Is the FHA Creating Sustainable Homeownership?[1]

Andrew Caplin, Anna Cororaton, and Joseph Tracy[2]

March 2013

[1] We thank Damien Weldon and CoreLogic for providing us with access to crucial data. We also thank Diego Aragon, Gunnar Blix, Scott Frame, Charles Freeman, Ahu Gemici, Joseph Gyourko, Andy Haughwout, Steven Laufer, John Leahy, Roy Lowrance, Daniel Martin, and Ruth Wyatt for their invaluable input. Joshua Abel provided capable research assistance.

[2] Affiliations: Caplin, NYU: Cororaton, Tracy; FRBNY. Contact authors: andrew.caplin@nyu.edu and Joseph.Tracy@ny.frb.org. The opinions expressed herein are those of the authors alone, not of New York University, the Federal Reserve Bank of New York, the Federal Reserve System, or CoreLogic.

I. Introduction

Following the collapse of the private mortgage market, the Federal Housing Administration (FHA) has become the only major issuer of low down-payment mortgages. In 2012 the FHA had an estimated $1 trillion dollars of insurance in force – roughly three times its 2005 level (IFE [2011], HUD [2005]). Former FHA Commissioner David Stevens in a speech on 12 December 2009 described the FHA's mission as follows. "As a mission-driven organization, FHA's goal is to provide sustainable homeownership options for qualified borrowers." A minimum condition for a sustainable homeownership experience is that the borrower successfully pays off the FHA in full by either selling the house or refinancing into a non-FHA mortgage. A key question is how many of the FHA borrowing experiences initiated since the downturn in housing markets will end in default. The answer is central to whether the FHA is meeting its goal of sustainable homeownership and also to the FHA's financial health.

In this paper, we use a new data set to provide first answers on sustainability for recent FHA-insured borrowers. Since FHA borrowers can refinance their existing FHA mortgages into new FHA mortgages, it is important to measure sustainability in terms of borrowers and not mortgages, as the credit exposures follow the borrowers. We find for the 2007 vintage of FHA borrowers that more than 40 percent have already been 90 or more days delinquent. By way of contrast, less than 12 percent of these borrowers have demonstrated sustainable homeownership by fully paying off all FHA mortgages. We project for this vintage that the proportion who have been 90 or more days delinquent will rise above 50 percent within five years, while only 30 percent will have been fully paid off without a loss to the FHA. The 2008 vintage of FHA borrowers is performing only marginally better with nearly 30 percent already having reached 90 or more days delinquent, and with the projection that this will increase to 40 percent or more. Our projected ten year default rate for the 2009 vintage of FHA borrowers is close to 25 percent, even with half of these borrowers still projected to have active mortgages.

The FHA is designed to be a self-financing program with any credit losses covered by mortgage insurance premiums paid by the borrowers. As we will show, these projected cumulative default rates are well

in excess of the levels where the premiums can cover the credit losses for these vintages of FHA borrowers. Any excess of credit losses over and above these premiums must be paid out of the FHA's capital. If for a sustained period of time FHA lending is not generating sustainable homeownership experiences at a rate sufficient to insure that premiums cover all credit losses, then the FHA's capital will be depleted and additional losses will have to be coverd by taxpayers.

The high delinquency rates that we identify are easy to understand. Most FHA borrowing experiences that were initiated between 2007 and 2009 began with mortgages having a loan-to-value (LTV) of nearly 100 percent. Many of the homes bought with these mortgages subsequently fell in value. Hence, many of these FHA borrowers are underwater making it difficult for them to successfully exit the FHA system either by selling their homes or by refinancing into a non-FHA insured mortgages. In the meantime, the continuing high unemployment rates leave many of these FHA borrowers exposed to the risk of job loss and therefore the risk of a default.

Accurate assessments of the expected credit losses to the FHA are also important. As shown in Figure 1, for the last several Annual Reports to Congress the one year ahead projection for the status of the insurance fund has had to be revised downward in the subsequent year (Gyourko [2011]). In the 2012 report, the insurance fund was estimated to have a value of negative $16.3 billion, a downward revision of $27.7 billion from the prior year's projection, and $40 billion below the 2 percent required capitalization rate. What explains this consistent pattern of underestimation of credit risk even in 2012 when the housing market has been recovering?

The data structure used by the FHA likely contributes to this underestimation of risk (Aragon et al. [2010], Caplin ([2010], [2011]). The FHA counts as a successful termination any mortgage that is paid off without an insurance claim, even when this payoff reflects an internal refinance of one FHA mortgage into another FHA mortgage. So, a borrower who defaults after internally refinancing, is treated by the FHA's data structure as creating one success (payoff of the first mortgage) and one failure (default of the second mortgage). Looked at from the perspective of the borrower rather than the mortgage correctly identifies that

there was no success and only the default.[3] The situation with the GSE's (Fannie Mae and Freddie Mac) is the same. They treat as successful any payoff of an agency mortgage that refinances into another agency mortgage.

We will demonstrate that this overcounting of success makes the FHA's risk assessment overly optimistic reflecting an under estimate of the forcasted credit losses. We show below that the over-counting of success is important in the recent period not only in qualitative but also in quantitative terms. Of those mortgages paid off by FHA borrowers in the 2007-2009 cohorts, more than two-thirds have been internally refinanced. These neither represent successful transitions to sustainable homeownership, nor resolutions of credit risk to the FHA.

We describe the CoreLogic data in Section 2. We develop and estimate our model of borrower performance in Sections 3 and 4. We forecast default and prepayment rates in Section 5. In section 6 we show that the methodology used in the Actuarial Reviews understates the risks that FHA-insured mortgages pose to borrowers and taxpayers alike. Section 7 concludes.

II. The FHA Sample

A. The FHA and the Borrower Experience

The Federal Housing Administration (FHA) was set up in 1934 to improve the functioning of mortgage markets. Since then it has become an important institution for implementing housing policy, providing strong support for housing affordability and homeownership. As a result of the recent financial crisis, the FHA has moved into the unprecedented position of being the low down-payment "lender of last resort." The FHA must ensure that alongside its increased market share of mortgage originations it still achieves its key goal of creating sustainable homeownership.

---

[3] We are not criticizing the FHA's internal refinance programs. In fact, reducing the frictions for high LTV mortgages to refinance has been shown to have important macroeconomic benefits (Caplin et. al. 1997). Rather, our criticism is with how these programs are treated in measuring FHA default rates and projected default losses.

While FHA mortgages are fully backed by the federal government, the FHA is designed to be a self-financing program. Borrowers pay both an up-front mortgage insurance premium (u), and an annual premium (a) which is incorporated into the interest rate on the mortgage. The up-front premium is typically financed by the borrower by adding it into the mortgage balance. An implication is that the borrower pays the up-front premium only when the borrower experience is sustainable in that the mortgage balance if fully paid off. In addition, if the borrower pre-pays the mortgage over the initial five years, some of the up-front premium is rebated back on a sliding scale (r_t).[4] Similarly, if the borrower refinances the current FHA mortgage into a new FHA mortgage any rebate of the prior up-front premium is applied against the new up-front premium.

We can use this structure of the FHA's mortgage insurance premiums to approximate the minimum sustainability rate that is consistent with the premiums covering all credit losses. Let P(s) denote the sustainability rate for a vintage, and LGD denote the average loss-given-default rate. Assume for simplicity that a vintage of FHA borrowers will completely terminate in T years – with the rate of payoffs and defaults uniformly distributed over this period.[5] The break-even sustainability rate sets the expected insurance premiums equal to the expected credit losses. Ingnoring discounting and for the moment internal refinancing, this equality is given by.

$$P(s)u\left(1-\frac{1}{T}[\sum_{t=1}^{5} r_t]\right)+\frac{a}{T}[\sum_{t=1}^{T} t]=(1-P(s))LGD$$

Substituting the FHA up-front and annual premiums that were in place in 2007 through mid-2008 assuming that the borrower used the minimum downpayment and that T is set to 10 years, we get approximated break-even sustainability rates ranging from 92.2 percent if the loss-given-default is set to 50 percent to 94.4 percent

[4] The sliding scale consists of a rebate of 75 percent in the first year, 55 percent in the second year, 35 percent in the third year, 17.5 percent in the fourth year, and finally 0 percent after 5 years.

[5] This will understate the break-even sustainability rate since terminations tend to occur earlier rather than later in the borrower experience. Earlier exits imply that fewer annual premiums are paid as well as more rebates of up-front premiums.

if the loss-given-default rate is set to 70 percent.[6] If an FHA borrower does an internal refinance, then this lowers the break-even sustainability rate since those borrowers with sustainable homeownership experiences end up paying a higher combined up-front premium. For example, if we assume that the borrowering experience begins with a purchase mortgage in 2007 and the borrower internally refinances in 2011 with all other assumptions as above, then the break-even sustainability rates decreases to 86.4 percent if the loss-given-default is 50 percent, and to 90.2 percent if the loss-given-default is 70 percent.[7]

What little is known about recent FHA performance with regard to sustainable homeownership is contained in the FHA's mandated annual actuarial reviews and its contemporaneous reports to Congress (see for example IFE [2009, 2010, 2011]). As pointed out by Aragon et. al., the FHA's current mortgage-based analysis is inappropriate for the measurement of sustainability. Instead, one must focus on the borrower by constructing borrower experiences which link together consecutive FHA mortgages taken out by the same borrower and are secured on the same property. A borrower experience begins with the initial credit exposure to the FHA fund through a purchase mortgage or a refinance from a non-FHA mortgage and ends either through a full payoff to the FHA or through a default. We now turn to the construction of a borrower based FHA data sample that will permit measurement of the sustainability of recent FHA vintages.

B. Sampling Method

Using borrower experiences to measure sustainability in homeownership requires data in which contiguous FHA mortgages on the same home and with the same borrower are linked together.[8] In

---

[6] The 2011 Actuarial Review (Exhibit E-1, page E-2)reports that the average loss rate for FHA claims in 2009 was 63.67 percent.

