a liquidity shock following the loan's origination. The typical approach to building predictive models of mortgage performance used in underwriting – one that we have utilized in Section 3 – is to use measurements of variables that are observable or at least forecastable at the time of origination to predict the likelihood that the trigger event occurs over the performance horizon.

To understand how model instability can arise when using this approach to predict mortgage performance, note that we can decompose $Event_{i,s,s+t}$ as follows

$$Event_{i,s,s+t} = E\left[Event_{i,s,s+t}|x_{i,s}\right] + \nu_{i,s,s+t} \tag{10}$$

where $E\left[Event_{i,s,s+t}|x_{i,s}\right]$ denotes the conditional expectation of the arrival of the event at time $s+t$ conditional on the information available at time s, denoted $x_{i,s}$.

If we express the conditional expectations as a linear function of $x_{i,s}$, then we can write Equation 10 as

$$Event_{i,s,s+t} = \delta_s^0 + x_{i,s}\delta_s^1 + \nu_{i,s,s+t} \tag{11}$$

The parameters in the vector δ_s^1 in Equation 11 have an interesting interpretation. Consider, for instance the parameter – call it δ_s^{LTV} – on the component of $x_{i,s}$ that measures LTV as of period s. δ_s^{LTV} captures the relationship between leverage at the time of origination for loans in origination cohort s and the likelihood that the loan experiences negative equity between s and $s+t$. δ_s^{LTV} will vary significantly over the economic cycle. In a regime where all property values are increasing, δ_s^{LTV} will be small because regardless of a loan's original LTV, very few loans will experience negative equity. When property values are declining, on the other hand, δ_s^{LTV} will be large as loans with low levels of initial equity will be highly likely to experience negative equity.

We can study the impact of changes in the economic environment on the statistical relationship between the elements in $x_{i,s}$ and default risk by substituting Equation 11 into Equation 9

$$D_{i,s,s+t} = \beta_0 + \beta_1 \underbrace{Event_{i,s,s+t}}_{=\delta_s^0 + x_{i,s}\delta_s^1 + \nu_{i,s,s+t}} + \varepsilon_{i,s,t+s} \tag{12}$$

$$D_{i,s,s+t} = \left(\delta_s^0 + \beta_0\right) + \beta_1 x_{i,s}\delta_s^1 + \varpi_{i,s,t+s} \tag{13}$$

$$D_{i,s,s+t} = \alpha_s^0 + x_{i,s}\alpha_s^1 + \varpi_{i,s,t+s} \tag{14}$$

where $\varpi_{i,s,t+s} = \beta_1\nu_{i,s,s+t} + \varepsilon_{i,s,t+s}$, $\alpha_s^0 = \delta_s^0 + \beta_0$, and $\alpha_s^1 = \beta_1\delta_s^1$.

Within this framework, even if the structural relationship described in Equation 9 is time-invariant in the sense that β_0 and β_1 do not change across loan cohorts, the estimated intercept ($\widehat{\alpha}_s^0$) and vector of slope parameters on the covariates ($\widehat{\alpha}_s^1$) based on the "plug-in" model will not be intertemporally stable. The instability in the regression parameters arises because the parameters being estimated are clearly a function of the economic environment from which the training data were drawn. Specifically, if the data used to estimate Equation 14 is drawn from a period of rising home values, declining unemployment, and few defaults, then the absolute magnitude of the components of δ_s^1 will be small, which will in turn attenuate the components of α_s^1, implying a low sensitivity between the components of $x_{i,s}$ and default risk. When the training data is drawn from a period of declining home values, worsening economic conditions, and rising default rates, on the other hand, the components of δ_s^1 will be large, as will the components of α_s^1, implying that the factors in $x_{i,s}$ are strong predictors of mortgage default.

While the toy model discussed here is far simpler than the Logit model we discussed in Section 3, the implications from this framework mirror our finding that the estimated average marginal effects on the LTV terms vary significantly with the rate of post-origination price appreciation experienced by loans in the training data. The simple linear probability regression example also suggests that predictive models of mortgage performance built using variables that proxy for the future arrival of equity and liquidity shocks will be inherently unstable in terms of their predictive accuracy and that the magnitude of out-of-sample prediction errors will grow as the macroeconomic conditions in the training data become increasingly dissimilar from those in the validation sample. One mechanism likely driving the deterioration in the predictive performance of mortgage default models is thus the fact that such models are typically constructed using proxies for the arrival of trigger events, and the relationship between these proxies and the actual occurrence of these trigger events changes significantly over time.

A second potential factor underlying the poor performance of predictive models of mortgage default concerns the Lucas critique. Lucas [1983] argued that observational data on economic phenomena reflects optimal decision rules of forward-looking economic agents and that changes in the economic environment – policy or otherwise – would alter expectations and result in changes in behavior. These changes in behavior, in turn, would then alter the joint distribution of the outcome variable and the predictors. In the case of parametric models, changes in the statistical relationship between the dependent variable and the regressors will manifest themselves in estimated parameters that vary significantly when the model is estimated using different training data. For non-parametric forecasting methods such as the RF classifier, changes in the joint distribution between the outcome variables and predictors will result in changes in the predicted outcome for an observation with a fixed set of values for the predictors.[38,39]

[38] In essence, the Lucas critique concerns the the parameter constancy and invariance of an econometric model. The idea of parameter constancy and invariance can be traced back to Frisch [1938] and Haavelmo [1944]. Hendry [1988] and Engle, Hendry and Richard [1983] suggested approaches for testing these parameter properties.

[39] The Lucas critique has found relatively little empirical support, especially among those studies using the concept of superexogeneity for testing the critique. Ericsson and Irons [1995] provides an extensive discussion of previous work on the topic. In addition to papers studying forward-looking models, there is also a branch of the literature that analyzes the performance of backward-looking models. See, for example, Ball [1999], Svensson [1997], Rudebusch and Svensson [1999], Taylor [1999], Estrella and Fuhrer [1999], and Fuhrer [1997]. As an exception, Lindé [2001] identified the relevance of the Lucas critique in practice by using a real-business cycle model. He further suggested that the superexogeneity test does not have enough power

Although Lucas [1983] focused his critique on the use of macroeconometric models for predicting policy responses, the core of his critique can be easily applied to microeconometric models such as those used to predict mortgage performance.[40] Consider, for example, the estimated impact of the LTV ratio – a variable used in virtually all predictive models of mortgage performance – on default. Recent work [Bailey et al., 2019] has argued theoretically and demonstrated empirically that mortgage borrower leverage is a function of borrowers' expectations for future price appreciation.[41,42] *If household expectations are leading indicators of future realized house price appreciation,* LTV measured as of time period s will be correlated with realized home price changes between s and $s+t$; this implies that borrower house price expectations as of period s will in part determine whether the borrower is "underwater" in $s+t$.[43] As discussed above, negative equity plays a critical role in determining whether or not a loan goes into default in both the "double trigger" and "strategic" models. Thus under both schools of thought concerning the determinants of mortgage default, Bailey et al.'s [2019] framework, when paired with the assumption that borrower price expectations are predictive of subsequent house price changes, implies that LTV measured as of time period s will be correlated with the arrival of default trigger events.[44]

Importantly, in addition to finding that leverage is a function of borrower price expectations, Bailey et al. [2019] also provide evidence that the relationship between home price expectations and leverage changes

to test the critique in small samples. Similar to Lindé [2001], Lubik and Surico [2010] found evidence of a small-sample bias which tends to hide the instability of backward-looking specifications. They demonstrated that when the instability of the reduced-form error variances is accounted for, the Lucas critique is found to be empirically relevant.

[40]In the context of the Lucas critique, the "deep parameters" in the linear probability model of default discussed above are β_0 and β_1.

[41]Bailey et al. [2019] model a household's leverage choice decision when purchasing a home. A loan's LTV in their theoretical framework is determined by borrower expectations about the future path of housing values and the extent to which households can adjust housing consumption. Specifically, Bailey et al. [2019] analyze leverage choice for two different types of households: households that are free to adjust both their leverage and their level of housing consumption ("variable house size") and households whose level of housing consumption is driven by non-financial factors such as family size and can be treated as independent of leverage choice ("fixed house size"). For variable-house-size households, their theoretical model predicts that higher levels of expected house price appreciation *increase* LTV. In contrast, fixed-house-size households *reduce* LTV when they have greater expectations for home price gains.

[42]We also provide empirical evidence linking price expectations to leverage in Appendix A using data from the National Survey of Mortgage Originations.

[43]Using survey data, Case, Shiller and Thompson [2012] provide evidence that between 2003 and 2006, household price expectations did in fact predict post-survey movements in housing values. From 2007 forward, however, household expectations regarding home price appreciation and actual changes in housing values diverged significantly.

[44]More formally, Bailey et al.'s [2019] analysis implies that $LTV_{i,s}$ is a function of borrower price expectations. That is,

$$LTV_{i,s} = f_s\left(\text{Price Expectations}_i^{s,s+t}\right) \tag{15}$$

where $f_s\,()$ is indexed by s to allow for the relationship between leverage and expectations to vary over time as in Bailey et al.'s [2019] empirical work and our findings in Appendix A. Case, Shiller and Thompson's [2012] findings that expectations are predictive of house price appreciation – denoted $HPA_{i,s,s+t}$ – and that the relationship between expectations and $HPA_{i,s,s+t}$ varies over time imply that

$$HPA_{i,s,s+t} = g_s\left(\text{Price Expectations}_i^{s,s+t}\right) \tag{16}$$

Inverting Equation 15 allows us to write

$$\text{Price Expectations}_i^{s,s+t} = f_s^{-1}\left(LTV_{i,s}\right) \tag{17}$$

Substituting Equation 17 into Equation 16, we can express realized house price appreciation as a function of $LTV_{i,s}$

$$HPA_{i,s,s+t} = g_s\left(f_s^{-1}\left(LTV_{i,s}\right)\right) \tag{18}$$

Lastly, because $Event_{i,s,s+t}$ is in part determined by whether a borrower has negative equity and negative equity status is dependent upon house price appreciation, we have

$$Event_{i,s,s+t} = \phi_s\left(HPA_{i,s,s+t}\right) \tag{19}$$

The arrival of a trigger event in this framework is thus a time-varying function of $LTV_{i,s}$

$$Event_{i,s,s+t} = \phi_s\left(g_s\left(f_s^{-1}\left(LTV_{i,s}\right)\right)\right) \tag{20}$$

over time with the state of the economy.[45,46] This finding is critical for our analysis as it implies that the relationship between LTV, price expectations, realized house price changes, and thus default risk should vary over the economic cycle.

For the sake of parsimony, Bailey et al.'s [2019] theoretical model abstracts away from borrower uncertainty about shocks to income and non-housing wealth. In richer models where borrowers also incorporate affordability considerations into the decision calculus, variables such as DTI and non-housing assets would likely also reflect borrower expectations for future changes in economic conditions.[47]

Thus in addition to expectations about the future path of housing values, a loan's observed LTV and DTI can be viewed as a function of a borrower's expectations about future income streams and other post-origination changes in other measures of financial wellbeing that several papers have documented are important determinants of default risk [Gerardi et al., 2017; Fuster and Willen, 2017; Hsu, Matsa and Melzer, 2018]. Intertemporal changes in the relationship between borrowers' *ex ante* expectations about the evolution of these factors and the *ex post* realized values of these variables as in the example above would generate intertemporal variation in the estimated relationship between LTV, DTI, and mortgage performance. Taken as a whole, Bailey et al.'s [2019] framework suggests that the estimated relationship between mortgage default and variables commonly used in credit risk models will depend heavily on the macroeconomic conditions that prevailed in the training data; when such models are used to predict loan performance exposed to much different macroeconomic conditions, the predictive accuracy of these models will be poor.

The foregoing discussion raises an important question that is straightforward to answer empirically: does the correlation structure between risk factors observed at the time of origination and the arrival of post-origination trigger events such as negative equity and income shocks vary over time? This question cannot be addressed using loan-level servicing records like those we utilized in Section 3 because servicing records do not report information on post-origination changes in a household's financial situation. This question can, however, be addressed using data the Panel Study of Income Dynamics (PSID). We turn now to an analysis of the intertemporal variation in the correlation between variables that are used to predict default and the subsequent materialization of adverse financial shocks using the PSID. In the context of our linear probability framework, this analysis provides direct evidence on the mechanisms causing intertemporal instability in δ_s^0 and δ_s^1.

4.2 Panel Study of Income Dynamics (PSID) Analysis

As discussed above, the variables that are traditionally used in models of mortgage performance can be viewed as predictors of the likelihood that a borrower experiences an adverse event that triggers default at some point in the future. Administrative servicing records like the McDash files do not report information on whether borrowers receive adverse shocks following loan origination, so it is not possible to use loan-level servicing data to study the relationship between the variables used in predictive models of loan performance and the subsequent arrival of trigger events. For that reason, we turn to the PSID to study the possibility that intertemporal heterogeneity in the relationship between these proxies and the actual arrival rate of such shocks could potentially be one of the mechanisms driving model failure and the extreme swings in the marginal effects discussed in Section 3.6.2.[48]

[45] Specifically, in a cross-sectional analysis, Bailey et al. [2019] find that the relationship between house price beliefs and leverage is stronger for more pessimistic borrowers, a finding that is consistent with the "fixed house size" model of leverage choice. In the time series, they find a stronger relationship between beliefs and leverage when home values are declining.

[46] We also provide empirical evidence that the relationship between price expectations and LTV changes over time in Appendix A using data from the National Survey of Mortgage Originations.

[47] Developing such a model is outside of the scope of this paper. We could imagine, however, that borrowers that are more optimistic regarding future income growth would be more apt to take out loans with high DTIs. On the supply side of the market, lenders offer "physician mortgage" programs based on this very idea. In such programs, borrowers are offered attractive terms on mortgages with high LTVs and DTIs because the borrowers have high expected income potential. In an economic sense, expected changes in wealth – for example, because of an inheritance – are simply a component of permanent income. That said, we would also expect borrowers that anticipate an increase in wealth to take out loans that are riskier along traditional underwriting dimensions.

[48] Our use of survey data to study loan performance is not unique in the literature. Several post-crisis papers have utilized similar data sources to demonstrate that adverse shocks to a household's financial health such as a sudden reduction in income, an increase in mortgage payments, or an increase in other financial obligations are key determinants of mortgage default. See, for example, Gerardi et al. [2017], Fuster and Willen [2017], and Hsu, Matsa and Melzer [2018].

The PSID is a nationally representative household survey that has been administered since 1968. The questions in the PSID data are designed to collect information on a wide variety of factors that affect the well being of U.S. households, such as health indicators, spending behavior, income, wealth.[49] Most importantly for our study, unlike other social surveys that are cross-sectional nature, PSID respondents are followed over time. It is the panel structure of the PSID that allows us to study the relationship between variables used in predictive models of mortgage performance and the *subsequent* arrival of delinquency trigger events. Due to restrictions on the availability of several key variables, we limit our analysis to the PSID survey waves from 1999 and 2017. During this time period, the PSID was administered on a biennial basis. To align as closely as possible with the model performance analysis in Section 3, we study the relationship between responses to survey questions about a household's finances in a particular survey year (t) and variables characterizing negative financial shocks in the next survey wave ($t + 2$).[50] In this portion of our analysis, we "observe the unobservables" in the sense that, while the trigger events that result in default are not generally observable in the administrative data that econometricians use to build credit risk models, these events are reported in the PSID [Gerardi et al., 2017]. Our approach can be contrasted with methodologies that, instead of trying to collect data on typically unobservable risk factors, attempt to extract information on latent risk factors by modeling unobservables (e.g., Duffie et al. [2009]).

Our analysis focuses on two variables that are commonly considered in mortgage underwriting: the combined loan-to-value ratio and the payment-to-income ratio. In what follows, we will refer to these measures as CLTV^{PSID} and PTI^{PSID} to avoid any possible confusion with the LTV and DTI variables in the McDash data that we utilized in our model performance analysis. We define CLTV^{PSID} as the sum of unpaid principal balances on first- and second-lien mortgages divided by the survey respondent's estimated value of his or her home as of the survey date. This CLTV^{PSID} measure should closely mirror the LTV values reported in loan-level servicing records such as those used in Section 3. We construct the PTI^{PSID} measure by summing the annual payments for first- and second-lien mortgages and dividing this sum by annual family income; this PTI variable is similar to the front-end debt-to-income ratio that is commonly used in mortgage underwriting.

Unfortunately, because the PSID does not report required payments on non-mortgage debts, we are unable to construct a variable that is comparable to the back-end debt-to-income ratio that is commonly used as a measure of a borrower's debt-servicing capacity. The PSID data does, however, report the outstanding balances on non-mortgage debts such as automobile loans, student loans, medical debt, and credit card balances. As a compromise, we construct a balance-to-income (BTI^{PSID}) measure by aggregating all mortgage and non-mortgage debt balances and dividing this sum by the annual income reported in the survey year. While our BTI^{PSID} measure does arguably capture variation in households' ability to service their debts, this variable is not directly comparable to the DTI measure in the McDash data that we utilize in our prediction accuracy analysis.

To study intertemporal variation in the relationship between variables that are typically used in credit risk models and the subsequent arrival of trigger events, we ran a series of simple linear regressions of CLTV^{PSID}, PTI^{PSID} and BTI^{PSID} as measured in period $t+2$ and the value of these variables in period t. Because theory suggests that negative equity is a critical trigger of mortgage default, in addition to studying the relationship between CLTV_t^{PSID} and CLTV_{t+2}^{PSID} over time, we also estimated a series of regressions in which an indicator variable that is equal to one if the household reported having a CLTV_{t+2}^{PSID} in excess of one in period $t + 2$ was regressed on CLTV_t^{PSID}. We estimated all of these regressions using a sample of observations pooled from 1999 through 2015 as well as on a survey-wave-by-survey-wave basis.

While the discussion in Section 4 focused on changes in the relationship between proxies for trigger events measured as of the prediction date and the subsequent arrival of trigger events, an additional potential source of model instability concerns the correlation between these proxies and unobservables that influence a borrower's ability to repay the loan. One such unobservable is a borrower's amount of non-housing wealth: when a borrower experiences liquidity issues, for example, this wealth can be used to continue to pay the mortgage even if the borrower is currently "underwater." If a borrower's leverage decision depends on expected changes in non-housing wealth, then CLTV_t^{PSID} would be correlated with future wealth for reasons

[49] More information on the PSID can be found at https://psidonline.isr.umich.edu/.

[50] For example, we study the relationship between the CLTV reported by survey respondents in 1999 and the incidence of negative equity in 2001. Because the 2019 PSID data was not available at the time of this writing, the 2017 PSID wave was only used to construct the trigger event variables associated with the 2015 survey responses.

similar to those generate the link between CLTV_t^{PSID} and future trigger events. We study this by regressing a borrower's non-housing net worth as of period $t+2$ on the CLTV_t^{PSID}.

We use two distinct samples from the PSID to conduct our analysis: an "all mortgages" sample and an "originations" sample. To construct the "all mortgages" sample, we identify household heads between the ages of 24 and 75 in the PSID data that report being homeowners with a positive mortgage balance in period t. Unlike mortgage servicing data that is structured to follow particular loans over time, the PSID data is structured to follow *people* over time. That said, if a household moves between t and $t+2$, the house value reported by the respondent in period $t+2$ may not correspond to the home that the respondent occupied in period t. The reported CLTV^{PSID} in period $t+2$ plays a critical role in our analysis, and studying changes in the CLTV^{PSID} measure when households are simultaneously changing residences would significantly complicate the interpretation of findings. To address this issue, we further limit the sample to households that reported in period $t+2$ that they had not moved since the previous survey. Lastly, we limited the sample to those observations with CLTV^{PSID} values between 0.1 and 2.5, PTI^{PSID} values less than 0.8, and BTI^{PSID} ratios less than 8.[51] The second sample that we construct (the "originations" sample) is comprised of loans from the "all mortgages" sample that are further limited to observations where the household reported that its mortgage was originated in the same year that the survey was administered. While a borrower's financial health can certainly change over the course of a year, we feel that the financial data reported in the PSID data for borrowers in the " originations" sample should be quite similar to the financial data that was used to underwrite the loans.

