HONG KONG INSTITUTE FOR MONETARY AND FINANCIAL RESEARCH

PRIVATE MORTGAGE SECURITIZATION AND ADVERSE SELECTION - NEW EVIDENCE FROM EXPECTED LOAN LOSSES

Abdullah Yavas and Shuang Zhu

HKIMR Working Paper No.05/2024

June 2024

HONG KONG ACADEMY OF FINANCE

金融學院

Hong Kong Institute for Monetary and Financial Research
香港貨幣及金融研究中心
(a company incorporated with limited liability)

Private Mortgage Securitization and Adverse Selection - New Evidence from Expected Loan Losses

Abdullah Yavas[1]
Robert E. Wangard Chair
School of Business
University of Wisconsin - Madison
Madison, WI 53706

and

Shuang Zhu
Associate Professor
Department of Finance
Kansas State University
Manhattan, KS 66506

We thank Editor Geert Bekaert and two anonymous reviewers for their very helpful comments. We thank Brent Ambrose, Michael LaCoure-Little, Qi Min, Lan Shi, and the participants at the 2019 ASSA AEA/AREUEA joint session, 2019 Multinational Finance Society Conference, 2019 International Conference on Applied Theory, Macro and Empirical Finance, 2019 Asian Real Estate Society Conference, 2018 Annual Meeting of the Financial Management Association, 2018 Annual Meeting of American Real Estate Society, and the seminar participants at the Office of the Comptroller of the Currency, the National University of Singapore, Xiamen University, Tsinghua University, Renmin University, Hitotsubashi University, California State University-Fullerton, and Hong Kong Institute of Monetary Research for their valuable comments. Shuang Zhu thanks the generous summer research support for this project from Kansas State University College of Business Administration. Abdullah Yavas gratefully acknowledges financial support from Hong Kong Institute for Monetary and Financial Research. This paper represents the views of the authors, which are not necessarily the views of the Hong Kong Monetary Authority, Hong Kong Academy of Finance Limited, or Hong Kong Institute for Monetary and Financial Research. The above-mentioned entities except the authors take no responsibility for any inaccuracies or omissions contained in the paper.

[1] Corresponding author. Tel: 1-608-692-5430 Email: yavas@wisc.edu

Abstract

This paper studies expected loan loss and adverse selection in private mortgage securitization. The research extends the previous literature on securitization that has focused on default probability. Expected loan loss incorporates both the probability of default and loss given default and represents a comprehensive measure of loan quality that has the ultimate impact on lenders and investors. This new measure of loan quality reverses some of the findings in the previous literature. The disparity in results between the alternative measures is more pronounced among loans with higher expected loss given default. Our results provide new evidence of adverse selection in prime loans. This cherry-picking behavior does not hold for subprime loans.

Keywords: Securitization, Default, Loan Loss, Adverse Selection
JEL Codes: G01, G21

1 Introduction

In the years leading up to the Great Recession, private securitization of residential mortgages experienced dramatic expansions. From 2001 to 2006, non-agency originations increased from \$680 billion to \$1.480 trillion and non-agency MBS issuance increased from \$240 billion to \$1.033 trillion. What is more striking is that, while private label mortgages constituted about 15% of all outstanding mortgages in 2009, they made up more than half of the foreclosure starts (Piskorski et al., 2010). There is also evidence that the surge in private label mortgage securitization before the financial crisis fueled a large expansion in mortgage credit supply (Mian and Sufi, 2019).

Given the critical role that mortgage default loss played in the recent financial crisis and given the rapid growth of private securitization in the years prior to the financial crisis, it is not surprising that the role of private securitization in the financial crisis has been studied extensively. However, due to data constraints, earlier studies have used a narrow definition of loan quality: probability of default.[1] While default probability reflects one aspect of loan quality, loan loss has the ultimate impact on lenders' and investors' balance sheets. Lenders, investors, and policymakers are concerned about both the probability that a loan goes into default and the loss suffered in the case of a default. Lending institutions estimate the expected loan loss and record the estimation under the loan loss reserve in their balance sheets.[2] In addition, the loss severity rate is a well-known and commonly used measure in

[1] A large number of previous studies use default probability as a proxy for loan quality (e.g., Ambrose et al., 2005; Mian and Sufi, 2009; Keys et al., 2010a, 2010b, 2012; Demyanyk and Van Hemert, 2011; Agarwal et al., 2012; Adekino et al., 2013; Elul, 2016; Ambrose et al., 2016).

[2] Loan loss reserve represents the money that banks set aside to offset future loan losses. The Financial Accounting Standards Board (FASB) sets the standards on expected credit loss accounting.

mortgage risk analysis. Thus, there is no question that unconditional loan loss, which captures both the probability that a loan goes into default and the loss incurred in the case of a default, rather than the probability of default alone, is a more comprehensive measure of loan performance, and plays a key role in lenders' securitization and origination decisions.[3,4]

This paper contributes to the existing literature by utilizing expected loan loss to study lender's mortgage securitization decisions for both prime and subprime loans. Specifically, we investigate whether lenders securitize loans with different expected quality to sell in the secondary market than those retained in their portfolios. Literature has documented mixed findings of adverse selection in private securitization. We aim to shed additional light on lender's securitization decisions by using a more comprehensive measure of loan quality.

To investigate whether lenders cherry-pick and keep loans with different qualities on their books, we need to estimate lenders' ex ante perceived loan quality, i.e. their expected loan loss. Since loan origination and securitization decisions could be jointly determined, to infer the potential causal relationship, the expected loan loss needs to be estimated without the impact of the securitization status. We adopt a structured approach as proposed by Ambrose et al. (1995) and Agarwal et al. (2012) to address this challenge. The basic idea of this method is to construct an out-of-sample expected loan loss variable that is estimated regardless of the securitization status, and then regress the securitization decision on the

---

[3] This paper uses the terms "loan loss" and "unconditional loan loss" interchangeably.