[7] The 2009 vintage of FHA borrowers were charged an up-front premium of 175 basis points (compared to 100 basis points earlier) and an annual premium of 55 basis points (compared to 50 basis points earlier). These higher premiums reduce the break-even sustainability rates to 91.3 percent with a loss-given-default of 50 percent and to 93.8 percent with a loss-given-default of 70 percent.

[8] Gerardi *et al* (2007) is the only other example we are aware of in which the frame of reference is the borrower rather than the mortgage. They examine the sustainability of borrower experiences that begin with a subprime mortgage.

particular, to follow a borrower experience we need to match FHA mortgages that are the two sides of an internal FHA refinance. To execute this we specified an FHA borrower data set to CoreLogic, which made available anonymized data from two of its largest databases: (i) the mortgage servicing database containing monthly FHA mortgage performance history data from servicers, and (ii) the property record database containing property record information sourced from county assessors and recorders. Use of the latter data enables us to identify mortgages from the servicing database that are tied to the same property, enabling us to link the data. The CoreLogic servicing database includes monthly performance information for a large percent of the FHA's outstanding insured mortgage portfolio. More recent vintages have a higher coverage of FHA loans, due in part to large servicers becoming more involved in FHA servicing. This database covers more than 7.5 million FHA loans insured between 2003 and 2011. CoreLogic's property record database contains information for real estate sales and mortgage transactions covering 97 percent of the US population.[9]

CoreLogic was able to identify 282,000 FHA to FHA refinance pairs. CoreLogic also provided FHA loans that were not internally refinanced – either purchase or external refinance mortgages. To construct a 5 percent representative sample for estimation, we worked backwards in time. In each year, we aligned with published FHA data the sample composition of originations across the purchase, external FHA refinance, and internal FHA refinance categories. To standardize the sample, we only selected first-lien, owner-occupied mortgages. Starting in 2010, we randomly selected purchase and refinance loans to match the published FHA data scaled down to a 5 percent sample. For the internal FHA refinances that were randomly selected we pulled in the full chain of linked prior FHA mortgages as well. We then adjusted the required number of mortgages of each type for prior years based on any linked mortgages that were already selected into the sample by virtue of being part of a borrower experience linked to the sampled mortgage for 2010. We then

[9] CoreLogic's coverage is 99.9 percent (3,120 counties) of the US population for county assessor-sourced data such as tax valuations and 97 percent (1,800 counties) of the US population for title and mortgage recordings sourced from county recorders.

repeated the exercise for 2009 and proceeded in this manner back to 2007.[10] Our final sample is a 5 percent random sample of FHA originations from 2007 through mid-2010. These loans have performance data up to September 2011.[11]

The sample generated from the CoreLogic linked data closely matches published FHA data in key respects. Appendix Tables A1 to A3 provide comparisons between the published FHA data and our sample for average FICO scores, origination LTV levels, and early delinquency rates. These show that the sample tracks published values of key risk measures quite closely. Since the origination balance reported in the CoreLogic data incorporates any up-front mortgage premium that was financed by the borrower, we back out estimates of these origination fees from the initial loan balance in order to make our origination LTV measures comparable to the published FHA data.[12] We track the measured LTV trends quite closely up to the third quarter of 2010. On average however, our early delinquency rates by year tend to be slightly lower than the published rates.

C. Internal Refinancing

The CoreLogic data confirm the significance of the distinction between the borrower experience and the mortgage experience reflecting the high rate of FHA internal refinancing in recent years. Our sample indicates that FHA-to-FHA refinancing in 2009 has been 44 percent as large as the number of purchase originations, similar to the HUD estimate of 42 percent.

---

[10] These refer to FHA fiscal years (which run from Q4 of the prior year through Q3 of the stated year) so that the last loan for any borrower must be originated no earlier than calendar year 2006 Q4. However, prior loans in a chain for any borrower can have earlier origination dates.

[11] We include in the hazard estimation all of the data including the borrower experiences that begin in 2010, but we censor all monthly observations prior to January 2007. In addition, we restrict our cohort level analysis to FHA borrowers from calendar years 2007 to 2009 for whom a longer history of performance is available.

[12] The upfront premium charged for the 2009 fiscal year through March 2010 was 1.75% for fully underwritten loans and 1.5% for streamline refinance loans. The premium rose to 2.25% in April 2010, for all loans. For loans originated prior to the 2009 fiscal year, the rule of thumb for fully underwritten loans from 2009-2010 (1.75%) was applied. Upfront premiums can be refunded according to the following schedule: http://portal.hud.gov/FHA-Handbooks/collections/current/print/4155-2_7.pdf. We adjust the upfront premiums subtracted from the LTV accordingly.

The FHA's internal streamline refinance program, in particular, has had a high take-up rate in recent years. This program is directed towards FHA borrowers with high LTVs that exceed the maximum allowed for new FHA mortgages. Without necessarily requiring a new appraisal, the program allows these borrowers to refinance into new FHA mortgages with lower monthly payments even in cases in which borrowers no longer have enough equity to meet the standard down payment amount.[13] The FHA has put in place several requirements including that the new balance cannot exceed the prior balance and that the borrower must have a sufficiently clean payment history. Since many of the FHA borrowers using the internal refinancing program are underwater, it will be difficult for them to exit the FHA system by either selling the house or refinancing into a non-FHA mortgage. As such, these borrowers will remain at risk of default for many years.

D. Delinquency and Sustainability

Whether an FHA borrower experience is sustainable or generates a claim against the FHA is not definitively known until the borrower experience finally terminates. This claim represents the last stage of a process that begins with an initial delinquency, is followed by serious delinquency, the onset of the foreclosure process, and finally the sale of the property and settlement of the FHA claim. Given the long delay between the initial delinquency and the settlement of an FHA claim, it is useful to have an earlier indicator that a borrower experience will not be sustainable.

The time between the events that precipitate a borrower's delinquency and the claim is long and highly variable. Table 1 illustrates these time lags derived from our FHA sample for various default triggers, where the time lag is defined as the number of months in an unbroken string of delinquencies from 30-days delinquency until the onset of the specified trigger. On average it takes 3.9 months from an initial delinquency, with a standard deviation of 2.4 months, to reach the 90-days delinquency trigger. Pushing back the definition of the default trigger to be defined as the beginning of a foreclosure process, the time lag

[13] Since no new underwriting is required for a streamline refinance, the origination loan-to-value, debt-to-income, and borrower credit score is not reported.

increases to a mean of 9.6 months with a standard deviation of 6.0 months. Lastly, defining the default trigger to be the start of the REO process, the time lag increases to 15.8 months with a standard deviation of 7.2 months. A high variability of the gap between the initial delinquency and the default trigger poses a challenge for specifying an empirical model of the default event.[14]

Our assessment of FHA performance requires estimating sustainability for FHA borrower experiences that are still ongoing. Using 90-plus day delinquency as a default trigger allows us to gain early insight into what fraction of recent FHA vintages are likely to generate final claims on the FHA, even if these claims may not materialize until much later. Table 2 shows the fraction of FHA mortgages in our data that terminate in a claim conditional on the mortgage ever reaching a given stage of delinquency. The table shows that nearly 13 percent of mortgages that reach 90-days delinquency have terminated by the end of our sample, 65 percent of which have generated a claim. We can estimate the "cure" rate for each default trigger by adding the number of terminated borrower experiences that fully pay off to the number of active borrower experiences that are no longer delinquent, and then dividing by the number of borrower experiences that ever reach the default trigger. The estimated cure rate declines from 24.5 percent for the 60-plus default trigger, to 20.7 percent for the 90-plus trigger, and down to 15.4 percent for the foreclosure start trigger. The choice of the appropriate definition of the default trigger needs to balance the aims of a low variability of time since the initial delinquency with a low cure rate. We define the default event to be the first time the mortgage reaches 90-days delinquent.

We can use our FHA sample to evaluate the sustainability of recent vintages of FHA borrowers. The first step is to simply tabulate the fraction of FHA borrowers in each vintage who default (that is, ever reach 90 days delinquent) and who fully pay off their FHA mortgages. We can contrast this to the same tabulations constructed for vintages of FHA mortgages instead of FHA borrowers. These tabulations are provided in Table 3. As of the end of our sample in September 2011, the default rate for the 2007 vintage of FHA borrowers is 43 percent, and for the 2008 vintage of borrowers is 30 percent In contrast, the prepayment rate

[14] The increase in the mean time lag is not as serious a problem since the time-varying control variables can be lagged appropriately.

(reflecting a full payment to the FHA) is only 11 percent for both vintages. For the 2007 and 2008 vintages, organizing the data by mortages rather than borrowers triples the measured prepayment rates. In addition, using mortgages rather than borrowers as the data structure results in an under count of the default rates for the 2007 and 2008 vintages by between 24 and 29 percent respectively. To understand this result, consider an FHA borrower who takes out a purchase mortgage in 2007, does an internal refinance in 2009, and defaults in 2010. This default is properly assigned to the 2007 vintage in the borrower data, but is mis-assigned to the 2010 vintage in the mortgage data. In addition, the mortgage data records a prepayment of the 2007 mortgage as well as the default of the 2010 mortgage, whilc the borrower data correctly records only the default for the 2007 borrower. This underscores the point that borrower based data is required to correctly measure the sustainability of FHA homeownership experiences.

The breakeven sustainability rates derived earlier were based on the assumption that each vintage of FHA borrowers fully have terminated their borrower experiences within 10 years. It is of interest, then to project the sustainability rates for the 2007 to 2009 vintages of FHA borrowers out 10 years following their origination dates. In order to do this forecast exercise, we need to estimate a statistical model of borrower performance.

