Credit Risk of Low Income Mortgages

Hamilton Fout, Grace Li, and Mark Palim

Economic and Strategic Research, Fannie Mae

3900 Wisconsin Avenue NW, Washington DC 20016

May 2017

Abstract

Using Fannie Mae loan-level data on fixed-rate owner occupied purchase mortgage acquisitions, we examine the role of tightened underwriting standards on the default risk of low and moderate income (LMI) homebuyers. In three distinct underwriting regimes and subsequent housing market environments – 2002-2004, 2005-2007 and 2011-2013 – we find that loan performance improves as a borrower's income relative to area median income rises – for both actual performance and the marginal predicted performance after controlling for standard credit risk measures, vintage and region. Second, for pre-crisis loans applying the tighter underwriting standards of the post-crisis period dramatically reduces the performance differences across relative income, indicating the importance of underwriting standards for sustainable low and moderate income lending and home ownership. Finally, for all but very-low income borrowers (<= 50% area median income), credit risk is well accounted for by the usual risk factors considered in the underwriting process along with vintage and regional controls.

JEL Codes: G21, R31, R38

Key Words: Low Income Lending, Credit Risk, Underwriting Standards

I. INTRODUCTION

There has been longstanding interest in understanding the risks associated with extending mortgage credit to low and moderate income (LMI) households, defined for the purposes of this study as households with income less than or equal to area median income. Given the changes in underwriting and economic environment over the last 15 years, Fannie Mae's loan-level data contains useful information to assist lenders and policy makers in understanding the relative importance of borrower attributes, underwriting, and the economic cycle on loan performance. Over this time period, household income has grown less than home prices and asset prices (hourly wages have grown at 2.7% annually, while home prices and the S&P 500 have grown at 3.2% and 3.6%, respectively, over the period 2001 – 2015), potentially constraining the ability of low and moderate income buyers to find affordable housing and to come up with sufficient down payments.[1]

We are interested in exploring the role of underwriting standards in sustainable homeownership for low and moderate income borrowers and the extent to which the credit risk associated with extending mortgage credit to low and moderate income borrowers is predictable in different credit and economic environments. Therefore, in this research, we seek to answer the following three questions: (1) does credit risk increase as relative income of borrowers declines; (2) to what extent have tighter eligibility standards reduced the risk of extending mortgage credit to low and moderate income homebuyers relative to the pre-crisis period; and (3) to what extent can standard underwriting factors (e.g. FICO and LTV) account for the credit risk on conforming purchase loans without directly controlling for relative income? [2] We implement several specifications to

---

[1] The wage measure here is the BLS non-farm non-supervisory hourly wage, and the home price measure is the FHFA Purchase-Only HPI.
[2] It is worth noting that, the primary interest of our paper is in the role of underwriting in mitigating the additional risk associated with low and moderate income loans and the extent any such risk is

determine the marginal risk of low and moderate income lending – the additional risk of default after controlling for standard risk factors.[3] We also examine whether the pattern of low and moderate income lending risk changes before, during and after the housing crisis period. Dramatic changes in the underwriting and lending environment following the crisis period could potentially alter the linkage of various risk factors, including relative income, to default behavior. Finally, we estimate the importance of underwriting by comparing the performance of pre-crisis loans that would and would not meet post-crisis underwriting standards.

We find that low and moderate income lending is associated with higher DTIs and lower FICOs across all analyzed periods (before, during and after the housing crisis). Furthermore, DTI, LTV and FICO values vary more for low and moderate income loans relative to their higher income counterparts. Consistent with that credit profile, we find that the actual default rate of low and moderate income loans (before any additional controls) are higher as relative income declines in each period. The rank ordering of relative income stays the same across the periods even as we see meaningful differences in overall default rates and in rates by low and moderate income category. In the period with the lowest overall default rates, the post-crisis period, the difference in default rates by relative income are the smallest. We further find that the default rates for low and moderate income lending in the two historical periods (pre-crisis and crisis) dramatically decline once we restrict loans to those eligible for delivery to Fannie Mae today, indicating the importance of underwriting standards for sustainable low and moderate income

---

stable across different economic and underwriting environments and, therefore, can be effectively modeled and predicted. Thus, we are interested in the information available at the time of underwriting in predicting early default behavior.

[3] We refer to the marginal effect as the low-income credit risk not accounted for by standard risk factors and fixed effects, which may or may not be correlated with relative income measures. We refer to the average effect as the estimated effect of low and moderate income measures on loan performance with no additional controls.

lending. Finally, we find that standard underwriting factors along with region and vintage controls are sufficient for measuring the risk associated with most low and moderate income categories. Very low income loans, however, present additional risk not fully captured by these factors.

The rest of the paper is organized as follows: Section 2 provides an overview of the relevant economic research in this area, Section 3 explains our data and methodology, Section 4 presents our empirical results, and Section 5 concludes.

II.   BACKGROUND

There are two strands of economic research that are most relevant to the current paper. The first deals with the design of credit risk models and their usage as a tool for gauging the risks of mortgages at the time of underwriting. The second deals with the credit and prepay performance of low and moderate income mortgages.

Credit Risk Models

There is a rich literature regarding credit risk management models and how they are used in the underwriting process of a mortgage loan (for instance Quercia, 1992 and Avery et al., 1996). Information to assess credit risk is collected, verified, and evaluated in the underwriting process of the loan. During this process, financial institutions carefully assess credit risk using information on a range of risk factors that are thought to affect or predict repayment behavior. These factors include the current and past payment behavior of the borrower (e.g. FICO scores), the type and purpose of the loan, loan characteristics including the loan-to-value (LTV) and debt-to-income (DTI) ratios and the characteristics and value of the property serving as collateral for the loan (Avery et al., 1996, Haughwout et al., 2008 and Mayer et al., 2009). Credit risk models are used to quantify the expected

future performance of mortgage and other loans based on the information available at origination.

A typical approach in modeling mortgage credit risk involves estimating a logit model to explain some binary outcome of loan performances (e.g. 90 or more days delinquent within two years since origination based on the data observed at underwriting). The general predictive accuracy of the estimated model can be evaluated by the Gini coefficient, which measures the rank-ordering power of the logit model to separate those loans that went delinquent versus those that did not (Mays, 2001 and Crook et al., 2007). Higher Gini coefficients for a credit risk model indicate better model fit, and adding significantly predictive variables, reducing measurement errors in the existing variables, and accounting for non-linear relationships between risk factors and default outcomes can contribute to higher Gini coefficients.

There are a number of attributes entering the logit model for prediction of default, such as FICO, LTV, and DTI. Borrower income, other than its use as an input into the calculation of DTI, is not typically considered as a direct input in risk models, either as a level or a ratio relative to area median income (AMI).[4] This omission is potentially due to the possibility of disparate impacts on protected classes from including direct income controls in the underwriting process.[5] One of the cases where income is used in underwriting is with loans guaranteed by the US Department of Veterans Affairs (VA), which uses a "residual income" (net after fixed obligations) measure as one of the underwriting factors

---

[4] For instance, Fannie Mae's selling guide lists the following risk factors evaluated as part of the automated underwriting process: credit history, delinquent accounts, installment loans, revolving credit utilization, public records, foreclosures, collection accounts, inquiries, borrower's equity and LTV, liquid reserves, loan purpose, loan term, loan amortization type, occupancy type, DTI, property type, co-borrowers and self-employment (See https://www.fanniemae.com/content/guide/sel053116.pdf, p. 316).
[5] See https://www.federalreserve.gov/boarddocs/caletters/2009/0906/09-06_attachment.pdf (p. iv) for a description of disparate impact.

(Goodman, 2015).[6] However, some of the risk factors typically captured during the underwriting process and used to predict loan performance may be correlated with borrower income. For example, higher-income borrowers are more often able to make a larger down payment. Therefore, the LTV on loans to low and moderate income borrowers tend to be higher, and higher LTV loans have higher default risk (Kelly, 2008, An et al., 2012 and Lam et al., 2013).[7] Additionally, lenders typically include DTI when assessing credit risks. So for any nominal level of debt, lower income borrowers would have a higher DTI, which is associated with greater credit risk (Avery et al., 1996 and Haughwout et al., 2008).

Therefore, standard risk management models, although they do not control directly for a borrower's income relative to the area median, still account for part of the additional credit risk of low-income borrowers because of the correlation of relative income with other risk characteristics typically used in underwriting. In this paper, among other things, we investigate whether directly controlling for relative income in addition to the usual credit risk factors improves our understanding of past mortgage performance and the extent to which including relative income increases explanatory power in different underwriting and economic environments.

Loan Performance of Low and Moderate Income Mortgages

In addition to the literature on credit risk models, there is a separate strand of economic literature that focuses on the issue of loan performance among the low and moderate

---

[6] This residual income measure can be thought of as similar to DTI, in that it looks at income relative to debt and other obligations.

[7] Lam et al. presented both empirical and simulation results to show that the lifetime default and foreclosure rates increase monotonically with original LTV.

income population of mortgages.[8] Archer et al. (1996) represent one of the earliest works in this literature, finding that low and moderate income borrowers are less likely to sell their property and move when an income or life event shock hits. Van Order and Zorn (2003) finds that default responses to negative equity are similar across higher income and low and moderate income neighborhoods and the differences in defaults are small and can be explained by omitted variables such as those measuring credit history. Deng et al. (1996, 2000), Deng and Gabriel (2006) and Firestone et al. (2007) all find slower voluntary prepayment speed among low income borrowers. Although these four works are part of the broader (call) option pricing literature with a focus on prepayment behavior, they do provide some insight into the credit risk of low and moderate income borrowers.