[4] It is important to differentiate unconditional loan loss from loss given default (LGD). For LGD, as the name implies, it is a loss 'conditional' on default. In contrast, unconditional loan loss is not conditional on default, as both current loans (with zero loan loss) and defaulted/liquidated loans (with realized loan loss) are included in this measure.

expected loan loss. We also estimate the expected loan loss as if loans were all treated as portfolio loans or loans were all treated as securitized loans. This further alleviates any concerns in the expected loan loss estimation that might reflect variation in servicing treatments of portfolio versus sold loans.

Using a nationwide dataset, the empirical results show that, in the years leading to the Great Recession, lenders were more likely to securitize prime loans with greater expected losses. Our findings survive various robustness tests. We also show that omitted variables are unlikely to weaken our estimates. This cherry-picking behavior does not hold for subprime loans: we find no significant difference in expected loan loss between sold loans and portfolio loans for the overall subprime sample. On the contrary, in some subprime loan subsamples, lenders were less likely to securitize subprime loans with greater expected losses.

Our analysis of prime loans provides new evidence of adverse selection in mortgage securitization; lenders were more likely to sell lower quality prime loans into securitization and retain higher quality prime loans for their own portfolios. Our finding is in contrast to Ambrose et al. (2005) who find that observably riskier loans are more likely to be retained by the lender, and Agarwal et al. (2012) who find no significant difference between portfolio loans and sold loans. A comparison of our results to those of Ambrose et al. (2005) and Agarwal et al. (2012), who use a similar empirical methodology as the current paper, shows that using the probability of default as a measure of loan quality can lead to very different conclusions about the impact of expected loan quality on securitization decision.[5] The

[5] Elul (2016) uses a different methodology and reports mixed results for loans originated in 2005 and 2006.

disparity in results between the alternative measures is more pronounced among loans with higher expected loss given default.

It is important to note that even though the expected loan loss is constructed using "observable" loan characteristics, asymmetric information between lenders and MBS investors can still exist. For example, lenders have proprietary models that they use in mortgage valuation and loan performance predictions (DeMarzo, 2005). More importantly, many MBS investors rely on credit ratings to make their investment decisions, and credit rating agencies failed to do their due diligence in examining the mortgages in the MBS pools during the pre-crisis years, hence leading to distortions in the perceived loan quality by investors.[6]

Contrary to prime loans, we do not find significant evidence that lenders cherry-pick better quality subprime loans to retain in their portfolio. On the contrary, in some subsamples, lenders do the opposite where they sell higher quality subprime loans and retain lower quality ones. This seemingly counterintuitive result is also reported for commercial mortgages in Black et al. (2020) who show that higher quality commercial mortgages are more likely to be securitized and lower quality loans are more likely to be in lenders'

The paper compares loan performances of sold loans to observably similar portfolio loans. He shows that privately securitized prime FRM loans are less likely to default while privately securitized prime ARM loans are more likely to default, compared to portfolio loans. LaCour-Little and Zhang (2014) find that, ex post, securitized home equity loans incur a higher default rate and larger loss given default than portfolio loans. Hwang et al. (2017) find that an increase in FHA loan limits results in riskier borrowers 'cherry-picking' against FHA. Those newly qualified loans have a higher default rate and larger loss given default.

[6] Other evidence of cherry-picking using observable characteristics is reported in Downing, Jaffee, and Wallace (2009) who show that Freddie Mac sells more lower-credit-quality residential mortgage-backed securities to bankruptcy-remote special purpose securitization vehicles than it retains in its portfolio. Similarly, An, Deng and Gabriel (2011) report empirical evidence for the presence of adverse selection problems based on observable characteristics in the market for commercial mortgage loans.

portfolios.[7]

Why do we observe adverse selection with respect to prime loans but not subprime loans? One explanation is that the buyers of prime MBS are typically institutional investors (e.g., pension funds) who are less financially sophisticated than MBS underwriters and often rely on rating agencies for the assessment of these securities. It has been well established that rating agencies have failed in their duties to accurately assess the riskiness of these securities in the years leading up to the financial crisis (Benmelech and Dlugosz, 2010; Ashcraft et al., 2011; Becker and Milbourn, 2011; Griffin, 2019).[8] Since the valuation of MBS relies on the credit ratings and the credit ratings failed to convey the accurate valuation of MBS during the pre-crisis time, this created incentives for lenders to cherry-pick higher quality prime loans to retain in their own portfolios and sell lower quality prime loans still with triple-A ratings. We also provide one piece of direct evidence on lenders' cherry-picking based on loan observables: within the same loan-to-value (LTV) pricing bracket, securitized loans are more likely to have an LTV ratio close to the upper limit of the LTV bracket. As loans in the same LTV pricing bracket (with otherwise similar characteristics) carry the same interest rate, this result confirms that lenders cherry-pick less risky mortgages among those with the same rate of returns to keep in their portfolios.

In contrast, subprime mortgage-backed securities are generally purchased by more

---

[7] In contrast, Agarwal et al. (2012) find no statistically significant difference in the probability of default between securitized and portfolio subprime loans while Elul (2016) reports that privately securitized subprime loans, originated in 2005 and 2006, perform worse than portfolio loans.