III. The Proportional Hazard Model

A. Econometric Specification

We use a standard competing-risk model to analyze the impact of borrower risk characteristics, mortgage and property characteristics, and economic factors on prepayment and default outcomes. We use a proportional hazard framework assuming independent risks. The prepayment (*p*) and default (*d*) hazard rates since origination at duration *t* are given by:

$$h^p\left(t|X_t^p\right) = \exp\left(g^p(t)\right)\exp\left(X_t^p\beta^p\right) \qquad (1a)$$

$$h^d\left(t\middle|X_t^d\right) = \exp\left(g^d(t)\right)\exp\left(X_t^d\beta^d\right) \quad (1b)$$

where *g(t)* is the baseline hazard function of the time since the mortgage was originated. We approximate the baseline hazard using a monthly step-function.[15] The key assumption is that the explanatory variables X_t shift the baseline hazard proportionally.[16]

As detailed above, we define the default event as the borrower experience first reaching 90-days delinquent and censor any remaining payment history on the borrower experience.[17] We use borrower experiences as the frame of reference, hence prepayment only occurs when the borrower pays off the FHA mortgage by either selling the house or refinancing into a non-FHA mortgage. We follow each borrower experience starting in the third month after its origination date up until either the borrower fully prepays the FHA, the default event occurs, or the sample ends.[18] Our data are at a monthly frequency.

With the estimated prepayment and default hazards we can calculate the probability that any active FHA borrower will default over a specified horizon and consequently, the probability that the borrower experience survives this horizon. Let *S* denote the estimated joint survivor function, given by:

$$S(t) = \exp(-\sum_{j=1}^{t}(h^p(t) + h^d(t))) \quad (2)$$

where *t* indexes the number of months into the forecast period and *S*(0) equals one.

---

[15] See Meyer (1990) for an early example of using step-function approximations to the baseline hazard.

[16] See Kalbfleisch and Prentice (2002) for details. More complicated hazard models can be estimated on the borrower experience data where the hazard functions are allowed to vary across each of the mortgages in a given borrower experience. For a discussion of these modeling issues see Flinn and Heckman (1983).

[17] That is, we do not model whether a seriously delinquent mortgage cures or the time it takes to work through the various stages of foreclosure to a final claim.

[18] The third month is the first time period that the borrower is at risk of going 90 days delinquent.

For forecasting default and prepayment probabilities, we need to specify the path of the dynamic variables over the forecast horizon. The estimated probability that an active FHA borrower experience with current duration *t* will default and prepay over the next *T* months is given by:

$$Pr^D(t,T) = \sum_{j=t+1}^{T} (S(j)\, h^d(j)) \tag{3a}$$

$$Pr^P(t,T) = \sum_{j=t+1}^{T} (S(j)\, h^p(j)) \tag{3b}$$

We compute these default and prepayment projections for all active FHA borrower experiences starting at the end of our sample to generate an overall prepayment and default rate.

While the default and prepayment hazards condition on a large list of observable factors that could affect default, there may still be unobservable factors that are important for determining default rates. These unobservable factors may be correlated with the "vintage" of the mortgage when it was underwritten. We test for the presence of unobserved determinants by estimating the model allowing for a parametric distribution of the unobserved heterogeneity. The data indicate that this distribution was degenerate.[19]

B. Variables in the Specification

The CoreLogic data provide information on a variety of borrower, loan, and property risk characteristics. We supplement these with data on economic factors and state legal requirements that may impact underlying mortgage performance. For a borrower with more than one FHA mortgage, we use reported characteristics of the refinances as opposed to carrying over characteristics from the previous loan. In the case of streamline refinances however, where origination data are often missing, we carry over loan characteristics from the

[19] That is, the variance of the distribution of unobserved heterogeneity was converging to zero in the estimation.

previous loan. Summary statistics of the variables in the estimation are given in Table 4. The first panel presents summary statistics of static variables for underwritten loans (not streamline refinances) whose values do not change over the life of the underlying mortgage. The second panel presents dynamic variables whose values are time-varying. For the categorical variables, the left-out group is selected to be either the high-quality or the relatively more common category.

A borrower's current LTV likely impacts the probability of default and prepayment. We calculate this using the current mortgage balance – reflecting amortization, any accelerated payments, and inclusive of any upfront mortgage insurance premium financed by the borrower – and an estimate of the current value of the home. To estimate the house value each month, we update the property appraisal value using the CoreLogic metro area overall repeat-sale house price indices. Thus, the current LTV is a dynamic variable that changes over time to reflect both debt amortization as well as house price changes. For streamline refinances, we impute the origination LTV using the updated LTV at the end of the prior FHA mortgage. In the estimation, we include intervals for this current LTV variable beginning with an indicator for 80-84 LTV and ending with an indicator for 120 LTV or higher. The left-out category is a current LTV below 80.

Two other important borrower-specific risk factors are the credit score (FICO) and the debt-to-income ratio (DTI). For underwritten mortgages, we observe the borrower's FICO score at the origination date, so in our data FICO is a static variable that does not change over the life of a mortgage. However, it can change between mortgages for a given borrower experience if more than one underwritten mortgage is involved. For internal FHA refinances, we use the new FICO score if the loan is underwritten, while we carry over the FICO score from the prior FHA mortgage for streamline refinances. In the estimation, we include a series of indicator variables for FICO score ranges from below 580 to 680-719. The left-out category is a FICO score of 720 or higher. We include a separate indicator if the FICO score is missing and the mortgage is not a streamline refinance.

We also include the borrower's back-end DTI which is a measure of the borrower's ability to meet monthly mortgage payments. The numerator of the back-end ratio is the sum of the annual mortgage

payments, property taxes, house insurance and any other annual recurring debt payments such as student loans, auto loans and minimums on credit card balances. The denominator is the borrower's annual income.[20] We include indicators for DTI intervals from 28 – 35, 36 – 43, and 44 or higher. The left-out category is a back-end DTI of less than 28. We include an indicator if the DTI is not recorded and the mortgage is not a streamline refinance.

Additional indicators are included for the type of mortgage (fixed rate mortgages (FRM) are the left-out type), term of the loan (30-year is the left-out term), reason for the mortgage (purchase loan is the left-out reason), level of documentation (full documentation is the left out status), and property type (single family residence is the left-out property type). In addition, we control for the size of the FHA mortgage at the origination. Models of strategic default predict that the default risk should increase with the mortgage balance since the cost of default is mostly fixed while the benefit tends to increase with the size of the remaining balance.[21]

Several local economic variables are included. To help ensure that the coefficients on the LTV indicators reflect the effect of the borrower-specific equity position on observed default behavior, we include the 12-month change in local house prices to capture the effect of non borrower-specific factors. FHA mortgages with high LTVs are likely to be located in local housing markets that have suffered more serious house price declines and also have faced many other economic challenges as well. While we attempt to control for local dynamics, there are other stress variables not completely captured in our data. Including the change in local house prices helps isolate any of these left-out factors that are correlated with declining house prices, which in turn impact borrower behavior.

---

[20] In cases where there is a co-applicant for the mortgage the sum of the two annual incomes is used.

[21] See for example Haughwout *et al* (2010).

The effect of an unemployment spell on a borrower's behavior is difficult to capture since we do not observe these spells.[22] We use the MSA unemployment rate as a proxy, although it likely is not highly correlated with the unobserved borrower-specific unemployment. We include the MSA unemployment rate reported by the U.S. Bureau of Labor Statistics lagged 6 months to take into account the time lag between the onset of unemployment and when the borrower reaches the 90-day delinquency status. The unemployment variable is dynamic at a monthly frequency.[23]

We capture the incentive to prepay a mortgage by including a variable that reflects any decline in mortgage interest rates since origination. This incentive is calculated each month by taking the difference between the average 30-year fixed-rate mortgage interest rate at the origination month and its average value at that month, then setting it to zero if the difference is positive, i.e. interest rates have increased.[24]

Local contagion may affect a borrower's decision to default as well. A recent survey (Fannie Mae, [2010]) finds that borrowers who know someone who has experienced a foreclosure respond that they are more than twice as likely to seriously consider default compared to borrowers who do not.[25] We control for possible contagion effects by including the number of distressed sales per 10,000 households in the MSA. We calculate a three-month moving average of distress sales for each MSA using CoreLogic data. The number of households in the MSA is taken from the 2008 American Community Survey.[26]

Finally, the legal environment that governs how mortgage delinquencies are handled varies by state and we include two variables to capture this variation. First, mortgages originated in a state with a judicial

---

[22] See Foote *et al* (2008) and Gerardi, Shapiro, and Willen (2007) for discussions of the "double-trigger" hypothesis that the combination of a borrower being in negative equity and then suffering an income shock leads to a higher likelihood of default and foreclosure.

[23] For purchase mortgages, we zero out the first four observations (months 3 – 6 following origination) on the unemployment variable since the lag value is pointing to a time before the mortgage is underwritten. This is based on the premise that an unemployed borrower would not be approved for a new FHA mortgage.

[24] We use the mortgage rate reported in the Freddie Mac Primary Mortgage Market Survey.

[25] A similar result is also found in Guiso *et al* (2009).

[26] http://factfinder.census.gov/servlet/DTSelectedDatasetPageServlet?_lang=en&_ts=286380818796)

foreclosure process can expect longer delays in completing any foreclosure, which could incentivize borrowers to strategically default.[27] An indicator for judicial foreclosure is included in the estimation. Second, mortgages are considered as recourse depending on the state, meaning the lender has the ability to pursue a defaulted borrower with a deficiency judgment. Recourse loans potentially provide more security to the lender and as a result, borrowers may be less likely to default. We also include an indicator for recourse.