In particular, Deng et al. (1996) investigate a set of loans purchased by Freddie Mac between 1976 and 1983 and create a loan-level set of low and moderate income indicators (<=60% of area median income, >60 and <=100% of area median income, >100 and <=150% of area median income, >150% of area median income). They present evidence within a competing hazards framework that default risks decline as household income rises and also that low and moderate income households are more likely to default when faced with negative equity than higher income households. Deng and Gabriel (2006) also use a proportional hazard model to quantify the prepayment and default risks among Federal Housing Agency (FHA) mortgage loans originated between 1992 and 1996, controlling directly for household income. They find a significant negative effect on default probabilities as income rises, after controlling for a number of borrower and market-level measures. Firestone et al. (2007) analyze Freddie Mac deliveries from 1993 to 1997 and

---

[8] Past research on low-income lending has also focused on the equity-building potential of low-income home ownership as well as the relationship between borrowing constraints and homeownership. For example, Painter et al. (2001) examine the determinants of housing tenure choices by racial and ethnic groups. Duca and Rosenthal (1994) and Barakova et al. (2014) analyze the effect of borrowing constraints on homeownership.

using a set of low and moderate income indicators find that default probability increases as income relative to area median income declines.[9] Quercia et al. (2002) focus on the performance of a small number (N=1,017) of CRA loans (loans made by banks to satisfy the Community Reinvestment Act requirement that banks serve the local communities where they obtain deposits) originated in 1998 using a variety of factors including income relative to area median income. They find an insignificant effect of income relative to area median on early delinquencies for the population they investigate.

Our research extends the literature in a number of ways. First, while the literature described above has focused on the historical period before the recent housing crisis of 2007, we take advantage of a rich dataset of Fannie Mae acquisitions originated between 2002 and 2013 with loan level household income and area income data to investigate relative low and moderate income performance under a variety of underwriting regimes and subsequent housing market environments. Second, we are interested in quantifying the additional default risk (as opposed to prepay behavior) associated with low and moderate income lending using a set of indicators that allow us to separately measure the relative risk of very low income (<=50% of area median income), low income (>50% and <=80% of area median income) and moderate income (>80% and <=100% of area median income) borrowers compared to higher income (>100% of area median income) borrowers after controlling for a variety of loan-level attributes as well as region and vintage fixed effects. Finally, we focus on the role of underwriting in low and moderate income lending along the following two dimensions: (1) the use of tightened underwriting standards in effectively mitigating the additional risks associated with low and moderate income lending and (2) the ability of standard underwriting factors such as FICO, LTV and DTI

---

[9] In an earlier version of this paper Van Order and Zorn (2002) present further evidence of increased default risks among the low and moderate income population using a similar set of indicators.

along with region and vintage controls to sufficiently explain the credit risks presented by low and moderate income loans without the need to explicitly control for relative income.

III. DATA AND METHODOLOGY

Data

The data set we analyze in this paper is the population of fixed-rate single family 1-4 unit[10] conventional owner occupied purchase loans acquired by Fannie Mae during the period from July 2002 to July 2013.[11] The data set contains details on each loan, including loan purpose (purchase or refinance), value of the property at origination, number of borrowers on the mortgage, borrower's and co-borrower's FICO scores at origination, loan balance(s) at origination, interest rate at origination, term of the loan, monthly payment amount, income documentation associated with the loan application, address of the property, number of units, borrower's income relative to area median, whether the loan is negatively amortizing, LTV, DTI and whether the loan is originated through a third party.
[12] The dataset also contains data on loan performance for each loan, including whether and when the loan was first 30, 60 or 90 days past due and whether/when the loan prepaid.

---

[10] We focus on single family $1 - 4$ unit properties and exclude condos and manufactured housing from the sample. These two property types may potentially be subject to increased unobservable regional risk exposures relative to other property types. We also control for the number of units in our modeling approach.

[11] Our focus in this paper is on low and moderate income loan performance within the context of conventional conforming lending (i.e. non-government mortgages with balances that conform to the conventional loan limits), and the reliance on Fannie data should not present a significant limitation in generalizing to the broader conventional conforming market.

[12] For the purposes of this research, LTV refers to the combined loan-to-value ratio which is the unpaid principle of all first and subordinate mortgages divided by the value of the property at origination.

We use explanatory variables from this set of data to model default, which we define for the purposes of this paper as the case when a loan goes 90 days or more delinquent within the first 24 months from the first payment date.[13] We use this definition of default outcome for two primary reasons. First, this is a standard approach to modeling credit risk which we believe simply and sufficiently captures the additional risks of low and moderate income lending.[14] Second, the primary interest of our paper is in the role of underwriting in mitigating the additional risk associated with low and moderate income loans and the extent any such risk is stable and predictable across different underwriting regimes and economic environments. Thus, we are interested in the information available at the time of underwriting in predicting early delinquencies as opposed to later delinquency outcomes that are just as likely to be driven by risk factors that accrue over time and are unavailable at origination (e.g. changes in FICO scores, borrower employment situation or the actual path of home prices).

Given the interest of public policy in sustainable homeownership, we focus on mortgages for primary owner-occupied residences in our empirical results. We exclude refinance loans and government loans from the sample. Thus our data is exclusively conventional, conforming owner-occupied fixed-rate purchase loans. Additionally, we exclude all long-term standby commitments and seasoned loans (first payment date at least one year prior to being acquired by Fannie Mae). In part of the analysis, in order to measure the impact of post-crisis tightening of Fannie Mae's eligibility criteria, we evaluate the performance of a subset of pre-crisis period loans that qualify under current eligibility standards.

---

[13] With this definition of default there is a potential that a portion of the sample may represent mortgage fraud and compromise the ability to model actual mortgage credit performance. In a separate robustness check we drop the loans in our sample that never made a payment (~2% of all defaults) and repeat the modeling analysis. We find in this case that there is no systematic relationship between low and moderate income and potentially fraudulent loans, and our results for both the average and marginal risks of low and moderate income loans stay the same.

[14] See for instance Haughwout et al. (2008).

Specifically, this drops loans that have LTV higher than 97 percent, FICO score less than 620, DTI ratio higher than 50 percent, Alt-A loans, interest-only loans, negative and balloon amortizations and loans with low or no documentation for income.

The loans we study span originations over the period 2002 – 2013 and thus cover a range of underwriting and economic environments. In particular, housing prices appreciated over the period 2002 – 2007, followed by a sharp decline of home prices during the second half of 2008, with prices continuing to decline through 2011. Home prices began their recovery in 2012, with the nominal national home price index surpassing its pre-crisis peak level by late 2015 (see for instance the FHFA Purchase-Only Index). Over this period, the unemployment rate declined to 4.5 percent in 2007 followed by a sharp increase to approximately 10 percent in 2010 and a subsequent decline to 5 percent by December 2015. Underwriting standards for conventional single family mortgages also differed significantly in each of these regimes. We examine two commonly cited measures of mortgage credit availability to understand the shift of regimes in the underwriting standards for mortgage contracts. The first, the Federal Reserve's Senior Loan Officer Opinion Survey, suggests that lending standards were loosening through the first half of 2007. Following this period, underwriting standards dramatically tightened, according to the survey. The second measure we use is the median borrower's credit score (based on CoreLogic servicer data), which has greatly increased since the start of 2007 and stayed elevated through the current period (Li, 2015), suggesting tighter underwriting in the recent period.

Based on differences in the macroeconomic and underwriting environment, we define the pre-crisis sample as owner occupied fixed-rate purchase money mortgages that have a first payment month from July 2002 to July 2004 (representing the regime of neutral underwriting followed by housing price increase), the crisis sample as mortgages that have

a first payment month from July 2005 to July 2007 (representing the regime of relaxed underwriting followed by housing price decline), and the post-crisis sample as mortgage loans that have a first payment month between July 2011 and July 2013 (representing the regime of tight underwriting followed by housing price increase).[15] Every loan in the sample is allowed 24 months after the first payment date as its performance window.

Variable definitions and summary statistics for all variables included in our analysis are listed in Table 1a. In our sample (owner occupied primary fixed-rate purchase loans), 2.0 percent of loans originated during 2002 to 2004 experienced an early default, 6.5 percent of loans originated between 2005 and 2007 experienced an early default, and 0.3 percent of loans originated between 2011 and 2013 experienced an early default. Comparing the mean of typical credit risk factors, we find that the post-crisis period acquisitions have higher FICO scores, lower LTVs, and lower DTIs compared with the earlier origination vintages. In the post-crisis period, more loans have a co-borrower present compared with the crisis period, more loans have no second liens attached, and more loans have fully documented income. This reflects the dramatic changes in the regulatory and lending environment during and after the financial crisis. In terms of the share of low and moderate income lending, in the pre-crisis period, 8 percent of loans were very low income, 22 percent were low income and 16 percent were moderate income.[16] For the crisis period, these values are similar at 8 percent, 22 percent and 15 percent, respectively. In

---

[15] Our interest in this paper is in isolating periods where there are meaningful differences in underwriting environments as well as in the subsequent economic environment. Inevitably, choosing starting and ending points for these periods will involve some degree of judgement. We have investigated the effects on our results of adjusting these time periods, for instance by extending the crisis period through 2008, and our major results were robust to these changes.

[16] One potential issue that needs to be considered is the mismeasurement of income, as this is the primary variable of interest in this paper. One source of inaccurate measurement of income, involves higher income borrowers who potentially only report enough income to qualify, which is not a significant concern for this paper, as we focus on the low and moderate income populations. Another segment with potentially significant mismeasurement of income are those mortgages with low or no documentation of income. As these loans no longer meet eligibility requirements for the GSEs, we remove these loans from much of our later analysis in the paper.

the post-crisis period, however, the share of low and moderate income loans is lower, with a 6 percent share of very low income, a 17 percent share of low income and a 12 percent share of moderate income borrowers. Our sample sizes are ~2.2 million loans in the pre-crisis period, ~2.0 million loans in the crisis period and ~1.1 million loans in the post-crisis period.

In Table 1b we provide a breakdown of the risk characteristics by relative income status of the borrowers for the 2011 – 2013 period. The early default rate is 0.7 percent for very low income borrowers for this period, higher than that of other relative income groups (0.4 percent for low income borrowers, 0.3 percent for moderate income borrowers and 0.2 percent for higher income borrowers). In general, credit risk factors improve with the relative income of borrowers in this period, for instance FICO rises with relative income and DTI and the share of single borrowers falls with relative income. LTVs, however, are the lowest for the lowest relative income group (77.7), rise for the next two income groups (80.8 and 81.7, respectively) and slightly decline for the highest income group (81.3). Table 2 provides a further breakdown of the three key credit risk characteristics of DTI, LTV and FICO by relative income group for each sample period. In general, the patterns shown in Table 1b are also true for the first two sample periods: DTI falls and FICO rises as relative income rises. The Pearson correlation coefficient estimates in Panel B confirm the negative correlation between relative income and DTI and the positive correlation with FICO. Furthermore, there is a significant negative relationship between relative income and LTV in each sample period. Note also that the very low income population has the lowest FICO score and highest DTI (although not the highest LTV) in each sample period. This income group generally has a higher standard deviation for these credit risk factors relative to other income groups (with the exception of DTI in the post-crisis period). This indicates greater dispersion of credit risk attributes across the lower income populations.