We report summary statistics for the "all mortgages" and "originations" samples in Tables 5 and 6, respectively, while we report the regression results based on these samples in Tables 7 and 8. As evidenced in Table 5, in the "all mortgages" sample, on average CLTV^{PSID} generally declined between 1999 and 2007, increased between 2007 and 2013, and declined between 2013 and 2015, a pattern that reflected the impact of national house price trends on measures of homeowner equity. Average PTI^{PSID}, in contrast, was remarkably stable over the course of our sample, ranging only between 0.15 and 0.18. The average of our BTI^{PSID} measure – which measures the amount of total debt outstanding to annual income – rose steadily between 1999 and 2011 before declining in the last two wave years of our sample. The dynamics of our BTI^{PSID} measure mirror the well-documented leveraging and de-leveraging of American households over the course of the most recent credit cycle. Based on the changes in our debt measures, these changes in household indebtedness were driven primarily by changes in mortgage debt, as household income and non-mortgage debt were quite stable by comparison throughout our sample.

Perhaps most importantly for our analysis, the transition between positive and negative equity varied enormously over time. In 2005, only 1.56 percent of the households in the "all mortgages" sample reported owing more on their home than it was worth. By the height of the housing crisis in 2011, 9.57 percent of households reported having negative equity, an astonishing 6-fold increase in the incidence of negative equity in a mere 6 years.

There are several similarities between the sample statistics for the "all mortgages" sample and those of the "originations" sample reported in Table 5. For instance, PTI^{PSID} in the "originations" sample was also very stable over time, while BTI^{PSID} rose and fell with the ebb and flow of the credit cycle. There are, however, important differences between the two samples. First, CLTV^{PSID} was, on average, uniformly higher in the "originations" sample, reflecting the gradual deleveraging that occurs through amortization. A second noticeable difference between the "originations" and the "all mortgages" samples is that the incidence of negative equity reported as of the survey date is lower in the former sample than in the latter. This finding is unsurprising as even during the height of the mortgage boom, mortgages with negative equity at the time of origination were quite rare.[52]

The most striking difference between the "originations" and "all mortgages" samples concerns the transition of households from positive to negative equity during the Great Recession. 2.17 percent of the "all mortgages" sample reported having negative equity in 2007, while 8.91 percent of this sample reported hav-

[51] The sample limitations discussed in the main text removed only a small number of observations with seemingly implausible values.

[52] Some loan products guaranteed by the Department of Veterans Affairs (VA) do allow homeowners to take out loans that exceed the value of their home. VA loans make up only a small part of the overall mortgage market. Most of the households in the "originations" sample that report having negative equity as of the survey date thus likely had zero or positive equity at the time that their loan was originated but experienced a decline in the value of their property between the origination of the loan and the survey date.

ing negative equity when they completed the survey in 2009. In the "originations" sample, in contrast, while only 2.15 percent of the households reported having negative equity in 2007, a shocking 24.58 percent of households that reported a newly-originated loan in 2007 had negative equity as of 2009.

The results reported in Table 7 and Table 8 confirm our expectation that variables typically used in credit risk models of mortgage performance are strongly predictive of *future* values of these variables. More importantly, these regressions also reveal that the correlation between credit risk factors observed as of period t and the arrival of trigger events in $t+2$ varies significantly over the economic cycle.

Consider first the case of the relationship between CLTV_t^{PSID} and the reported incidence of negative equity in $t+2$. When we use the "all mortgages" data from all survey waves between 1999 and 2015 to estimate this regression, a ten-percentage-point increase in CLTV^{PSID} in period t – say, from 0.8 to 0.9 – is associated with a 2.77 percent higher chance of experiencing negative equity in $t+2$.[53] As only 5.46 percent of households in this sample report experiencing negative equity in the subsequent survey wave, this correlation between CLTV_t^{PSID} and the the negative equity indicator is economically as well as statistically significant. When we estimate the regression using wave-by-wave samples, we find that the magnitude of the correlation between these two variables exhibited significant variation over time. In the waves from the late 1990s and early 2000s, we estimate that a ten-percentage-point increase in CLTV^{PSID} increased the risk of subsequently experiencing negative equity by roughly 1 percentage point. The nationwide decline in home values that began in 2007 fundamentally altered the nature of this correlation: for the 2007, 2009, and 2011 wave years, we find that the same ten-percentage-point increase in CLTV^{PSID} increased the likelihood of subsequently experiencing negative equity by 3.84 percent, 5.32 percent, and 4.34 percent, respectively. As expected, the recovery of the housing market that started in 2012 resulted in the coefficient on the CLTV_t^{PSID} term in our negative equity regressions declining significantly in the final two waves of our sample. These findings are consistent with our hypothesis that the relationship between CLTV_t^{PSID} and the likelihood of *subsequently* experiencing negative equity – the δ_s^1 term in the linear probability model in Section 4 – is highly unstable over time, particularly when housing price dynamics are changing quickly. It is this kind of instability that can give rise to the types of fluctuations in the estimated relationship between trigger-event proxies and realized defaults discussed in Section 3.

In addition to the instability of the relationship between CLTV_t^{PSID} and the negative equity indicator, we also find evidence that the relationship between CLTV_{t+2}^{PSID} and CLTV_t^{PSID} varied significantly across survey waves. If CLTV_t^{PSID} is a perfect proxy for CLTV_{t+2}^{PSID}, then a regression of the latter variable on the former would have an intercept of approximately zero and a slope of approximately one. The reported coefficients in the second panel of Table 7 suggest that the actual relationship between CLTV_t^{PSID} and CLTV_{t+2}^{PSID} is quite far from this "perfect proxy" benchmark. To put context around the magnitude of the wave-to-wave variations in the correlation structure between CLTV_t^{PSID} and CLTV_{t+2}^{PSID}, consider a loan with a CLTV_t^{PSID} value of 0.7, which is roughly the average value of this variable in the "all mortgages" sample. The regression parameters predict that CLTV_{t+2}^{PSID} for such a loan would fall to around 0.6 for the 1999 through 2005 PSID waves. Thus, even though mortgage balances in this sample were, on average, increasing during this time period, home values were increasing so rapidly that they more than offset the run-up in household leverage, resulting in a decline in CLTV_{t+2}^{PSID} relative to $\text{CLTV}_{t.}^{PSID}$ The estimated parameter coefficients for the 2007 and 2009 survey waves reveal that relationship between CLTV_{t+2}^{PSID} and CLTV_t^{PSID} changed radically with the onset of the Great Recession, and for these two waves, a CLTV_t^{PSID} value of 0.7 was predicted to *increase* to between 0.71 and 0.74. Finally, as home prices recovered in the final waves in our sample, the estimated relationship between CLTV_{t+2}^{PSID} and CLTV_t^{PSID} was more similar to what was seen in the pre-crisis period. In light of the discussion in Section 4.1, these findings are consistent with the large changes in the average marginal effects of the LTV variable reported in Table 3 and Table 4 being driven by significant intertemporal variation in the relationship between current and subsequent values of LTV.

We report in the third panel of Table 7 the results from regressions where CLTV_t^{PSID} is the lone independent variable and the dependent variable is the household's reported non-housing wealth in period $t+2$. For every sample wave, we find statistically significant correlation between CLTV_t^{PSID} and future non-housing wealth. The negative coefficient on the CLTV_t^{PSID} terms in these models implies that homeowners that are more highly leveraged have fewer non-housing assets that could be used to continue to service debts in the

[53]The reader is reminded that our CLTV measure takes values between 0 and 2.5.

Table 5: Characteristics of Family Units with Mortgages in the PSID ("All Mortgages Sample")

	Sample: All Active Loans in Survey Year [1] Survey Year									
Variable	All Years	1999	2001	2003	2005	2007	2009	2011	2013	2015
Mean Income [2]	$119,602	$113,423	$121,518	$118,403	$119,390	$122,182	$125,741	$115,965	$116,715	$121,738
Median Income	$100,759	$93,463	$99,361	$99,652	$99,898	$102,537	$106,512	$100,144	$99,850	$103,826
Mean CLTV^{PSID} [3]	0.61	0.60	0.57	0.58	0.56	0.56	0.64	0.67	0.67	0.62
Median CLTV^{PSID}	0.61	0.60	0.58	0.59	0.56	0.55	0.63	0.67	0.68	0.63
% Negative Equity in t	4.41%	2.26%	1.63%	2.10%	1.56%	2.17%	7.45%	9.57%	8.22%	3.70%
% Negative Equity in $t+2$	5.46%	2.61%	2.33%	1.94%	2.36%	8.91%	12.32%	9.65%	4.61%	2.58%
Mean PTI^{PSID} [4]	0.17	0.16	0.17	0.17	0.17	0.18	0.17	0.17	0.16	0.15
Mean BTI^{PSID} [5]	1.38	1.41	1.61	1.74	1.84	1.82	1.84	1.85	1.77	1.71
Mean Monthly Mortgage Payment [6]	$1,429	$1,268	$1,403	$1,416	$1,442	$1,567	$1,585	$1,455	$1,340	$1,344
Mean House Value	$297,210	$221,149	$259,145	$288,574	$344,670	$368,098	$320,800	$287,485	$278,328	$287,158
Mean Net Worth without Home Equity	$293,870	$299,934	$276,116	$262,842	$323,680	$345,883	$330,855	$268,856	$236,469	$292,367
Median Net Worth without Home Equity	$58,049	$61,806	$64,668	$62,245	$61,719	$60,554	$57,353	$53,584	$50,662	$47,675
Mean Net Worth with Home Equity	$427,705	$396,732	$399,370	$398,258	$498,240	$530,869	$465,192	$381,062	$349,582	$414,646
Median Net Worth with Home Equity	$163,842	$149,240	$170,249	$173,085	$195,699	$209,516	$160,353	$140,658	$132,351	$153,620
Mean Mortgage Debt	$163,362	$124,456	$135,836	$153,089	$170,061	$183,270	$186,421	$175,141	$165,213	$164,867
Median Mortgage Debt	$128,135	$102,508	$106,361	$122,852	$128,580	$145,329	$147,794	$138,425	$133,661	$127,134
Mean Non-Mortgage Debt	$16,887	$8,582	$10,718	$11,955	$12,009	$14,016	$15,404	$14,598	$31,870	$30,667
Median Non-Mortgage Debt	$3,404	$1,809	$2,836	$2,457	$2,572	$3,270	$4,682	$3,349	$5,390	$5,509
Mean Liquid Assets [7]	$26,056	$19,651	$23,260	$23,149	$26,774	$26,141	$28,290	$26,879	$26,923	$31,704
Median Liquid Assets	$6,825	$6,030	$7,091	$6,825	$6,429	$7,266	$7,023	$6,698	$6,467	$6,886
Mean Illiquid Assets [8]	$213,862	$222,053	$205,470	$185,574	$235,379	$239,390	$224,363	$195,299	$180,921	$232,392
Median Illiquid Assets	$36,284	$40,702	$42,544	$38,221	$38,574	$39,965	$36,284	$33,490	$32,337	$31,783
Observations	18632	1943	2023	2025	2060	2217	2251	2148	2017	1948

[1] The sample is restricted to family units that: own their own homes; reported having a mortgage; and did not change residences between consecutive surveys. All variables are defined at the family unit level. All dynamic variables are calculated by following the household head between survey waves. All sample statistics are constructed using the PSID's cross-sectional sample weights.
[2] All dollar-denominated variables are expressed in 2018 dollars.
[3] Combined Loan-to-Value ($CLTV^{PSID}$) is defined as the sum of outstanding first- and second-lien mortgage debt divided by the respondent's estimate of the home's value as of the survey.
[4] Payment-to-Income (PTI^{PSID}) is defined as the sum of annual payments for first- and second-lien mortgages divided by annual income.
[5] Balance-to-Income (BTI^{PSID}) is defined as the ratio of all outstanding mortgage and non-mortgage debts divided by annual family income.
[6] This variable is the sum of reported first- and second-lien monthly mortgage payments.
[7] Liquid assets are defined as the sum of checking or savings accounts, money market funds, certificates of deposit, government bonds, and treasury bills.
[8] Illiquid assets include equity in vehicles, retirement accounts, annuities, stocks, and business equity.
All statistics are constructed using the PSID cross-sectional weights.
All monetary values are expressed in 2018 dollars.

Table 6: Characteristics of Family Units with Mortgages Originated in the PSID Survey Year ("Originations Sample")

	Sample: All Loans Originated in Survey Year [1] Survey Year									
Variable	All Years	1999	2001	2003	2005	2007	2009	2011	2013	2015
Mean Income [2]	\$138,761	\$121,965	\$144,061	\$144,504	\$126,131	\$120,838	\$158,034	\$134,922	\$139,215	\$139,377
Median Income	\$115,162	\$98,890	\$123,583	\$116,858	\$107,107	\$100,761	\$140,455	\$113,928	\$119,147	\$118,938
Mean CLTV^{PSID} [3]	0.66	0.70	0.65	0.59	0.64	0.69	0.67	0.73	0.69	0.66
Median CLTV^{PSID}	0.68	0.73	0.69	0.62	0.64	0.71	0.68	0.73	0.71	0.69
% Negative Equity in t	1.97%	2.04%	2.74%	0.67%	1.34%	2.15%	2.71%	4.59%	3.24%	0.19%
% Negative Equity in $t+2$	4.79%	3.66%	0.94%	1.82%	2.18%	24.58%	7.67%	3.54%	4.02%	0.26%
Mean PTI^{PSID} [4]	0.17	0.18	0.16	0.16	0.18	0.21	0.17	0.15	0.15	0.16
Mean BTI^{PSID} [5]	1.93	1.83	1.57	1.77	1.97	2.49	2.13	1.75	2.00	2.04
Mean Monthly Mortgage Payment [6]	\$1,691	\$1,529	\$1,578	\$1,753	\$1,684	\$1,926	\$1,971	\$1,501	\$1,484	\$1,575
Mean House Value	\$378,243	\$267,258	\$312,419	\$417,721	\$407,331	\$413,851	\$439,267	\$283,687	\$375,352	\$382,224
Mean Net Worth without Home Equity	\$393,659	\$468,501	\$256,965	\$457,699	\$586,593	\$155,071	\$366,173	\$229,128	\$341,212	\$492,504
Median Net Worth without Home Equity	\$85,587	\$63,314	\$102,107	\$113,297	\$57,733	\$52,076	\$102,415	\$113,866	\$133,661	\$98,529
Mean Net Worth with Home Equity	\$548,979	\$556,977	\$385,749	\$655,868	\$764,221	\$309,774	\$528,221	\$318,357	\$502,458	\$642,041
Median Net Worth with Home Equity	\$209,395	\$142,607	\$195,705	\$253,894	\$180,398	\$165,917	\$223,558	\$192,679	\$250,075	\$235,197
Mean Mortgage Debt	\$222,997	\$178,782	\$183,635	\$219,490	\$229,703	\$260,336	\$277,219	\$194,458	\$214,106	\$232,687
Median Mortgage Debt	\$183,245	\$158,284	\$155,996	\$177,453	\$167,155	\$230,104	\$243,456	\$174,148	\$183,245	\$180,106
Mean Non-Mortgage Debt	\$14,300	\$9,853	\$9,098	\$8,979	\$8,910	\$10,595	\$12,725	\$16,241	\$36,385	\$21,401
Median Non-Mortgage Debt	\$2,422	\$2,713	\$1,560	\$1,229	\$3,215	\$3,997	\$3,511	\$2,233	\$5,390	\$3,496
Mean Liquid Assets [7]	\$29,562	\$16,465	\$28,452	\$37,434	\$17,718	\$14,179	\$40,702	\$29,361	\$31,558	\$30,019
Median Liquid Assets	\$9,689	\$7,537	\$7,800	\$10,920	\$5,786	\$4,844	\$11,705	\$8,931	\$14,013	\$7,416
Mean Illiquid Assets [8]	\$281,912	\$330,083	\$193,793	\$276,285	\$491,390	\$97,174	\$269,942	\$185,039	\$232,197	\$440,816
Median Illiquid Assets	\$51,740	\$40,702	\$54,599	\$61,426	\$42,432	\$36,332	\$58,523	\$66,980	\$62,519	\$37,081
Observations	1306	163	106	333	122	108	173	80	141	80

[1] The sample is restricted to family units that: own their own homes; reported having a mortgage *that was originated in the same year as the survey*; and did not change residences between consecutive surveys. All variables are defined at the family unit level. All dynamic variables are calculated by following the household head between survey waves. All sample statistics are constructed using the PSID's cross-sectional sample weights.
[2] All dollar-denominated variables are expressed in 2018 dollars.
[3] Combined Loan-to-Value ($CLTV^{PSID}$) is defined as the sum of outstanding first- and second-lien mortgage debt divided by the respondent's estimate of the home's value as of the survey.
[4] Payment-to-Income (PTI^{PSID}) is defined as the sum of annual payments for first- and second-lien mortgages divided by annual income.
[5] Balance-to-Income (BTI^{PSID}) is defined as the ratio of all outstanding mortgage and non-mortgage debts divided by annual family income.
[6] This variable is the sum of reported first- and second-lien monthly mortgage payments.
[7] Liquid assets are defined as the sum of checking or savings accounts, money market funds, certificates of deposit, government bonds, and treasury bills.
[8] Illiquid assets include equity in vehicles, retirement accounts, annuities, stocks, and business equity.
All monetary values are expressed in 2018 dollars.

face of liquidity pressures. As with the other results, the magnitude of the estimated parameters from these models varies significantly across the different PSID waves. These findings imply that an additional source of model instability could be a time-varying relationship between the CLTV^{PSID} terms used in predictive models and subsequent changes in a borrower's non-housing wealth.

In contrast to the estimated relationship between CLTV_t^{PSID} and CLTV_{t+2}^{PSID}, we find that the relationship between PTI_t^{PSID} and PTI_{t+2}^{PSID} was quite stable over time. At the grand PTI_t^{PSID} average of 0.17, for example, across the different wave-by-wave models, the predicted value of PTI_{t+2}^{PSID} was in a narrow range between 0.152 and 0.164. Thus while CLTV_t^{PSID} appears to be an unstable proxy for a household's future home equity, payment-to-income ratios measured as of the prediction date are likely strong predictors of a borrower's debt-servicing capacity in the near term.

The results in the final panel of Table 7 indicate that the relationship between BTI_t^{PSID} and BTI_{t+2}^{PSID} also varies over time, though the intertemporal heterogeneity in the BTI^{PSID} relationship is less extreme than what we observe for the CLTV^{PSID} variables. Though our BTI^{PSID} results cannot speak directly to how well the back-end DTI measure typically used in mortgage underwriting captures variation in a borrower's future ability to pay his or her debts, the intertemporal heterogeneity in the relationship between BTI_t^{PSID} and BTI_{t+2}^{PSID} suggests that the relationship between current and future back-end DTI measures also vary over time.

The regression results for the "originations" sample reported in Table 8 are qualitatively similar to the results of the "all mortgages" analysis. In the CLTV^{PSID} analysis, the estimated coefficient on the CLTV_t^{PSID} variable exhibits large changes across the different sample waves in the models where the negative equity indicator, CLTV_{t+2}^{PSID}, and non-housing wealth serve as dependent variables. In some cases, the estimated coefficients are no longer statistically significant. The reduction in statistical significance as we move from the "all mortgages" to the "originations" sample is unsurprising given the much smaller size of the "originations" sample. Also like the "all mortgages" results, the coefficients in the PTI^{PSID} models are generally stable across waves, while the coefficients in the BTI^{PSID} models exhibit more heterogeneity.