A concern with loan-level data is that we only have information on first lien mortgages. If the borrower takes out a second lien, then the actual combined LTV will exceed the measured LTV used in the analysis. This will bias the hazard coefficient estimates on the LTV intervals to the extent that second liens are prevalent in the data. To explore this we examined a unique panel data set which links credit files across household members over time constructed by Equifax for the Federal Reserve Bank of New York. For FHA borrowers, we calculate the fraction who have second liens. The data indicated that second-liens is not a significant source of concern.[28]

IV. Estimation Results

We report the exponentiated hazard coefficients from the estimation in Table 5. A reported hazard above (below) one indicates a higher (lower) default or prepayment risk relative to the baseline borrower experience consisting of a high-quality, fully-documented, 30-year, fixed-rate purchase mortgages with a current LTV below 80, a FICO score above 720, a DTI below 28, and where the property is secured by a single family residence.

---

[27] See for example discussions of strategic default in Foote *et al* (2008) and Haughwout *et al* (2010).

[28] The Equifax data follows a 5 percent random sample of households over time by linking credit files across household members. As of December 2009, the data indicate that only 6.7 percent of FHA borrowers had some form of a second lien. In contrast, 27.6 percent of prime borrowers and 25.9 percent of non-prime borrowers hold at least one second lien. Conditional on an FHA borrower having a second lien, the average balance on the second lien is 22.4 percent of the existing combined balances across all liens. Hence not having data on second liens poses less of a problem for FHA borrower experiences. We also looked at households with an FHA mortgage and no other first lien mortgage. While it is possible that the borrower has two homes, with an FHA mortgage on one house and only a second lien on the second house, we do not expect many such cases. It is then highly likely that any existing second-liens are tied to the same property.

We begin by examining the impact of the current LTV on the default and prepayment risks. [29] The results show high LTV to be an important driver of FHA default risk, which rises monotonically with the estimated current LTV of the underlying mortgage. Holding other factors constant, a borrower with an estimated current LTV between 100 and 104 is more than twice as likely to default compared to a borrower with an estimated current LTV below 80. As we raise the current LTV to 120 or higher, the relative default risk increases to over three and a half times higher than the baseline. Recall that these estimated LTV effects reflect changes in default risks holding constant the change in local house prices over the past year.

A borrower's LTV is also a determinant of prepayment. Recall that prepayment in our borrower-based data involves either paying off the mortgage due to a sale of the house or refinancing to a new non-FHA mortgage. Finding the required resources to pay off a mortgage in full is particularly hard for borrowers in negative equity (LTV more than 100) who need to make up the difference between the property value and the mortgage balance in cash.[30] Comparing a borrower with an estimated current LTV of 100 – 104 to one with an estimated LTV below 80, the prepayment rate is reduced to 64 percent of the baseline. As we raise the estimated current LTV to 120 or higher, the prepayment rate declines further to 50 percent of the baseline. The data indicate that a borrower with an underwater mortgage faces a combination of a high default risk and a low prepayment rate. This combination implies a high cumulative probability that the borrower experience will be non-sustainable since the borrower will be exposed to the high default risk for a prolonged period of time.

The FHA has stressed the improvement in the FICO scores for its new originations over the past couple of years as a factor that will hold down credit losses compared to earlier vintages.[31] This improvement in

---

[29] Note, however, that the measured current LTV for a borrower is a noisy measure of the true current LTV. As a result, some borrowers with measured LTVs above 100 in fact have sufficient equity to sell their house and pay off the mortgage balance from the proceeds of the sale, while others with measured LTVs below 100 may in fact be in negative equity.

[30] The effect of this "collateral constraint" on prepayments has previously been observed in other mortgage products (see Caplin *et al* (1997)). The FHA streamline refinance program is designed precisely to prevent the collateral constraint imposed by high LTVs from limiting borrowers' ability to refinance.

[31] See Secretary Donovan's written testimony before the House Committee on Financial Services, December 1, 2011.

FICO scores can be seen in Table A1 where the average FICO score for all new originations went from the low to mid 600s in 2007 to close to 700 in 2010. Our results confirm that the credit score at origination is a strong predictor of default. The default risk rises dramatically as the borrower's FICO score is lowered. Relative to a borrower with a FICO score above 720, the default hazard is over seven times higher for borrowers with FICO scores between 580 and 619, and over ten times higher for borrowers with FICO scores below 580.[32]

The affordability of the FHA mortgage as gauged by the borrower's back-end DTI is also a determinant of the sustainability of an FHA borrower experience. Recall that the back-end DTI captures not only the mortgage and related housing costs (such as property taxes and insurance), but also other recurring debt obligations. As we move from a borrower with considerable budget leeway (a back-end DTI below 28) to a borrower who is more cash-flow constrained in making all required payments (a back-end DTI of 44 or higher), the relative default hazard increases by more than 50 percent.

Turning to the other mortgage-specific factors, we find that FHA borrower experiences that begin with a cash-out refinance are 28 percent more likely to default and 47 percent more likely to prepay (holding constant the current LTV and the mortgage balance). In terms of documentation, the results indicate that borrower experiences involving mortgage originations with less than full documentation are associated with lower default risks than the baseline. However, the data also indicate that borrower experiences involving mortgages with missing LTV, FICO and DTI information all have significantly higher default risks. The mortgages with these missing origination characteristics are concentrated in the low- and no-doc loans. Taken together, this indicates that mortgages that are not subject to full underwriting pose higher credit risks to the FHA. Adjustable rate mortgages are over three times more likely to prepay than FRMs. Lastly, borrower experiences with higher mortgage origination balances have higher prepayment and default risks. The higher

[32] Deng and Gabriel (2006) also find that FICO scores are a strong predictor of FHA defaults using data covering from 1992 to 1996. Their results indicate a smaller effect than our findings. This may reflect that they can control for other borrower characteristics such as age, sex, race and number of dependents.

default risk is consistent with simple models of strategic default where the incentive to strategic default is increasing in the mortgage balance controlling for the current LTV (see Haughwout *et al* (2010)).

Housing market-specific variables are meant to capture local economic determinants of prepayment and default. As noted earlier, the local unemployment rate is included as a proxy for whether the borrower experiences an unemployment spell. While measurement error will attenuate the estimated impact, the data still suggest that the lagged MSA unemployment rate is significantly related to the default risk on an FHA borrower experience. Controlling for the current LTV, declining house prices over the past year are associated with rising prepayment and default risks. This could reflect a variety of other economic factors in local housing markets that are correlated with declining house prices. The data do not indicate the presence yet of a contagion effect from distress sales. With regard to the interest rate incentive to prepay, the interest rate spread is a strong predictor of prepayment for borrower experiences involving fixed-rate FHA mortgages.[33]

Finally, the legal environment governing mortgage lending which varies by state affects default risk. The data indicate that judicial foreclosure is associated with higher default risk on FHA borrower experiences, supporting the hypothesis that borrowers may strategically default in response to long foreclosure delays. As noted earlier, mortgages with recourse should raise the costs of default and therefore lower the default risk on a mortgage. The data indicate, however, that default risk is 11 percent higher for borrower experiences located in states with recourse lending.

V. Forecasts

We use the estimated model to generate default and prepayment forecasts for the portfolio of active FHA borrower experiences as of the end of our estimation sample. The forecast is performed over a horizon producing a ten year window between the origination year and the end of the forecast period. This is done to match our earlier calculations of the break even sustainability rates which assumed a ten year payoff period.

[33] Deng and Gabriel (2006) also find this result for a sample of FHA mortgages originated between 1992 and 1996.

We use updated dynamic variables based on actual values from the end of our sample to current and going forward consensus forecasts for four underlying data series – house prices, unemployment rate, 30-year fixed-rate mortgage interest rates and distressed sales ratios. In particular, the unemployment rate is projected to decrease by a total of 1.8 percentage points by the end of 2015 and then flatten at the median long-run value given in the Survey of Economic Projections.[34] Mortgage rates are projected to stay flat through 2015 (following the FOMC's statements) and then increase by 20 basis points in each of the five years.[35] The percent of distressed sales is projected to stay at current levels.[36] From the end of our sample until July 2012, we update house values using CoreLogic MSA repeat-sale house price indices. For the remainder of the forecast period, we update house values using the average house price expectations from the Macro Markets Survey.[37]

The projected ten year performance record for the 2007 to 2009 vintages are provided in Table 6. Overall, 31 percent of borrowers from the 2007 to 2009 vintages are expected to default by the tenth year anniversary dates. Looking across the three vintages, the cohort of 2007 FHA borrowers is projected to perform the worst with over 50 percent of these borrowers defaulting on their loans. While the 2009 vintage is projected to perform better, the default rate is still estimated to be nearly 25 percent with 50 percent projected to still be active.

In addition to the continuing high rate of default, the other striking factor is how few borrowers are projected finally to demonstrate sustainable homeownership by paying off their FHA-insured mortgages in full. Across all three vintages, the estimated maximum sustainability rates are well below the breakeven levels

---

[34] Unemployment forecast taken from Survey of Economic Projections
http://www.federalreserve.gov/monetarypolicy/files/fomcprojtabl20120125.pdf

[35] Author's assumption

[36] Author's assumptions

[37] See https://pulsenomics.com/Sept2011-HPE-Survey.html

presented earlier. The final column of Table 6 presents the projected maximum sustainability rate for each vintage. We use the conservative assumption that all borrower experiences that are still active at the end of the ten years will successfully terminate without a clam to the FHA. The projections indicate that all three vintages of FHA borrowers will negatively impact the capital in the FHA insurance fund.

There is a simple reason why recent FHA borrowers may not fit the historical pattern in which after a few years default rates decrease rapidly over time. In the past, generally rising house prices and the availability of non-FHA mortgages have made it possible for borrowers to pay off their FHA mortgages in full either through a sale or a refinance to a non-FHA mortgage. Neither of these routes is currently available for most FHA borrowers. There has been little or no increase in prices since these borrowers purchased their homes – in fact, we estimate that 40 percent of the active borrowers at the end of our sample period are in negative equity, and another 20 percent are near negative equity with an LTV of 95 to 100. Additionally, house prices are not expected to significantly increase over the next five years. As a consequence, many recent FHA borrowers are likely to remain vulnerable to economic shocks so that defaults are projected to continue at a high rate.