Figure 1 presents the share of very low income purchase loans in the post-crisis sample over all purchase loans by state. Because the low and moderate income status of a loan is determined by the relative income of borrower to the MSA median, the share of very low income loans varies significantly across states based on income distributions within each state, ranging from 3 percent (Hawaii) to 12 percent (Idaho) in this period, with an average of 6 percent. Low income loan distributions (not shown) have a similar pattern across states, ranging from 12 percent (Texas and Hawaii) to 24 percent (Wisconsin) in the post-crisis sample period, with an average of 17 percent.

Figure 2 highlights the changing composition of the low and moderate income loans over time. In particular the share of low and moderate income loans falling into the lower income category has increased in the most recent period (2011-2013) compared with the previous time periods (2002-2004 and 2006-2008). Furthermore, the share of loans less than approximately two-thirds of area median income has increased at each point of the distribution for the later period versus the two earlier periods and decreased for every point of the distribution greater than approximately two-thirds of area median income. This relative shift to lower income borrowers in the low and moderate income purchase segment has occurred as home prices have risen post-crisis at approximately 5.5 percent[17] per year while wages have grown only at approximately 2.5 percent[18] per year over the period 2011 to 2015, potentially undermining the ability of low and moderate income homebuyers to save for a down payment and find affordable housing.

[17] Source: https://www.fhfa.gov/DataTools/Downloads/pages/house-price-index.aspx
[18] Source: https://frbatlanta.org/chcs/wage-growth-tracker/?panel=1

Methodology

In addition to the usual risk factors considered in a traditional credit risk model, we introduce three low and moderate relative income categories: very low income, low income and moderate income loans. These are used as additional explanatory variables in the model and are defined as follows: very low income is defined as one if income is less than or equal to 50 percent of area median income and zero otherwise; low income is defined as one if income is greater than 50 percent but less than or equal to 80 percent of area median income and zero otherwise; moderate income is defined as income greater than 80 percent but less than or equal to 100 percent of area median income or zero otherwise.[19] In our empirical results the additional credit risk of these low and moderate income groups is compared to the omitted group of high income, defined as income greater than 100 percent of area median income.

We use logit models to estimate the relationship between various risk factors and the early default outcome at the loan level. We fit separate logit models for the sample periods 2002-2004, 2005-2007 and 2011-2013 to investigate the additional risks of low and moderate income loans in different underwriting regimes and subsequent housing market environments. For each sample period, we begin by linking the default outcome to the low and moderate income indicators and calculate an otherwise uncontrolled default rate across low and moderate income categories. If the contribution of each low income categorical indicator is positive and significant, then we can reach the conclusion that low and moderate income lending is on average more risky than loans to borrowers with an income above 100 percent of area median income.

---

[19] Our definitions of very low income, low income and moderate income are consistent with FHFA's current definitions of low and moderate income categories used in the measurement of 2015 – 2017 Enterprise Housing Goals (see http://www.gpo.gov/fdsys/pkg/FR-2015-09-03/pdf/2015-20880.pdf).

In order to examine whether low and moderate income lending provides additional risk beyond factors usually used to model credit risk, we add the low and moderate income indicators into a traditional credit risk model. In theory, the usual risk factors should have accounted for some, or perhaps all, of the additional risks associated with low and moderate income lending because of the correlation of these characteristics with relative income category (see Table 2). Therefore, the estimated effect from low and moderate income indicators should be smaller in this extended model. If the estimated marginal contributions from low and moderate income indicators are statistically significant, then this suggests that there is a positive, marginal contribution to risk from the relative income status of the borrower after controlling for other factors. In general the traditional logistic credit model we analyze takes on the form:

$$\Pr(90 \text{ days delinquent in } 24 \text{ months}_i)$$
$$= f(X_i\beta + \gamma_1 VLI_i + \gamma_2 LI_i + \gamma_3 MI_i + fixed\ effects_i) \quad (1)$$

Here, X_i refers to the vector of loan-level characteristics of loan i including, for instance, LTV, FICO and DTI; β is a vector of parameters; VLI_i, LI_i and MI_i are loan-level indicators for whether a loan is very low income, low income or moderate income, respectively; γ_1, γ_2 and γ_3 are our coefficients of interest, measuring the increase in risk for the respective low and moderate income categories; and $fixed\ effects_i$ represents the set of fixed effects we use to control for state and month of acquisition.

We rely on a final modeling approach to answer the question of the extent to which standard underwriting measures adequately account for differences in credit risks across income groups. Specifically, we start by regressing the default outcome on the usual risk characteristics of the loan without controlling for the low and moderate income indicators in the first step:

$$\Pr(90 \text{ days delinquent in 24 months}_i) = f(X_i \varphi + fixed\ effects_i) \quad (2)$$

The estimated coefficient vector is denoted as $\hat{\varphi}$. We then treat $X_i \hat{\varphi}$ as an offset in the second step regression and introduce the relative income indicators (including HI_i to indicate loans with relative income greater than 100% of area median income) as additional controls. The estimated odds ratio shows the marginal risk of low and moderate income loans after fully controlling for the standard risk characteristics and state and time fixed effects. [20]

$$\Pr(90 \text{ days delinquent in 24 months}_i) = f(X_i \hat{\varphi} + \theta_1 VLI_i + \theta_2 LI_i + \theta_3 MI_i + \theta_4 HI_i) \quad (3)$$

The general rank-ordering power of the estimated model for the loan sample in this study is evaluated by the Gini coefficient. Gini coefficients compare a given credit risk model's ranking of loans by probability of default with that of a random rank ordering. The higher the value of the Gini coefficient, the better the rank-ordering power of the model. We also include AIC in addition to the Gini coefficient as a standard measure of the goodness-of-fit for each of the models.

IV.   EMPIRICAL RESULTS

Our empirical results focus on the actual default rates and odds ratios of default rates of a given low and moderate income category relative to those of the high income category. [21] We begin with a comparison of raw default rates across low and moderate income

---

[20] For more technical details about this two-step approach, see
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statu g_logistic_sect010.htm.
[21] The odds ratio we report measures the ratio of the default rate for a given low and moderate income category over the default rate for the high income category. For instance, in the post-crisis period, we find an overall default rate of 0.7% for very low income loans and 0.2% for high income loans. The odds ratio in this case is calculated as 0.7/0.2 ≈ 4.

categories and time periods, before analyzing the marginal additional risks from low and moderate income lending across time periods after controlling for other observable risk factors. We are also interested in the extent to which tighter eligibility standards in the post-crisis period have helped limit the risks across low and moderate income categories. In particular, following the 2008 financial crisis, Fannie Mae tightened underwriting standards by no longer considering as eligible for purchase new loans with the following characteristics: LTV greater than 97 percent; FICO scores less than 620; DTI ratios greater than 50 percent; loans with features such as negative amortization and/or low documentation of income. We analyze the effectiveness of this tighter underwriting on sustainable home ownership for the low and moderate income borrowers in a subset of the results that follow by applying today's eligibility standards to historical loans from the 2002 – 2004 and 2005 – 2007 sample periods and tracking the default behavior of the restricted population across relative income groups.[22]

Average Low and Moderate Income Risks

In our first empirical specification, we use only low and moderate income indicators to explain defaults in order to capture the average risk across low and moderate income categories before controlling for any other credit risk factors. The results for this specification are shown in the top panel of Table 3. We list the estimated default odds relative to the high income population for all originations in a given sample period in the first column and the actual default rate for each relative income group in the second column. Consistent with past findings in the literature (see Firestone et al., 2007 and Avery and Brevoort, 2015), mortgage loans to borrowers whose income is below area median income experience higher rates of default, and the lower the borrower's income

---

[22] All conventional loans eligible for purchase by Fannie Mae during the post-crisis sample conform to this underwriting/eligibility criteria.

relative to area median income, the higher the default rate. Before considering changes in eligibility standards or risks explained by other observable factors, the absolute and relative risk of low and moderate income groups vary greatly over different time periods due to differences in underwriting regime and subsequent housing market experience.

Table 3 shows that for the 2002 – 2004 sample period, when underwriting was relatively loose and home prices were appreciating, the default rate ranges from 4.4 percent to 1.5 percent as relative income increases. For the period with more lax underwriting and a subsequent housing price decline (2005 – 2007), actual default rates range from 11.1 percent to 4.2 percent. Due to the sharp decline of home prices during this period, even the higher income group experienced a default rate almost triple that of the higher income group in the first time period. The most recent period is characterized by tighter underwriting relative to the other two periods followed by an improving macroeconomic and housing market environment. The default rate by area median income category ranges from 0.7 percent to 0.2 percent, lower for all income groups compared to the first two time periods.

Importance of Tighter Eligibility Standards in Limiting Low and Moderate Income Risk

As shown in the second and third panels of Table 3, loans ineligible for delivery to Fannie Mae under today's underwriting standards have a higher default rate than eligible loans for each relative income group before introducing any additional controls. For example, for the 2005 – 2007 time period, non-eligible very low income loans have a default rate of 16.8 percent, higher than the 2.9 percent for their counterparts in the eligible-loan category. The default odds ratio is 15.4 for non-eligible very low income borrowers relative to high income borrowers in the eligible loans category, while this odds ratio is 2.6 for the eligible very low income loans relative to the eligible high income loans in this period. For

the 2002-2004 period, the ineligible very low income to eligible high income default odds ratio is 12.3 versus 3.5 for the eligible very low income to eligible high income default odds ratio. We find a similar pattern for the low income, moderate income and high income categories, indicating the potential importance of tighter eligibility criteria in lowering default rates across all income groups for both appreciating and depreciating home price environments.