Table 7: Analysis of PSID Mortgage Variables: All Mortgages

	Survey Wave Year [1]									
Variable	All Years	1999	2001	2003	2005	2007	2009	2011	2013	2015
	Dependent Variable: *Negative Equity in* $t+2$									
CLTV_t^{PSID} [2]	0.277***	0.110***	0.109***	0.0981***	0.145***	0.384***	0.532***	0.434***	0.230***	0.149***
	(0.0104)	(0.0253)	(0.0204)	(0.0177)	(0.0314)	(0.0328)	(0.0267)	(0.0262)	(0.0299)	(0.0294)
Intercept	-0.114***	-0.0392***	-0.0397***	-0.0378***	-0.0577***	-0.126***	-0.217***	-0.197***	-0.108***	-0.0661***
	(0.00529)	(0.0137)	(0.00947)	(0.00772)	(0.0149)	(0.0146)	(0.0139)	(0.0144)	(0.0165)	(0.0155)
R^2	0.114	0.031	0.033	0.030	0.056	0.110	0.238	0.221	0.114	0.064
	Dependent Variable: $CLTV_{t+2}^{PSID}$									
CLTV_t^{PSID}	0.831***	0.754***	0.814***	0.793***	0.845***	0.968***	0.966***	0.873***	0.702***	0.740***
	(0.0119)	(0.0229)	(0.0229)	(0.0278)	(0.0189)	(0.0370)	(0.0357)	(0.0308)	(0.0369)	(0.0344)
Intercept	0.0509***	0.0640***	0.0433***	0.0352**	0.0201*	0.0629***	0.0300	0.0458**	0.0947***	0.0655***
	(0.00668)	(0.0149)	(0.0134)	(0.0154)	(0.0105)	(0.0187)	(0.0208)	(0.0188)	(0.0223)	(0.0201)
R^2	0.529	0.516	0.552	0.550	0.594	0.483	0.529	0.593	0.538	0.544
	Dependent Variable: *Real Non-Housing* $Wealth_{t+2}$ [5]									
CLTV_t^{PSID}	-615672.6***	-364017.6***	-467746.6***	-550122.3***	-822763.5***	-1299965.2***	-585582.2***	-396557.8***	-560479.6***	-581855.5***
	(62805.5)	(69127.9)	(86660.6)	(172438.7)	(177364.6)	(488152.8)	(107461.9)	(78816.4)	(114211.5)	(100905.6)
Intercept	705237.1***	501313.3***	560082.9***	677832.5***	875204.0***	1133705.1***	691500.4***	537338.0***	673764.5***	685346.9***
	(49338.0)	(54867.3)	(59431.0)	(127185.4)	(128343.3)	(345190.6)	(93034.6)	(69316.4)	(98208.8)	(84406.0)
R^2	0.012	0.008	0.016	0.007	0.016	0.010	0.025	0.023	0.030	0.023
	Dependent Variable: PTI_{t+2}^{PSID}									
PTI_t^{PSID} [3]	0.612***	0.578***	0.573***	0.567***	0.674***	0.639***	0.651***	0.557***	0.594***	0.643***
	(0.0127)	(0.0461)	(0.0370)	(0.0414)	(0.0352)	(0.0327)	(0.0370)	(0.0366)	(0.0364)	(0.0355)
Intercept	0.0555***	0.0600***	0.0661***	0.0647***	0.0498***	0.0553***	0.0536***	0.0577***	0.0512***	0.0454***
	(0.00195)	(0.00674)	(0.00593)	(0.00643)	(0.00540)	(0.00538)	(0.00571)	(0.00570)	(0.00527)	(0.00498)
R^2	0.349	0.287	0.280	0.315	0.375	0.417	0.384	0.349	0.345	0.366
	Dependent Variable: BTI_{t+2}^{PSID}									
BTI_t^{PSID} [4]	0.690***	0.615***	0.692***	0.617***	0.733***	0.744***	0.762***	0.683***	0.634***	0.658***
	(0.0111)	(0.0405)	(0.0324)	(0.0321)	(0.0329)	(0.0293)	(0.0261)	(0.0358)	(0.0326)	(0.0327)
Intercept	0.415***	0.421***	0.460***	0.547***	0.330***	0.364***	0.346***	0.465***	0.462***	0.402***
	(0.0167)	(0.0500)	(0.0424)	(0.0487)	(0.0506)	(0.0464)	(0.0413)	(0.0567)	(0.0541)	(0.0511)
R^2	0.458	0.359	0.362	0.361	0.474	0.524	0.540	0.450	0.451	0.473
Observations	18632	1943	2023	2025	2060	2217	2251	2148	2017	1948

[1] The sample is restricted to family units that: own their own homes; reported having a mortgage; and did not change residences between consecutive surveys. All variables are defined at the family unit level. All dynamic variables are calculated by following the household head between survey waves. All sample statistics are constructed using the PSID's cross-sectional sample weights.
[2] All dollar-denominated variables are expressed in 2018 dollars.
[3] Combined Loan-to-Value ($CLTV^{PSID}$) is defined as the sum of outstanding first- and second-lien mortgage debt divided by the respondent's estimate of the home's value as of the survey date.
[4] Payment-to-Income (PTI^{PSID}) is defined as the sum of annual payments for first- and second-lien mortgages divided by annual income.
[5] Balance-to-Income (BTI^{PSID}) is defined as the ratio of all outstanding mortgage and non-mortgage debts divided by annual family income.
[6] This variable is the sum of reported first- and second-lien monthly mortgage payments.
[7] Liquid assets are defined as the sum of checking or savings accounts, money market funds, certificates of deposit, government bonds, and treasury bills.
[8] Illiquid assets include equity in vehicles, retirement accounts, annuities, stocks, and business equity.
All monetary values are expressed in 2018 dollars.

Table 8: Analysis of PSID Mortgage Variables: New Originations

	Survey Wave Year [1]									
Variable	All Years	1999	2001	2003	2005	2007	2009	2011	2013	2015
	Dependent Variable: *Negative Equity in* $t+2$									
CLTV_t^{PSID} [2]	0.230***	0.184**	-0.0209	0.139***	0.112*	0.712***	0.390***	0.305*	0.306**	0.0158
	(0.0347)	(0.0858)	(0.0240)	(0.0526)	(0.0649)	(0.206)	(0.119)	(0.169)	(0.133)	(0.0162)
Intercept	-0.103***	-0.0929*	0.0231	-0.0638**	-0.0497	-0.243*	-0.184***	-0.186*	-0.170**	-0.00790
	(0.0186)	(0.0503)	(0.0246)	(0.0252)	(0.0318)	(0.127)	(0.0657)	(0.106)	(0.0762)	(0.00817)
R^2	0.055	0.040	0.002	0.049	0.028	0.115	0.079	0.127	0.125	0.005
	Dependent Variable: $CLTV_{t+2}^{PSID}$									
CLTV_t^{PSID}	0.889***	0.875***	0.687***	0.892***	0.681***	1.216***	0.987***	0.718***	0.834***	0.920***
	(0.0298)	(0.0666)	(0.109)	(0.0388)	(0.0800)	(0.208)	(0.115)	(0.113)	(0.0713)	(0.0540)
Intercept	0.0216	0.0185	0.130*	-0.0208	0.146***	0.0316	0.0146	0.128	-0.00137	-0.0446
	(0.0183)	(0.0475)	(0.0720)	(0.0226)	(0.0550)	(0.132)	(0.0708)	(0.0810)	(0.0484)	(0.0404)
R^2	0.484	0.524	0.385	0.648	0.489	0.332	0.487	0.514	0.596	0.784
	Dependent Variable: *Real Non-Housing* $Wealth_{t+2}$ [5]									
CLTV_t^{PSID}	-1121130.2***	-1130827.5	-451757.3***	-1736331.2	-366308.7	-613441.9	-537750.8***	-999854.0***	-1014413.0***	-1431608.0*
	(349642.1)	(724545.2)	(160515.6)	(1109177.3)	(315327.4)	(382341.4)	(184627.4)	(294628.2)	(319823.9)	(833627.5)
Intercept	1128904.7***	1175066.8*	551776.7***	1660165.7**	555782.6**	601070.0**	710798.7***	973538.8***	1009480.7***	1326103.1**
	(279458.6)	(618077.1)	(105771.5)	(840644.4)	(218387.5)	(298951.8)	(149114.9)	(253485.2)	(255154.9)	(663284.9)
R^2	0.019	0.041	0.051	0.013	0.010	0.080	0.026	0.264	0.167	0.115
	Dependent Variable: PTI_{t+2}^{PSID}									
PTI_t^{PSID} [3]	0.667***	0.678***	0.589***	0.697***	0.648***	0.638***	0.690***	0.411***	0.757***	0.444**
	(0.0309)	(0.0653)	(0.117)	(0.0760)	(0.112)	(0.0717)	(0.0683)	(0.139)	(0.0701)	(0.190)
Intercept	0.0541***	0.0618***	0.0813***	0.0554***	0.0587***	0.0669***	0.0462***	0.0788***	0.0263**	0.0712**
	(0.00512)	(0.0152)	(0.0205)	(0.0120)	(0.0173)	(0.0165)	(0.0106)	(0.0189)	(0.0106)	(0.0298)
R^2	0.398	0.395	0.212	0.360	0.374	0.545	0.563	0.201	0.609	0.157
	Dependent Variable: BTI_{t+2}^{PSID}									
BTI_t^{PSID} [4]	0.691***	0.798***	0.620***	0.695***	0.628***	0.666***	0.714***	0.724***	0.628***	0.675***
	(0.0289)	(0.0729)	(0.0872)	(0.0635)	(0.133)	(0.0854)	(0.0700)	(0.174)	(0.0667)	(0.0950)
Intercept	0.495***	0.300**	0.711***	0.514***	0.606***	0.736***	0.415***	0.427	0.462***	0.491**
	(0.0534)	(0.143)	(0.179)	(0.106)	(0.226)	(0.201)	(0.134)	(0.260)	(0.131)	(0.205)
R^2	0.457	0.533	0.236	0.444	0.374	0.492	0.624	0.358	0.448	0.390
Observations	1306	163	106	333	122	108	173	80	141	80

[1] The sample is restricted to family units that: own their own homes; reported having a mortgage *that was originated in the same year as the survey*; and did not change residences between consecutive surveys. All variables are defined at the family unit level. All dynamic variables are calculated by following the household head between survey waves. All sample statistics are constructed using the PSID's cross-sectional sample weights.
[2] Combined Loan-to-Value ($CLTV^{PSID}$) is defined as the sum of outstanding first- and second-lien mortgage debt divided by the respondent's estimate of the home's value as of the survey date.
[3] Payment-to-Income (PTI^{PSID}) is defined as the sum of annual payments for first- and second-lien mortgages divided by annual income.
[4] Balance-to-Income (BTI^{PSID}) is defined as the ratio of all outstanding mortgage and non-mortgage debts divided by annual family income.
[5] Non-housing wealth is defined as the sum of the value of seven types of assets – business equity, checking and savings accounts, non-residential real estate, stocks, vehicles, other assets, and retirement accounts – net of debt.
Robust standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

5 Leave-One-Out Analysis

Previous research has shown that negative equity and adverse financial shocks are the primary determinants of mortgage delinquency. Because the post-origination path of a borrower's home value and income are never known at the time that a loan is originated, credit risk models are typically constructed using variables such as LTV and DTI measured at the time the loan is originated. In the context of both leading schools of thought concerning mortgage default, these variables are, in essence, predictors of the likelihood that a default trigger event occurs at some point in the future. In Section 3 we demonstrated that the estimated relationship between two of these variables – LTV and FICO score – and default risk varied wildly over time.

The discussion and results in Section 4 suggested that this intertemporal instability is likely attributable two factors. First, we found one of the most widely-used variables in mortgage credit risk modeling – the LTV ratio – is strongly correlated with borrowers' expectations for the future path of variables known to influence mortgage delinquency, such as housing values. Second, in our PSID analysis we demonstrated that the relationship between several key credit risk predictors – variables which are functions of borrower expectations – and the subsequent arrival of trigger events exhibits significant intertemporal heterogeneity. In the context of our model performance analysis in Section 3, both of these factors imply that the out-of-sample predictive performance models will be poor when the economic conditions characterizing the validation sample differ significantly from those experienced by the development sample.

In light of these modeling limitations, it is natural to ask how the predictive performance of these models might be improved. One possible solution would be to try to train models using data that experienced similar economic conditions to those that would occur in the out-of-time sample where the predictions are applied. This solution is, of course, fundamentally infeasible as the economic conditions for the out-of-time sample will never be known at the time a predictive model is being built. An alternative approach to addressing model failure that is feasible would be ensuring that the training data contains observations from a mix of economic conditions: if the poor performance of the predictive models is primarily driven by the intertemporal instability in the estimated parameters in the regression models, then pooling observations from many different origination cohorts could possibly stabilize out-of-sample predictive performance.[54] To investigate this possibility, we repeated the prediction exercise described in Section 3 but, instead of using only data from one origination cohort to construct the training sample, we estimated the Logistic regression models and trained the RF classifier using the data from every origination cohort in our data *except for the validation or "holdout" cohort.* In what follows, we refer to this exercise as the "leave-one-out" (LOO) analysis.

The LOO analysis clearly predicts performance using information that would not have been available at the time the loans in the holdout cohort were originated. The performance of loans originated between 2006 and 2016, for example, would not be known in 2005, but in the LOO analysis the performance of the 2006 through 2016 origination cohorts is used to train the models that predict the behavior of loans originated in 2005. We constructed the sampling for the LOO analysis in this way to allow for the predicted default behavior of each of the holdout samples to depend on loans originated throughout the credit cycle, not just those that happened to pre-date the holdout sample. That said, we interpret the results of this analysis as evidence of the accuracy of predictive models of mortgage performance when such models are trained on a sample of loans drawn from throughout the credit cycle.

We estimate the Logistic regression models in the LOO analysis using the complete sample of all loans in all years in our sample, except for loans originated in the holdout year k. We then use the parameters from the Logit model to predict the probability that each loan in the validation sample defaults, and we construct our accuracy measure $EvA_{-k,k}$ as in Equation 6, where $-k$ denotes that the training data was comprised of all origination cohorts except for cohort k.

When constructing the training sample for the LOO analysis using the RF classifier, we first retain all defaults and non-defaults from all origination cohorts except for cohort k. Let S^D_{-k} and S^{ND}_{-k} denote the set of defaults and non-defaults in this sample, respectively. When growing each tree b of the $2,000$ trees in our random forest, we randomly sample 50,000 observations from S^D_{-k} and 50,000 observations from S^{ND}_{-k} with replacement. To remain consistent with the aggregation rule used in the year-by-year analysis, we classify a

[54]In the context of the Basel regulatory capital framework, our LOO sampling scheme is similar to the use of "through-the-cycle" (TTC) data for risk parameter quantification, while our initial analysis is similar to using a "point-in-time" (PIT) approach to quantification.

loan in the holdout data as a default if at least 70 percent of the tree-specific classifications were default.[55,56]

We summarize the predictive accuracy results of this analysis in Tables 9 and 10.[57] For each of the three different models, we report under the $EvA_{-k,k}$ column the accuracy measure associated with using the LOO model to predict mortgage performance in a particular holdout year. To serve as a point of comparison, we also report for each of these models the average, minimum and maximum values of the $EvA_{j,k}$ measures from the analysis based on year-by-year (YBY) training samples described in Section 3. The LOO $EvA_{-k,k}$ measures for both versions of the Logit model are almost uniformly bounded between the minimum and maximum values from the $EvA_{j,k}$ measures for both purchase and refinance loans. The lone exceptions to this bounding occur for the LOO perfect foresight Logit model when used to predict the performance of 2014 and 2015 refinance loans. When we alternatively exclude the 2014 and 2015 refinance originations data from the training sample, the LOO Logit models overpredict defaults in these cohorts by 303 and 563 percent, respectively. Notably, the 2014 and 2015 refinance origination cohorts had two of the lowest default rates in our sample. In many cases we also find that $EvA_{-k,k}$ is quite close to the average of the $EvA_{j,k}$ values. We interpret our results as evidence that using training data from a mix of economic conditions typically bounds out-of-sample prediction errors for Logit models away from the extremes associated with using training data drawn from a particular point in the economic cycle.

The behavior of the LOO RF classifier differs significantly from that of the LOO Logit models. Most notably, for purchase loans, $EvA_{-k,k}$ from the RF LOO model frequently exceeded the maximum value of $EvA_{j,k}$ associated with training the RF classifier on the YBY data in the 2007 through 2010 validation cohorts, while in the refinance population $EvA_{-k,k}$ from the RF LOO model exceeded the largest overpredictions from the YBY models between 2006 and 2010. In some cases, the difference between the RF LOO and YBY results was quite extreme. For instance, consider the accuracy of the predictions for 2007 purchase loan cohort. The LOO RF model overpredicts defaults by roughly 216 percent. When the RF model was iteratively trained using the YBY samples, in contrast, $EvA_{j,k}$ ranged from -95.65 to 55.36.

Introducing macroeconomic heterogeneity into the training data for the RF classifier does, however, appear to have some benefits regarding underpredictions. $EvA_{j,k}$ is bounded below at -100, a bound which is realized when the model predicts no defaults in a particular validation cohort. Several of the minimum values of $EvA_{j,k}$ for the RF classifier reported in Tables 9 and 10 are close to this lower bound of -100 when the RF classifier was trained using data from a high-default regime and predictions were made in a low-default regimes.[58] All of the $EvA_{-k,k}$ measures from the LOO RF model exceed the minimum values of $EvA_{j,k}$. Thus while we find no evidence that training the RF classifier on data from a mix of macroeconomic conditions reduces the likelihood of experiencing extreme overpredictions of default activity, our exercise does suggest that including data from a wide sample of time periods bounds RF prediction errors away from the most severe underpredictions.

To get a better sense of what might be driving the improved predictive performance of the LOO models relative to the YBY models, we estimated the average marginal effects (AMEs) associated with the LTV and FICO variables that enter our Logit model. The AMEs for each of the holdout years are reported in Tables 11 and 12.[59] Relative to the AMEs reported for the YBY models in Tables 3 and 4 that varied enormously across different training samples, the AMEs from the LOO analysis are quite stable across different samples. This stability is unsurprising given the nature of the LOO exercise. The composition of the training data changes little as we change the holdout years.

Taken as a whole, the results of our LOO exercise suggest that when a Logit model is trained using data drawn from a wide variety of economic conditions, the out-of-sample forecast errors will be bounded away from some of the extreme overpredictions and underpredictions that we witnessed when the Logit model was

[55] See Appendix D for more detail on how this threshold was selected.

[56] Note that we do not use the SMOTE procedure in the construction of the LOO training samples for two reasons. First, $S_{-k}^{Default}$ for the LOO analysis contained far more defaults than any of the year-by-year samples for each training year k. There was thus limited value in adding synthetic defaults to any of the training samples. Second, the sheer size of the training sample for the LOO analysis rendered the SMOTE procedure computationally infeasible.

[57] We report the full regression results for these Logit models in Appendix C

[58] For instance, the RF model that was trained using purchase loans originated in 2007 –the cohort with the highest default rate in our sample –predicted that there would be only 543 defaults in the 2015 origination cohort, which is the cohort with the lowest default rate in our sample. There were 9619 defaults in the 2015 validation sample, and the prediction error measure in this case was $EvA_{2007,2015} = 100 * \left(\frac{543-9619}{9619}\right) = -94.36$.

[59] We report the full set of estimated parameters for the Logit LOO analysis in Appendix C.