An important policy question is whether the significant expansion of FHA lending following the onset of the housing was a success. An important element to answering this question is the degree to which the FHA's increased lending activity was producing sustainable homeownership experiences. To date, no data has been provided by the FHA addressing this key question. The bottom line from our analysis is that even if all of the FHA borrowers in our sample that are projected to be still active at the end of the vintage's ten year anniversary subsequently successfully pay off – an improbable event –the cumulative default rates for the 2007 and 2008 vintages of FHA borrowers will both be projected to exceed 40 percent. A remaining question is whether at the time these vintages were originated could it have reasonably been expected that their performance would have been much better than our estimates indicate? This also seems unlikely given the existing weakness in housing markets at the time combined with the low downpayments for these mortgages and the prevailing high unemployment rate. The evidence suggests that using the FHA as a lender-of-last

resort during times of housing market stress may be costly both for borrowers, their neighborhoods, and for taxpayers.

VI. Underestimation of Risk

The FHA is designed to be a self-funded mortgage insurance program and not to need direct taxpayer support. To gauge the FHA's performance and to evaluate its financial condition, an external audit of its insurance fund is conducted each year. A key part of this audit review involves projecting the value of the FHA's insurance fund based on currently active FHA mortgages. As shown in Figure 1, for the past three yearsthe forecast for the FHA's insurance fund from the prior year has been subsequently revised lower.[38]

The recent pattern of underestimation of future credit losses may be due to deficiencies in the audit analysis (Gyourko [2011). Aragon et al [2010] argued that an important flaw is the data framework used in the analysis – it is based on FHA mortgages rather than on FHA borrower experiences. In the mortgage-based data structure an internal refinance is treated as a successful exit as if the credit risk had been eliminated. In contrast, in the borrower experience data structure, subsequent FHA mortgages taken out by the same borrower are tracked and the borrower's risk only leaves the FHA system when the borrower pays off the final FHA mortgage in full. This distinction between the data structures is quantitatively important since recent reductions in mortgage interest rates have dramatically increased the volume of internal refinancing from one FHA mortgage into another. In our sample where we track borrowers and their possibly multiple FHA mortgages, we estimate that only 6.4 percent of FHA borrowers since 2007 have successfully exited the FHA program. In contrast, if we use the mortgage data framework, the apparent success rate is three times as high at 19.4 percent.[39] This reflects the fact that a majority of terminated mortgages have in fact immediately been refinanced back into new FHA mortgages.

---

[38] IFE 2009 (page i), 2010 (page iii) and 2011 (page iii).

[39] Around 18 percent of our borrowers have two or more FHA mortgages comprising their current borrower experience. The time in months between a mortgage origination (either a purchase mortgage or a refinance from a non-FHA mortgage) to an internal refinance peaks at around 12 months.

The intuition that an analysis using mortgages rather than borrowers overstates successes to date and consequently leads to understating future losses has not resonated with everyone for a couple of reasons. First, it is standard academic and industry practice to perform credit risk analysis at the mortgage level. In this sense, the external audit seems to be implementing what is considered the "best practice". Second, the FHA have argued that there is no understatement of risk because for internal refinances the credit risk is carried forward in the data as the FHA mortgage that prepays is replaced by the new FHA mortgage. Specifically, the FHA states in an online publication that: "Streamline refinance loans do not disappear but are counted as newly insured loans, and with the risk characteristics that are appropriate for them.".[40] The FHA's explanation confuses what takes place in-sample and out-of-sample. As we have noted, in-sample an FHA mortgage that prepays through a streamline refinance is replaced by the new FHA mortgage. In this sense, the credit risk in the estimation sample is not eliminated from the data. However, the same does not necessarily occur out-of-sample over the forecast horizon. To see this, consider the example of a new FHA purchase mortgage taken out in 2011 that is still active as of the end of the data. In the forecast exercise, the estimated competing risk model based on the mortgage data will predict a prepayment rate at each future date for this purchase mortgage based on an estimated prepayment hazard that reflects both real prepayments (where the credit risk terminates) and internal refinances (where the credit risk continues). As a consequence, the predicted prepayment rate for this purchase mortgage over the forecast horizon that consists of a mix of true prepayments and internal refinances. For the fraction of the predicted prepayments that represent future internal refinances, the analysis does not replace this mortgage with its successor mortgage. Instead, in the forecast all predicted prepayments are assumed to reflect a termination of the credit risk of the borrower to the FHA. Furthermore, the same problem also exists to a degree for mortgages that start out in the estimation sample as streamline refinances. The FHA allows a borrower to streamline refinance multiple times. So, a forecasted prepayment for an active streamline refinance mortgage will with some probability also reflect a subsequent streamline refinance and not a payoff of the credit risk to the FHA.

---

[40] See pages 5-6: http://portal.hud.gov/hudportal/documents/huddoc?id=MythsandFactsLoanPortfolio.pdf

Our customized FHA data allow us to investigate the implications of the data structure on projected credit losses. We already presented the hazard estimates based on analyzing the data organized by borrower experiences in Table 5. For comparison, Appendix Table A4 presents the hazard estimates based on analyzing the data organized by mortgages. It is important to highlight that the same data is used in both cases, the same variables are used to explain defaults and prepayments, and the same statistical model is used in the estimation. The only difference is how the unit of observation is defined – a borrower experience versus a mortgage.

The relative hazard coefficients between the two approaches are similar in many respects. The key difference is seen in the coefficients of the high LTV indicators in the prepayment hazards and in the baseline hazards. Data organized according to borrower experiences indicate that those with a current LTV of 120 or higher have a prepayment rate that is only 50 percent of the rate for borrowers with a current LTV of 80 or less. In contrast, data organized according to mortgages produces a relative prepayment rate of 78 percent of the baseline. The same pattern exists for each of the high LTV indicators which indicate that the FHA's streamline refinance program does help to reduce the friction that high LTVs otherwise would create for refinancing.

Figures 2 and 3 contrast the estimated default and prepayment baseline hazards between the borrower and mortgage data frameworks. The baseline default hazard curves are very similar for both data frameworks for the first two and a half years. Afterwards, the borrower baseline default hazard is above the mortgage baseline default hazard. However, the differences are more apparent for the baseline prepayment hazards. The mortgage data approach significantly raises the estimated baseline prepayment hazards relative to the borrower-centered data approach particularly in the first two years.[41] The striking difference in the estimated baseline prepayment hazards meaningfully impacts default projections going forward by underestimating the amount of time for which the borrower will remain at risk of default. This is illustrated in Figure 4 where we

[41] While the direction of the effect on the baseline prepayment baseline hazards should be robust, the magnitude of the difference between the borrower and mortgage based hazards will depend on the degree to which the borrower data base spans periods where the internal refinancing activity has been less active than of recent.

show the estimated prepayment survivor functions over the five year forecast horizon for the active mortgages at the end of our sample. At any point in the forecast horizon, the survivor functions as we have constructed them indicate the fraction of the borrowers (or mortgages) that would be predicted to still be active assuming that there are no exits due to serious delinquency. Using borrower data, 92 percent of the borrowers are predicted to still be active in five years. In contrast, using mortgage data, only 47 percent of the mortgages are predicted to still be active in five years – a decline of 45 percentage points.

Table 7 provides 5-year projections of the estimated hazard models based on the two different frameworks for the same set of mortgages that are still active at the end of our sample. To reiterate, the only difference is due to the data structure since all variables, including forecast paths of time-dependent variables, are identical. Using the data organized by borrower experiences, we estimate that 24.6 percent of the currently active borrowers will prepay over the next five years. In contrast using the data organized by mortgages, we estimate that the prepayment rate will be much higher at 80.1 percent. Switching the analysis from FHA borrowers to mortgages increases the forecasted prepayment rate by 55 percentage points due to the recent heavy use of the FHA's internal refinance programs. Similarly, using the estimates from the borrower data we project that 14.6 percent of the active mortgages at the end of our sample will reach 90 days delinquent over the five year forecast horizon. In contrast, using the estimates from the mortgage data we project that only 9.0 percent will reach 90 days delinquent – a 36 percent underestimate of the five year ahead default risk.

As noted above, in the estimation model used in the 2011 actuarial review mortgages that begin as a streamline refinance are allowed to have separate estimated default and prepayment hazards.[42] This raises the question of how well this strategy does at approximating the results one would obtain from switching to the borrower data framework. The results in Table 7 indicate that in our sample of FHA data this strategy does not move the mortgage based estimates closer to the borrower based estimates. In our data, allowing streamline refinances to have their own competing risk specifications is an inadequate substitute for using borrower based analysis. In the 2012 actuarial review, streamline refinances are treated as a separate type of

[42] See Integrated Financial Engineering (2010), page A-11.

exit. In the forecast, mortgages that are predicted to prepay with a streamline refinance are probabilistically included back in the estimate of future volumes of FHA originations.[43] However, the description of the origination volume model is silent on how the higher credit risk associated with these internal refinance borrowers is preserved. Given that the FHA can organize their data by borrower, the FHA can easily compare the impact on their expected credit losses from moving from a mortgage to a borrower based analysis.

VII. Concluding Remarks

We produce first results on the sustainability of homeownership for FHA borrowers between 2007 and 2009. More than 18 percent of these borrowers have already been 90 days or more delinquent, with less than half that number demonstrating sustainable homeownership by paying off all FHA mortgages. We project that the proportion of FHA borrowers from these three vintages who have been 90 days or more delinquent will rise to 31 percent by the end of their ten year anniversaries. For the 2007 vintage, the default rate is projected to exceed 50 percent.

We show that accurately measuring sustainability requires a new data structure organized around borrowers rather than mortgages. In the presence of an active internal refinance program, a mortgage based data structure results in future defaults being under-estimated and prepayments overestimated. This combination artificially lowers loss estimates to the FHA's insurance fund. This same data structure is used by the GSE's (Fannie Mae and Freddie Mac), even after the introduction of their high LTV refinance program – the Home Affordable Refinance Program (HARP).