Importance of Standard Underwriting Factors in Explaining Low and Moderate Income Risk

Table 4 shows the comparison of the estimated default odds ratios for each income group after adding standard controls for risk characteristics including DTI, FICO, LTV, subordinate financing indicators, number of borrowers, third party originator indicators, term and number of units associated with each loan, as well as state and time period (acquisition month) fixed effects. One important takeaway is that after controlling for these standard risk characteristics, the relative risk of low and moderate income lending versus high income lending significantly declines. For loans eligible under current underwriting standards (bottom panel Figure 4) during the 2002 – 2004 sample period, the default odds ratio for very low income default is 1.7 relative to the same-period high income loan default rate after controlling for the usual risk characteristics, a large reduction from 3.5 which is the corresponding default odds without standard controls for the same period. For the 2005 – 2007 sample period, the odds ratio for this group of loans decreases from 2.6 to 1.4 after adding standard controls. The most recent period shows a reduction of the odds ratio for the very low income group from 4.0 to 1.8 after controlling for standard risk characteristics. Results are similar for loans not eligible for delivery under current underwriting standards (top panel of Figure 4). This indicates that standard

21

risk characteristics considered in the underwriting process explain much of the additional credit risk associated with low and moderate income lending.

Remaining Risks in Low and Moderate Income Lending After Controlling for Standard Underwriting Factors and Tighter Eligibility Standards

The analysis above suggests that there is additional average and marginal risk associated with the relative income of a borrower in explaining default behavior of owner-occupied purchase mortgages, even after controlling for standard credit risk factors and region and vintage fixed effects and assuming today's tighter eligibility standards. In this section, we employ a two-step approach to examine how much additional risk is present among low and moderate income borrowers after we allow for default risks to be explained by standard underwriting factors. In this approach, we first calculate the residual from the model including standard risk factors and state and vintage fixed effects, and then we estimate the contribution of the relative income categories in explaining this residual. Table 5 provides the results for this exercise for eligible loans only across the three time periods and reveals that low and moderate income loans have between 1.1 to 1.6 times unexplained default risk after accounting for underwriting, geographic and time controls relative to high income loans in the post-crisis period (similar for the crisis period with 1.0 to 1.4 times the unexplained default risk and for the pre-crisis period with 1.1 to 1.5 times the unexplained default risk). Furthermore, the additional risks for the moderate income borrowers are insignificant (at the 5% level) for all periods and those for the low income borrowers are insignificant (at the 1% level) for the post-crisis period. Note, however, that for the pre-crisis and crisis periods, low income borrowers do have statistically significant additional risk (at the 1% and 5% levels, respectively). This additional risk, however, is small in terms of percent of total risk for these low income loans (6% for the pre-crisis period and 4% for the crisis period) as explained further below. For very low income

borrowers, on the other hand, the remaining risks are significant (at the 1% level) ranging from 1.4 to 1.6 times those of high income borrowers across all three time periods.

We also provide Gini coefficients for the models in Table 5 to measure the additional value from including low and moderate income indicators in sorting out the credit risks at the time of underwriting.[23] The models with direct controls including low and moderate income indicators, standard underwriting factors and region and vintage controls have much higher Gini coefficients (0.640 to 0.732) than models with just low and moderate income indicators (0.112 to 0.234), indicating the relative importance of factors outside of low and moderate income indicators in explaining credit risk. Furthermore, the Gini coefficients do not change much when adding relative income controls (increases range from 0.001 to 0.002), indicating the limited additional gain in including low and moderate income indicators in ranking overall credit risks. Table 5 also provides a breakdown of the Gini coefficient in the model without income controls by relative area median income group. These results show that the model's rank-ordering power deteriorates as relative income declines for each of the time periods, likely driven by the relatively large share of loans in the higher income categories across time (see Table 1a).

Figure 3 presents additional results from the two-step modeling approach that reinforces the findings in Table 5. In particular, the right bar in Table 3 for each time period and relative income group shows the predicted average default rate using a traditional credit risk model that includes the standard underwriting factors and vintage and region controls. The left bar shows the actual average default rate for the corresponding group of loans. The difference of the two bars is the default risk that is unexplained by the model

---

[23] Gini coefficients are commonly used by the housing finance industry to compare the rank order of alternative credit risk models on the same populations. Crook et. al, 2007 surveys the most used credit scoring models and provides a detailed explanation of this metric.

and is listed above the bars in terms of levels and below each figure as a ratio of actual over predicted default rates. Note that for all sample periods the very low income group has a larger portion of credit risk unexplained by the model.[24] For the 2002 – 2004 period, the actual default rate for very low income loans is 23 percent greater than predicted, for the 2005 – 2007 period the corresponding value for very low income loans is 21 percent and for the post-crisis period, the value for very low income loans is 33 percent. Thus, despite substantial variation in default rates, the ratio is reasonably stable across different underwriting regimes and subsequent housing market environments, with the lowest percentage of unexplained risk coming in the crisis period and the highest percentage of unexplained risk in the recent low default post-crisis period.[25] For the remaining low and moderate income categories, this ratio is either insignificant or small as a percentage of additional unexplained risk relative to total risk. For instance, the additional unexplained risk for the low income borrowers ranges from 4 percent to 7 percent and is insignificantly different from zero in the post-crisis period. Coefficient estimates for the models used in constructing Table 5 and Figure 3 are shown in Tables A1 – A3 in the appendix.

V. CONCLUSION

This research uses Fannie Mae's conventional conforming purchase mortgage loan level acquisitions data to examine the following three questions: (1) does credit risk increase as

---

[24] This additional risk can potentially arise from the correlation of relative income with other risk factors not typically controlled for in the underwriting process, for instance first-time home buyer status or industry of employment. The former would imply part of the increased default with low income borrowers may relate to being relatively less experienced with mortgage debt. The latter could potentially result in low and moderate income borrowers being more vulnerable to shocks that result in a decline in income and an increased difficulty in making mortgage payments. We reserve the important topics of the performance of loans to first-time homebuyers and the effects on credit risk of industry of employment for future research.

[25] The higher relative risks of the very low income group in the recent period is consistent with the observed shift we have seen to lower relative income buckets within the low and moderate income borrower segment in the post-crisis period compared with other periods as highlighted in Figure 2.

the relative income of a borrower declines; (2) to what extent have tighter eligibility standards reduced the risk of extending mortgage credit to low and moderate income home buyers relative to the pre-crisis period; and (3) to what extent can standard underwriting factors (e.g. DTI, FICO and LTV) account for the credit risk on conforming purchase loans without directly controlling for relative income?

We find first that, in general, credit risk increases as relative income falls. Low and moderate income loans typically have higher LTVs, lower FICOs and higher DTIs. Consistent with these attributes, the rate of early defaults suggests that low and moderate income lending is on average more risky for all three sample periods (pre-crisis, crisis and post-crisis), with the lowest income group of borrowers carrying the highest default risk.

Second, we find that eligibility standards are extremely important for sustainable lending to low and moderate income borrowers. In particular, tighter eligibility standards after the crisis greatly improved the credit performance of low and moderate income loans acquired by Fannie Mae in the recent period. The default rate for non-eligible loans for the very low income group in the pre-crisis period is 6.6 percent, compared to 1.9 percent for the eligible very low income loans. In the crisis period, the non-eligible very low income loans have a default rate of 16.8 percent, compared to 2.9 percent for the very low income loans in the eligible-loan category in the same period.

We find that standard credit variables are good predictors of default across relative income groups, with the exception of very low income borrowers. In particular, standard underwriting risk factors (e.g. DTI, FICO and LTV) along with state and vintage fixed effects sufficiently control for additional risks for currently eligible moderate income and low income loans for all sample periods, with remaining risk unexplained by standard underwriting factors either insignificantly different from zero or small in terms of percent

of total risk. However, the lowest income group displays additional default risks that cannot be fully explained by standard underwriting risk factors. The unexplained risk among very low income loans, however, after applying today's underwriting standards and including standard risk factors and state and vintage controls is reasonably stable across underwriting regimes and subsequent housing market environments. In particular, this additional risk for very low income loans ranges from approximately 21 to 33 percent across our three sample periods.

Table 1a: Variable Definitions and Summary Statistics by Sample Period

Variables	Description	2002-2004 Sample	2005-2007 Sample	2011-2013 Sample
Dependent Variable (Binary, %)				
Default	The loan is 90 days past due within 24 months after first payment date	2.04%	6.52%	0.27%
Independent Variable (Splines)				
LTV	The mortgage's combined total loan to value ratio at origination, %	83.5	87.4	81.0
FICO	The lower of borrower and co-borrower's FICO score at origination	713.4	712.7	757.4
DTI	The total monthly debt to combined monthly income ratio	0.36	0.40	0.33
Independent Variable (Continuous)				
Difference of FICOs	The absolute difference of borrower and co-borrower FICO score	29.0	28.4	23.3
Independent Variable (Categorical, %)				
Sub Finance	Subfin:<=10pct: the loan has less than 10% of sub-ordinate financing	5%	6%	2%
	Subfin:15pct: the loan has more than 10% but less than 15% of sub-ordinate financing	6%	6%	1%
	Subfin:>=20pct: the loan has more than 20% of sub-ordinate financing	3%	11%	0%
	OtherSubfin: the loan has other type of sub-ordinate financing	2%	3%	2%
	NoSubfin: the loan has no sub-ordinate financing	84%	74%	96%
Number of Borrower	One Borrower	47%	53%	48%
	Two borrowers or more	53%	47%	52%
Third Party Origination	Broker: The mortgage is initiated through a broker	20%	19%	8%
	Correspondent: The mortgage is initiated through a correspondent	31%	38%	40%
	Other: Not third party origination	49%	42%	51%
Loan Type	FRM 15: 15 - 20 year Fixed rate mortgage	10%	6%	11%
	FRM 30: 30 - 40 year Fixed rate mortgage	90%	94%	89%
Income Documentation	Missing: No income documentation	1%	1%	0%
	NonMiss - low doc: Part of the income documentation is missing	12%	16%	0%
	Full Doc: The mortgage contract has full documentation on borrower income	86%	83%	100%
Number of Units	Single family dwelling units = 1	98%	99%	99%
	Single family dwelling units = 2 or more	2%	1%	1%
Income to Area Median	Income/AMI below or equal to 50%	8%	8%	6%
	Income/AMI greater than 50% and less than or equal to 80%	22%	22%	17%
	Income/AMI greater than 80% and less than or equal to 100%	16%	15%	12%
	Income/AMI greater than 100%	54%	55%	64%
Population size	Number of Loans	2.2 million	2.0 million	1.1 million

Note: 1. Population is Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans only.