Table 9: Results of Leave-One-Out Analysis for Purchase Loans

	Purchase Loans **Model Type**											
	Logit Without Perfect Foresight				Logit With Perfect Foresight				Random Forest With Perfect Foresight			
	LOO[1]	YBY $EvA_{j,k}$			LOO	YBY $EvA_{j,k}$			LOO	YBY $EvA_{j,k}$		
Holdout Year	$EvA_{-k,k}$	Average	Minimum	Maximum	$EvA_{-k,k}$	Average	Minimum	Maximum	$EvA_{-k,k}$	Average	Minimum	Maximum
2004	93.83	14.38	-71.15	276.37	-14.44	-36.72	-73.84	10.09	113.52	82.57	-48.30	291.18
2005	47.31	-14.05	-76.40	168.23	-6.99	-35.87	-74.45	17.39	118.71	52.91	-51.52	151.67
2006	-29.40	-47.90	-82.77	47.30	-16.42	-45.41	-79.72	0.09	157.44	-33.66	-88.71	81.17
2007	-55.42	-61.23	-84.76	0.00	-3.56	-50.63	-90.02	0.00	216.34	-56.80	-95.65	55.36
2008	-45.04	-47.28	-77.13	12.17	-5.55	-35.01	-78.66	16.37	177.08	-50.16	-78.88	-2.44
2009	17.02	10.34	-45.91	119.45	31.52	26.74	-25.80	117.01	189.17	-36.09	-91.63	23.11
2010	30.04	23.49	-40.07	147.57	17.51	33.88	-26.84	157.45	135.47	-40.51	-81.09	36.24
2011	42.14	32.61	-40.43	177.89	-12.52	8.42	-38.70	92.23	-3.02	-33.44	-87.89	52.63
2012	67.67	54.65	-39.97	225.45	-1.10	13.08	-38.68	97.91	4.37	-8.94	-96.49	454.91
2013	84.39	73.45	-12.92	289.96	8.82	35.18	-14.05	131.30	-50.68	-21.64	-94.10	115.46
2014	111.51	95.55	-1.55	352.43	27.74	53.35	-5.99	165.95	-51.12	-32.70	-88.56	204.92
2015	110.57	95.89	0.00	362.89	32.24	46.06	-3.11	135.61	-26.76	-25.37	-97.50	156.09
2016	61.60	46.76	-28.09	263.00	5.42	7.55	-37.18	85.77	-86.17	-56.62	-99.10	158.96

[1] LOO denotes the results from the leave-one-out estimation described in the text.

Table 10: Results of Leave-One-Out Analysis for Refinance Loans

	Refinance Loans **Model Type**											
	Logit Without Perfect Foresight				Logit With Perfect Foresight				Random Forest With Perfect Foresight			
	LOO[1]	Year-By-Year $EvA_{j,k}$			LOO	Year-By-Year $EvA_{j,k}$			LOO	Year-By-Year $EvA_{j,k}$		
Holdout Year	$EvA_{-k,k}$	Average	Minimum	Maximum	$EvA_{-k,k}$	Average	Minimum	Maximum	$EvA_{-k,k}$	Average	Minimum	Maximum
2004	200.75	86.88	-56.17	596.25	99.06	-21.64	-67.13	103.77	76.22	275.44	-51.62	688.40
2005	145.78	40.63	-64.78	405.05	33.40	-14.00	-59.93	136.26	111.67	237.90	-27.98	560.99
2006	-49.83	-36.89	-81.70	113.61	-49.83	-34.49	-81.12	76.40	264.32	17.77	-90.60	379.34
2007	-79.20	-66.49	-87.36	0.00	-79.20	-52.69	-93.31	11.62	375.80	-66.05	-97.05	-5.06
2008	-52.54	-59.55	-83.08	1.89	-70.28	-49.58	-87.52	0.00	237.29	-56.33	-87.73	-3.74
2009	18.42	-7.52	-61.21	140.63	1.41	-4.44	-45.13	93.06	155.93	26.12	-88.58	160.60
2010	68.13	32.51	-47.02	282.15	53.08	33.36	-27.90	153.81	87.04	-12.37	-71.12	169.54
2011	82.47	47.87	-48.39	356.64	6.56	6.11	-37.26	101.68	-53.99	73.54	-75.09	256.34
2012	151.26	110.56	-27.53	538.84	-12.43	9.84	-48.47	122.00	-62.32	182.33	-84.75	1115.88
2013	177.87	119.83	-18.87	552.13	22.53	36.61	-17.71	163.15	38.25	216.39	-74.33	572.21
2014	199.43	138.46	-15.33	639.45	303.18	59.18	0.00	213.67	1.63	137.81	-56.30	608.36
2015	263.76	194.23	0.00	857.61	563.20	62.78	0.00	274.31	-16.12	227.63	-79.73	743.20
2016	151.69	103.71	-31.32	557.62	-22.60	0.30	-64.51	149.46	-84.67	43.28	-90.26	565.54

[1] LOO denotes the results from the leave-one-out estimation described in the text.

Table 11: Average Marginal Effects of LTV and FICO Variables for Leave-One-Out "Perfect Foresight" Logit Models: Purchase Mortgages

	Dependent Variable: Default within 24 Months of Origination[1] **Average Marginal Effects**[2] *Purchase Loans* Holdout Data Cohort												
Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
LTV Buckets[3]													
$60 < LTV \leq 70$	0.020	0.020	0.018	0.015	0.018	0.023	0.021	0.021	0.021	0.021	0.021	0.021	0.022
$70 < LTV \leq 80$	0.035	0.034	0.030	0.028	0.033	0.039	0.037	0.037	0.037	0.038	0.037	0.037	0.038
$80 < LTV \leq 90$	0.048	0.046	0.041	0.039	0.044	0.052	0.049	0.050	0.050	0.050	0.050	0.050	0.051
$90 < LTV \leq 95$	0.053	0.051	0.046	0.043	0.050	0.057	0.054	0.055	0.055	0.056	0.055	0.055	0.056
$95 < LTV \leq 99$	0.056	0.055	0.049	0.046	0.053	0.061	0.058	0.059	0.059	0.059	0.059	0.059	0.060
$LTV > 99$	0.065	0.063	0.058	0.054	0.063	0.070	0.067	0.068	0.068	0.068	0.068	0.068	0.068
FICO Buckets													
$550 < FICO \leq 650$	-0.040	-0.039	-0.037	-0.035	-0.038	-0.040	-0.041	-0.042	-0.042	-0.042	-0.041	-0.041	-0.041
$650 < FICO \leq 700$	-0.072	-0.070	-0.068	-0.064	-0.067	-0.073	-0.074	-0.076	-0.076	-0.077	-0.076	-0.076	-0.075
$700 < FICO \leq 750$	-0.103	-0.101	-0.097	-0.093	-0.096	-0.103	-0.105	-0.107	-0.108	-0.108	-0.107	-0.107	-0.107
$750 < FICO \leq 800$	-0.141	-0.138	-0.132	-0.126	-0.130	-0.141	-0.143	-0.145	-0.147	-0.147	-0.146	-0.146	-0.146
$FICO > 800$	-0.161	-0.158	-0.151	-0.143	-0.146	-0.161	-0.164	-0.166	-0.168	-0.168	-0.167	-0.167	-0.169

[1] A loan is classified as defaulting if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] To estimate the average marginal effect (AME) associated with moving from the reference category of a categorical variable to some category of interest x, we proceed as follows. First, for every observation in the training sample we set the categorical variable equal to the reference category and use the estimated parameters from the Logit model to predict the probability of default while holding the other covariates at their observed values. Second, we change the value of x to one and predict the probability default. Third, we take the difference between these two predicted probabilities. Lastly, to construct the AME we average over all of these observation-specific differences.
[3] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.

Table 12: Average Marginal Effects of LTV and FICO Variables for Leave-One-Out “Perfect Foresight” Logit Models: Refinance Mortgages

	Dependent Variable: Default within 24 Months of Origination[1] **Average Marginal Effects**[2] *Refinance Loans* Holdout Data Cohort												
Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
LTV Buckets[3]													
$60 < LTV \leq 70$	0.018	0.018	0.016	0.013	0.014	0.018	0.019	0.019	0.019	0.019	0.018	0.018	0.018
$70 < LTV \leq 80$	0.030	0.030	0.027	0.021	0.024	0.031	0.032	0.032	0.032	0.032	0.030	0.030	0.031
$80 < LTV \leq 90$	0.041	0.041	0.036	0.030	0.035	0.042	0.043	0.043	0.043	0.043	0.041	0.041	0.042
$90 < LTV \leq 95$	0.046	0.046	0.041	0.034	0.040	0.047	0.049	0.048	0.049	0.049	0.047	0.047	0.047
$95 < LTV \leq 99$	0.050	0.049	0.044	0.037	0.043	0.050	0.052	0.052	0.053	0.053	0.050	0.050	0.051
$LTV > 99$	0.057	0.057	0.051	0.043	0.049	0.058	0.059	0.059	0.059	0.061	0.057	0.058	0.058
FICO Buckets													
$550 < FICO \leq 650$	-0.026	-0.026	-0.024	-0.022	-0.023	-0.027	-0.028	-0.028	-0.028	-0.028	-0.027	-0.027	-0.027
$650 < FICO \leq 700$	-0.045	-0.045	-0.041	-0.037	-0.038	-0.048	-0.048	-0.048	-0.049	-0.048	-0.046	-0.046	-0.047
$700 < FICO \leq 750$	-0.063	-0.062	-0.058	-0.052	-0.054	-0.065	-0.067	-0.067	-0.068	-0.067	-0.064	-0.064	-0.065
$750 < FICO \leq 800$	-0.090	-0.090	-0.083	-0.074	-0.076	-0.093	-0.095	-0.096	-0.096	-0.096	-0.091	-0.092	-0.093
$FICO > 800$	-0.109	-0.108	-0.101	-0.087	-0.090	-0.111	-0.114	-0.116	-0.116	-0.116	-0.110	-0.111	-0.112

[1] A loan is classified as defaulting if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] To estimate the average marginal effect (AME) associated with moving from the reference category of a categorical variable to some category of interest x, we proceed as follows. First, for every observation in the training sample we set the categorical variable equal to the reference category and use the estimated parameters from the Logit model to predict the probability of default while holding the other covariates at their observed values. Second, we change the value of x to one and predict the probability default. Third, we take the difference between these two predicted probabilities. Lastly, to construct the AME we average over all of these observation-specific differences.
[3] For refinance mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the property's appraised value.

trained using only data from one particular origination cohort. Based on the results from our marginal effects analysis, it appears that the less extreme forecast errors for the LOO models stems from the stability of the estimated parameters across the different LOO models. In Section 4 we posited that one of the mechanisms underlying the poor out-of-sample predictive performance of mortgage default models is likely due to a highly unstable relationship between variables typically used in credit risk models and the arrival rate of trigger events following origination. In the context of this argument, the results of our LOO analysis suggest that training predictive Logit models on data drawn from a wide variety of macroeconomic conditions smooths over the volatility in the parameter estimates seen in the YBY analysis, and this increased parameter stability bounds the forecasts of mortgage performance away from extreme underpredictions and overpredictions. If we view the parameters estimated using data from throughout the economic cycle as a type of averaging of the parameters associated with using only particular cohorts to train the data, then our finding that using "average" parameters reduces forecast volatility is conceptually similar to results in the time series literature suggesting that averaging over many different forecasts can improve forecast accuracy [Timmerman, 2006]. It is worth noting that the averaging associated with using through-the-cycle data does not produce out-of-sample forecasts that are uniformly superior to those based on training data from a specific point in time. It may thus be possible to improve out-of-sample model performance further by limiting the training data to loans originated during time periods most similar in terms of economic conditions to those experienced by the validation data.

Notably, our findings in the LOO analysis that using data from a wide variety of macroeconomic conditions bounds prediction errors does not appear to generalize to the RF classifier. While we find that training the RF classifier using data from throughout the economic cycle reduces the frequency of extreme underpredictions of delinquency, in many cases using this through-the-cycle data to train the RF classifier resulted overpredictions that exceeded those produced by the RF model using the YBY training samples. The non-parametric nature of the RF classifier makes identifying the mechanisms underlying the differences in the performance of the LOO Logit model and LOO RF classifier extremely difficult and beyond the scope of this paper.

6 Conclusion

While components of the mortgage approval process have long been automated, recent changes in lending regulations have rapidly expanded the portion of the first-lien mortgage market that could qualify for nearly completely automated underwriting; this automation relies heavily on the outputs of statistical and machine learning models that are designed to predict loan performance. Understanding the risk associated with using outputs from these models to make decisions (i.e.,"model risk") will thus be of paramount importance for ensuring safe and sound mortgage lending practices in the years to come. Previously published work on the extent of model risk in the mortgage market was limited to the subprime lending boom and did not utilize machine learning methods that have become increasingly popular for predictive analytics in consumer finance in recent years. One of the primary contributions of this paper is thus providing evidence on the performance of both traditional statistical models and machine learning methods in the post-crisis lending environment. For both types of models, we find that predictive accuracy deteriorates rapidly when models that are trained in one type of macroeconomic environment are used to predict loan performance in out-of-time samples characterized by much different economic conditions.

The second major contribution of our paper is studying *why* this deterioration in predictive accuracy occurs. Using data from the Panel Study of Income Dynamics reveals that the relationship between predictors typically used to predict mortgage performance and the *subsequent* arrival of mortgage default trigger events has exhibited significant intertemporal heterogeneity. These findings imply that the failure of models documented in previous research was not simply a by-product of information asymmetries in boom-era mortgage lending; rather, our results suggest that the poor out-of-time performance of both statistical and machine learning methods is structural in nature. While systematically investigating possible "fixes" for this model failure is beyond the scope of this paper, we do provide evidence that model instability can be at least partially alleviated for traditional statistical models through the use of data from a wide mix of economic conditions. Interestingly, diversifying the macroeconomic conditions from which the training data are sampled did not generally improve the out-of-time performance of the machine learning methods that we considered. We leave further investigation of how the out-of-time performance of statistical models and machine learning models alike might be improved to future work.

References

An, X., Y. Deng, E. Rosenblatt and V. Yao. 2012. "Model Stability and the Subprime Mortgage Crisis." *The Journal of Real Estate Finance and Economics* 45(3):545–568.

Anderson, R. 2007. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. New York, NY: Oxford University Press.

An Overview of Enterprise Appraisal Waivers. 2018. Federal Housing Finance Agency Office of the Inspector General White Paper.

Athey, S. and G. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11:685–725.

Avery, R., M. Bilinski, A. Clement, T. Critchfield, I. Keith, I. Mohamed, S. Patrabansh, C. Redmer, J. Schultz and R. Sullivan. 2018. "National Survey of Mortgage Originations Technical Documentation." Federal Housing Finance Agency NMDB Technical Report 2.

Avery, R., M. Bilinski, A. Clement, T. Critchfield, I. Keith, I. Mohamed, S. Patrabansh, C. Redmer, J. Schultz and R. Sullivan. 2019. "National Survey of Mortgage Originations Public Use File Codebook and Tabulations." Federal Housing Finance Agency NMDB Technical Report 2.

Bailey, M., E. Dávila, T. Kuchler and J. Stroebel. 2019. "House Price Beliefs And Mortgage Leverage Choice." *The Review of Economic Studies* pp. 1–50.

Ball, L. 1999. Policy Rules for Open Economies. In *Monetary Policy Rules*, ed. J. Taylor. Chicago, IL: University of Chicago Press pp. 127–156.

Berg, T., V. Burg, A. Gombović and M. Puri. 2018. "On the Rise of FinTechs–Credit Scoring Using Digital Footprints." *National Bureau of Economic Research Working Paper* .

Breiman, L. 2001. "Random Forests." *Machine Learning* 45(1):5–32.

Buchak, G., G. Matvos, T. Piskorski and A. Seru. 2018. "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks." *Journal of Financial Economics* 130(3):453–483.

Case, K., R. Shiller and A. Thompson. 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity* pp. 265 – 315.

Chawla, N., N. Japkowicz and A. Kotcz. 2004. "Special Issue on Learning from Imbalanced Data Sets." *ACM Sigkdd Explorations Newsletter* 6(1):1–6.

Chawla, N. V, K. Bowyer, L. Hall and W. Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research* 16:321–357.

Chen, C., A. Liaw and l. Breiman. 2004. "Using Random Forest to Learn Imbalanced Data." University of California Department of Statistics Technical Report 666.

Duffie, D., A. Eckner, G. Horel and L. Saita. 2009. "Frailty Correlated Default." *The Journal of Finance* 64(5):2089–2123.

Engle, R., D. Hendry and J. Richard. 1983. "Exogeneity." *Econometrica* 51(2):277–304.

Ericsson, N. and J. Irons. 1995. The Lucas Critique in Practice. In *Macroeconometrics*, ed. K. Hoover. Vol. 46 Dordrecht: Springer pp. 263–324.

Estrella, A. and J. Fuhrer. 1999. "Are "Deep" Parameters Stable?: The Lucas Critique as an Empirical Hypothesis." *Federal Reserve Bank of Boston Working Paper* (99-4).

Final Rule: Real Estate Appraisals. 2019. *Federal Register* 84(195):53579–53598.

Frame, W., L. Wall and L. White. 2018. "Technological Change and Financial Innovation in Banking: Some Implications for Fintech." *Federal Reserve Bank of Atlanta Working Paper* .

Frisch, R. 1938. "Statistical versus Theoretical Relations in Economic Macrodynamics." League of Nations Memorandum.

Fuhrer, J. 1997. "The (Un)importance of Forward-Looking Behavior in Price Specifications." *Journal of Money, Credit, and Banking* 29(3):338–350.

Fuster, A., M. Plosser, P. Schnabl and J. Vickery. 2018. "The Role of Technology in Mortgage Lending." *National Bureau of Economic Research Working Paper* .

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai and A. Walther. 2018. "Predictably Unequal The Effects of Machine Learning on Credit Markets." *Working Paper* .

Fuster, A. and P. Willen. 2017. "Payment Size, Negative Equity, and Mortgage Default." *American Economic Journal: Economic Policy* 9(4):167–91.

Gerardi, K., K. Herkenhoff, L. Ohanian and P. Willen. 2017. "Can't Pay or Won't Pay? Unemployment, Negative Equity, and Strategic Default." *The Review of Financial Studies* 31(3):1098–1131.

Greenwell, B. 2017. "PDP: An R Package for Constructing Partial Dependence Plots." *The R Journal* 9(1):421–436.

Haavelmo, T. 1944. "The Probability Approach in Econometrics." *Econometrica* 12:1–115.

Harney, K. 2017. "Fannie and Freddie Say Appraisals Are Not Always Necessary." *The Washington Post* . `https://www.washingtonpost.com/realestate/fannie-and-freddie-say-appraisals-are-not-always-necessary/2017/08/29/cd4c96e4-8c1d-11e7-8df5-c2e5cf46c1e2_story.html?utm_term=.92bb6dcce72a` Accessed 1/23/2019.

Hastie, T., R. Tibshirani and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2 ed. New York, NY: Springer.

Hau, H., Y. Huang, H. Shan and Z. Sheng. 2019. "How FinTech Enters China's Credit Market." *American Economic Association Papers and Proceedings* 109:60–64.

Hendry, D. 1988. "The Encompassing Implications of Feedback versus Feedforward Mechanisms in Econometrics." *Oxford Economic Papers* 40(1):132–149.

Hosmer, D., S. Lemeshow and R. Sturdivant. 2013. *Applied Logistic Regression.* Hoboken, NJ: John Wiley & Sons.

Housing Finance at a Glance: A Monthly Chartbook. 2018. Urban Institute Housing Finance Policy Center Report.

Housing Finance at a Glance: A Monthly Chartbook. 2019. Urban Institute Housing Finance Policy Center Report.

Hsu, J., D. Matsa and B. Melzer. 2018. "Unemployment Insurance as a Housing Market Stabilizer." *American Economic Review* 108(1):49–81.

Interagency Appraisal and Evaluation Guidlines. 2010. *Federal Register* 75(237):77450–77473.

Jagtiani, J. and C. Lemieux. 2018. "The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform." *Federal Reserve Bank of Philadelphia Working Paper* .

Karikari, J., I. Voicu and I. Fang. 2011. "FHA vs. Subprime Mortgage Originations: Is FHA the Answer to Subprime Lending?" *The Journal of Real Estate Finance and Economics* 43(4):441–458.

Khandani, A., A. Kim and A. Lo. 2010. "Consumer Credit-Risk Models via Machine-Learning Algorithms." *Journal of Banking & Finance* 34(11):2767–2787.

Kim, Y., S. Laufer, K. Pence, R. Stanton and N. Wallace. 2018. "Liquidity Crises in the Mortgage Market." *Brookings Papers on Economic Activity* .

Kirasich, K., T. Smith and B. Sadler. 2018. "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets." *SMU Data Science Review* 1(3):9.

Lindé, J. 2001. "Testing for the Lucas Critique: A Quantitative Investigation." *American Economic Review* 91(4):986–1005.

Lubik, T. and P. Surico. 2010. "The Lucas Critique and the Stability of Empirical Models." *Journal of Applied Econometrics* 25(1):177–194.