[8] According to the Financial Crisis Inquiry Commission "there was a clear failure of corporate governance at Moody's, which did not ensure the quality of its ratings on tens of thousands of mortgage-backed securities and CDOs." By the end of year 2010, 73% of the mortgage-backed securities that Moody's had rated triple-A in 2006 were downgraded to junk bond rating (Financial Crisis Inquiry Report, 2011).

sophisticated buyers (e.g., hedge funds) who scrutinize the detailed loan characteristics disclosed by the sellers before their purchase. The lack of adverse selection in subprime loans might also be due to the fact that lenders are in a better position to execute modifications/loss mitigations, which are more likely to arise for lower quality mortgages (Black et al., 2020). Piskorski et al. (2010) and Kruger (2016) show that securitized loans are less likely to be modified and more likely to be foreclosed than portfolio loans. This indicates that portfolio loans are serviced more efficiently than sold loans. If MBS investors' tendency to avoid extremely low quality subprime loans leads to deep discounts in their valuation, lenders would be better off keeping and servicing those very low quality loans in their books. In addition, compared to prime loans where about 37 percent in our sample are kept in banks' balance sheets, only about nine percent of subprime mortgages are kept as portfolio loans. Since lenders seem to securitize as many subprime loans as possible, another possible explanation is that the remaining nine percent of subprime mortgages might be loans with very low quality that fail to meet securitization parameters and are difficult to securitize. Another explanation is that subprime loans have a higher default rate. If a large number of loans defaulted soon after origination, that could hurt the lender's reputation. Furthermore, the Servicing and Pooling Agreement could be used to require lenders to repurchase loans that default soon after securitization. Since lower quality subprime loans are likely to default early, lenders may choose to sell relatively better quality subprime loans to investors.

This paper is structured as follows: Section 2 discusses the estimation methodology and data, Section 3 presents the main empirical results, Section 4 conducts robustness checks, and Section 5 summarizes the paper.

2 Estimation Methodology and Data

This section first discusses the estimation methodology in Section 2.1 and then introduces the data set and the sample in Section 2.2.

2.1 Estimation Methodology

Mortgage credit characteristics vary in many dimensions such as a borrower's credit score, loan-to-value ratio, and documentation status, etc. In determining what kind of loans are sold into securitization or kept on a bank's books, previous literature typically compares individual risk characteristics between portfolio loans and securitized loans (e.g., Krainer and Laderman, 2014), or regresses the probability of loan sale on the individual risk characteristics (e.g., Jiang et al., 2013). While these approaches are intuitive, two potential issues might arise. First, individual loan characteristics may point in different directions in terms of loan quality. For example, in Jiang et al. (2013) paper, sold loans have a lower loan-to-value ratio while at the same time exhibiting a higher proportion of low documentation status. Thus, often times this approach makes it difficult to infer the overall riskiness or loan quality of sold loans versus that of portfolio loans.

The second issue associated with the approach mentioned above is that securitization decisions and loan characteristics could be simultaneously determined. For example, a lender might ask for full documentation if the mortgage is intended to be kept on the bank's books. On the other hand, a lender might accept low or no documentation if the loan is originated with the intention to be sold (Bubb and Kaufman, 2014). If securitization and some loan risk characteristics are jointly determined, the reduced form regression might yield biased coefficient estimates.

To overcome the first challenge, we need to have a single comprehensive measure of loan quality that factors in the various risk dimensions to formally test whether lenders choose to sell loans with different quality than those kept in their portfolios. Earlier studies utilize the default event as the measure of loan quality to investigate adverse selection in securitization. In this paper, we proxy the loan quality by using the unconditional loan loss as a more comprehensive measure to capture loan quality. The unconditional loan loss takes both default probability and loss given default into consideration, and more closely represents the overall loan quality that matters to lenders. For example, a significant portion of defaults are later self-cured and result in no loan losses. While these loans incur no losses to the investors, using probability of default as the measure of loan performance would miss this point and would group these loans with loans that go into foreclosure. In addition, from an econometric perspective, as a continuous variable, loan loss helps reveal more information about loan quality than the dummy variable for the default event. As an example, if we assume that the source of default is a drop in property price, then default probability captures the threshold price level below which default is triggered while unconditional loan loss captures both the threshold price level and the actual drop in price below the threshold level.

To address the second concern, we adopt a structured approach as proposed by Agarwal et al. (2012) and Ambrose et al. (2005). The main idea is to infer the *ex ante* lenders' perceived loan quality, regardless of its securitization status. Specifically, the whole sample is divided into an estimation sample (75% of the whole sample) and a holdout sample (25% of the whole sample). Step I uses the estimation sample to estimate the loan loss equation (this is the first stage regression). The independent variables include only risk characteristics available at origination. Information available after the loan origination, such as the loan servicing and changes in housing market conditions, are excluded from the regression equation. The reason is that only information at origination is available to lenders when they form their ex ante loan quality estimations and make their securitization decisions. Note that the loan loss estimation equation excludes the securitization status as an independent variable. Step II applies the estimated coefficients from Step I to the holdout sample to form the out-of-sample predicted/expected loan losses. The expected loan loss is estimated regardless of the securitization status and represents a lender's rational expectation of loan loss at origination. Step III regresses the observed securitization status on the predicted/expected loan loss (this is the second stage regression). This main regression investigates how the expected loan loss affects a lender's securitization decision.

We also use portfolio loans only and securitized loans only to estimate the first stage regressions to address potential concerns about any servicing differences between sold loans and portfolio loans. Since mortgage servicing may have an impact on loan loss but is excluded from the first stage regression due to lack of data, the expected loss calculated in Step II may still carry securitization status information through inconsistent coefficient estimates from the pooled sample (the sample including both portfolio loans and securitized loans). This might happen in the case where both mortgage servicing and certain loan characteristics are correlated with securitization status. To address this issue, we estimate the expected loan loss as if all loans in the holdout sample were treated as portfolio loans, by using the coefficient estimates from the portfolio loans only first stage estimation. We also estimate the expected loan loss as if all loans in the holdout sample were treated as securitized loans by using the coefficient estimates from the securitized loans only first stage estimation. This further alleviates any concerns about variation in servicing treatments of the portfolio versus sold loans. The model construction thus helps alleviate the potential reverse causality between loan quality and securitization.[9]

The unconditional loan loss variable is left-censored since loan loss is observable only for liquidated loans. Otherwise, loan loss is censored at zero if the mortgage is current or still under the default/foreclosure process. However, loans in default or foreclosure process are likely to carry higher expected losses for the lender. For this reason, we also calculate the imputed loss to alleviate the censoring issue. The imputed loss is calculated as a linear interpolation between the first default and liquidation. We then conduct robustness checks by using the imputed loss in the first stage regressions.