2. Categorical percentages do not sum to 100% due to rounding.

3. Difference in FICOs is summarized for only those loans with FICO scores for both the borrower and the co-borrower.

Table 1b: Summary Statistics of 2011 – 2013 Sample by Relative Income Status

Variables	Description	Income/AMI <= 50%	Income/AMI > 50% and <= 80%	Income/AMI > 80% and <= 100%	Income/AMI > 100%
Dependent Variable (Binary, %)					
Default	The loan is 90 days past due within 24 months after first payment date	0.7%	0.4%	0.3%	0.2%
Independent Variable (Splines)					
LTV	The mortgage's combined total loan to value ratio at origination, %	77.7	80.8	81.7	81.3
FICO	The lower of borrower and co-borrower's FICO score at origination	750.5	754.6	756.6	759.0
DTI	The total monthly debt to combined monthly income ratio	0.4	0.4	0.3	0.3
Independent Variable (Continuous)					
Difference of FICOs	The absolute difference of borrower and co-borrower FICO score	26.1	24.7	24.7	22.3
Independent Variable (Categorical, %)					
Sub Finance	Subfin:<=10pct: the loan has less than 10% of sub-ordinate financing	1%	1%	1%	2%
	Subfin:15pct: the loan has more than 10% but less than 15% of sub-ordinate financing	0%	0%	0%	1%
	Subfin:>=20pct: the loan has more than 20% of sub-ordinate financing	2%	1%	0%	0%
	OtherSubfin: the loan has other type of sub-ordinate financing	1%	1%	0%	2%
	NoSubfin: the loan has no sub-ordinate financing	95%	97%	98%	95%
Number of Borrower	One Borrower	85%	73%	59%	36%
	Two borrowers or more	15%	27%	41%	64%
Third Party Origination	Broker: The mortgage is initiated through a broker	8%	8%	8%	8%
	Correspondent: The mortgage is initiated through a correspondent	41%	42%	41%	40%
	Other: Not third party origination	51%	50%	51%	52%
Loan Type	FRM 15: 15 - 20 year Fixed rate mortgage	8%	8%	9%	13%
	FRM 30: 30 - 40 year Fixed rate mortgage	92%	92%	91%	87%
Income Documentation	Missing: No income documentation	0%	0%	0%	0%
	NonMiss - low doc: Part of the income documentation is missing	0%	0%	0%	0%
	Full Doc: The mortgage contract has full documentation on borrower income	100%	100%	100%	100%
Number of Units	Single family dwelling units = 1	99%	99%	99%	99%
	Single family dwelling units = 2 or more	1%	1%	1%	1%
Population size	Number of Loans	~74k	~196k	~139k	~730k

Note: 1. Population is Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans for the years 2011 – 2013 only.
2. Categorical percentages do not sum to 100% due to rounding.
3. Difference in FICOs is summarized for only those loans with FICO scores for both the borrower and the co-borrower.

Table 2. Correlations between Relative Income and Other Risk Characteristics in Each Sample Period

	2002-2004			2005-2007			2011-2013		
	DTI	LTV	FICO	DTI	LTV	FICO	DTI	LTV	FICO
Panel A: Sample Mean of Variables									
Income/AMI <= 50%	**0.45**	**82.3**	**704.7**	**0.46**	**88.8**	**684.2**	**0.38**	**77.7**	**750.5**
	(0.15)	(17.8)	(80.7)	(0.13)	(16.8)	(76.3)	(0.08)	(16.3)	(49.5)
Income/AMI > 50% and <= 80%	**0.41**	**84.8**	**709.7**	**0.44**	**88.8**	**699.4**	**0.35**	**80.8**	**754.6**
	(0.14)	(16.0)	(75.7)	(0.13)	(15.0)	(73.1)	(0.09)	(14.1)	(44.4)
Income/AMI > 80% and <= 100%	**0.38**	**84.6**	**712.3**	**0.41**	**87.6**	**708.4**	**0.34**	**81.7**	**756.6**
	(0.13)	(15.3)	(67.2)	(0.12)	(15.0)	(67.7)	(0.09)	(13.4)	(43.2)
Income/AMI > 100%	**0.33**	**82.7**	**716.3**	**0.37**	**84.1**	**724.0**	**0.31**	**81.3**	**759.0**
	(0.13)	(15.4)	(61.7)	(0.12)	(15.8)	(59.8)	(0.10)	(13.4)	(44.7)
Panel B: Correlation Coefficient of Variables with Relative Income									
Borrower Relative Income to AMI	**-0.31**	**-0.08**	**0.06**	**-0.31**	**-0.15**	**0.17**	**-0.29**	**-0.04**	**0.03**

Note: 1. Top panel: standard deviations are in parentheses.
2. Bottom Panel: all correlation coefficients are significant at the 0.1% level.
3. Population: Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans only.

Table 3: Early Default Odds and Rates by Income Group --- Eligible vs. Non-eligible Loans

	2002-2004		2005-2007		2011-2013	
	Relative Risk to hi-inc. Elig. 02-04	Default Rate	Relative Risk to hi-inc. Elig. 05-07	Default Rate	Relative Risk to hi-inc. Elig. 11-13	Default Rate
Overall						
Income/AMI <= 50%	8.1	4.4%	10.2	11.1%	4.0	0.7%
Income/AMI > 50% and <= 80%	5.6	3.0%	7.4	8.0%	2.3	0.4%
Income/AMI > 80% and <= 100%	4.3	2.3%	5.8	6.3%	1.8	0.3%
Income/AMI > 100%	2.7	1.5%	3.8	4.2%	1.0	0.2%
Non-eligible Loans Under Current Underwriting Standards						
Income/AMI <= 50%	12.3	6.6%	15.4	16.8%		
Income/AMI > 50% and <= 80%	8.8	4.8%	11.7	12.7%		
Income/AMI > 80% and <= 100%	7.1	3.9%	9.5	10.4%		
Income/AMI > 100%	4.8	2.6%	6.7	7.3%		
Eligible Loans Under Current Underwriting Standards						
Income/AMI <= 50%	3.5	1.9%	2.6	2.9%	4.0	0.7%
Income/AMI > 50% and <= 80%	2.3	1.3%	1.8	2.0%	2.3	0.4%
Income/AMI > 80% and <= 100%	1.7	0.9%	1.4	1.5%	1.8	0.3%
Income/AMI > 100%	1.0	0.5%	1.0	1.1%	1.0	0.2%
Number of Observations	2.2 million		2.0 million		1.1 million	

Note: 1. Table shows the odds ratio of 90+day delinquency rate within the first 24 month post origination relative to higher income group (Income/Area Median Income above 100%) within the specified time period.
2. Default Rates shown are the actual rate of 90+day delinquency within the first 24 month post origination for each income group within the same time period.
3. Population: Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans only.

Table 4: Early Default Odds by Income Group with Controls, Eligible Loans vs. Non-eligible Loans

Variable	2002-2004 Relative Risk to hi-inc. Elig. 02-04		2005-2007 Relative Risk to hi-inc. Elig. 05-07		2010-2013 Relative Risk to hi-inc. Elig. 10-13	
	No Additional Controls	With Direct Controls	No Additional Controls	With Direct Controls	No Additional Controls	With Direct Controls
Non-Eligible Loans Under Current Underwriting Standards						
Income/AMI <= 50%	12.3***	2.2***	15.4***	2.0***		
Income/AMI > 50% and <= 80%	8.8***	1.9***	11.7***	1.9***		
Income/AMI > 80% and <= 100%	7.1***	1.9***	9.5***	1.8***		
Income/AMI > 100%	4.8***	1.7***	6.7***	1.8***		
Eligible Loans Under Current Underwriting Standards						
Income/AMI <= 50%	3.5***	1.7***	2.6***	1.4***	4.0***	1.8***
Income/AMI > 50% and <= 80%	2.3***	1.3***	1.8***	1.1***	2.3***	1.2***
Income/AMI > 80% and <= 100%	1.7***	1.2***	1.4***	1.0	1.8***	1.2***
Income/AMI > 100%	1.0	1.0	1.0	1.0	1.0	1.0
Number of Observations	2.2 million		2.0 million		1.1 million	

Note: 1. No additional control results correspond to the default odds for each income group relative to Income/Area Median Income above 100% without controlling for other risk characteristics or state and time fixed effects.
2. Direct control results correspond to the default odds ratio for a given income group relative to Income/Area Median Income above 100%, when controlling for DTI, LTV, FICO, subordinate financing indicators, number of borrowers, third party originator indicators, term, number of units and state and vintage fixed effects in the logit regression.
3. Population: Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans only.
4. *10%, **5%, ***1% Significance Levels.

Table 5: Early Default Odds by Income Group – Direct Control vs. Two Step Control Results, Eligible Loans Only

Variable	2002 - 2004				2005 - 2007				2011 - 2013			
	LMI Only	Direct Control	Two Step Control	GINI without Income Control	LMI Only	Direct Control	Two Step Control	GINI without Income Control	LMI Only	Direct Control	Two Step Control	GINI without Income Control
Income/AMI <= 50%	3.5***	1.7***	1.5***	**0.664**	2.6***	1.4***	1.4***	**0.647**	4.0***	1.8***	1.6***	**0.540**
Income/AMI > 50% and <= 80%	2.3***	1.3***	1.2***	**0.680**	1.8***	1.1***	1.1**	**0.676**	2.3***	1.2***	1.1	**0.563**
Income/AMI > 80% and <= 100%	1.7***	1.2***	1.1	**0.711**	1.4***	1.0	1.0*	**0.705**	1.8***	1.2***	1.1	**0.619**
Income/AMI > 100%	1.0	1.0	1.0	**0.743**	1.0	1.0	1.0	**0.747**	1.0	1.0	1.0	**0.659**
Gini Coefficient	**0.234**	**0.732**	**0.732**	**0.731**	**0.112**	**0.719**	**0.719**	**0.717**	**0.231**	**0.640**	**0.640**	**0.638**
-2LogL	119,406	98,976	99,017		120,545	98,337	98,352		41,833	36,615	36,629	
AIC	119,414	99,184	99,225		120,553	98,545	98,560		41,841	36,823	36,837	
Number of Observations	1.2 million				0.8 million				1.1 million			

Note: 1. Direct control results are the odds ratio of default for that income group relative to Income/Area Median Income above 100%, when including other risk characteristics including DTI, LTV, FICO, subordinate financing indicators, number of borrowers, third party originator indicators, term, number of units and state and vintage fixed effects in the logit regression.
2. Two step control results are the odds ratio of default for that income group after using risk characteristics including DTI, LTV, FICO, subordinate financing indicators, number of borrowers, third party originator indicators, term, number of units and state and vintage fixed effects to model default in the first step, and model the residuals as a function of the low and moderate income indicators.
3. Gini without income control results are the Gini coefficients of the model without relative income controls applied to each income category and/or the total population within the same period.
4. *10%,**5%,***1% Significance Levels
5. Population: Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans only.