Lucas, R. 1983. Econometric Policy Evaluation: A Critique. In *Theory, Policy, Institutions: Papers from the Carnegie-Rochester Conferences on Public Policy*, ed. K. Brunner and A. Meltzer. Amsterdam: North Holland pp. 257–284.

Niculescu-Mizil, A. and R. Caruana. 2005. Predicting Good Probabilities with Supervised Learning. In *Proceedings of the 22nd International Conference on Machine Learning*. ACM pp. 625–632.

Rajan, U., A. Seru and V. Vig. 2010. "Statistical Default Models and Incentives." *American Economic Review* 100(2):506–10.

Rajan, U., A. Seru and V. Vig. 2015. "The Failure of Models That Predict Failure: Distance, Incentives, and Defaults." *Journal of Financial Economics* 115(2):237–260.

Rodrigues, K., B. Gebre and J. Liu. 2016. "The Future of Technology in Mortgage Originations." Oliver Wyman White Paper.

Rudebusch, G. and L. Svensson. 1999. Policy Rules for Inflation Targeting. In *Monetary Policy Rules*, ed. J. Taylor. Chicago, IL: University of Chicago Press pp. 203–262.

Supervisory Guidance on Model Risk Management. 2011. Office of the Comptroller of the Currency Bulletin 2011-12.

Svensson, L. 1997. "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets." *European Economic Review* 41(6):1111–1146.

Taylor, J. 1999. "The Robustness and Efficiency of Monetary Policy Rules as Guidelines for Interest Rate Setting by the European Central Bank." *Journal of Monetary Economics* 43(3):655–679.

Thomas, L. 2009. *Consumer Credit Models: Pricing, Profit and Portfolios: Pricing, Profit and Portfolios*. New York, NY: Oxford University Press.

Timmerman, A. 2006. Forecast Combinations. In *Handbook of Economic Forecasting: Volume 1*, ed. G. Elliott, C. Granger and A. Timmerman. North Holland, Amsterdam pp. 136–194.

Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

For Online Publication

A National Survey of Mortgage Originations (NSMO) Analysis

Previous work suggests that one of the mechanisms potentially driving the failure of predictive models of mortgage performance concerns the forward-looking nature of consumers. Specifically, Bailey et al. [2019] provide evidence that loan-to-value ratios (LTV) reflect borrowers' expectations for future house-price growth and that the relationship between price expectations and leverage changes over time with the state of the economy. If the relationship between expected house price growth and realized house price appreciation also changes over the course of the economic cycle, then predictive models constructed using variables that reflect borrower expectations will be inherently unstable.[60]

More formally, Bailey et al.'s [2019] analysis implies that $LTV_{i,s}$ is a function of borrower price expectations between time period s and $s+t$. That is,

$$LTV_{i,s} = f_s\left(\text{Price Expectations}_i^{s,s+t}\right) \tag{A.1}$$

where i indexes borrowers. $f_s\left(\right)$ is indexed by s to allow for the relationship between leverage and expectations to vary over time as in Bailey et al.'s [2019] empirical work. Case, Shiller and Thompson's [2012] findings that expectations are predictive of house price appreciation – denoted $HPA_{i,s,s+t}$ – and that the relationship between expectations and $HPA_{i,s,s+t}$ varies over time imply that

$$HPA_{i,s,s+t} = g_s\left(\text{Price Expectations}_i^{s,s+t}\right) \tag{A.2}$$

Inverting Equation A.1 allows us to write

$$\text{Price Expectations}_i^{s,s+t} = f_s^{-1}\left(LTV_{i,s}\right) \tag{A.3}$$

Substituting Equation A.3 into Equation A.2, we can express realized house price appreciation as a function of $LTV_{i,s}$

$$HPA_{i,s,s+t} = g_s\left(f_s^{-1}\left(LTV_{i,s}\right)\right) \tag{A.4}$$

Lastly, because default trigger events, denoted $Event_{i,s,s+t}$, are in part determined by whether a borrower has negative equity and negative equity status is dependent upon house price appreciation, we have

$$Event_{i,s,s+t} = \phi_s\left(HPA_{i,s,s+t}\right) \tag{A.5}$$

The arrival of a trigger event in this framework is thus a time-varying function of $LTV_{i,s}$

$$Event_{i,s,s+t} = \phi_s\left(g_s\left(f_s^{-1}\left(LTV_{i,s}\right)\right)\right) \tag{A.6}$$

The framework above suggests two approaches for studying the mechanisms that drive model failure. First, Equation A.6 can be estimated using data on borrower leverage choice and the subsequent arrival of trigger events; we conduct this analysis in the main text in Section 4.2. Second, Equation A.2 implies that a model's predictive accuracy will deteriorate if the relationship between borrower home-price expectations and LTV is unstable over time. Testing this hypothesis is the goal of this appendix. Because servicing records do not report information on borrower expectations, this analysis cannot be conducted using traditional administrative mortgage performance records like those that comprise the McDash database. This intertemporal stability of the relationship borrower leverage and price expectations can, however, be studied using data from the National Survey of Mortgage Originations (NSMO).

The Housing and Economic Recovery Act (HERA) of 2008 required that the Federal Housing Finance Agency (FHFA) "must, through a monthly survey of the mortgage market, collect data on the characteristics

[60] The most recent boom-bust cycle in the housing market provides a nice case in point of how the relationship between expected and realized house prices moves can change over time. Using survey data, Case, Shiller and Thompson [2012] document that between 2003 and 2006, household price expectations generally aligned with realized post-survey price appreciation. Starting in 2007, however, price expectations and actual price movements diverged significantly.

of individual mortgages, including those eligible for purchase by Fannie Mae and Freddie Mac and those that are not, and including subprime and nontraditional mortgages. In addition, FHFA must collect information on the creditworthiness of borrowers, including a determination of whether subprime and nontraditional borrowers would have qualified for prime lending"[Avery et al., 2018, p. 1]. In response to this requirement, the FHFA, jointly with the Consumer Financial Protection Bureau (CFPB), constructed the National Mortgage Database (NMDB).

There are three primary components to the NMDB. The first component is a database that reports information derived from credit bureau records on the characteristics of a nationally representative sample of first-lien mortgage originations. The second component of the NMDB is the American Survey of Mortgage Borrowers (ASMB), which was designed to collect information on borrowers' experiences repaying their mortgage loans.

The third component of the NMDB is the NSMO data.[61] Beginning in 2013, roughly 1 in every 15 loan originations that were included in the NMDB were selected to receive a survey soliciting information on borrowers related to their financial health, mortgage shopping behavior, and, most importantly for our purposes, borrower expectations for future house price growth. We use the NSMO data to test the hypothesis that the LTV variables used in mortgage performance models are systematically correlated with borrowers' expectations. The specific version of the NSMO data that we use in our analysis contains responses from borrowers that took our first-lien mortgages between January of 2013 and December of 2016.

While the NSMO records contain an incredibly rich set of variables describing a borrower's creditworthiness, loan characteristics, and shopping behavior, our analysis focuses on 3 specific variables. The variable that we use to measure borrower expectations for house price growth is based on the following question in NSMO: "What do you think will happen to the prices of homes in this neighborhood over the next couple of years?" Respondents were given the option of choosing one of the following five responses: increase a lot; increase a little; remain about the same; decrease a little; decrease a lot [Avery et al., 2019]. The variable from NSMO that we use to measure household leverage is the combined loan-to-value (CLTV) as of origination. To measure a borrower's creditworthiness, which in turn determines lender CLTV limits, we use a field in NSMO that reports the respondent's VantageScore 3.0 credit score at the time that the loan was originated.[62]

We limit the data in our analysis to purchase and refinance loans with a valid VantageScore value and a back-end debt-to-income ratio of less than 55. For the summary statistics and regression results, observations were weighted using the NSMO "analysis weight" field. We report in Table 13 the sample means of the variables that we used in our analysis both for the full NSMO sample as well as by loan origination year. On average, CLTV and VantageScore changed little over the course of our sample. Borrowers' expectations for post-origination property price changes, however, did shift over the course of our data frame. In 2013, for example, 15 percent of borrowers expected property values in their neighborhood to increase "a lot over the next couple of years." The fraction of borrowers in this category increased monotonically over the course of the sample, peaking at 22 percent in 2016. This increase in expectations for home price growth primarily came from fewer borrowers indicating that they expected property prices to "remain about the same" or "increase a little." In every year in the NSMO sample, less than 5 percent of borrowers indicated that they expected property values to decline at all.

We use the full sample of originations to study whether CLTV is correlated with borrower house price expectations and thus potentially a source of model instability by estimating a regression of the CLTV variable on the price expectations variables, dummy variables created using the VantageScore variables that mimic the FICO score buckets utilized in Section 3, and an indicator variable that is equal to one if the loan was used to refinance existing debt on the borrower's home. The results of this regression are reported in the first column in Table 14. Then, to get a sense of whether the nature of the correlation between leverage and expectations changes over time, we estimated the same regression on a year-by-year basis; the results of this analysis are reported in the second through fifth columns of Table 14.

In the regression on the pooled sample, price expectations cleanly rank order CLTV.[63] Specifically, we

[61] More information on the structure of the NSMO data is available at https://www.fhfa.gov/DataTools/Downloads/Pages/National-Survey-of-Mortgage-Originations-Public-Use-File.aspx.

[62] VantageScore 3.0 ranges from 300 to 850, which is the same range as the Base FICO scores that we use in our mortgage performance analysis.

[63] This finding is consistent with Bailey et al.'s [2019] theoretical model for "fixed house size" households whose demand for

Table 13: Summary Statistics for the National Survey of Mortgage Originations Data

Origination Year	Obs.[1]	CLTV[2]	Vantage Score	Loan Type Purchase	Refinance	Average Values Price Expectations[3] Increase A Lot	Increase A Little	Stay The Same	Decrease A Little	Decrease A Lot
2013	5411	76.68	743.05	0.40	0.60	0.15	0.63	0.18	0.03	0.01
2014	5583	78.69	734.07	0.60	0.40	0.17	0.62	0.18	0.02	0.01
2015	5599	77.51	732.12	0.53	0.47	0.20	0.60	0.18	0.02	0.00
2016	5742	76.46	734.05	0.51	0.49	0.22	0.61	0.15	0.02	0.00
All Years	22335	77.19	736.22	0.50	0.50	0.18	0.62	0.17	0.02	0.01

[1] While we utilize the analysis weight field to construct the sample means, the observation count reported here corresponds to the unweighted number of cases from the survey that were used.
[2] CLTV denotes the combined loan-to-value ratio at the time of origination.
[3] The Price Expectations variables are based on respondents' answers to the following question: "What do you think will happen to the prices of homes in this neighborhood over the next couple of years?"

Table 14: Borrower Price Expectations and CLTV Regression Results

Variable	Dependent Variable: CLTV[1] All Years	Origination Year 2013	2014	2015	2016
Price Expectations[2]					
Increase a Lot	-8.050	-10.59	-4.597	-5.620	-6.256
	(2.074)***	(3.849)***	(3.534)	(2.893)*	(3.312)*
Increase a Little	-6.570	-8.605	-3.623	-4.196	-5.353
	(2.061)***	(3.806)**	(3.499)	(2.865)	(3.287)
Remain About the Same	-5.622	-7.040	-2.824	-3.307	-5.165
	(2.076)***	(3.842)*	(3.529)	(2.899)	(3.323)
Decrease a Little	-4.477	-6.367	-0.0285	-2.531	-5.592
	(2.265)**	(4.146)	(4.128)	(3.205)	(3.949)
VantageScore Buckets					
$550 < VantageScore \leq 650$	5.114	3.732	22.42	3.139	-3.034
	(3.143)	(4.697)	(5.965)***	(4.877)	(6.390)
$650 < VantageScore \leq 700$	2.801	0.979	22.05	0.481	-6.202
	(3.134)	(4.654)	(5.934)***	(4.862)	(6.379)
$700 < VantageScore \leq 750$	-0.344	-2.128	17.68	-1.813	-9.499
	(3.129)	(4.639)	(5.928)***	(4.851)	(6.370)
$750 < VantageScore \leq 800$	-5.065	-6.951	12.52	-6.741	-13.78
	(3.127)	(4.631)	(5.927)**	(4.849)	(6.364)**
$VantageScore > 800$	-12.97	-16.23	6.451	-14.55	-21.55
	(3.132)***	(4.643)***	(5.936)	(4.863)***	(6.371)***
Refinance Loan?	-13.11	-11.29	-13.00	-13.81	-14.92
	(0.249)***	(0.515)***	(0.480)***	(0.458)***	(0.476)***
Intercept	93.25	97.52	71.95	92.56	100.8
	(3.736)***	(5.966)***	(6.902)***	(5.618)***	(7.133)***
Observations	22335	5411	5583	5599	5742

[1] CLTV denotes the combined loan-to-value ratio at the time of origination.
[2] The Price Expectations variables are based on respondents' answers to the following question: "What do you think will happen to the prices of homes in this neighborhood over the next couple of years?" The reference category for the Price Expectations variables is "Decline a Lot."
All regressions utilize the "analysis weights" reported in the NSMO data.
Robust standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

find that as expectations for price growth increase, CLTV declines. These findings support the primary

housing consumption is primarily driven by non-financial factors.

hypothesis from Bailey et al.'s [2019] theoretical framework.

The estimated relationship between expectations and leverage is economically as well as statistically significant: relative to borrowers that expect property prices to " decrease a lot," the CLTV for borrowers that expect home values to "increase a lot" is predicted to be 8 percentage points lower, a sizable decline relative to the average CLTV of 77 in the sample. When we estimate the regression model on a year-by-year basis, the estimated parameters change significantly across origination cohorts in terms of both their magnitude and statistical significance. We interpret these results as evidence that borrower expectations are correlated with a measure of leverage that is universally used in predictive models of mortgage default and that the relationship between CLTV and price expectations changes over time. Intertemporal heterogeneity in the relationship between borrower house price expectations and leverage may thus be one of the mechanisms driving the failure of predictive models of mortgage performance.

B Full Logistic Regression Estimation Results: Year-by-Year

Table 15: Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, Year-by-Year Estimation

	Dependent Variable: D_i [1]						
	Training Year						
Variable	2004	2005	2006	2007	2008	2009	2010
ΔHPI_i [2]	-0.0207	-0.0227	-0.0355	-0.0397	-0.0346	-0.0226	-0.0184
	(0.0009)***	(0.0007)***	(0.0005)***	(0.0003)***	(0.0004)***	(0.0009)***	(0.0010)***
$\Delta Rate_i$ [3]	-0.0001	-0.0133	0.0034	-0.0039	0.0081	0.0161	-0.0041
	(0.0040)	(0.0008)***	(0.0014)**	(0.0024)*	(0.0005)***	(0.0008)***	(0.0024)*
$\Delta Unemp_i$ [4]	-0.0053	-0.0005	0.0023	0.0015	0.0020	0.0002	-0.0009
	(0.0009)***	(0.0004)	(0.0002)***	(0.0001)***	(0.0001)***	(0.0004)***	(0.0006)
Non-Owner Occupied? [5]	-0.2065	-0.4900	-0.2205	0.0579	-0.1691	-0.1519	-0.6731
	(0.0634)***	(0.0297)***	(0.0223)***	(0.0158)**	(0.0139)***	(0.0276)***	(0.0485)***
Adjustable Rate? [6]	0.1632	0.7169	1.0959	0.9546	0.3363	0.0722	0.0949
	(0.0244)***	(0.0141)***	(0.0120)***	(0.0111)**	(0.0201)***	(0.0593)	(0.0373)**
Product Type [7]							
Conventional with PMI	-0.1584	-0.2095	-0.0604	0.3763	0.4253	0.2106	-0.3509
	(0.0355)***	(0.0200)***	(0.0155)***	(0.0103)**	(0.0132)***	(0.0260)***	(0.0525)***
FHA	0.2118	0.3070	0.0372	-0.0735	0.0254	0.4005	0.3745
	(0.0388)***	(0.0228)***	(0.0186)***	(0.0136)**	(0.0136)***	(0.0204)***	(0.0237)***
VA	-0.6422	-0.5964	-0.6799	-0.7728	-0.6043	-0.0729	-0.2308
	(0.0702)***	(0.0381)***	(0.0283)***	(0.0211)**	(0.0210)***	(0.0242)***	(0.0229)***
DTI Buckets [8]							
$0 < DTI \leq 20$	-0.5751	-0.7196	-0.7158	-0.1983	-0.1496	-0.6186	-0.8373
	(0.0391)***	(0.0283)***	(0.0233)***	(0.0183)***	(0.0186)***	(0.0330)***	(0.0457)***
$20 < DTI \leq 30$	-0.4577	-0.3620	-0.5169	-0.0294	-0.0087	-0.5473	-0.6014
	(0.0364)***	(0.0219)***	(0.0179)***	(0.0153)***	(0.0132)	(0.0205)***	(0.0239)***
$30 < DTI \leq 40$	-0.0882	-0.1192	-0.2831	0.1764	0.1114	-0.3125	-0.3002
	(0.0359)***	(0.0194)***	(0.0157)***	(0.0140)***	(0.0115)***	(0.0161)***	(0.0179)***
$DTI > 40$	0.2975	0.1826	-0.0616	0.3321	0.3399	0.0567	0.0416
	(0.0341)***	(0.0178)***	(0.0147)***	(0.0132)***	(0.0103)***	(0.0139)	(0.0153)***
Observations	306651	743497	734438	906161	1157604	1189008	1095221

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 15 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, Year-by-Year Estimation

	Dependent Variable: D_i[1]					
	Training Year					
Variable	2011	2012	2013	2014	2015	2016
ΔHPI_i[2]	-0.0145	-0.0146	-0.0295	-0.0290	-0.0188	-0.0157
	(0.0008)***	(0.0008)***	(0.0014)***	(0.0021)***	(0.0022)***	(0.0020)***
$\Delta Rate_i$[3]	-0.0002	-0.0144	0.0019	0.0046	0.0045	-0.0013
	(0.0004)	(0.0022)***	(0.0010)*	(0.0015)***	(0.0018)**	(0.0010)
$\Delta Unemp_i$[4]	0.0019	0.0027	-0.0033	0.0005	-0.0001	-0.0100
	(0.0008)**	(0.0009)***	(0.0008)***	(0.0006)	(0.0008)	(0.0008)***
Non-Owner Occupied?[5]	-1.0475	-0.9600	-0.6652	-0.3707	-0.0877	-0.0244
	(0.0719)***	(0.1095)***	(0.1015)***	(0.1089)***	(0.1038)	(0.0614)
Adjustable Rate?[6]	0.0606	-0.3245	-0.4907	-0.4377	-0.5917	-1.2978
	(0.0389)	(0.0815)***	(0.0892)***	(0.0906)***	(0.1325)***	(0.1782)***
Product Type[7]						
Conventional with PMI	-0.1007	-0.4352	-0.7567	-0.2369	0.3449	0.3638
	(0.0350)***	(0.0433)***	(0.0477)	(0.0544)***	(0.0674)***	(0.0459)***
FHA	0.4030	0.2543	0.1314	0.5658	1.0874	0.7238
	(0.0253)***	(0.0284)***	(0.0283)	(0.0436)***	(0.0640)***	(0.0469)***
VA	-0.3133	-0.2816	-0.2636	0.1924	0.7366	0.2780
	(0.0247)***	(0.0299)***	(0.0298)	(0.0476)***	(0.0766)***	(0.0627)***
DTI Buckets[8]						
$0 < DTI \leq 20$	-0.9850	-0.9868	-0.4536	0.9055	-0.3742	-0.3274
	(0.0485)***	(0.0629)***	(0.0697)***	(0.1210)***	(0.0973)***	(0.1369)**
$20 < DTI \leq 30$	-0.6620	-0.6217	-0.1681	1.0845	-0.3008	-0.1078
	(0.0255)***	(0.0308)***	(0.0468)***	(0.1061)***	(0.0805)***	(0.1258)
$30 < DTI \leq 40$	-0.3247	-0.2181	0.1957	1.5103	0.1028	0.2253
	(0.0190)***	(0.0215)***	(0.0419)***	(0.1027)***	(0.0756)	(0.1236)*
$DTI > 40$	0.0724	0.1697	0.6213	1.8701	0.4543	0.6211
	(0.0162)***	(0.0181)***	(0.0404)***	(0.1019)***	(0.0746)***	(0.1231)***
Observations	1046435	945750	944053	685955	678931	690784