Opening up the FHA and the GSE's data to external researchers would promote transparency and facilitate external validation of the risk analysis undertaken by these institutions. Even without making their

[43] See Appendix F of the 2012 Review.

data publicly available, the FHA and the GSE's could readily adopt the borrower-based data structure developed herein. This would better inform policy-makers charged with oversight of these institutions.

References:

Aragon, Diego, Andrew Caplin, Sumit Chopra, John V. Leahy, Yann LeCun, and Mario Scoffier. "Reassessing FHA Risk." NBER Working Paper No. 15802, March 2010.

Brevoort, Kenneth P. and Cheryl R. Cooper. "Foreclosure's Wake: The Credit Experience of Individuals Following Foreclosure." Finance & Economics Discussion Series 2010-59, Board of Governors of the Federal Reserve System, 2010.

Caplin, Andrew. "The FHA Reform Act of 2010." Testimony before House Committee on Financial Services, Subcommittee on Housing and Community Opportunity, March 11, 2010. http://www.house.gov/apps/list/hearing/financialsvcs_dem/caplin.pdf

Caplin, Andrew. "Perspectives on the Health of the FHA Single-Family Insurance Fund." Testimony before the House Committee on Financial Services, December 1, 2011.

Caplin, Andrew, Charles Freeman, and Joseph Tracy. "Collateral Damage: Refinancing Constraints and Regional Recessions." *Journal of Money, Credit, and Banking* 29 (November 1997): 496-516.

Deng, Yongheng, and Stuart Gabriel. "Risk-Based Pricing and the Enhancement of Mortgage Credit Availability Among Underserved and Higher Credit-Risk Populations." *Journal of Money, Credit and Banking* 38 (September 2006): 1431-1460.

Fannie Mae Foundation. *The Fannie Mae National Housing Survey*, 2010.

Flinn, Christopher, and James Heckman. "The Likelihood Function for the Multistate-Multiepisode Model in "Models for the Analysis of Labor Force Dynamics"." In *Advances in Econometrics*, edited by R. Bassman and G. Rhodes, 225-231, 1983.

Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen. "Negative Equity and Foreclosure: Theory and Evidence." *Journal of Urban Economics* 6, no. 2 (2008): 234-245.

Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." Working Paper No. 07-15. Federal Reserve Bank of Boston, December, 2007.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. "Moral and Social Constraints to Strategic Default on Mortgages." Working Paper No. 15145. National Bureau of Economic Research, July, 2009.

Gyourko, Joseph. *Is FHA the Next Housing Bailout?* A Report Prepared for the American Enterprise Institute, 2011.

Haughwout, Andrew, Ebiere Okah, and Joseph Tracy. "Second Chances: Subprime Mortgage Modification and Re-Default." Staff Report No. 417. Federal Reserve Bank of New York, August, 2010.

Integrated Financial Engineering, Inc. *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund (Excluding HECMs) for Fiscal Year 2009*. Prepared for U.S. Department of Housing and Urban Development. Washington, DC, November 6, 2009.

Integrated Financial Engineering, Inc. *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund (Excluding HECMs) for Fiscal Year 2010*. Prepared for U.S. Department of Housing and Urban Development. Washington, DC, 2010.

Integrated Financial Engineering, Inc. *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund Forward Loans for Fiscal Year 2011*. Prepared for U.S. Department of Housing and Urban Development. Washington, DC, October 12, 2011.

Kalbfleisch, Jack D., and Ross L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley, 2002.

Meyer, Bruce D. "Unemployment Insurance and Unemployment Spells." *Econometrica* 58 (July 1990): 757-782.

U.S. Department of Housing and Urban Development, *Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund for Fiscal Year 2005*. Washington, DC, October 14, 2005.

U.S. Department of Housing and Urban Development. *Annual Report to Congress Regarding the Financial Status of the FHA Mutual Mortgage Insurance Fund Fiscal Year 2010*. Washington, DC, November 15, 2010.

U.S. Department of Housing and Urban Development. *Written Testimony of Secretary Shaun Donovan. Hearing before the House Committee on Financial Services.* "Status of the FHA Mutual Mortgage Insurance Fund and the FY 2011 Actuarial Report." Washington, DC, December 1, 2011.

Table 1. Time from 1st Missed Payment to "Default" Trigger for Loans Ending in REO

Trigger Definition	Mean	Std Dev	25th	50th	75th	Maximum
60+	2.2	1.4	1	1	2	35
90+	3.9	2.4	2	2	5	40
Foreclosure start	9.6	6.0	5	7	12	55
REO start	15.8	7.2	10	14	20	61

Source: CoreLogic FHA Data

Table 2. Definition of "Default" and Likelihood of a Claim

	Total	Still Active		Censored	Terminated			
Trigger Definition	Reached Trigger	Last Obs Delinq	Last Obs Current	Servicing Transferred	Paid Off or REO	Pct Paid Off or REO	Pct REO or Short-Sale	Est. "Cure" Rate
60+	42,992	25,945	8,724	3,643	4,680	10.9	60.9	24.5
90+	35,131	21,801	5,728	3,154	4,448	12.7	64.9	20.7
Foreclosure start	17,045	9,037	1,601	2,479	3,928	23.0	73.6	15.4

Notes: Percent REO is conditional on a loan being paid off. Estimated cure rate = (number paid off w/o claim + number w. last observation current) / number that reach trigger.
Source: CoreLogic FHA Data

Table 3. FHA Sustainability – borrower vs. mortgage

	Borrower data			Mortgage data		
Vintage	Active	Default	Prepay	Active	Default	Prepay
2007	46.4	42.7	10.9	35.0	32.3	32.7
2008	59.3	29.6	11.2	42.0	21.0	37.0
2009	86.9	9.5	3.6	84.8	9.3	5.8
Combined	75.0	18.6	6.4	64.5	16.1	19.4

Notes: Default is defined as the borrower ever reaching 90-days delinquent. Prepay for the borrower data takes place when the borrower either sells the house or refinances into a non-FHA mortgage. Prepay for the mortgage data includes internal refinances as well.

Table 4. Summary Statistics

	a) Static Variables for Underwritten Loans			
Variable	Mean	Std Dev	Minimum	Maximum
Credit Score (FICO):	673	64	300	900
Less than 580	0.061	0.238	0	1
580 – 619	0.120	0.325	0	1
620 – 679	0.353	0.478	0	1
680 – 719	0.177	0.382	0	1
720 or higher	0.237	0.425	0	1
Missing	0.053	0.224	0	1
Debt-to-income (DTI):	40.5	9.8	10	75
Less than 28	0.072	0.259	0	1
28 – 35	0.133	0.339	0	1
36 – 43	0.193	0.394	0	1
44 of higher	0.255	0.436	0	1
Missing	0.347	0.476	0	1
Loan purpose:				
Cash out refinance	0.139	0.346	0	1
Non-cash out refinance	0.170	0.376	0	1
Unknown refinance	0.050	0.218	0	1
Purchase	0.641	0.480	0	1
Other loan-specific:				
Not full documentation	0.405	0.491	0	1
Adjustable rate	0.015	0.122	0	1
Not 30-year term	0.087	0.282	0	1
Origination balance ($10k)	17.8	9.1	1.5	77.9
State-specific				
Judicial foreclosure	0.371	0.483	0	1
Recourse	0.761	0.427	0	1
Borrower Experience	1.2	0.404	1	5

Table 4. Summary Statistics (continued)

	b) Dynamic Variables			
Variable	Mean	Std Dev	Minimum	Maximum
Loan-to-value (LTV):				
Less than 80	0.056	0.229	0	1
80 – 84	0.038	0.192	0	1
85 – 89	0.073	0.260	0	1
90 – 94	0.145	0.352	0	1
95 – 99	0.211	0.408	0	1
100 – 104	0.154	0.361	0	1
105 – 109	0.090	0.286	0	1
110 – 114	0.046	0.210	0	1
115 – 119	0.025	0.156	0	1
120 or higher	0.044	0.206	0	1
Missing	0.117	0.321	0	1
Economic determinants:				
Lag unemployment rate change	7.52	4.07	0.0	32.6
House price change, 12 month (10 percent)	−0.48	0.63	−3.9	2.4
Distress sales share (1 percent)	5.47	5.79	0	56.3
Interest rate differential (100 bp)	0.73	0.63	0	4.7
Duration at risk (months)	19	16	2	276

Notes: Summary statistics for the dynamic variables vary slightly between the prepayment and default as well as unlinked and linked data files. We report the statistics from the linked default data.