Figure 1: Very Low Income Borrower Distributions across States, Post Crisis Period (2010 – 2013)
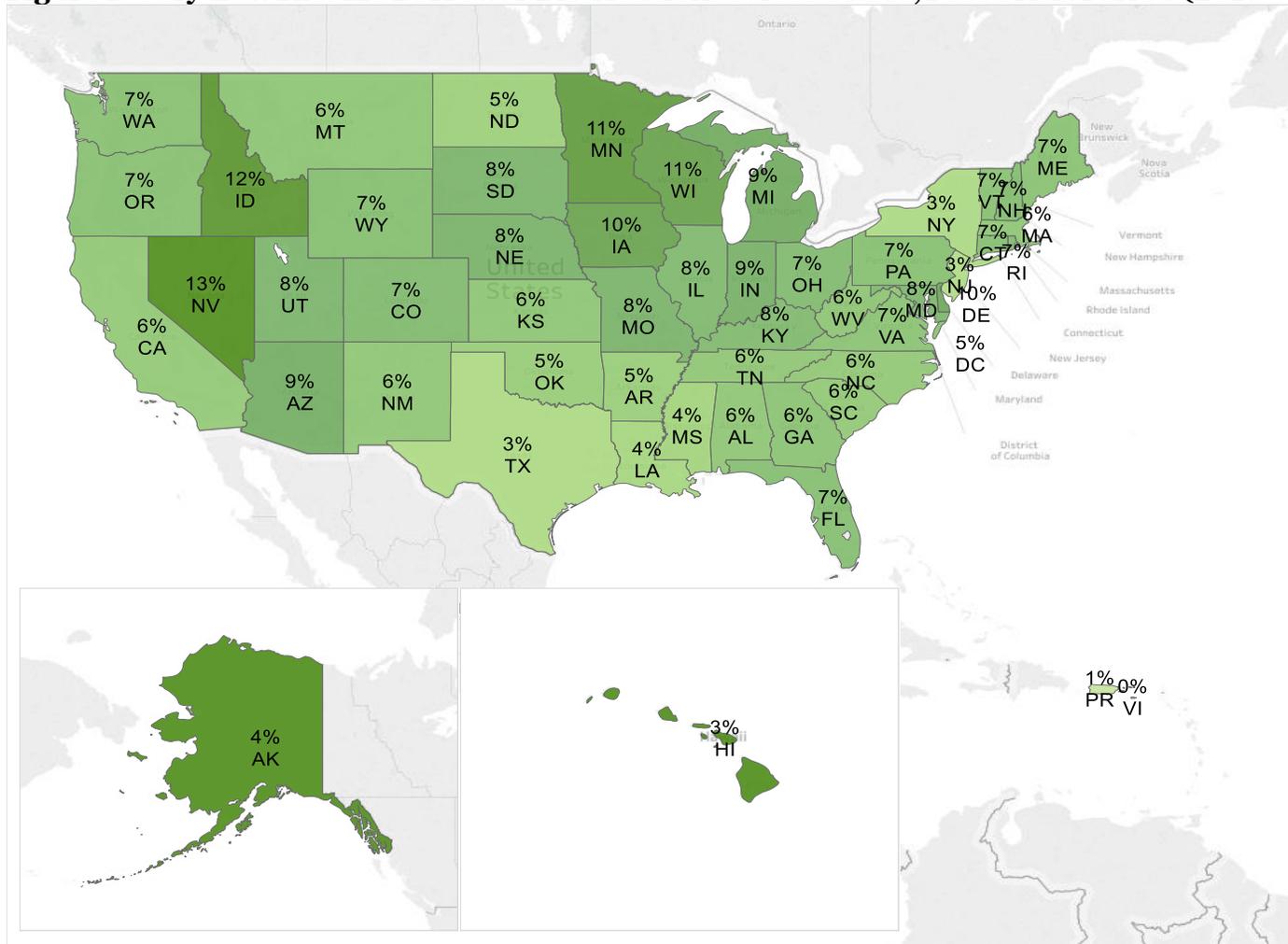
Note: Very low income shares vary significantly across States with a range of 1% to 11%. The states with lowest share of very low income borrowers (except PR and VI) are HI, NJ, and NY (3%). The States with highest share of very low income borrowers are ID (12%), WI (11%) and MN (11%).
Population: Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans only.

Figure 2: Distribution of Low and Moderate Income Borrowers by Relative Income
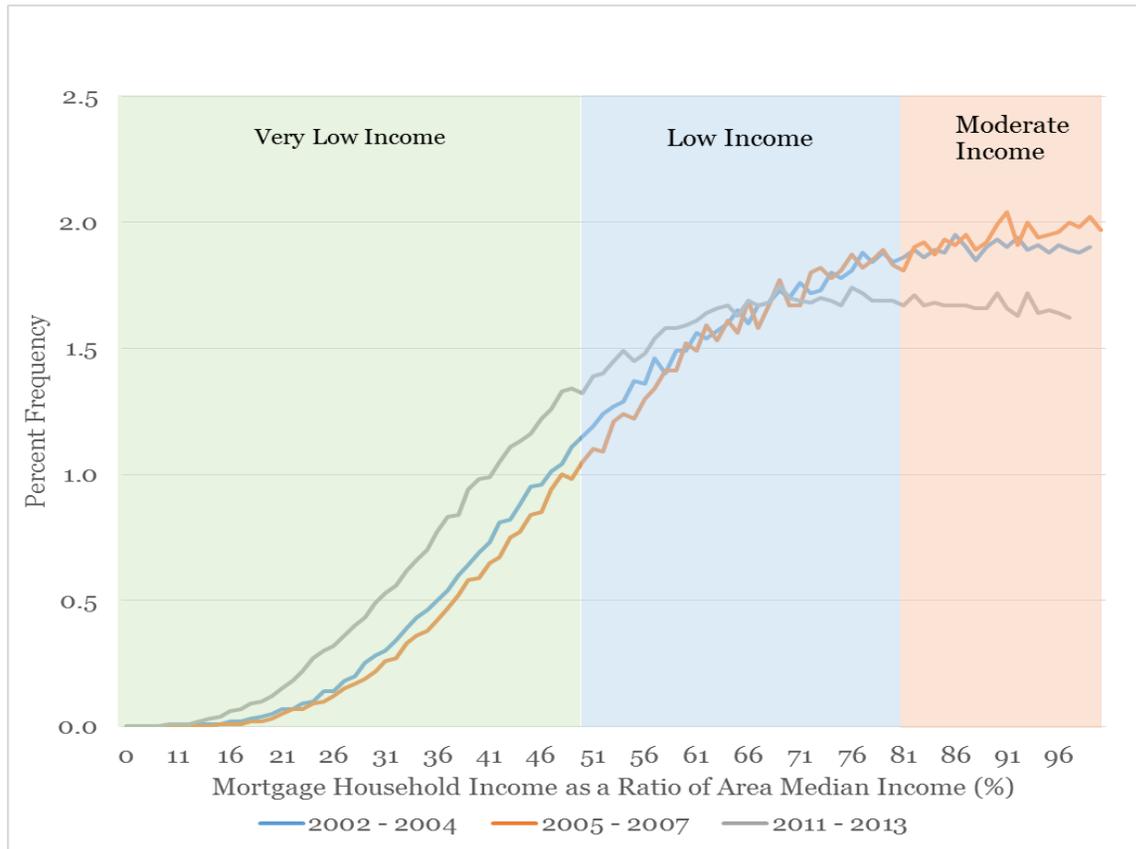
Note: X-axis is the mortgage household's income as a ratio of area median income (%) up to 100% of area median income (i.e. high income group >100% of area median income is excluded). Y-axis is the percent frequency, as a share of total low and moderate income mortgages (i.e. <=100% of area median income) for the sample period.