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 15 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, Year-by-Year Estimation

	Dependent Variable: D_i [1]						
	Training Year						
Variable	2004	2005	2006	2007	2008	2009	2010
LTV Buckets [2]							
$60 < LTV \leq 70$	0.8125	0.6027	0.7491	0.8344	0.5680	0.0150	0.2431
	(0.1150)***	(0.0638)***	(0.0522)***	(0.0371)***	(0.0340)***	(0.0516)	(0.0871)***
$70 < LTV \leq 80$	1.2343	1.1894	1.5030	1.2646	0.8130	0.0885	0.4991
	(0.0964)***	(0.0523)***	(0.0438)***	(0.0318)***	(0.0292)***	(0.0420)**	(0.0715)***
$80 < LTV \leq 90$	1.5651	1.5932	1.9188	1.5472	1.1064	0.3184	0.8310
	(0.0965)***	(0.0524)***	(0.0440)***	(0.0319)***	(0.0288)***	(0.0400)***	(0.0689)***
$90 < LTV \leq 95$	1.9514	1.7810	2.0412	1.7906	1.2351	0.3972	1.0415
	(0.1012)***	(0.0556)***	(0.0467)***	(0.0329)***	(0.0294)***	(0.0404)***	(0.0690)***
$95 < LTV \leq 99$	2.0823	1.8468	2.1533	1.9160	1.2903	0.4353	1.0851
	(0.1000)***	(0.0550)***	(0.0459)***	(0.0325)***	(0.0291)***	(0.0391)***	(0.0678)***
$LTV > 99$	2.4221	2.2548	2.3824	2.0245	1.3625	0.7928	1.6257
	(0.1018)***	(0.0552)***	(0.0458)***	(0.0325)***	(0.0301)***	(0.0394)***	(0.0672)***
LTV Missing	2.0823	1.7836	0.7462	0.1403	0.2974	0.4016	0.7865
	(0.2496)***	(0.1223)***	(0.1409)***	(0.2108)	(0.1439)**	(0.2526)***	(0.2418)***
FICO Buckets							
$550 < FICO \leq 650$	-1.0043	-0.8946	-0.9993	-1.1143	-0.8422	-0.7249	-0.6317
	(0.0372)***	(0.0232)***	(0.0220)***	(0.0174)***	(0.0219)***	(0.0725)***	(0.1507)***
$650 < FICO \leq 700$	-2.0799	-1.9058	-1.9253	-1.9750	-1.7265	-1.6344	-1.4843
	(0.0420)***	(0.0251)***	(0.0232)***	(0.0185)***	(0.0226)***	(0.0727)***	(0.1508)***
$700 < FICO \leq 750$	-2.9039	-2.7015	-2.6285	-2.5330	-2.3575	-2.5107	-2.3717
	(0.0492)***	(0.0285)***	(0.0247)***	(0.0194)***	(0.0235)***	(0.0735)***	(0.1513)***
$750 < FICO \leq 800$	-3.7909	-3.6548	-3.6214	-3.4530	-3.2747	-3.4792	-3.3100
	(0.0670)***	(0.0373)***	(0.0297)***	(0.0217)***	(0.0255)***	(0.0757)***	(0.1526)***
$FICO > 800$	-3.7707	-3.8048	-4.0873	-4.1448	-4.0193	-4.0272	-3.5205
	(0.1933)***	(0.0908)***	(0.0702)***	(0.0468)***	(0.0488)***	(0.1023)***	(0.1652)***
$FICO$ Missing	-1.7524	-1.2893	-1.5194	-1.6478	-1.4817	-1.2788	-1.2052
	(0.0577)***	(0.0360)***	(0.0342)***	(0.0321)***	(0.0363)***	(0.0903)***	(0.1657)***
Intercept	-2.5002	-2.4804	-2.5987	-2.7727	-2.2265	-1.8776	-2.8921
	(0.1121)***	(0.0581)***	(0.0502)***	(0.0607)***	(0.0389)***	(0.0815)***	(0.1736)***
Observations	306651	743497	734438	906161	1157604	1189008	1095221

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 15 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, Year-by-Year Estimation

	Dependent Variable: D_i[1]					
	Training Year					
Variable	2011	2012	2013	2014	2015	2016
LTV Buckets[2]						
$60 < LTV \leq 70$	0.3918	0.3540	0.4135	0.3783	0.2289	-0.0468
	(0.0936)***	(0.1182)***	(0.1373)***	(0.1505)**	(0.1431)	(0.0911)
$70 < LTV \leq 80$	0.4736	0.3529	0.4981	0.4955	0.2198	0.0621
	(0.0795)***	(0.0993)***	(0.1149)***	(0.1251)***	(0.1182)*	(0.0709)
$80 < LTV \leq 90$	0.7388	0.8767	1.2128	0.8907	0.4071	0.0971
	(0.0767)***	(0.0953)***	(0.1111)***	(0.1231)***	(0.1184)***	(0.0737)
$90 < LTV \leq 95$	0.9523	1.1479	1.5358	1.1149	0.6348	0.2363
	(0.0765)***	(0.0951)***	(0.1111)***	(0.1235)***	(0.1199)***	(0.0766)***
$95 < LTV \leq 99$	1.0316	1.1829	1.6692	1.2848	0.7486	0.2911
	(0.0755)***	(0.0939)***	(0.1097)***	(0.1218)***	(0.1181)***	(0.0753)***
$LTV > 99$	1.6184	1.6374	1.9293	1.3856	0.7377	0.5378
	(0.0748)***	(0.0934)***	(0.1094)***	(0.1236)***	(0.1285)***	(0.0899)***
LTV Missing	0.6114	1.7075	2.6402	0.8592	1.8433	2.9534
	(0.2796)**	(0.3816)***	(0.3879)***	(1.0212)	(0.5377)***	(0.6816)***
FICO Buckets						
$550 < FICO \leq 650$	-0.8448	-0.6858	-1.2452	-1.5964	-1.4889	-0.6247
	(0.2785)***	(0.3157)***	(0.3600)***	(0.3818)***	(0.4591)***	(0.7437)
$650 < FICO \leq 700$	-1.6412	-1.4434	-1.9348	-2.2139	-2.1355	-1.2948
	(0.2785)***	(0.3157)***	(0.3599)***	(0.3816)***	(0.4589)***	(0.7436)*
$700 < FICO \leq 750$	-2.6046	-2.4256	-2.9365	-3.2058	-3.1219	-2.0736
	(0.2789)***	(0.3161)***	(0.3604)***	(0.3823)***	(0.4594)***	(0.7437)***
$750 < FICO \leq 800$	-3.5681	-3.4509	-3.9708	-4.1637	-4.0060	-2.7716
	(0.2797)***	(0.3173)***	(0.3617)***	(0.3841)***	(0.4606)***	(0.7439)***
$FICO > 800$	-3.9597	-3.8227	-4.5592	-4.2956	-4.2914	-2.7883
	(0.2900)***	(0.3296)***	(0.3784)***	(0.3966)***	(0.4718)***	(0.7457)***
$FICO$ Missing	-1.3135	-1.0965	-1.7000	-1.9228	-1.9608	-2.1711
	(0.2868)***	(0.3252)***	(0.3693)***	(0.3926)***	(0.4686)***	(0.7630)***
Intercept	-2.4060	-2.5371	-3.0486	-3.9624	-2.8047	-3.0643
	(0.2885)***	(0.3305)***	(0.3784)***	(0.4120)***	(0.4749)***	(0.7565)***
Observations	1046435	945750	944053	685955	678931	690784

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 16: Full "Perfect Foresight" Logistic Regression Model Results: Refinance Loans, Year-by-Year Estimation

	Dependent Variable: D_i[1]						
	Training Year						
Variable	2004	2005	2006	2007	2008	2009	2010
ΔHPI_i[2]	-0.0243	-0.0181	-0.0238	-0.0327	-0.0366	-0.0311	-0.0308
	(0.0011)***	(0.0008)***	(0.0005)***	(0.0003)***	(0.0004)***	(0.0010)***	(0.0014)***
$\Delta Rate_i$[3]	-0.0106	-0.0116	0.0029	-0.0433	-0.0010	0.0071	-0.0068
	(0.0045)**	(0.0010)***	(0.0013)**	(0.0021)***	(0.0005)**	(0.0008)***	(0.0029)***
$\Delta Unemp_i$[4]	-0.0007	-0.0010	0.0038	0.0010	0.0004	0.0028	-0.0031
	(0.0010)	(0.0005)**	(0.0001)***	(0.0001)***	(0.0001)***	(0.0004)***	(0.0008)***
Non-Owner Occupied?[5]	-0.0314	0.0445	0.3465	0.3813	-0.0705	-0.0440	-0.2073
	(0.0611)	(0.0303)	(0.0195)***	(0.0126)***	(0.0129)***	(0.0338)	(0.0451)***
Adjustable Rate?[6]	0.0747	0.4518	0.7494	0.9126	0.4970	0.5340	0.3433
	(0.0250)***	(0.0140)***	(0.0102)***	(0.0082)***	(0.0148)***	(0.0432)***	(0.0397)***
Product Type[7]							
Conventional with PMI	0.0562	-0.1929	-0.1932	0.2568	0.3601	0.2459	0.0742
	(0.0412)	(0.0238)***	(0.0183)***	(0.0108)***	(0.0112)***	(0.0275)***	(0.0390)*
FHA	0.2230	0.2330	-0.1116	-0.1822	-0.0875	0.5522	0.2329
	(0.0579)***	(0.0374)***	(0.0246)***	(0.0143)***	(0.0117)***	(0.0193)***	(0.0211)***
VA	0.1001	0.0867	-0.3138	-0.2779	-0.3097	0.1138	-0.2064
	(0.1525)	(0.1005)	(0.0817)***	(0.0551)***	(0.0362)***	(0.0315)***	(0.0399)***
DTI Buckets[8]							
$0 < DTI \leq 20$	-0.2728	-0.3821	-0.4316	-0.4270	-0.3548	-0.9985	-1.0811
	(0.0444)***	(0.0298)***	(0.0235)***	(0.0171)***	(0.0178)***	(0.0312)***	(0.0413)***
$20 < DTI \leq 30$	-0.2711	-0.3823	-0.4566	-0.4126	-0.2383	-0.8071	-1.0397
	(0.0438)***	(0.0256)***	(0.0185)***	(0.0141)***	(0.0137)***	(0.0221)***	(0.0321)***
$30 < DTI \leq 40$	-0.1649	-0.2357	-0.2846	-0.1946	-0.0030	-0.4840	-0.6672
	(0.0436)***	(0.0227)***	(0.0159)***	(0.0127)***	(0.0124)	(0.0189)***	(0.0276)***
$DTI > 40$	0.0502	-0.0039	-0.1246	0.0464	0.3400	0.0376	-0.1124
	(0.0393)	(0.0211)	(0.0150)***	(0.0121)***	(0.0114)	(0.0159)**	(0.0239)***
Observations	483952	1022559	883143	1043423	1199856	1680750	1439271

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 16 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Refinance Loans, Year-by-Year Estimation

Variable	Dependent Variable: D_i[1]					
	Training Year					
	2011	2012	2013	2014	2015	2016
ΔHPI_i[2]	-0.0170	-0.0078	-0.0227	-0.0425	-0.0333	-0.0236
	(0.0013)***	(0.0012)***	(0.0016)***	(0.0038)***	(0.0045)***	(0.0034)***
$\Delta Rate_i$[3]	-0.0062	-0.0060	0.0033	-0.0053	-0.0127	-0.0005
	(0.0006)***	(0.0032)*	(0.0013)**	(0.0028)*	(0.0036)***	(0.0017)
$\Delta Unemp_i$[4]	-0.0052	-0.0009	-0.0041	-0.0012	0.0043	-0.0117
	(0.0011)***	(0.0013)	(0.0012)***	(0.0014)	(0.0018)**	(0.0014)***
Non-Owner Occupied?[5]	-0.0099	0.0552	0.0601	0.1914	-0.0183	-0.2041
	(0.0380)	(0.0418)	(0.0326)*	(0.0602)***	(0.0865)	(0.0785)***
Adjustable Rate?[6]	0.1596	-0.1535	-0.8106	-0.5851	-0.6047	-1.5039
	(0.0406)***	(0.0764)**	(0.1112)***	(0.1341)***	(0.1537)***	(0.1812)***
Product Type[7]						
Conventional with PMI	0.0046	-0.0473	-0.0328	0.0532	-0.0293	0.0778
	(0.0376)	(0.0410)	(0.0346)	(0.0664)	(0.0999)	(0.0743)
FHA	0.1710	0.5146	0.2426	0.1963	0.5898	0.4048
	(0.0254)***	(0.0329)***	(0.0341)***	(0.0628)***	(0.0664)***	(0.0589)***
VA	-0.5471	0.1230	0.1393	-0.1356	-0.2071	-0.1530
	(0.0452)***	(0.0455)***	(0.0415)***	(0.0813)*	(0.0985)**	(0.0748)**
DTI Buckets[8]						
$0 < DTI \leq 20$	-1.4317	-0.2974	-0.2088	0.3035	-0.0908	-0.6862
	(0.0483)***	(0.0427)***	(0.0407)***	(0.0879)***	(0.1107)	(0.0875)***
$20 < DTI \leq 30$	-1.1997	-0.2061	-0.1726	0.1774	-0.1487	-0.6271
	(0.0341)***	(0.0386)***	(0.0399)***	(0.0862)**	(0.1056)	(0.0782)***
$30 < DTI \leq 40$	-0.7547	-0.0639	-0.1155	0.3278	0.0236	-0.3037
	(0.0278)***	(0.0367)*	(0.0392)***	(0.0828)***	(0.1018)	(0.0729)***
$DTI > 40$	-0.1957	0.4644	0.4252	0.8827	0.5224	0.1002
	(0.0225)***	(0.0321)***	(0.0342)***	(0.0768)***	(0.0971)*	(0.0701)
Observations	1250577	1222559	1159307	367232	455509	537799

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 16 (Continued): Full “Perfect Foresight” Logistic Regression Model Results: Refinance Loans, Year-by-Year Estimation

	Dependent Variable: D_i[1]						
	Training Year						
Variable	2004	2005	2006	2007	2008	2009	2010
LTV Buckets[2]							
$60 < LTV \leq 70$	0.5233	0.3766	0.6671	0.7905	0.7109	0.5752	0.5832
	(0.0464)***	(0.0278)***	(0.0214)***	(0.0145)***	(0.0154)***	(0.0319)***	(0.0404)***
$70 < LTV \leq 80$	0.7557	0.7391	1.1903	1.2943	1.1110	0.9723	0.8548
	(0.0414)***	(0.0242)***	(0.0188)***	(0.0129)***	(0.0136)***	(0.0275)***	(0.0350)***
$80 < LTV \leq 90$	1.2518	1.2295	1.7584	1.6278	1.3108	1.3072	1.4137
	(0.0426)***	(0.0240)***	(0.0185)***	(0.0130)***	(0.0140)***	(0.0275)***	(0.0335)***
$90 < LTV \leq 95$	1.5776	1.6904	2.0904	1.8148	1.4177	1.5142	1.6889
	(0.0587)***	(0.0326)***	(0.0254)***	(0.0171)***	(0.0168)***	(0.0312)***	(0.0392)***
$95 < LTV \leq 99$	1.9185	1.8813	2.1683	1.9316	1.5111	1.7086	1.8080
	(0.0753)***	(0.0444)***	(0.0321)***	(0.0198)***	(0.0166)***	(0.0296)***	(0.0391)***
$LTV > 99$	2.1753	2.1601	2.2870	2.0468	1.6127	1.7822	2.1889
	(0.1127)***	(0.0671)***	(0.0550)***	(0.0364)***	(0.0261)***	(0.0320)***	(0.0375)***
LTV Missing	1.6602	1.0138	-0.3286	0.5396	0.1305	0.6682	0.2675
	(0.1872)***	(0.1125)***	(0.1562)***	(0.0951)***	(0.1071)	(0.2049)***	(0.2393)
FICO Buckets							
$550 < FICO \leq 650$	-1.4267	-1.1136	-0.9994	-0.9297	-0.7208	-1.0769	-1.3220
	(0.0372)***	(0.0230)***	(0.0200)***	(0.0157)***	(0.0179)***	(0.0493)***	(0.0782)***
$650 < FICO \leq 700$	-2.3330	-1.9865	-1.7455	-1.5905	-1.4460	-1.6019	-1.8538
	(0.0411)***	(0.0245)***	(0.0206)***	(0.0164)***	(0.0187)***	(0.0491)***	(0.0770)***
$700 < FICO \leq 750$	-3.1343	-2.8453	-2.3960	-2.1069	-2.0338	-2.3559	-2.4431
	(0.0491)***	(0.0294)***	(0.0225)***	(0.0172)***	(0.0195)***	(0.0503)***	(0.0770)***
$750 < FICO \leq 800$	-3.9861	-3.7853	-3.2873	-2.9064	-2.9636	-3.4190	-3.4160
	(0.0712)***	(0.0441)***	(0.0299)***	(0.0197)***	(0.0215)***	(0.0525)***	(0.0783)***
$FICO > 800$	-4.3382	-3.9937	-3.7097	-3.6070	-3.7465	-4.1638	-4.0432
	(0.2532)***	(0.1283)***	(0.0886)***	(0.0502)***	(0.0463)***	(0.0765)***	(0.0921)***
FICO Missing	-2.3327	-1.7665	-1.7362	-1.7021	-0.8244	-1.4648	-1.8644
	(0.0607)***	(0.0495)***	(0.0511)	(0.0369)***	(0.0301)***	(0.0562)***	(0.1231)***
Intercept	-1.8858	-2.3050	-2.6775	-3.2925	-2.3045	-2.6677	-2.7067
	(0.0756)***	(0.0356)***	(0.0291)***	(0.0477)***	(0.0277)***	(0.0564)***	(0.1074)***
Observations	483952	1022559	883143	1043423	1199856	1680750	1439271

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For refinance mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 16 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Refinance Loans, Year-by-Year Estimation

	Dependent Variable: D_i[1]					
	Training Year					
Variable	2011	2012	2013	2014	2015	2016
LTV Buckets[2]						
$60 < LTV \leq 70$	0.4409	0.3739	0.2388	0.1599	0.1610	0.2660
	(0.0427)***	(0.0556)***	(0.0427)***	(0.0703)**	(0.0754)**	(0.0540)***
$70 < LTV \leq 80$	0.7720	0.6053	0.4216	0.3282	0.3699	0.3729
	(0.0365)***	(0.0481)***	(0.0378)***	(0.0617)***	(0.0666)***	(0.0490)***
$80 < LTV \leq 90$	1.3294	1.1782	0.7875	0.7309	0.7595	0.6056
	(0.0347)***	(0.0452)***	(0.0366)***	(0.0616)***	(0.0692)***	(0.0540)***
$90 < LTV \leq 95$	1.6399	1.3979	0.9517	0.9542	1.0724	0.8376
	(0.0416)***	(0.0516)***	(0.0440)***	(0.0835)***	(0.1035)***	(0.0886)***
$95 < LTV \leq 99$	1.7408	1.4902	1.0862	0.8929	1.0064	0.8483
	(0.0422)***	(0.0530)***	(0.0502)***	(0.1125)***	(0.1245)***	(0.1071)***
$LTV > 99$	2.2972	2.0869	1.4742	1.3633	1.4041	1.1039
	(0.0381)***	(0.0442)***	(0.0368)***	(0.0753)***	(0.1095)***	(0.1045)***
LTV Missing	-0.0512	1.9877	1.5045	-9.8467	0.4123	0.5192
	(0.3356)	(0.4580)***	(0.3451)***	(111.2448)	(1.0101)	(0.2367)**
FICO Buckets	*					
$550 < FICO \leq 650$	-0.9585	-1.1189	-1.0121	-1.1162	-1.2680	-1.0173
	(0.0726)***	(0.0665)***	(0.0431)***	(0.0848)***	(0.1213)***	(0.1488)***
$650 < FICO \leq 700$	-1.5416	-1.7134	-1.6936	-1.7870	-1.9929	-1.6969
	(0.0706)***	(0.0646)***	(0.0426)***	(0.0838)***	(0.1187)***	(0.1469)***
$700 < FICO \leq 750$	-2.1101	-2.4105	-2.3216	-2.4861	-2.8163	-2.3321
	(0.0705)***	(0.0649)***	(0.0440)***	(0.0878)***	(0.1218)***	(0.1475)***
$750 < FICO \leq 800$	-3.0879	-3.5222	-3.4465	-3.2952	-3.7816	-2.9632
	(0.0721)***	(0.0678)***	(0.0497)***	(0.0955)***	(0.1289)***	(0.1489)***
$FICO > 800$	-3.6887	-4.0072	-3.8018	-3.7721	-3.8565	-3.1813
	(0.0869)***	(0.0903)***	(0.0782)***	(0.1534)***	(0.1663)***	(0.1645)***
$FICO$ Missing	-1.7353	-2.2321	-1.7889	-1.6948	-1.7029	-1.1243
	(0.1533)***	(0.1408)***	(0.1747)***	(0.4202)***	(0.2955)***	(0.1896)***
Intercept	-2.7142	-3.1890	-2.9641	-3.1056	-2.6259	-2.6203
	(0.0773)***	(0.0955)***	(0.0661)***	(0.1291)***	(0.1582)***	(0.1635)***
Observations	1250577	1222559	1159307	367232	455509	537799