Table 5. Borrower Based Prepayment and Default Hazard Estimates:

Variable	Default	Prepayment
Loan-to-Value:		
80 – 84	1.29***	0.96**
	(0.064)	(0.061)
85 – 89	1.28***	0.65***
	(0.055)	(0.038)
90 – 94	1.53***	0.68***
	(0.058)	(0.034)
95 – 99	1.79***	0.64***
	(0.065)	(0.031)
100 – 104	2.22***	0.64***
	(0.081)	(0.032)
105 – 109	2.57***	0.60***
	(0.097)	(0.032)
110 – 114	2.85***	0.58***
	(0.115)	(0.036)
115 – 119	3.13***	0.53***
	(0.139)	(0.041)
120 or higher	3.72***	0.50***
	(0.152)	(0.035)
Missing	1.73***	0.57***
	(0.073)	(0.032)
Credit Score (FICO):		
Less than 580	10.82***	0.43***
	(0.319)	(0.023)
580 – 619	7.52***	0.59***
	(0.214)	(0.022)
620 – 679	3.91***	0.70***
	(0.106)	(0.019)
680 – 719	2.01***	0.84***
	(0.064)	(0.027)
Missing	5.28***	0.57***
	(0.180)	(0.029)
Back end Debt-to-Income (DTI):		
28 – 35	1.17***	1.05
	(0.039)	(0.054)
36 – 43	1.41***	1.10*
	(0.044)	(0.054)
44 or higher	1.59***	1.10**
	(0.049)	(0.052)
Missing	1.65***	0.97
	(0.052)	(0.047)

Table 5. Borrower Based Prepayment and Default Hazard Estimates (continued)

Variable	Default	Prepayment
Loan purpose:		
Cash-out refinance	1.28***	1.47***
	(0.024)	(0.045)
Non-cash-out refinance	1.22***	1.15***
	(0.021)	(0.034)
Unknown refinance	1.15***	1.31***
	(0.029)	(0.060)
Other loan-specific:		
Not full documentation	0.77***	1.80***
	(0.013)	(0.048)
Adjustable rate	1.06	3.16***
	(0.049)	(0.21)
Not 30-year term	1.98***	0.98
	(0.039)	(0.045)
Origination balance ($10k)	1.02***	1.04***
	(0.0007)	(0.001)
State-specific:		
Judicial foreclosure	1.09***	0.81***
	(0.015)	(0.019)
Recourse	1.11***	1.13***
	(0.020)	(0.033)
Economic determinants:		
Lag unemployment rate	1.04***	0.99**
	(0.003)	(0.005)
House price change, 12 month (10 percent)	0.94***	0.92***
	(0.011)	(0.019)
Distress sales share (1 percent)	1.00	0.99***
	(0.001)	(0.002)
Interest rate differential (1 percent)		3.09***
		(0.055)
Number of months at risk	4,171,301	4,097,331
Number of subjects	160,939	159,169

Note: The left out group is the set of high-quality FHA fully documented 30-year fixed rate purchase mortgages with current LTV below 80, FICO score above 720, DTI below 28, secured by a single family residence. Six property type indicators are included. Borrower experiences are left-censored at January 2007.

*** significant at the 1% level ** significant at the 5% level * significant at the 10% level

Table 6. Vintage projections at 10 years from origination

	Still Active	Default	Prepay	Maximum Sustainability Rate
Combined: 2007-2009	39.0	31.0	29.2	74.8
2007	17.8	51.8	30.4	58.9
2008	23.5	40.5	36.0	67.9
2009	50.4	24.4	25.3	80.7

Note: Vintages represent calendar years. Maximum sustainability rate = Still Active + Prepay + Default*cure rate.

Table 7. Borrower vs. Mortgage: Five-Year Projections for Active Loans

	Still Active	Default	Prepay
Borrower Experience	60.6	14.6	24.6
Mortgage			
No interactions	10.9	9.0	80.1
Streamline refinance interactions	10.5	8.7	80.8

Note: Streamline refinance interactions allow all mortgages that begin as a streamline refinance to have separate default and prepayment hazards from mortgages that begin otherwise.

Figure 1. FHA Insurance Fund Projections & Revisions, ($M)