To estimate the loan loss equation, we choose the commonly used Tobit model to deal with the censored dependent variable in the first stage regression.[10,11] The estimation

[9] There remains a possibility that loan contractual terms might induce reverse causality. When including only loan/borrower characteristics that the lender does not influence, our main results remain.

[10] As Tobit may be sensitive to model specifications, we conduct a robustness check by using OLS for the first stage regression.

[11] In the standard empirical framework in the literature, especially when dealing with binary mortgage outcomes like default or prepayment, the prevailing approach is to adopt the competing risk framework. However, the measurement of the mortgage outcome in this paper is different - it is a continuous variable representing loan loss. In this context, prepayment (and current) is associated with zero loan loss, while default is associated with a realized loan loss. This characteristic renders the Tobit model where zero loan loss is treated as censored more appropriate for loan loss regression than the competing risk framework that is typically used for binary outcomes.

equation is in Equation (1) below, where y_i, the dependent variable, is the loan loss. The variable y^* is a latent variable that has a linear relationship with independent variables x_i. Independent variables include loan characteristics at origination, state fixed effects, and closing quarter fixed effects. States have different foreclosure laws governing foreclosure procedures such as judicial versus non-judicial, foreclosure delay, and deficiency judgments. These differences in state foreclosure laws could have an impact on loan losses (Qi and Yang, 2009; Pennington-Cross, 2003; Kahn and Yavas, 1994, Zhu and Pace, 2021). The state fixed effects help control the potential differences in loan losses due to varying state foreclosure laws. Lending standards and housing markets change over time (Zhang, Ji, and Liu (2010)). All else equal, mortgages originated under a lax lending environment tend to incur higher loan losses. Thus, we include the loan closing quarter fixed effects to control the impact of changing lending standards. The error term follows a normal distribution $\varepsilon_i \sim N(0, \sigma^2)$.

$$y_i = \begin{cases} y_i^* \; if \; y_i^* = x_i'\beta + \varepsilon_i > 0 \\ 0 \; if \; y_i^* = x_i'\beta + \varepsilon_i \leq 0 \end{cases} \quad (1)$$

The coefficient estimates β from Equation (1) are then plugged into Equations (2) and (3) to calculate a lender's rationally expected loan loss for the holdout sample. $\Phi(\cdot)$ is the normal cumulative probability function. $\lambda(\cdot)$ is the Inverse Mill's ratio. $\phi(\cdot)$ is the normal probability density function.

$$E(y_i|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right) * (x_i'\beta + \sigma\lambda(x_i'\beta)) \quad (2)$$

where

$$\lambda(x_i'\beta) = \left(\frac{\phi(x_i'\beta)}{1-\Phi(x_i'\beta)}\right) \quad (3)$$

The second stage regression uses the Logit model to investigate whether a lender considers her expectation of loan loss in making the securitization decision.[12] The model specification is in Equation (4) and (5) below. As constructed, loan loss is a function of the control variables in the first stage regression. Hence, we omit those variables in the main regressions (see Agarwal et al., 2012; Jiang et al., 2014), other than those noted below. In addition to expected loan loss, a lender might also consider market yield information in making the securitization decision. Following the literature, we include various yield variables in the lender securitization regression (Agarwal et al., 2012). Yield spread measures the difference between the original mortgage coupon rate and the 10-year Treasury bond rate at origination. Credit spread is measured as the difference between the AAA bond index and the Baa bond index. The yield curve is defined as the ratio between the 10-year risk-free rate and the one-year risk-free rate. Interest rate volatility (Sigma Int) is estimated as the standard deviation of the one-year risk-free rate during the fifteen months before the origination. A jumbo loan dummy (Jumbo=1, Conforming=0) is also included as a control variable as jumbo loans can be sold only to the privately securitized market (confirming loans can be sold to both the private market and GSEs). Lenders might consider mortgage interest rate risk in making loan sale decisions, so an adjustable-rate mortgage dummy is included to capture a lender's loan sale preference in regard to interest rate risk.

$$Prob(Securtization = 1) = \left(\frac{e^{x'\beta}}{1+e^{x'\beta}}\right) \quad (4)$$

where

$$x'\beta = \beta_0 + \beta_1 * Expct\ Loss + \beta_2 * Controls \quad (5)$$

[12] We also conduct a robustness check by using OLS for the second stage regression.

2.2 Data and Sample

The main data source of this study comes from Black Knight Financial Services, Inc (BKFS). BKFS provided us with the McDash Core Data,[13] the McDash Property Module, and the McDash Resolution Module. We also utilize the treasury interest rates from the US Department of Treasury, the Corporate Bond Indexes from the S&P 500, house price indexes from FHFA, and lender information from the RealtyTrac data.

McDash Core Data includes residential mortgages serviced by nine out of the ten largest US mortgage servicers. This data set contains detailed mortgage-level information at origination, such as a borrower's credit score, loan-to-value ratio, documentation status, etc. The data set also reports the subsequent monthly loan activities, such as payment, default, foreclosure, etc. The McDash Property Module was created by BKFS utilizing their proprietary methodology matching the McDash Core Data with the nationwide county-level Recorder's data set. The Property Module reports the real estate transactions associated with both mortgage originations and terminations, such as the transaction dates and the transaction prices, etc. Monthly zip code-level house price index (HPI) updated property values are also reported in the Property Module. McDash Resolution Module tracks the transactions of foreclosed properties until liquidation. The Resolution Module was created by BKFS using their proprietary methodology by merging the McDash Core Data with the nationwide

[13] McDash Core Data was previously called LPS data, which was provided by LPS Applied Analytics. LPS Applied Analytics later was acquired by Black Knight Financial Service. The LPS data has been used for academic research such as Piskorski et al. (2010) and Agarwal et al. (2012).

county-level Recorder's data set. The McDash Resolution Module provides distressed property liquidation details such as liquidation prices and dates.