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For refinance mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

C Full Logistic Regression Estimation Results: Leave-One-Out

Table 17: Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, "Leave-One-Out" Estimation

	Dependent Variable: D_i[1]						
	Holdout Year						
Variable	2004	2005	2006	2007	2008	2009	2010
ΔHPI_i[2]	-0.0332	-0.0337	-0.0311	-0.0290	-0.0317	-0.0352	-0.0333
	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***
$\Delta Rate_i$[3]	0.0047	0.0049	0.0048	0.0057	0.0062	0.0045	0.0036
	(0.0001)***	(0.0001)***	(0.0001)***	(0.0001)***	(0.0001)***	(0.0001)***	(0.0001)***
$\Delta Unemp_i$[4]	0.0054	0.0054	0.0057	0.0065	0.0059	0.0048	0.0050
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
Non-Owner Occupied?[5]	-0.2324	-0.1970	-0.2148	-0.3306	-0.2328	-0.2336	-0.2155
	(0.0080)***	(0.0083)***	(0.0085)***	(0.0093)***	(0.0098)***	(0.0083)***	(0.0081)***
Adjustable Rate?[6]	0.9229	0.9196	0.7806	0.8642	0.9653	0.8760	0.9110
	(0.0053)***	(0.0058)***	(0.0060)***	(0.0059)***	(0.0055)***	(0.0053)***	(0.0053)***
Product Type[7]							
Conventional with PMI	0.1850	0.2091	0.2375	0.0752	0.1137	0.1666	0.1799
	(0.0054)***	(0.0056)***	(0.0058)***	(0.0065)***	(0.0060)***	(0.0055)***	(0.0055)***
FHA	0.0386	0.0359	0.0811	0.0581	0.0622	0.0316	0.0295
	(0.0055)***	(0.0056)***	(0.0058)***	(0.0061)***	(0.0061)***	(0.0057)***	(0.0057)***
VA	-0.4916	-0.4819	-0.4554	-0.4581	-0.4872	-0.5332	-0.5197
	(0.0077)***	(0.0078)***	(0.0080)***	(0.0083)***	(0.0082)***	(0.0081)***	(0.0081)***
DTI Buckets[8]							
$0 < DTI \leq 20$	-0.5156	-0.4876	-0.4814	-0.5860	-0.6038	-0.5007	-0.4892
	(0.0087)***	(0.0089)***	(0.0091)***	(0.0098)***	(0.0095)***	(0.0088)***	(0.0086)***
$20 < DTI \leq 30$	-0.3429	-0.3448	-0.3220	-0.3990	-0.4340	-0.3258	-0.3170
	(0.0062)***	(0.0063)***	(0.0065)***	(0.0067)***	(0.0069)***	(0.0064)***	(0.0063)***
$30 < DTI \leq 40$	-0.1203	-0.1207	-0.0999	-0.1663	-0.1846	-0.1024	-0.0968
	(0.0052)***	(0.0054)***	(0.0055)***	(0.0056)***	(0.0058)***	(0.0055)***	(0.0054)***
$DTI > 40$	0.1782	0.1782	0.2111	0.1570	0.1231	0.1911	0.1969
	(0.0047)***	(0.0049)***	(0.0050)***	(0.0050)***	(0.0053)***	(0.0050)***	(0.0049)***
Observations	10,817,837	10,380,991	10,390,050	10,218,327	9,966,884	9,935,480	10,029,267

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 18: Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, "Leave-One-Out" Estimation (Continued)

	Dependent Variable: D_i[1]				
	Holdout Year				
Variable	2012	2013	2014	2015	2016
ΔHPI_i[2]	-0.0335	-0.0326	-0.0322	-0.0324	-0.0328
	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***
$\Delta Rate_i$[3]	0.0050	0.0049	0.0044	0.0052	0.0053
	(0.0001)***	(0.0001)***	(0.0001)***	(0.0001)***	(0.0001)***
$\Delta Unemp_i$[4]	0.0053	0.0054	0.0053	0.0054	0.0054
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
Non-Owner Occupied?[5]	-0.2215	-0.2292	-0.2341	-0.2328	-0.2332
	(0.0080)***	(0.0080)***	(0.0080)	(0.0080)***	(0.0080)***
Adjustable Rate?[6]	0.8994	0.8976	0.8992	0.8908	0.8956
	(0.0052)***	(0.0052)***	(0.0052)***	(0.0052)***	(0.0052)***
Product Type[7]					
Conventional with PMI	0.1927	0.1939	0.1897	0.1910	0.1862
	(0.0054)***	(0.0054)***	(0.0054)***	(0.0054)***	(0.0054)***
FHA	0.0291	0.0294	0.0232	0.0283	0.0272
	(0.0055)***	(0.0055)***	(0.0055)***	(0.0054)***	(0.0055)***
VA	-0.5148	-0.5156	-0.5118	-0.5040	-0.5114
	(0.0079)***	(0.0079)***	(0.0078)***	(0.0077)***	(0.0078)***
DTI Buckets[8]					
$0 < DTI \leq 20$	-0.4801	-0.5073	-0.5113	-0.5057	-0.5070
	(0.0086)***	(0.0085)***	(0.0085)***	(0.0085)***	(0.0085)***
$20 < DTI \leq 30$	-0.3125	-0.3370	-0.3394	-0.3330	-0.3369
	(0.0063)***	(0.0062)***	(0.0061)***	(0.0061)***	(0.0061)***
$30 < DTI \leq 40$	-0.0962	-0.1188	-0.1196	-0.1126	-0.1158
	(0.0054)***	(0.0052)***	(0.0052)***	(0.0052)***	(0.0052)***
$DTI > 40$	0.1956	0.1693	0.1717	0.1800	0.1757
	(0.0049)***	(0.0047)***	(0.0047)***	(0.0047)***	(0.0047)***
Observations	10,178,738	10,180,435	10,438,533	10,445,557	10,433,704

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 17 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, "Leave-One-Out" Estimation

	Dependent Variable: D_i[1]						
	Holdout Year						
Variable	2004	2005	2006	2007	2008	2009	2010
LTV Buckets[2]							
$60 < LTV \leq 70$	0.5324	0.5347	0.5114	0.4510	0.5254	0.6091	0.5484
	(0.0177)***	(0.0182)***	(0.0185)***	(0.0198)***	(0.0203)***	(0.0187)***	(0.0178)***
$70 < LTV \leq 80$	0.9403	0.9254	0.8342	0.8497	0.9905	1.0345	0.9554
	(0.0146)***	(0.0150)***	(0.0154)***	(0.0162)***	(0.0167)***	(0.0156)***	(0.0147)***
$80 < LTV \leq 90$	1.2683	1.2437	1.1461	1.1921	1.3122	1.3709	1.2849
	(0.0145)***	(0.0150)***	(0.0152)***	(0.0161)***	(0.0167)***	(0.0155)***	(0.0147)***
$90 < LTV \leq 95$	1.4034	1.3847	1.2968	1.3010	1.4668	1.5177	1.4208
	(0.0150)***	(0.0154)***	(0.0156)***	(0.0166)***	(0.0172)***	(0.0160)***	(0.0152)***
$95 < LTV \leq 99$	1.4978	1.4824	1.3915	1.3840	1.5800	1.6258	1.5238
	(0.0147)***	(0.0151)***	(0.0154)***	(0.0163)***	(0.0169)***	(0.0158)***	(0.0149)***
$LTV > 99$	1.7300	1.6986	1.6328	1.6474	1.8700	1.8556	1.7508
	(0.0148)***	(0.0152)***	(0.0154)***	(0.0163)***	(0.0169)***	(0.0158)***	(0.0150)***
LTV Missing	0.7276	0.5845	0.8507	0.8558	0.8899	0.8548	0.7857
	(0.0617)***	(0.0701)***	(0.0658)***	(0.0622)***	(0.0661)***	(0.0616)***	(0.0617)***
FICO Buckets							
$550 < FICO \leq 650$	-1.0494	-1.0438	-1.0565	-1.0657	-1.1162	-1.0615	-1.0652
	(0.0100)***	(0.0106)***	(0.0108)***	(0.0117)***	(0.0107)	(0.0098)***	(0.0097)***
$650 < FICO \leq 700$	-1.9154	-1.8972	-1.9180	-1.9436	-1.9923	-1.9337	-1.9397
	(0.0101)***	(0.0108)***	(0.0109)***	(0.0118)***	(0.0109)***	(0.0099)***	(0.0098)***
$700 < FICO \leq 750$	-2.7294	-2.7158	-2.7388	-2.8082	-2.8454	-2.7338	-2.7419
	(0.0106)***	(0.0112)***	(0.0114)***	(0.0123)***	(0.0114)***	(0.0104)***	(0.0103)***
$750 < FICO \leq 800$	-3.7328	-3.7229	-3.7306	-3.8239	-3.8637	-3.7318	-3.7393
	(0.0117)***	(0.0123)***	(0.0125)***	(0.0136)***	(0.0128)***	(0.0116)***	(0.0115)***
$FICO > 800$	-4.2716	-4.2757	-4.2659	-4.3088	-4.3329	-4.2638	-4.2924
	(0.0219)***	(0.0225)***	(0.0230)***	(0.0246)***	(0.0242)***	(0.0224)***	(0.0224)***
FICO Missing	-1.5836	-1.6245	-1.6005	-1.5869	-1.6241	-1.5990	-1.5974
	(0.0158)***	(0.0168)***	(0.0170)***	(0.0175)***	(0.0167)***	(0.0155)***	(0.0154)***
Intercept	-2.3725	-2.3714	-2.3235	-2.2023	-2.3229	-2.4251	-2.3671
	(0.0176)***	(0.0184)***	(0.0186)***	(0.0197)***	(0.0198)***	(0.0184)***	(0.0176)***
Observations	10,817,837	10,380,991	10,390,050	10,218,327	9,966,884	9,935,480	10,029,267

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 17 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Purchase Loans, "Leave-One-Out" Estimation

	Dependent Variable: D_i[1]					
	Holdout Year					
Variable	2011	2012	2013	2014	2015	2016
LTV Buckets[2]						
$60 < LTV \leq 70$	0.5417	0.5434	0.5428	0.5418	0.5437	0.5602
	(0.0178)***	(0.0176)***	(0.0176)***	(0.0176)***	(0.0176)***	(0.0178)***
$70 < LTV \leq 80$	0.9573	0.9570	0.9529	0.9513	0.9543	0.9731
	(0.0147)***	(0.0146)***	(0.0145)***	(0.0145)***	(0.0145)***	(0.0147)***
$80 < LTV \leq 90$	1.2847	1.2815	1.2759	1.2774	1.2801	1.3034
	(0.0146)***	(0.0145)***	(0.0145)***	(0.0145)***	(0.0145)***	(0.0147)***
$90 < LTV \leq 95$	1.4195	1.4168	1.4121	1.4170	1.4164	1.4412
	(0.0151)***	(0.0150)***	(0.0150)***	(0.0149)***	(0.0149)***	(0.0151)***
$95 < LTV \leq 99$	1.5178	1.5127	1.5007	1.5045	1.5068	1.5350
	(0.0149)***	(0.0148)***	(0.0147)***	(0.0147)***	(0.0147)***	(0.0149)***
$LTV > 99$	1.7435	1.7429	1.7316	1.7353	1.7360	1.7642
	(0.0150)***	(0.0148)***	(0.0148)***	(0.0147)***	(0.0147)***	(0.0149)***
LTV Missing	0.7792	0.7654	0.7537	0.7787	0.7709	0.7967
	(0.0612)***	(0.0605)***	(0.0605)***	(0.0599)***	(0.0601)***	(0.0601)***
FICO Buckets						
$550 < FICO \leq 650$	-1.0746	-1.0693	-1.0651	-1.0628	-1.0620	-1.0633
	(0.0097)***	(0.0097)***	(0.0096)***	(0.0096)***	(0.0096)***	(0.0096)***
$650 < FICO \leq 700$	-1.9598	-1.9521	-1.9474	-1.9404	-1.9388	-1.9411
	(0.0098)***	(0.0098)***	(0.0098)***	(0.0098)***	(0.0098)***	(0.0098)***
$700 < FICO \leq 750$	-2.7584	-2.7517	-2.7433	-2.7428	-2.7398	-2.7498
	(0.0103)***	(0.0103)***	(0.0103)***	(0.0103)***	(0.0103)***	(0.0103)***
$750 < FICO \leq 800$	-3.7550	-3.7460	-3.7362	-3.7409	-3.7368	-3.7608
	(0.0115)***	(0.0115)***	(0.0114)***	(0.0114)***	(0.0114)***	(0.0115)***
$FICO > 800$	-4.2921	-4.2825	-4.2660	-4.2827	-4.2726	-4.3439
	(0.0222)***	(0.0220)***	(0.0219)***	(0.0219)***	(0.0219)***	(0.0227)***
$FICO$ Missing	-1.6082	-1.6035	-1.6043	-1.6022	-1.5992	-1.5943
	(0.0154)***	(0.0153)***	(0.0153)***	(0.0153)***	(0.0153)***	(0.0152)***
Intercept	-2.3750	-2.3653	-2.3360	-2.3379	-2.3435	-2.3671
	(0.0175)***	(0.0175)***	(0.0174)***	(0.0174)***	(0.0174)***	(0.0175)***
Observations	10,078,053	10,178,738	10,180,435	10,438,533	10,445,557	10,433,704

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 19: Full "Perfect Foresight" Logistic Regression Model Results: Refinance Loans, "Leave-One-Out" Estimation

	Dependent Variable: D_i[1]						
	Holdout Year						
Variable	2004	2005	2006	2007	2008	2009	2010
ΔHPI_i[2]	-0.0313	-0.0316	-0.0336	-0.0270	-0.0304	-0.0306	-0.0310
	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***
$\Delta Rate_i$[3]	-0.0113	-0.0100	-0.0092	-0.0084	-0.0079	-0.0127	-0.0145
	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***
$\Delta Unemp_i$[4]	0.0053	0.0053	0.0052	0.0069	0.0056	0.0052	0.0046
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
Non-Owner Occupied?[5]	0.1506	0.1648	0.1187	0.0351	0.2137	0.1543	0.1492
	(0.0069)***	(0.0071)***	(0.0073)***	(0.0083)***	(0.0081)***	(0.0070)***	(0.0069)***
Adjustable Rate?[6]	0.6899	0.7118	0.6671	0.5393	0.7366	0.6705	0.6752
	(0.0047)***	(0.0050)***	(0.0053)***	(0.0057)***	(0.0050)***	(0.0047)***	(0.0047)***
Product Type[7]							
Conventional with PMI	0.2214	0.2507	0.2617	0.1873	0.1001	0.2162	0.2107
	(0.0060)***	(0.0061)***	(0.0063)***	(0.0072)***	(0.0070)***	(0.0061)***	(0.0060)***
FHA	0.0631	0.0755	0.0776	0.0995	0.0657	-0.0128	0.0447
	(0.0056)***	(0.0056)***	(0.0058)***	(0.0063)***	(0.0065)***	(0.0061)***	(0.0058)***
VA	-0.2729	-0.2695	-0.2592	-0.2641	-0.3132	-0.2986	-0.2703
	(0.0129)***	(0.0130)***	(0.0131)***	(0.0134)***	(0.0138)***	(0.0145)***	(0.0137)***
DTI Buckets[8]							
$0 < DTI \leq 20$	-0.5571	-0.5546	-0.5582	-0.5758	-0.6106	-0.4925	-0.5177
	(0.0084)***	(0.0086)***	(0.0088)***	(0.0095)***	(0.0093)***	(0.0086)***	(0.0084)***
$20 < DTI \leq 30$	-0.5038	-0.5026	-0.5026	-0.5117	-0.5849	-0.4496	-0.4668
	(0.0068)***	(0.0070)***	(0.0072)***	(0.0077)***	(0.0077)***	(0.0071)***	(0.0069)***
$30 < DTI \leq 40$	-0.2680	-0.2641	-0.2625	-0.2775	-0.3503	-0.2272	-0.2414
	(0.0060)***	(0.0062)***	(0.0065)***	(0.0069)***	(0.0068)***	(0.0064)***	(0.0061)***
$DTI > 40$	0.0768	0.0837	0.1041	0.1024	-0.0196	0.0954	0.0906
	(0.0056)***	(0.0057)***	(0.0059)***	(0.0062)***	(0.0063)***	(0.0059)***	(0.0056)***
Observations	12,261,985	11,723,378	11,862,794	11,702,514	11,546,081	11,065,187	11,306,666

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 20: Full "Perfect Foresight" Logistic Regression Model Results: Refinance Loans, "Leave-One-Out" Estimation (Continued)

	Dependent Variable: D_i[1]					
	Holdout Year					
Variable	2011	2012	2013	2014	2015	2016
ΔHPI_i[2]	-0.0310	-0.0314	-0.0307	-0.0305	-0.0307	-0.0309
	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***
$\Delta Rate_i$[3]	-0.0119	-0.0116	-0.0111	-0.0114	-0.0107	-0.0119
	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***	(0.0002)***
$\Delta Unemp_i$[4]	0.0052	0.0052	0.0052	0.0053	0.0053	0.0053
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***
Non-Owner Occupied?[5]	0.1537	0.1470	0.1560	0.1475	0.1480	0.1497
	(0.0070)***	(0.0070)***	(0.0070)***	(0.0069)***	(0.0069)***	(0.0069)***
Adjustable Rate?[6]	0.6879	0.6833	0.6750	0.6775	0.6747	0.6871
	(0.0047)***	(0.0047)***	(0.0047)***	(0.0046)***	(0.0046)***	(0.0047)***
Product Type[7]						
Conventional with PMI	0.2262	0.2289	0.2261	0.2193	0.2174	0.2189
	(0.0060)***	(0.0060)***	(0.0060)***	(0.0059)***	(0.0059)***	(0.0059)***
FHA	0.0622	0.0564	0.0479	0.0508	0.0560	0.0519
	(0.0057)***	(0.0056)***	(0.0057)***	(0.0056)***	(0.0056)***	(0.0056)***
VA	-0.2433	-0.2804	-0.2990	-0.2680	-0.2697	-0.2835
	(0.0134)***	(0.0135)***	(0.0137)***	(0.0130)***	(0.0130)***	(0.0131)***
DTI Buckets[8]						
$0 < DTI \leq 20$	-0.5039	-0.5488	-0.5556	-0.5531	-0.5474	-0.5461
	(0.0084)***	(0.0084)***	(0.0085)***	(0.0083)***	(0.0083)***	(0.0083)***
$20 < DTI \leq 30$	-0.4628	-0.5031	-0.5093	-0.5036	-0.4979	-0.5000
	(0.0069)***	(0.0068)***	(0.0068)***	(0.0067)***	(0.0067)***	(0.0068)***
$30 < DTI \leq 40$	-0.2384	-0.2699	-0.2736	-0.2705	-0.2649	-0.2696
	(0.0062)***	(0.0061)***	(0.0061)***	(0.0060)***	(0.0060)***	(0.0060)***
$DTI > 40$	0.0977	0.0684	0.0637	0.0693	0.0755	0.0722
	(0.0057)***	(0.0056)***	(0.0056)***	(0.0055)***	(0.0055)***	(0.0055)***
Observations	11,495,360	11,523,378	11,586,630	12,378,705	12,290,428	12,208,138