Figure 3: Comparison of Actual Default and Forecast Default – Eligible Loans Only

	Income/AMI <= 50%	Income/AMI > 50% and <= 80%	Income/AMI > 80% and <= 100%	Income/AMI > 100%
Actual Default	1.92%	1.27%	0.91%	0.54%
Forecast Default	1.48%	1.19%	0.92%	0.60%
Unexplained Risk	23%	6%	-2%	-12%

	Income/AMI <= 50%	Income/AMI > 50% and <= 80%	Income/AMI > 80% and <= 100%	Income/AMI > 100%
Actual Default	2.88%	1.98%	1.50%	1.09%
Forecast Default	2.28%	1.90%	1.58%	1.13%
Unexplained Risk	21%	4%	-5%	-4%
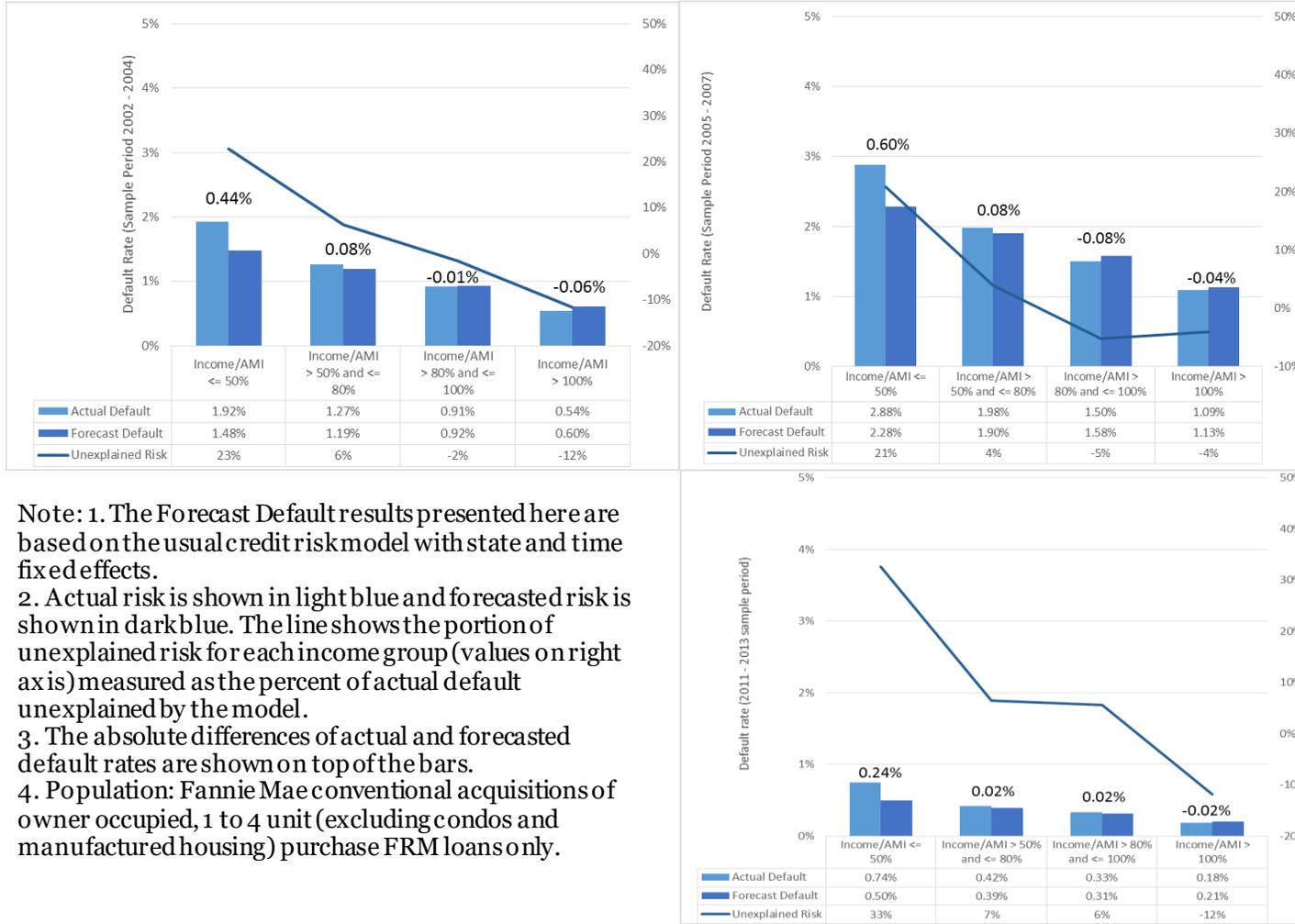
Note: 1. The Forecast Default results presented here are based on the usual credit risk model with state and time fixed effects.

2. Actual risk is shown in light blue and forecasted risk is shown in dark blue. The line shows the portion of unexplained risk for each income group (values on right axis) measured as the percent of actual default unexplained by the model.

3. The absolute differences of actual and forecasted default rates are shown on top of the bars.

4. Population: Fannie Mae conventional acquisitions of owner occupied, 1 to 4 unit (excluding condos and manufactured housing) purchase FRM loans only.

	Income/AMI <= 50%	Income/AMI > 50% and <= 80%	Income/AMI > 80% and <= 100%	Income/AMI > 100%
Actual Default	0.74%	0.42%	0.33%	0.18%
Forecast Default	0.50%	0.39%	0.31%	0.21%
Unexplained Risk	33%	7%	6%	-12%

35

REFERENCES

An, X., Deng, Y., Rosenblatt, E., & Yao, V. W. 2012. "Model stability and the subprime mortgage crisis." *The Journal of Real Estate Finance and Economics* 45(3), 545-568.

Archer, W. R., Ling, D. C., & McGill, G. A. 1996. "The effect of income and collateral constraints on residential mortgage terminations." *Regional Science and Urban Economics,* 26(3), 235-261.

Avery, R. B., Bostic, R. W., Calem, P. S., & Canner, G. B. 1996. "Credit risk, credit scoring, and the performance of home mortgages." *Federal Reserve Bulletin* 82, 621.

Avery, R. B., & Brevoort, K.P. 2015. "The subprime crisis: is government housing policy to blame?" *Review of Economics and Statistics* 352-363.

Barakova, I., Calem, P. S., & Wachter, S. M. 2014. "Borrowing constraints during the housing bubble." *Journal of Housing Economics* 4-20.

Crook, J. N., Edelman, D. B., & Thomas, L. C. 2007. "Recent developments in consumer credit risk assessment." *European Journal of Operational Research* 183.3, 1447-1465.

Deng, Y.H., & Gabriel, S. A. 2006. "Risk-based pricing and the enhancement of mortgage credit availability among underserved and higher credit-risk populations." *Journal of Money, Credit, and Banking* 38(6), 1431-1460.

Deng, Y.H., Quigley, J.M., & Van Order, R. 1996. "Mortgage default and low downpayment loans: the costs of public subsidy." *Regional Science and Urban Economics* 263-285.

Deng, Y.H., Quigley, J. M., & Van Order, R. 2000. "Mortgage terminations, heterogeneity and the exercise of mortgage options." *Econometrica* 275-307.

Duca, J. V., & Rosenthal S. S. 1994. "Borrowing constraints and access to owner-occupied housing." *Regional Science and Urban Economics* 24(3), 301-322.

Firestone, S., Van Order, R., & Zorn, P. 2007. "The performance of low-income and minority mortgages." *Real Estate Economics* 35(4), 479-504.

Goodman, L. S., Seidman E., & Zhu, J. 2015. "VA loans outperform FHA loans—why? and what can we learn?" *The Journal of Fixed Income* 39-51.

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

Kelly, A. 2008. "" Skin in the game": Zero downpayment mortgage default." *Journal of Housing Research* 17(2), 75-99.

Lam, K., Dunsky, R. M., & Kelly, A. 2013. "Impacts of down payment underwriting standards on loan performance — Evidence from the GSEs and FHA portfolios." *Office of Policy Analysis and Research, Federal Housing Finance Agency* Working Paper.

Li, W., and Goodman, L. 2015. "Measuring Mortgage Credit Availability Using Ex-Ante Probability of Default." *Washington, DC: Urban Institute*.

Mayer, C., Pence, K., & Sherlund, S. M. 2009. "The rise in mortgage defaults." *The Journal of Economic Perspectives* 27-50.

Mays, E. 2001. "Handbook of Credit Scoring." *Book, Global Professional Publishing* 382.

Painter, G., Gabriel, S., & Myers, D. 2001. "Race, immigrant status, and housing tenure choice." *Journal of Urban Economics* 49(1), 150-167.

Quercia, R. G., Stegman, M. A., Davis, W. R., & Stein, E. 2002. "Performance of community reinvestment loans: implications for secondary market purchases." *Low-Income Homeownership: Examining the Unexamined Goal* 348-74.

Quercia, R.G., & Stegman, M. A. 1992. "Residential mortgage default: a review of the literature." *Journal of Housing Research* 341.

Van Order, R., & Zorn, P. 2002. "Performance of low-income and minority mortgages." In *Low-income homeownership: Examining the unexamined goal*, by R., & Zorn, P. Van Order, 322-347.

Van Order, R., & Zorn, P. 2003. "Income, location and default: Some implications for community lending." *Real Estate Economics* 385-404.

APPENDIX: COEFFICIENT ESTIMATES FOR CREDIT RISK

The coefficients for the early default models we estimate in this paper are shown in Tables A1 – A3 for eligible conventional purchase loans for each of our three respective sample periods. The first column of each table shows the results with only the low and moderate income indicators entering the estimation. The second column shows the results with both the low and moderate income indicators and other standard risk characteristics, such as FICO, LTV and DTI used as explanatory variables. The third column shows the two step results: In the first step only the standard credit risk factors without relative income status are used as explanatory variables, and in the second step the residuals from the first step are regressed on the relative income indicators.

Most of the standard credit risk characteristics in the regressions are significant at the 1% level in each sample in both columns (2) and (3). Examining the coefficient estimate results for the standard controls (column (2)), the estimated effects of these factors to the loan's default risk for all three periods conform to intuition: for instance, loans with single borrowers and originated with low documentation are more likely to experience early default, and first liens that have secondary liens attached are less likely to default. The difference in FICOs has a negative impact on default probability, because the model uses the lower of the two FICO scores for loans with multiple borrowers as the direct control for credit score. Thus, the difference variable measures the creditworthiness of the higher scoring borrower and increases as the higher scoring borrower's credit score grows relative to the lower scoring borrower's score. The default risk profile is also differentiated by third party origination. In particular, if the loan is originated through a broker or correspondent, it is more likely to default than if it were originated through other channels. We transform the key continuous risk variables of LTV, FICO and DTI using linear splines. In general, there is a positive relationship between LTV and default risk,

with lower down payments as a proportion of a home's value associated with higher risks. Borrower's FICO score is negatively correlated with default risk, and DTI is positively correlated with default risk.

For the most part, the estimated coefficient values on the standard risk factors are consistent over the three time periods, with the exception being that a number of post-crisis variables are insignificant, potentially due to the lower default rate in this period.

When the low and moderate income indicators are used as direct controls (column (2)), they are estimated to be significant, with risk rising as relative income falls. The exception in this case being the second time period when the moderate income group enters positively but insignificantly different from the high income group. In the offset regression results (column (3)), the relative income variables enter only to explain the residual credit risk after the standard controls have been used to model default. In this case unexplained risk falls as relative income rises for the first and second time periods, but only the very low income group has significant residual unexplained risk in the last time period. Also note that, model coefficients for the standard controls and model fit are largely unchanged between columns (2) and columns (3), providing further support for the sufficiency of standard underwriting controls in predicting credit performance for all but the very low income group.