Mortgages are heterogeneous financial products. To reduce the heterogeneity, we restrict the mortgages included in our sample to conventional, single-family, first lien, new purchase loans with a mortgage term of either thirty or forty years. Our sample includes both portfolio loans and privately securitized loans (also called non-agency securitized loans). Loan origination time ranges from January 2005 to December 2006. We focus on mortgages originated from the beginning of the year 2005 since McDash data does not have comprehensive coverage before the year 2005 as some of the critical risk factors such as documentation status and debt-to-income ratio were not reported for the years prior to 2005. We include the loans originated by the end of the year 2006 since there was a structural change in the private mortgage securitization market starting from the beginning of the year 2007. The structural change of the private securitization market makes it difficult to identify lenders' original intentions with regard to the securitization decision of the mortgages (e.g., Kruger, 2016). To avoid potential data errors, mortgages are further limited to having the underlying property value between $5K and $1.5M, loan amount between $5K and $1.5M, and the original loan-to-value ratio lower than 1.5. We also require the observations to have valid values for each variable. To control survival bias, we require loans to enter the data set within four months of origination. We also exclude loans leaving the dataset for unknown reasons since mortgage outcomes for those loans cannot be properly measured.

To investigate the effect of securitization on loan quality, the intended securitization status, whether the mortgage is intentionally held on a bank's balance sheet or sold to investors, needs to be identified. We start with the final securitization status, either at liquidation or at the end of our sample period. Since a mortgage may end up retained by a bank on its book or sold to an investor for reasons other than the lender's original intention, we make several adjustments to ensure that the originally intended securitization status is correctly identified. First, a mortgage might fall into default too early to get securitized. These portfolio loans may not have been originated with the intention to be kept on the bank's books. Thus, we exclude mortgages that default within six months of origination for the main analysis. Second, securitized mortgages might be repurchased by the lender and get back into the portfolio pool due to the MBS warranty clauses. Given this concern, the repurchased mortgages are also excluded from the sample for the main analysis. Third, in order to capture the intention of securitization at origination, loans are required to be securitized within six months of loan origination. Although mortgages can be sold years after origination, those loans are likely to be substitution loans or put into the MBS pools to help with certain parameters. It is unlikely that those loans were originated with the intention to be securitized.

Loan performances are tracked until three, four, and five years after origination. The various tracking times serve as robustness checks.[14] If a loan is liquidated within the specified tracking time frame, loan loss rate is defined in Equation (7). If a mortgage is in current status, prepaid, or under the foreclosure procedure by the end of the specified time frame, loan loss is treated as zero. Since defaulted loans might carry loan losses before liquidation, as an alternative, we have also tracked loan performances until June 2016 (the

[14] Predictive regression models might carry look-ahead bias.

end of our data) and conducted the analysis by calculating and using imputed loan losses before liquidation for defaulted mortgages. The long tracking time helps alleviate the concern for censoring issue in loan losses. To avoid the results driven by the extreme values or some possible data errors, mortgages with the top and bottom 0.5 percentile of loan losses are excluded from the sample.

$$Loan\ Loss\ Rate = \frac{Outstanding\ Loan\ Balance - Liquidation\ Price}{Outstanding\ Loan\ Balance} \tag{7}$$

The outstanding loan balance is the unpaid loan balance at the time of default. The liquidation price is the final property sale price that is recorded at the County Recorder's office. Our measure of loan loss rate does not represent the total loan loss rate, nor does it include items such as legal fees, servicing fees, property maintenance costs, selling expenses, etc.[15] Although these other fees contribute to the total loan losses, they are not likely to be related to the initial loan quality perceived by the lender at origination. Since this paper uses loan loss to infer the expected loan quality, the measure of loss in (7) serves the purpose better than the total loss.

Loan characteristics at origination include borrower's credit score (scaled by 100), a low documentation dummy that equals one for loans with no or limited documentation, loan-to-value ratio, debt-to-income ratio, owner-occupied status, a second lien dummy that equals one for loans with junior liens, a jumbo loan dummy that equals one for mortgages with the purchase price higher than the OFHEO guideline for jumbo loans, and a loan term dummy

[15] McDash data does not provide information on these other costs.

(term30), which equals one for mortgages with a 30-year loan term. [16]

Table 1 reports the summary statistics of the loan characteristics at origination. Compared to securitized loans, portfolio loans on average have a higher credit score, a lower proportion of second liens, a higher proportion of loans with full documentation status, a lower debt-to-income ratio, and a lower proportion of adjustable rate loans. Securitized loans have a slightly lower loan-to-value ratio and a higher percentage of loans with a 30-year term.

The main motivation of this paper is to use loan loss, rather than default probability, as a comprehensive measure of loan quality to investigate lenders' securitization decisions. However, if loan loss and default probability are highly correlated and contain the same or similar amount of information, the contribution of the paper would be limited. Therefore, we first examine the correlations between default and loan loss. Although both measures represent credit risk, they are significantly different in several ways. The probability of default reflects only one aspect of loan performance. Once a mortgage enters into default status, the realized loss from the liquidated mortgage can vary dramatically from one loan to another, hence having different effects on investors' and/or banks' returns. Furthermore, it is possible that while loan A has a higher probability of default than loan B, the actual loss is bigger for loan B than loan A. In addition, a significant portion of defaults are later self-cured where defaulted borrowers catch up with late payments and exit delinquent status,