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] ΔHPI_i denotes the 24-month rolling window growth rate in the unadjusted county-level Case-Shiller house price index from Moody's Analytics.
[3] $\Delta Rate_i$ denotes the 24-month rolling window growth rate in the 30-year mortgage rate reported by Freddie Mac.
[4] $\Delta Unemp_i$ denotes the 24-month rolling window growth rate in the county-level unemployment rate reported by the Bureau of Labor Statistics.
[5] "Owner Occupied?" is equal to one if the loan was used to finance a property that will be owner-occupied, zero otherwise.
[6] "Fixed Rate?" is equal to one if the loan's interest rate is fixed for the entire scheduled duration of the loan, zero otherwise.
[7] "PMI" stands for "Private Mortgage Insurance." The reference category product type is a conventional loan without PMI.
[7] DTI denotes the borrower's back-end debt-to-income ratio, which is defined as the sum of all required monthly debt payments,such as mortgage debt, automotive debt, and student loans - divided by monthly gross income. The reference category for our DTI variable is "DTI Missing."
[8] For purchase mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the lesser of the: (1) the property's sales price and (2) property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 19 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Refinance Loans, "Leave-One-Out" Estimation

	Dependent Variable: D_i[1]						
	Holdout Year						
Variable	2004	2005	2006	2007	2008	2009	2010
LTV Buckets[2]							
$60 < LTV \leq 70$	0.6385	0.6557	0.6335	0.5717	0.6098	0.6376	0.6365
	(0.0078)***	(0.0080)***	(0.0083)***	(0.0091)***	(0.0089)***	(0.0080)***	(0.0079)***
$70 < LTV \leq 80$	1.0674	1.0834	1.0371	0.9546	1.0479	1.0598	1.0629
	(0.0068)***	(0.0070)***	(0.0072)***	(0.0080)***	(0.0078)***	(0.0070)***	(0.0069)***
$80 < LTV \leq 90$	1.4484	1.4538	1.3864	1.3568	1.5057	1.4423	1.4426
	(0.0068)***	(0.0070)***	(0.0073)***	(0.0079)***	(0.0077)***	(0.0070)***	(0.0069)***
$90 < LTV \leq 95$	1.6435	1.6351	1.5762	1.5567	1.7511	1.6349	1.6336
	(0.0086)***	(0.0088)***	(0.0090)***	(0.0098)***	(0.0099)***	(0.0089)***	(0.0087)***
$95 < LTV \leq 99$	1.7622	1.7597	1.7002	1.6754	1.8785	1.7322	1.7427
	(0.0089)***	(0.0090)***	(0.0092)***	(0.0099)***	(0.0105)***	(0.0094)***	(0.0091)***
$LTV > 99$	2.0237	2.0137	1.9684	1.9551	2.1473	2.0025	1.9996
	(0.0103)***	(0.0104)***	(0.0105)***	(0.0110)***	(0.0112)***	(0.0111)***	(0.0108)***
LTV Missing	0.3205	0.2604	0.5078	0.4424	0.4653	0.3894	0.4148
	(0.0488)***	(0.0522)***	(0.0494)***	(0.0545)***	(0.0522)***	(0.0482)***	(0.0479)***
FICO Buckets							
$550 < FICO \leq 650$	-0.9190	-0.9134	-0.9239	-0.9913	-1.0020	-0.9489	-0.9374
	(0.0086)***	(0.0090)***	(0.0092)***	(0.0099)***	(0.0094)***	(0.0085)***	(0.0084)***
$650 < FICO \leq 700$	-1.5993	-1.5852	-1.5995	-1.6958	-1.6784	-1.6558	-1.6182
	(0.0087)***	(0.0091)***	(0.0094)***	(0.0100)***	(0.0095)***	(0.0086)***	(0.0085)***
$700 < FICO \leq 750$	-2.2277	-2.2123	-2.2339	-2.3653	-2.3358	-2.2566	-2.2408
	(0.0091)***	(0.0095)***	(0.0097)***	(0.0105)***	(0.0100)***	(0.0090)***	(0.0090)***
$750 < FICO \leq 800$	-3.2004	-3.1919	-3.2171	-3.3658	-3.3181	-3.2003	-3.1951
	(0.0102)***	(0.0105)***	(0.0108)***	(0.0117)***	(0.0113)***	(0.0103)***	(0.0102)***
$FICO > 800$	-3.8567	-3.8577	-3.8855	-3.9819	-3.9202	-3.8422	-3.8393
	(0.0202)***	(0.0205)***	(0.0209)***	(0.0221)***	(0.0223)***	(0.0210)***	(0.0213)***
$FICO$ Missing	-1.3702	-1.3784	-1.3791	-1.4309	-1.5886	-1.4520	-1.3959
	(0.0160)***	(0.0163)***	(0.0164)***	(0.0172)***	(0.0183)***	(0.0169)***	(0.0155)***
Intercept	-2.8503	-2.8551	-2.7892	-2.6472	-2.7535	-2.8137	-2.8115
	(0.0113)***	(0.0118)***	(0.0121)***	(0.0128)***	(0.0125)***	(0.0114)***	(0.0112)***
Observations	12,261,985	11,723,378	11,862,794	11,702,514	11,546,081	11,065,187	11,306,666

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For refinance mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 19 (Continued): Full "Perfect Foresight" Logistic Regression Model Results: Refinance Loans, "Leave-One-Out" Estimation

	Dependent Variable: D_i[1]					
	Holdout Year					
Variable	2011	2012	2013	2014	2015	2016
LTV Buckets[2]						
$60 < LTV \leq 70$	0.6430	0.6425	0.6516	0.6421	0.6412	0.6437
	(0.0078)***	(0.0078)***	(0.0078)***	(0.0078)***	(0.0078)***	(0.0078)***
$70 < LTV \leq 80$	1.0666	1.0679	1.0780	1.0655	1.0639	1.0701
	(0.0069)***	(0.0068)***	(0.0069)***	(0.0068)***	(0.0068)***	(0.0068)***
$80 < LTV \leq 90$	1.4443	1.4471	1.4630	1.4492	1.4471	1.4558
	(0.0069)***	(0.0068)***	(0.0069)***	(0.0068)***	(0.0068)***	(0.0068)***
$90 < LTV \leq 95$	1.6345	1.6385	1.6636	1.6437	1.6415	1.6522
	(0.0087)***	(0.0086)***	(0.0086)***	(0.0085)***	(0.0085)***	(0.0085)***
$95 < LTV \leq 99$	1.7512	1.7578	1.7806	1.7620	1.7608	1.7713
	(0.0090)***	(0.0089)***	(0.0090)***	(0.0088)***	(0.0088)***	(0.0088)***
$LTV > 99$	1.9768	1.9701	2.0577	2.0251	2.0176	2.0331
	(0.0106)***	(0.0107)***	(0.0107)***	(0.0103)***	(0.0102)***	(0.0102)***
LTV Missing	0.3883	0.3747	0.3911	0.3925	0.3878	0.3684
	(0.0474)***	(0.0472)***	(0.0473)***	(0.0469)***	(0.0470)***	(0.0480)***
FICO Buckets						
$550 < FICO \leq 650$	-0.9421	-0.9424	-0.9461	-0.9426	-0.9409	-0.9458
	(0.0084)***	(0.0084)***	(0.0085)***	(0.0084)***	(0.0084)***	(0.0084)***
$650 < FICO \leq 700$	-1.6278	-1.6328	-1.6312	-1.6265	-1.6241	-1.6318
	(0.0086)***	(0.0086)***	(0.0087)***	(0.0085)***	(0.0085)***	(0.0085)***
$700 < FICO \leq 750$	-2.2604	-2.2618	-2.2607	-2.2567	-2.2510	-2.2627
	(0.0090)***	(0.0090)***	(0.0091)***	(0.0089)***	(0.0089)***	(0.0089)***
$750 < FICO \leq 800$	-3.2302	-3.2260	-3.2264	-3.2319	-3.2221	-3.2423
	(0.0102)***	(0.0101)***	(0.0102)***	(0.0100)***	(0.0100)***	(0.0101)***
$FICO > 800$	-3.9000	-3.8952	-3.9048	-3.8946	-3.8901	-3.9244
	(0.0212)***	(0.0208)***	(0.0208)***	(0.0203)***	(0.0203)***	(0.0206)***
$FICO$ Missing	-1.3918	-1.3924	-1.4210	-1.4194	-1.4127	-1.4155
	(0.0155)***	(0.0155)***	(0.0156)***	(0.0154)***	(0.0154)***	(0.0155)***
Intercept	-2.8388	-2.8162	-2.8018	-2.8008	-2.8056	-2.8247
	(0.0113)***	(0.0112)***	(0.0113)***	(0.0111)***	(0.0111)***	(0.0111)***
Observations	11,495,360	11,523,378	11,586,630	12,378,705	12,290,428	12,208,138

[1] D_i is equal to one if the mortgage is ever 90 days-past-due or worse, enters the foreclosure process, or is associated with bankruptcy proceedings within 24 months of origination.
[2] For refinance mortgages, LTV is defined as the outstanding principal balance at the time of the loan's origination divided by the property's appraised value.
Standard errors are reported in parentheses.
* , **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

D Sampling and Parameter Tuning for the Random Forest Classifier

Aggregating the output from the Logistic regression models to predict the number of defaults in a particular validation sample is very straightforward: because the output from the Logit model is a predicted probability, to calculate the number of predicted defaults in a given set of loans, we simply need to sum over all of the predicted probabilities in that set.[64]

Using the random forest (RF) model to predict defaults, in contrast, is complicated by two issues. The first complication is that our data is "imbalanced" in the sense that non-default observations are far more common in our data than observations that default. Classification algorithms typically seek to minimize the rate of misclassification in the full sample. When there are few positive outcomes in the data – defaults in our case – classifiers will have a tendency to focus on predicting the outcome of non-positive outcomes as the non-positive outcomes have an outsized impact on the calculation of the overall error rate. Because of this issue, when classifiers are applied to extremely unbalanced data they will tend to correctly classify very few of the positive outcomes.

There are two common approaches for addressing the imbalance issue when implementing the RF classifier [Chen, Liaw and Breiman, 2004]. One approach is to reconfigure the objective function underlying the classifier to assign a high cost to misclassifying non-positive cases. A second approach is to either oversample the positive outcomes or downsample the non-positive outcomes when constructing the training data; under this type of sampling scheme, the classifier will naturally give more weight to correctly classifying positive cases than it would when trained using a random sample from the population. In our implementation of the RF classifier in Section 3, we use the Synthetic Minority Over-Sampling Technique (SMOTE) to address the imbalance problem [Chawla et al., 2002]. Specifically, if N_j^D and N_j^{ND} denote the default and non-default observations in origination cohort j, we use SMOTE to create $\widetilde{N_j^D}$ synthetic defaults such that the sum of the true and augmented defaults comprise 20 percent of the "SMOTEd" data for origination cohort j. That is

$$\frac{S_j^D}{S_j^{ND}} = \frac{N_j^D + \widetilde{N_j^D}}{N_j^D + \widetilde{N_j^D} + N_j^{ND}} = 0.2$$

where S_j^D and S_j^{ND} denote the set of defaulted and non-defaulted loans, respectively.

The SMOTEd data serves as the sampling base for the bootstrap samples of our RF algorithm. When defining the training sample for each tree grown using the RF algorithm, we first partition the SMOTEd data into loans from an origination cohort j that default and those that do not. When growing a tree that is indexed by b, we randomly sample 50,000 observations from S_j^D with replacement to create a sample of defaults denoted $S_j^{D,b}$. Similarly, we create a sample of non-defaults, denoted $S_j^{ND,b}$, by randomly 50,000 observations from S_j^{ND} with replacement. Lastly, we use a sample comprised of the observations in $S_j^{D,b}$ and $S_j^{ND,b}$ to grow tree b using the Classification and Regression Trees (CART) algorithm. The resulting dataset is perfectly balanced and small enough to render the algorithm computationally feasible. This process is repeated $B = 2,000$ times, where B denotes the total number of trees in the random forest. Chen, Liaw and Breiman [2004] refer to this approach as " balanced random forest" (BRF).

The second complication with the implementation of the RF algorithm concerns aggregation. For any given observation, the RF algorithm produces B different predictions of whether the loan is a default or non-default, and the user of the RF algorithm is tasked with determining how to aggregate these B different predictions to create a final prediction. To conduct this aggregation, we will use information on fraction of the B trees that classify a loan as being a default. Formally, define $\widehat{Y}_{bi}$ as

$$\widehat{Y}_{bi} = \begin{cases} 1 \text{ if loan } i \text{ is classified as a default by tree } b \\ 0 \text{ if loan } i \text{ is classified as a non-default by tree } b \end{cases} \tag{D.1}$$

Now let $\widetilde{Y}_i$ denote the final prediction from the RF procedure based on an aggregation of all B of the

[64] While estimates of the probability of default can in principle be constructed based on the Random Forest procedure, estimating these probabilities is not trivial. For a discussion of this issue, see Niculescu-Mizil and Caruana [2005].

tree-specific predictions. We construct $\widetilde{Y}_i$ as follows

$$\widetilde{Y}_i = \begin{cases} 1 \text{ (default) if } \frac{1}{B}\sum_{b=1}^{B}\widehat{Y}_{bi} > \tau \\ 0 \text{ (non-default) if } \frac{1}{B}\sum_{b=1}^{B}\widehat{Y}_{bi} \leq \tau \end{cases} \tag{D.2}$$

where $\frac{1}{B}\sum_{b=1}^{B}\widehat{Y}_{bi}$ measures the fraction of trees that classify a loan as being in default and τ is a cutoff threshold that is used in the aggregation. In the standard presentation of the RF model, the aggregation is conducted via " majority rule" in the sense that the modal prediction in the set of B predictions is the final classification that is applied to the loan [Hastie, Tibshirani and Friedman, 2009]. Because of the binary nature of the prediction in our application, this majority rule aggregation is equivalent to setting $\tau = 0.5$ in Equation D.2. In many applications, instead of simply using majority rule to choose the final classification, the users of the RF algorithm "tune" the τ parameter to optimize particular performance measures.

That said, because the goal of our prediction exercise is to get a sense of the RF procedure's predictive accuracy in a production environment where tuning is common, instead of simply using the majority voting rule, we tuned the τ parameter as follows. First, we created a master training dataset comprised of loans that were originated in 2004. Then, we defined the following set G of candidate thresholds

$$G = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$$

For each $\tau \in G$, we then trained the RF classifier on the 2004 data using τ to define the voting rule defined above. Then, we used the RF classifier to produce predictions of loan performance for every origination year (2004 through 2016) in our sample based on the voting rule under the threshold τ.

The results of the final predictions can be summarized in a " confusion matrix" as follows

		Actual Class	
		$Y_i = 1$	$Y_i = 0$
Predicted	$\widetilde{Y}_i = 1$	True Positives (TP)	False Positives (FP)
Class	$\widetilde{Y}_i = 0$	False Negatives (FN)	True Negatives (TN)

We use the entries in the confusion matrix to construct the following model performance measures separately for each year j in our data. Note that all of these measures are functions of the threshold τ.

$$\begin{aligned} \text{Recall}_j(\tau) &= \frac{TP_j}{TP_j + FN_j} \\ \text{Specificity}_j(\tau) &= \frac{TN_j}{TN_j + FP_j} \\ \text{Precision}_j(\tau) &= \frac{TP_j}{TP_j + FP_j} \\ \text{Balanced Accuracy}_j(\tau) &= \frac{\text{Specificity}_j(\tau) + \text{Recall}_j(\tau)}{2} \\ F_j(\tau) &= \frac{2\text{Precision}_j(\tau)\text{Recall}_j(\tau)}{\text{Precision}_j(\tau) + \text{Recall}_j(\tau)} \\ EvA_j(\tau) &= 100 * \left(\frac{(TP_j + FP_j) - (TP_j + FN_j)}{(TP_j + FN_j)}\right) \end{aligned}$$

Lastly, we averaged these performance measures over all J separate sample years to create

$$\overline{\text{Recall}(\tau)} = \frac{1}{J}\sum_{j=2004}^{2016} \text{Recall}_j(\tau)$$

$$\overline{\text{Specificity}(\tau)} = \frac{1}{J}\sum_{j=2004}^{2016} \text{Specificity}_j(\tau)$$

$$\overline{\text{Precision}(\tau)} = \frac{1}{J}\sum_{j=2004}^{2016} \text{Precision}_j(\tau)$$

$$\overline{\text{Balanced Accuracy}(\tau)} = \frac{1}{J}\sum_{j=2004}^{2016} \text{Balanced Accuracy}_j(\tau)$$

$$\overline{F(\tau)} = \frac{1}{J}\sum_{j=2004}^{2016} F_j(\tau)$$

$$\overline{EvA(\tau)} = \frac{1}{J}\sum_{j=2004}^{2016} EvA_j(\tau)$$

The results of this exercise are summarized in Table 21. For the purchase loans, the absolute value of the prediction error rate – $\overline{EvA(\tau)}$ – is minimized at $\tau = 0.7$, while for the refinance loans the absolute value of $\overline{EvA(\tau)}$ is minimized at $\tau = 0.9$. When looking at the year-by-year components of the $\overline{EvA(\tau)}$ measure for the refinance loans, we discovered that while $\overline{EvA(\tau)}$ was lowest in absolute value for $\tau = 0.9$ and $\tau = 0.8$, these averages were driven down by extreme underpredictions of defaults in several of the validation sample years. Because our primary analysis focuses on $\overline{EvA(\tau)}$ and this measure appears to be roughly maximized around a value of $\tau = 0.7$ while still avoiding extreme underpredictions, we chose $\tau = 0.7$ as the threshold used for the voting rule in our implementation of the random forest classifier.

Note that there is a mechanical relationship between τ, recall, and specificity: as τ increases, fewer observations are classified as defaults, decreasing the fraction of defaults that are correctly classified (recall) while increasing the fraction of non-defaults that are correctly classified (specificity).

Table 21: Parameter Tuning Results

τ[1]	$EvA(\tau)$	Recall(τ)	Specificity(τ)	Balanced Accuracy(τ)	$F(\tau)$
	Purchase Loans Performance Measure				
0.9	-87.10	1.36	99.67	50.52	2.34
0.8	-40.33	5.84	98.52	52.18	6.96
0.7	34.16	12.68	96.78	54.73	11.16
0.6	226.87	23.78	91.93	57.85	14.50
0.5	617.28	47.32	80.91	64.11	18.14
0.4	1299.90	74.09	61.24	67.66	15.27
0.3	2094.20	90.47	38.71	64.59	12.37
0.2	2742.76	97.69	20.46	59.08	10.28
0.1	3217.42	99.71	8.45	54.08	9.00
	Refinance Loans Performance Measure				
τ	$EvA(\tau)$	Recall(τ)	Specificity(τ)	Balanced Accuracy(τ)	$F(\tau)$
0.9	-26.84	2.28	99.28	50.78	2.17
0.8	106.26	6.43	97.85	52.14	4.35
0.7	300.70	12.08	95.53	53.81	6.38
0.6	711.54	25.14	90.77	57.95	10.14
0.5	1214.59	43.82	83.73	63.78	12.77
0.4	1769.09	60.69	75.28	67.99	12.34
0.3	2996.57	79.35	57.25	68.30	10.13
0.2	5307.29	92.61	26.81	59.71	7.42
0.1	6884.03	98.07	12.27	55.17	6.45

[1] τ denotes the cutoff used to train the Random Forest model and calculate the performance measures.