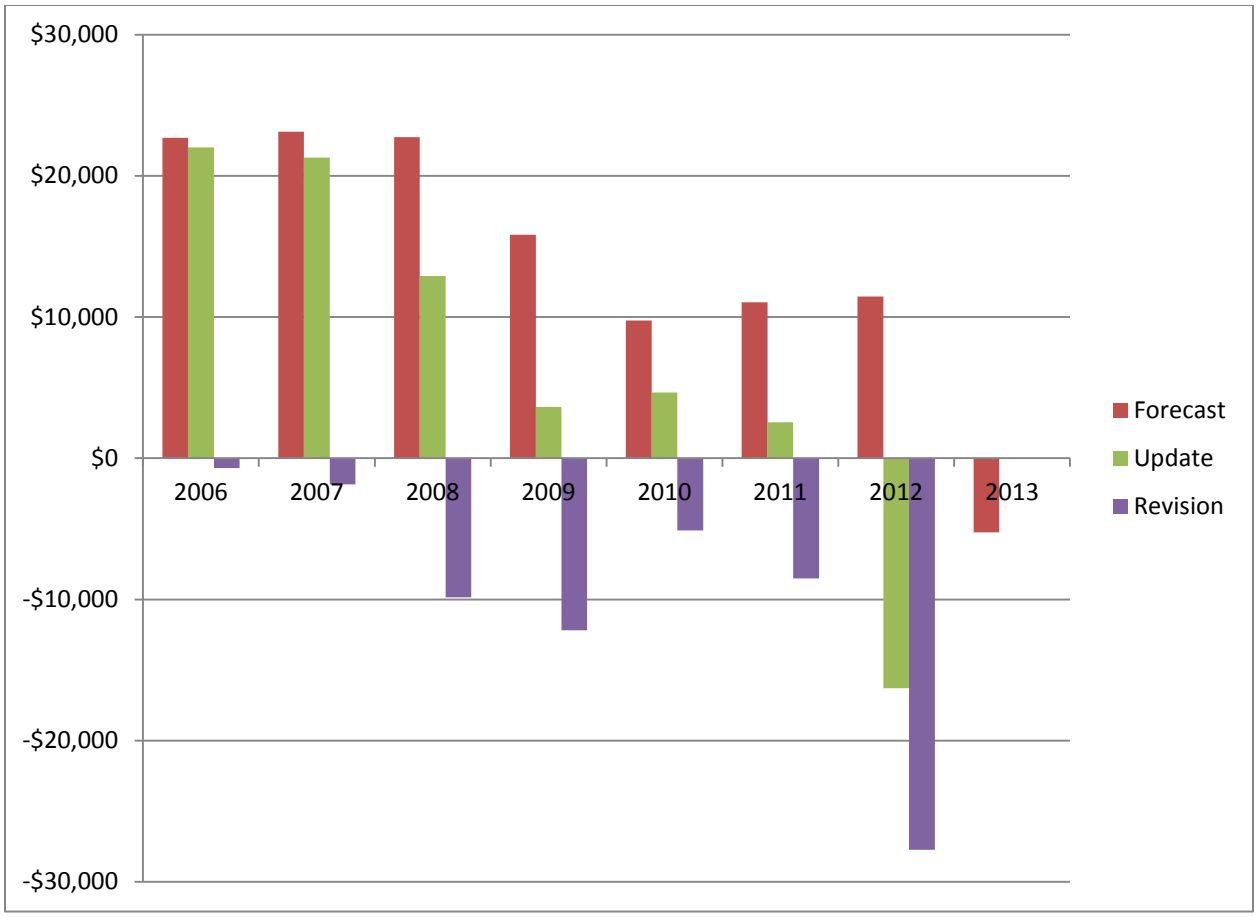


Figure 2. Baseline Default Hazards: Mortgage vs. Borrower

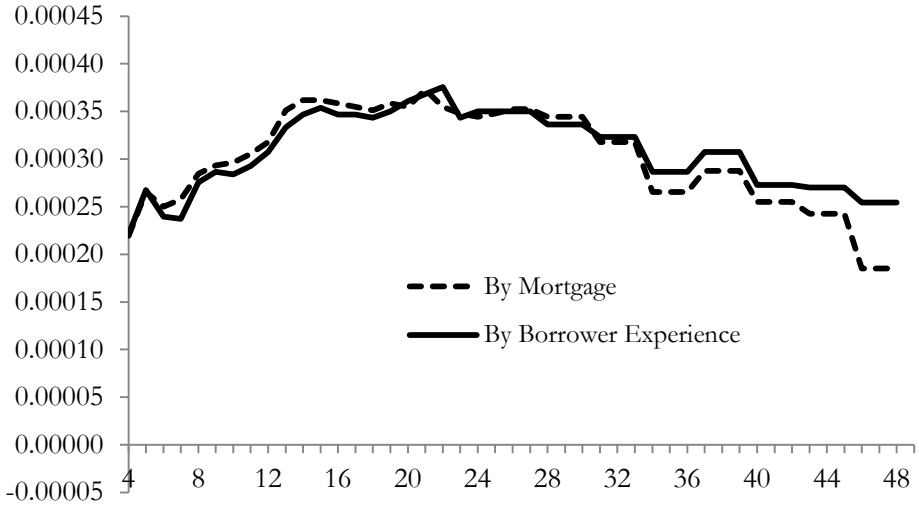


Figure 3. Baseline Prepayment Hazards: Mortgage vs. Borrower

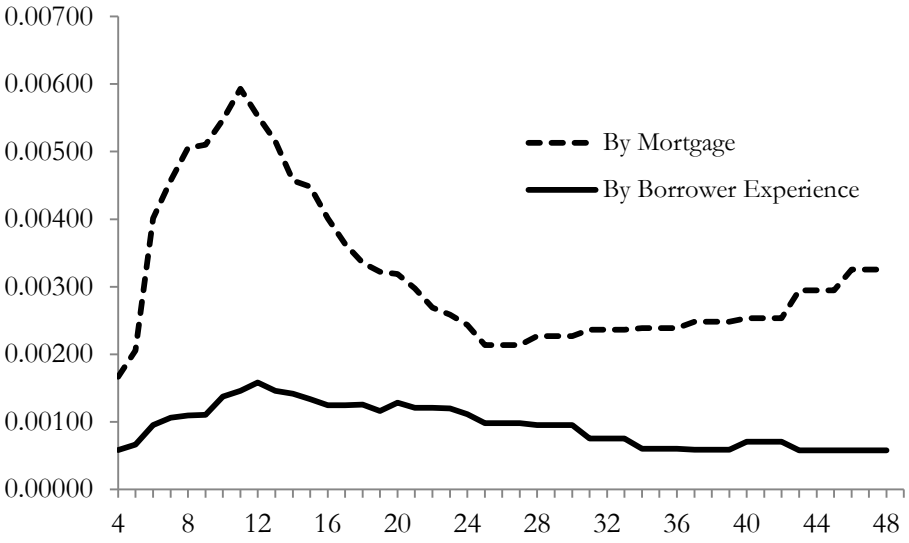
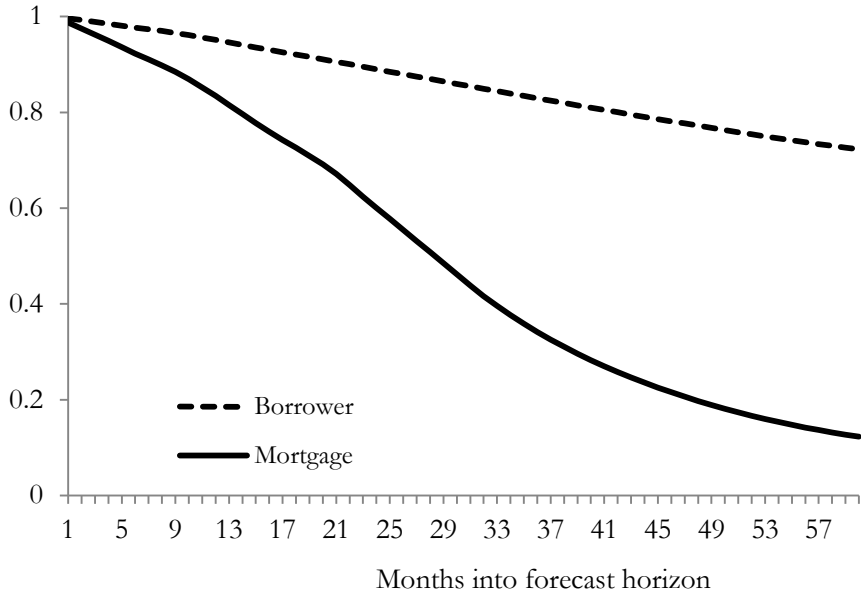


Figure 4. Survivor probabilities over the forecast horizon – borrower and mortgage



Note: The default hazards have been set to zero in constructing these survivor functions.

Table A1. Comparison between FHA Originations and CoreLogic Sample: Credit Score (FICO)

Fiscal	FY	Purchase		Conventional Refi		FHA-FHA Refi[a]		All[a]	
Year	Quarter	FHA	*CL*	FHA	CL	FHA	CL	FHA	CL
2007	1	633	639	619	620	557	625	607	634
	2	631	635	616	620	587	628	624	631
	3	628	632	613	618	613	628	623	628
	4	632	634	621	615	613	625	628	628
2008	1	631	633	620	615	612	626	625	626
	2	642	635	630	620	625	633	635	628
	3	667	655	644	637	622	643	656	648
	4	673	669	646	645	628	647	663	662
2009	1	676	673	655	652	648	649	668	666
	2	683	678	674	669	667	663	679	674
	3	696	688	689	685	676	676	691	687
	4	698	697	688	688	675	678	693	694
2010	1	697	697	692	690	681	680	694	695
	2	698	697	699	696	686	686	697	696
	3	698	698	695	699	693	689	698	698

Notes: FHA statistics are from the FHA Quarterly Reports. Fiscal years run from October of the prior year through September of the indicated year.

[a] Based on fully underwritten mortgages and exclude streamline refinances

Table A2. Comparison between FHA Originations and CoreLogic Sample: Loan-to-value (LTV)

Fiscal	FY	Purchase		Conventional Refi		FHA-FHA Refi		All	
Year	Quarter	FHA	CL	FHA	CL	FHA	CL	FHA	CL
2007	1	95.7	95.7	89.4	89.7	89.4	90.3	91.7	92.0
	2	95.6	95.4	89.2	89.7	90.0	90.6	93.4	93.4
	3	95.3	95.3	88.7	89.2	88.9	89.7	93.2	93.3
	4	95.7	95.5	89.1	89.6	89.4	90.1	93.6	93.6
2008	1	94.9	94.8	89.3	89.6	90.2	91.2	92.0	92.2
	2	95.0	94.7	89.8	90.4	90.8	91.4	92.1	92.4
	3	95.1	95.1	89.7	90.2	90.8	91.4	92.9	93.1
	4	95.2	95.2	89.4	89.8	89.8	90.5	93.1	93.3
2009	1	94.7	94.7	89.9	90.3	91.6	92.0	92.9	93.1
	2	95.1	95.0	90.9	91.2	93.0	93.7	93.1	93.2
	3	95.1	94.9	92.2	92.5	93.3	94.0	94.0	94.1
	4	95.2	94.9	90.6	90.9	91.5	92.3	93.7	93.7
2010	1	95.1	94.8	89.2	89.6	92.9	93.9	92.7	92.7
	2	95.0	94.8	89.5	89.9	90.4	91.3	93.0	93.1
	3	94.9	94.6	88.6	89.3	89.8	91.1	94.1	94.0

Notes: FHA statistics are from the FHA Quarterly Reports. Fiscal years run from October of the prior year through September of the indicated year. The FHA LTV averages exclude any financed up-front mortgage insurance premium from the loan balance. The up-front premium charged from FY09 through March 2010 was 1.75% for fully underwritten mortgages, and 1.5% for streamline refinanced mortgages. Starting in April 2010 the up-front premium was increased to 2.25% for all mortgages. Prior to FY09 the premiums varied. We adjust the CoreLogic LTV by backing out these up-front mortgage premiums.

Table A3. Comparison between FHA and CoreLogic Early Delinquency

		Purchase		Refinance		All	
Year	Quarter	FHA	CL	FHA	CL	FHA	CL
2007	1	0.6	0.4	0.8	1.3	0.7	2.2
	2	3.6	2.8	2.1	1.9	3.1	2.5
	3	3.4	2.6	3.1	2.0	3.3	2.4
	4	2.6	2.5	2.3	1.8	2.5	2.2
2008	1	2.1	2.3	3.3	1.7	2.8	2.2
	2	2.7	1.8	3.0	2.0	2.9	2.1
	3	2.3	1.5	4.0	2.1	3.1	1.8
	4	1.6	1.1	3.4	1.6	2.3	1.4
2009	1	1.2	0.9	2.2	0.9	1.6	1.3
	2	0.9	0.6	2.2	0.6	1.6	1.0
	3	0.7	0.4	2.0	0.6	1.5	0.7
	4	0.5	0.3	2.2	0.7	1.5	0.5
2010	1	0.4	0.4	0.9	0.3	0.7	0.4
	2	0.7	0.3	0.8	0.3	0.7	0.3
	3	0.7	0.5	0.7	0.3	0.7	0.5

Notes: FHA statistics are from the FHA Quarterly Reports. Fiscal years run from October of the prior year through September of the indicated year. Refinances include all fully underwriter conventional to FHA and FHA to FHA refinances. Early delinquency is defined to mean a mortgage that reaches 90-days delinquent within the first 6-months since origination.

Table A4. Mortgage Based Prepayment and Default Hazard Estimates

Variable	Default	Prepayment
Loan-to-Value:		
80 – 84	1.23***	0.93***
	(0.061)	(0.031)
85 – 89	1.21***	0.78***
	(0.053)	(0.022)
90 – 94	1.44***	0.75***
	(0.055)	(0.019)
95 – 99	1.72***	0.80***
	(0.063)	(0.019)
100 – 104	2.10***	0.86***
	(0.077)	(0.021)
105 – 109	2.44***	0.88***
	(0.093)	(0.023)
110 – 114	2.81***	0.88***
	(0.117)	(0.026)
115 – 119	3.11***	0.86***
	(0.144)	(0.030)
120 or higher	3.72***	0.78***
	(0.159)	(0.025)
Missing	1.67***	0.47***
	(0.064)	(0.013)
Credit Score (FICO):		
Less than 580	10.71***	0.85***
	(0.315)	(0.019)
580 – 619	7.41***	0.96**
	(0.211)	(0.017)
620 – 679	3.93***	0.97**
	(0.106)	(0.014)
680 – 719	2.02***	0.97
	(0.064)	(0.016)
Missing	4.71***	0.74***
	(0.162)	(0.018)
Back end Debt-to-Income (DTI):		
28 – 35	1.16***	1.18***
	(0.039)	(0.030)
36 – 43	1.39***	1.31***
	(0.044)	(0.031)
44 or higher	1.56***	1.32***
	(0.048)	(0.030)
Missing	1.68***	0.93**
	(0.052)	(0.022)

Table A4. Mortgage Based Prepayment and Default Hazard Estimates (continued)

Variable	Default	Prepayment
Loan purpose:		
Cash-out refinance	1.30***	1.38***
	(0.024)	(0.019)
Non-cash-out refinance	1.44***	0.78***
	(0.023)	(0.012)
Unknown refinance	1.28***	1.11***
	(0.031)	(0.025)
Other loan-specific:		
Not full documentation	0.84***	2.22***
	(0.014)	(0.027)
Adjustable rate	1.11**	4.42***
	(0.051)	(0.12)
Not 30-year term	1.97***	0.70***
	(0.038)	(0.018)
Origination balance ($10k)	1.02***	1.03***
	(0.0007)	(0.0005)
State-specific:		
Judicial foreclosure	1.08***	0.88***
	(0.015)	(0.010)
Recourse	1.10***	0.94***
	(0.020)	(0.013)
Economic determinants:		
Lag unemployment rate	1.03***	0.93***
	(0.003)	(0.002)
House price change, 12 month (10 percent)	0.95***	0.76***
	(0.011)	(0.007)
Distress sales share (1 percent)	1.00	1.00***
	(0.001)	(0.001)
Interest rate differential (1 percent)		2.70***
		(0.023)
Number of months at risk	4,052,279	4,014,051
Number of subjects	191,393	188,789

Note: The left out group is the set of high-quality FHA fully documented 30-year fixed rate purchase mortgages with current LTV below 80, FICO score above 720, DTI below 28, secured by a single family residence. Six property type indicators are included. Mortgages are right censored at January 2007

*** significant at the 1% level ** significant at the 5% level