[16] Private mortgage insurance (PMI) can have an impact on mortgage loss and may affect lender's securitization decision. However, the variable is not well populated in McDash data. For example, more than 99 percent of subprime portfolio mortgages have a missing value of PMI. This makes it impossible to include PMI in subprime loan analysis. Prime loans have better PMI coverage. When including PMI in the first and/or second stage regressions, the results remain similar.

resulting in no loan losses. For example, Adelino, Gerardi and Willen (2013) report that self-cure rates were as high as 70 percent in 2006 and dropped to 25% in 2009. While these loans incur no losses to the investors, using the probability of default as the measure of loan performance would miss this point and would group these loans with loans that go into foreclosure.[17]

Table 2 reports the Pearson correlation coefficients between default/default probability and loan loss rate. Loan losses are tracked 36 months, 48 months, and 60 months after origination (Loss36m-36 months after origination, Loss48m-48 months after origination, and Loss60m-60 months after origination). Loan loss rates are calculated using Equation (7) for liquidated loans and equal to zero for current or non-liquidated loans. Defaults are tracked one year and two years after origination (Default12m-12 months after origination, and Default24-24 months after origination), as those are the most commonly used measures of credit risk in academic research. Loan losses are tracked a longer time after origination than default since there is a delay from default to liquidation. We also calculate the default probability within one and two years after origination by running the multinomial logit estimation regression (a competing risk model).[18] The correlations range from 0.229 (between default within 12 months of origination and loss rate tracked until the end of 60 months) to 0.505 (between default within 24 months after origination and loan loss rate within 36 months of origination). Our data shows that over 40 percent of the loans that defaulted within 12 months of origination returned to current status (defined as three consecutive months in current status) or prepaid the outstanding balance within 60 months

[17] It is also possible for a loan to incur losses without technically going into default, such as through short sales.
[18] The independent variables of the default regression are the same as the control variables in Table 4.

after origination. Over 13 percent of loans that remained in current status within 12 months of origination ended in involuntary liquidations by 60 months after origination. In addition, the loss rates of liquidated mortgages vary dramatically from 2.23% in the bottom one percentile to 72.23% in the top one percentile. These contribute to the low correlations. The low correlations between default and loss rate indicate that the information content of those two measures and their assessments of the quality of a loan are different, and hence the inferences about the effects of securitization drawn from these two proxies are not necessarily the same or consistent with each other.

Table 3 reports the unconditional loan loss rates for the full sample and the various sub samples. The unconditional loan loss rate reported here is much lower than the loss given default rate found in the previous studies (e.g., Qi and Yang, 2009). The reason is that the loss given default rate is calculated using only defaulted loans while the calculation of unconditional loan loss includes both defaulted/liquidated loans and current loans. As for the average loan loss rates, securitized loans exhibit higher loan loss rates than portfolio loans for the full sample and for each of the sub samples. For example, when loan performance is tracked until 60 months after origination, the full sample shows that securitized loans have an average loan loss rate of 6.9%, while portfolio loans have an average loan loss rate of only 2.5%. The pattern is consistent across different tracking periods. Prime loans and loans with higher credit scores exhibit lower losses than the corresponding subprime loans and loans with lower credit scores. The difference in loan loss rate between sold loans and portfolio loans seems to be larger for better quality loans than lower quality loans. For instance, when tracked 60 months after origination, for the high credit score sample, the loan loss rate of sold loans is 2.91 times (0.067/0.023=2.91) of the loss rate of portfolio loans. The loan loss rate of sold loans, for the low credit score sample, is only 1.84 times (0.079/0.043=1.84) the corresponding loan loss rate of portfolio loans. The difference in loss rate between sold loans and portfolio loans might come from differences in loan quality and/or servicing treatments. The following sections conduct regression analyses to investigate whether lenders securitize loans with different expected quality than those kept in their balance sheets.

3 Main Empirical Results

This section presents our main empirical results. Section 3.1 empirically investigates whether expected loan loss has an impact on a lender's securitization decision. Section 3.2 uses the same sample to compare the regression estimates using expected loan loss versus the probability of default and investigates why the information contents are different between those two measures. Section 3.3 explores some potential economic reasoning for the findings.

3.1 Expected Loan Loss and Securitization Decision

This section follows the structured approach discussed in Section 2.1 to study whether lenders choose to sell loans with different perceived quality than those kept on their books. If lenders are able to sell lower quality loans to investors, their motivation to lessen the lending standard and originate lower quality loans is likely to increase.

The full sample is divided into prime and subprime sub samples according to a lender's original classification that is based on the credit quality of the mortgage.[19] Prime loans meet certain underwriting standards and have better credit risk profiles, while subprime loans do not meet prime loan underwriting standards and carry higher credit risk. Prime loans and subprime loans represent different mortgage market segments. The buyers of prime mortgage-backed securities are typically institutional investors (e.g., pension funds) who are considered to be less financially sophisticated than MBS underwriters and often rely on rating agencies for the assessment of these securities. In contrast, subprime mortgage-backed securities are generally purchased by more sophisticated buyers (e.g., hedge funds) who scrutinize the detailed loan characteristics disclosed by the sellers in their purchase decisions.

As discussed in Section 2.1, securitization decision could affect loan loss, and the perceived loan quality could affect the securitization decision. Thus, the reduced form regression might yield biased coefficient estimates. To overcome this challenge, we construct the *expected* loan loss variable, which reflects a lender's rationally expected loan quality at origination, regardless of the securitization status of the mortgage. We first utilize the Tobit model to estimate the loan loss rate equation according to Equation (1). The loss rate estimation model uses the estimation sample that consists of a random 75% of the mortgages in the full sample. The dependent variable is the loan loss rate. If a mortgage is liquidated within n months after origination, the loss rate is calculated as defined in Equation (7) by using the outstanding loan balance at default and the liquidation sale price. Otherwise, if a mortgage remains in current status, default but is not yet in the foreclosure procedure, or is still under the liquidation procedure, the loan loss rate is treated as censored at zero,

[19] McDash data provides lenders' mortgage credit classification as either prime or subprime. There is no classification for Alt-A loans where the risk profile falls between prime and subprime.

meaning there is no realized loan loss yet by the end of the tracking time period. The explanatory variables include loan characteristics at origination such as a borrower's credit score, loan-to-value ratio, and documentation status. State fixed effects are included in all the regressions to control the impact of the different state foreclosure laws on loan losses. Origination quarter fixed effects are included to control the potential effect of the time-varying lending standards and housing market conditions on loan losses.