Table A1: Early Default Regression Results, Conventional Eligible Purchase Loans, Sample Period 2002 - 2004

	Dependent Variable:			Early Default				
	Period:			**Sample Period 2002 - 2004**				
		(1)		(2)			(3)	
		Estimate	StdErr.	Estimate		StdErr.	Estimate	StdErr.
Intercept		-4.44 ***	0.07	-3.44	***	0.23	-3.39 ***	0.04
Baseline Controls (Splines)								
CLTV Spline	CLTV Less Than 60			-0.02	***	0.00	-0.02 ***	0.00
	CLTV between 60 and 70			0.04	***	0.01	0.04 ***	0.01
	CLTV between 70 and 80			0.03	***	0.01	0.03 ***	0.01
	CLTV between 80 and 90			0.05	***	0.00	0.05 ***	0.00
	CLTV Larger than 90			0.07	***	0.00	0.07 ***	0.00
FICO Spline	FICO between 620 and 660			-0.02	***	0.00	-0.02 ***	0.00
	FICO between 660 and 720			-0.02	***	0.00	-0.02 ***	0.00
	FICO between 720 and 760			-0.03	***	0.00	-0.03 ***	0.00
	FICO Larger than 760			-0.01	*	0.00	0.00 ***	0.00
DTI Spline	DTI Less than 36			0.47	*	0.20	1.17 *	0.17
	DTI between 36 and 45			1.38	***	0.38	1.65 ***	0.35
	DTI Larger than 45			0.79		0.79	1.10	0.61
Baseline Controls (Categorical and Continuous)								
Sub Finance	01:Subfin:<=10pct			-0.67	***	0.04	-0.02 ***	0.04
	02:Subfin:15pct			-0.67	***	0.03	-0.02 ***	0.03
	03:Subfin:>=20pct			-0.45	*	0.05	-0.03 ***	0.03
	04:OtherSubfin			-0.31	***	0.06	0.00 ***	0.06
	05:NoSubfin			0.00	.	.	0.00 .	.
Number of Borrower	One Borrower			0.92	***	0.03	1.03 ***	0.05
	Two borrowers or more			0.00	.	.	0.00 .	.
Difference of FICOs				-0.01	***	0.00	-0.01 ***	0.00
Third Party Origination	Broker			0.39	***	0.01	0.38 ***	0.01
	Correspondent			0.18	***	0.01	0.17 ***	0.01
	Other			0.00	.	.	0.00	.
Loan Type	FRM 15			-0.40	***	0.03	-0.42 ***	0.03
	FRM 30			0.00	.	.	0.00	.
Dwelling Units	One Unit			-0.68	***	0.03	-0.66 ***	0.03
	Two Units or More			0.00	.	.	0.00	.
Expanded Controls								
							Second Stage	
Income to Area Median	**Income/AMI <= 50%**	1.26 ***	-0.02	0.53	***	0.03	*0.28* ***	*0.03*
	Income/AMI > 50% and <= 80%	0.84 ***	0.04	0.27	***	0.02	*0.07* ***	*0.02*
	Income/AMI > 80% and <= 100%	0.55 ***	0.03	0.14	***	0.03	*-0.02*	*0.02*
	Income/AMI > 100%	0.00 .	.	0.00	.	.	*-0.11* ***	*0.02*
-2LogL		119,406		98,976			99,017	
AIC		119,414		99,184			99,225	
Number of Observations		1.2 million		1.2 million			1.2 million	

*10%, **5%, ***1% Significance Levels

Table A2: Early Default Regression Results, Conventional Eligible Purchase Loans, Sample Period 2005 - 2007

		Dependent Variable: Early Default					
		Period: Sample Period 2005 - 2007					
		(1)		(2)		(3)	
		Estimate	StdErr.	Estimate	StdErr.	Estimate	StdErr.
Intercept		-5.62 ***	0.04	-5.08 ***	0.23	-5.01 ***	0.04
Baseline Controls (Splines)							
CLTV Spline	CLTV Less Than 60			0.02 ***	0.00	0.01 ***	0.00
	CLTV between 60 and 70			0.06 ***	0.00	0.06 ***	0.00
	CLTV between 70 and 80			0.03 ***	0.00	0.03 ***	0.00
	CLTV between 80 and 90			0.06 ***	0.00	0.05 ***	0.00
	CLTV Larger than 90			0.04 ***	0.00	0.04 ***	0.00
FICO Spline	FICO between 620 and 660			-0.03 ***	0.00	-0.03 ***	0.00
	FICO between 660 and 720			-0.02 ***	0.00	-0.02 ***	0.00
	FICO between 720 and 760			-0.02 ***	0.00	-0.02 ***	0.00
	FICO Larger than 760			-0.02 ***	0.00	-0.02 ***	0.00
DTI Spline	DTI Less than 36			0.90 ***	0.04	1.06 ***	0.04
	DTI between 36 and 45			2.88 ***	0.06	3.02 ***	0.06
	DTI Larger than 45			2.60 ***	0.02	2.74 ***	0.02
Baseline Controls (Categorical and Continuous)							
Sub Finance	01:Subfin:<=10pct			-0.43 ***	0.04	-0.44 ***	0.01
	02:Subfin:15pct			-0.33 ***	0.05	-0.36 ***	0.01
	03:Subfin:>=20pct			-0.11 ***	0.04	-0.14 ***	0.01
	04:OtherSubfin			-0.52 ***	0.05	-0.51 ***	0.01
	05:NoSubfin			0.00 .	.	0.00 .	.
Number of Borrower	One Borrower			0.69 ***	0.02	0.74 ***	0.02
	Two borrowers or more			0.00 .	.	0.00	.
Difference of FICOs				-0.01 ***	0.00	-0.01 ***	0.00
Third Party Origination	Broker			0.39 ***	0.02	0.38 ***	0.00
	Correspondent			0.13 ***	0.02	0.12 ***	0.00
	Other			0.00 .	.	0.00 .	.
Loan Type	FRM 15			-0.62 ***	0.04	-0.63 ***	0.01
	FRM 30			0.00 .	.	0.00 .	.
Dwelling Units	One Unit			-0.63 ***	0.04	-0.63 ***	0.04
	Two Units or More			0.00 .	.	0.00 .	.
Expanded Controls						*Second Stage*	
Income to Area Median	Income/AMI <= 50%	0.96 ***	0.00	0.36 ***	0.03	*0.26* ***	*0.03*
	Income/AMI > 50% and <= 80%	0.61 ***	0.00	0.12 ***	0.02	*0.04* **	*0.02*
	Income/AMI > 80% and <= 100%	0.35 ***	0.00	0.01	0.02	*-0.05* *	*0.04*
	Income/AMI > 100%	0.00 .	.	0.00 .	.	*-0.04* ***	*0.01*
-2LogL		120,545		98,337		98,352	
AIC		120,553		98,545		98,560	
Number of Observations		0.8 million		0.8 million		0.8 million	

*10%, **5%, ***1% Significance Levels

Table A3: Early Default Regression Results, Conventional Eligible Purchase Loans, Sample Period 2011 - 2013

		(1) Estimate		StdErr.	(2) Estimate		StdErr.	(3) Estimate		StdErr.
Dependent Variable:					Early Default					
Period:					Sample Period 2011 - 2013					
Intercept		-6.45	***	0.01	-6.13	***	1.10	-6.07	***	0.47
Baseline Controls (Splines)										
CLTV Spline	CLTV Less Than 60				0.01		0.01	0.00		0.01
	CLTV between 60 and 70				0.03	*	0.01	0.02		0.01
	CLTV between 70 and 80				0.00		0.01	0.00		0.01
	CLTV between 80 and 90				0.03	***	0.01	0.03	***	0.01
	CLTV Larger than 90				0.08	***	0.01	0.09	***	0.01
FICO Spline	FICO between 620 and 660				-0.02	***	0.00	-0.02	***	0.00
	FICO between 660 and 720				-0.03	***	0.00	-0.03	***	0.00
	FICO between 720 and 760				-0.02	***	0.00	-0.02	***	0.00
	FICO Larger than 760				-0.01	***	0.00	-0.01	***	0.00
DTI Spline	DTI Less than 36				2.96	***	0.45	3.40	***	0.44
	DTI between 36 and 45				3.87	***	0.68	4.14	***	0.68
	DTI Larger than 45				-4.89		2.90	-4.51		2.89
Baseline Controls (Categorical and Continuous)										
Sub Finance	01:Subfin:<=10pct				-0.33		0.20	-0.31		0.20
	02:Subfin:15pct				-0.35		0.25	-0.34		0.25
	03:Subfin:>=20pct				-0.51	***	0.17	-0.34	**	0.16
	04:OtherSubfin				-0.31	*	0.18	-0.31	*	0.18
	05:No Subfin				0.00		.	0.00		.
Number of Borrower	One Borrower				0.82	***	0.06	0.92	***	0.06
	Two borrowers or more				0.00		.	0.00		.
Difference of FICOs					-0.01	***	0.00	-0.01	***	0.00
Third Party Origination	Broker				0.18	***	0.07	0.18	***	0.07
	Correspondent				0.16	***	0.04	0.17	***	0.04
	Other				0.00		.	0.00		.
Loan Type	FRM 15				-0.31	***	0.08	-0.34	***	0.08
	FRM 30				0.00		.	0.00		.
Dwelling Units	One Unit				0.25		0.21	0.27		0.21
	Two Units or More				0.00		.	0.00		.
Expanded Controls								*Second Stage*		
Income to Area Median	Income/AMI <= 50%	1.40	***	0.00	0.55	***	0.03	*0.35*	***	*0.04*
	Income/AMI > 50% and <= 80%	0.84	***	0.00	0.16	***	0.02	*0.00*		*0.04*
	Income/AMI > 80% and <= 100%	0.61	***	0.00	0.15	***	0.03	*0.01*		*0.05*
	Income/AMI > 100%	0.00	.	.	0.00	.	.	*-0.11*	***	*0.03*
-2LogL		41,833			36,615			36,629		
AIC		41,841			36,823			36,837		
Number of Observations		1.1 million			1.1 million			1.1 million		

*10%, **5%, ***1% Significance Levels