Table 4 reports the Tobit regression results for the loan loss equation. We report the coefficient estimates with the standard errors in parentheses. We track the loan performance 36, 48, and 60 months after origination, and report the corresponding regression results. Longer tracking time helps alleviate the censoring issue of loan losses. Regressions 1-3 report the results for prime mortgages with varying tracking times. Regressions 4-6 show the results for subprime mortgages with different loan performance tracking time frames.

The results show that, for both prime and subprime loans, higher loan loss rates are associated with risky loan features such as a lower credit score, a higher loan-to-value ratio, second lien status, a higher debt-to-income ratio, an adjustable interest rate, and a low documentation status. Jumbo loans carry lower loan loss rates except for subprime loans when tracking loan performance thirty-six months after origination.[20] Owner-occupied properties incur lower loan losses in all cases except for the prime loans tracking thirty-six months after origination. In sum, the results show that observably riskier loans incur higher loan losses.

Second, we apply the coefficient estimates from Table 4 to the remaining 25%

[20] One possible reason of jumbo loans lacking significance for subprime sample is that there are few observations of jumbo subprime loans.

holdout sample and use equations (2) and (3) to calculate the out-of-sample expected loan losses at origination. The expected loan losses incorporate only the information available at origination. Securitization status and loan servicing are not included in the first stage regression as explanatory variables. The estimated loan loss thus works as a proxy for a lender's rational expectation of loan quality, regardless of the securitization status of the mortgage.

Next, using the 25% holdout sample, we correlate the expected loss and securitization status and run the Logit model in Equation (4) to (5) to investigate a lender's securitization decision. The dependent variable is the securitization status, which equals one if a loan is privately securitized and equals zero if a loan is kept on a bank's own balance sheet. The independent variable of interest is the expected loan loss calculated in the previous steps. Other controls include a jumbo loan dummy, an ARM dummy, and yield variables as discussed in Section 2.1. We omit other control variables included in Table 4 for the second stage regression, as expected loss is a function of those variables (see Agarwal et al., 2012; Jiang et al., 2014). Table 5 Panel A reports the Logit regression results of expected loan losses on a lender's securitization decision. We report the coefficient estimates of expected loan loss with the standard errors in parenthesis. Similar to Table 4, we divide the sample into prime and subprime loans and track the loan performance for different time frames as robustness checks. Expected loan losses (Expct Loss) are estimated using the matching regression coefficients in Table 4. For example, the expected loan losses in regression 1 of Table 5 are estimated using the regression coefficient estimates from regression 1 of Table 4. For simplicity, we only report the results for the variable of interest, expected loan losses, from the second stage Logit regressions.

As a robustness check, Table 5 Panel B divides the full sample into high and low credit score subsamples as an alternative way to classify mortgages into high quality/easy-to-securitize and low loan quality/hard-to-securitize groups. Table 5 panel B reports the second stage Logit regression coefficient estimates and standard errors. The credit score cutoff line of 620 is chosen as mortgages with a credit score of 620 or higher are easier to get securitized than those with a credit score lower than 620 (Keys et al., 2010a). Other model specifications are the same as the regressions in Table 5 Panel A. Table 6 conducts additional robustness checks by further dividing the sample according to both loan quality and loan origination year.

The information available to the lender but not present in our data set represents unobservable loan characteristics in our analyses, and the unobservable characteristics might also affect a lender's securitization decision.[21] To further reduce heterogeneity associated with unobservable loan characteristics, Table 7 conducts robustness checks by using mortgages with full documentation status only. The reason to focus on full documentation mortgages is that previous studies found unobservable loan characteristics play a less important role in lender's screening for this kind of loan (e.g., Keys, et al, 2010a; Keys, et al, 2010b).

Focusing on prime loans, the results in Tables 5 to 7 consistently show that higher expected loan losses are associated with a higher probability of securitization. For the higher credit score sub sample and prime sub sample where loans have better quality and are easier

[21] Although, as long as the omitted loan characteristics are not correlated with the estimated expected loss, the coefficient estimate of expected loss should remain consistent.

to securitize, lenders choose to sell lower quality loans into securitization. The effect is mostly statistically and economically significant at a one percent level across different loan performance tracking times, different classifications of loan quality, and different loan origination years. As for marginal effects, according to Table 5, for prime loans, one standard deviation increase in expected loan loss increases the probability of securitization by 8.07 percent, 7.62 percent, and 6.56 percent, correspondingly to Regression 1 to 3. The marginal effects are calculated by holding other explanatory variables at their means. Since the majority of prime mortgage-backed securities received triple-A ratings during the sample time period, it is unlikely that the lower quality of sold prime loans is properly reflected in the sale price. Together, the results offer evidence of adverse selection that lenders choose to sell lower quality prime loans to the private securitization market and keep higher quality loans in their own portfolios.

Interestingly, for subprime loans, the overall results in Table 5 do not show any significant impact of expected loan loss on the probability of securitization. Loans with low credit scores show a marginally significant positive effect when tracking the performance 36 months after origination. The effect becomes insignificant when we divide the sample by loan origination year (Table 6). In fact, for the lower credit score subsample originated in 2006 (Table 6, Panel B), better quality subprime loans are sold into the secondary market. When we restrict the sample to full documentation loans to reduce heterogeneity (Table 7), subprime loans again show a negative relationship between expected loan loss and securitization.

We next perform additional robustness checks of model specifications for the main results in Tables 4 and 5. Regarding the origination times in the estimation sample versus the holdout sample, Agarwal et al. (2012) and Ambrose et al. (2005) employ different frameworks. Ambrose et al. (2005) adopt an "adaptive expectations" approach, where loans in the estimation sample were originated prior to those in the holdout sample. On the other hand, Agarwal et al. (2012) use a "rational expectations" approach, where loans in the estimation and the holdout samples were originated in the same time period.

The adaptive expectations approach might be more suitable in stable or slowly changing market conditions. In contrast, a rational expectations approach is deemed more reasonable during dramatically changing market conditions when prior experience may not be as informative. Given that the mortgage market in 2005 and 2006 is known to be atypical, we choose to follow Agarwal et al. (2012) and adopt the rational expectations approach. This assumes that lenders understand the contribution of loan characteristics to loan quality and that lenders' expected loan loss, as a measure of loan quality, is formed in the same way in both the estimation sample and the holdout sample.

Since loans originated in the years 2005 and 2006 share similarities, we group these loans in the estimation sample, as shown in Tables 4 and 5. As a robustness check, we further refine our analysis by considering loans originated in the same year in both the estimation sample and the holdout sample, and the results are presented in Table A.1. Additionally, Table 6 provides results separately for the years 2005 and 2006. In both Tables A.1 and 6, we apply the parameters estimated from loans originated in the same year to the holdout sample. The results remain robust across the various specifications.[22]

Since the timing of the securitization decision is also crucial for understanding the role of expected loan loss on the lender's securitization decision, we perform additional robustness checks by clustering standard errors at the origination quarter. The results are reported in Table A.1. Panel B. Across the various specifications of model standard errors, the results remain consistent with the previous findings.

In Table 5, the expected loan loss as a regressor is estimated from the first stage regression. Since generated regressors may cause the error-in-variable problem, we conduct robustness checks by using bootstrapped standard errors to mitigate this issue. The results are reported in Table A.2.

Lenders' tendency to securitize can vary and influence loan quality. For instance, consider a scenario where a lender adopts the originate-to-distribute model with lenient lending standards. In such cases, loans originated by this lender are more likely to be securitized, potentially leading to lower loan quality. To address this concern, it is necessary to identify the lender associated with each mortgage. Since McDash data lacks any financial institution information, we match the RealtyTrac data with the McDash data to acquire lender names and include lender fixed effects in the loan loss regressions.[23]

The time-varying local economic and housing conditions may influence the expected loan loss. As a robustness check, we consider the potential impact of these conditions on

[22] As a robustness check, we also perform the adaptive expectations approach by using loans originated in year 2005 in the estimation sample and loans originated in year 2006 in the holdout sample. The results remain consistent with our findings.

[23] The matching process is the same as in Yavas and Zhu (2023).

expected loan losses. To do so, we obtain the housing price index (HPI) at the Metropolitan Statistical Area (MSA) level from FHFA. We calculate house appreciation one year and three years prior to loan origination and include these two proxies of housing market conditions in the analysis.

For subprime loans, Keys et al. (2010) find that low-documentation loans with credit scores above 620 are more likely to be securitized compared to those with credit scores below 620. Additionally, full-documentation loans with credit scores above 580 are more likely to be securitized than loans with below 580 (Yavas and Zhu, 2023). To count for the FICO score threshold securitization effects, we perform robustness checks by including an additional credit score dummy in the analysis for subprime loans.

In Table A.3. Panel A, lender fixed effects are incorporated into the first stage regressions. Additionally, Panel B includes lender fixed effects, as well as house appreciation at the Metropolitan Statistical Area (MSA)-level one year and three years before loan origination in the first stage regressions. For subprime loans in Panel B, we introduce an additional FICO threshold dummy variable, taking the value of 1 if a borrower's credit score is greater than or equal to 620 or 580 for corresponding low or full documentation loans. Across the various specifications of model regressors, the main results consistently hold. As an additional robustness check, Table A.3 also reports the bootstrapped standard errors. Across the various model specifications, the results consistently align with the previous findings.

Our findings differ from the earlier results in the literature that use the probability of default as the proxy for loan quality. Among studies using a similar estimation methodology as in this paper, Agarwal et al. (2012) find no significant impact of expected loan quality on private securitization decisions, both for prime and subprime loans. Ambrose et al. (2005), focusing on prime loans, find that riskier prime loans are more likely to be retained by the lender. Thus, the two loan quality proxies, probability of default and unconditional loan losses yield significantly different results. The natural question is whether the differences in the results are due to differences in loan samples between our paper and the previous papers. In the next section, we address this concern by using the same mortgage samples to compare the results using expected loan loss versus probability of default. Additionally, we delve deeper into understanding the reasons behind the divergent information contents between these two measures.

3.2 Expected Loan Loss versus Probability of Default

To establish the significance of the loan quality proxy, we use the same mortgage samples in Table 7 and re-conduct the analysis using default probability as the measure of loan quality. The default probability is constructed using a Logit model with the same control variables as in Table 4, other than low documentation dummy. We focus on full documentation loans as in Table 7 to reduce the potential impact of unobservable characteristics. Default probabilities are calculated as the cumulative probability of default within 12 and 24 months after loan origination using a multinomial logit model.[24] The second

[24] Following the literature, including Agarwal et al. (2012) and Ambrose et al. (2005), we use cumulative default probabilities within 12 and 24 months after origination in Table 8. In addition to facilitating a better comparison with literature, this is also consistent with using loan performances after 36, 48 and 60 months after origination in Table 7 since loan losses are often not realized until one or more years after a loan goes into default.

stage regression results are reported in Table 8.

Comparing the coefficient estimates from Table 7 and Table 8, expected loan loss has a positive and significant impact on the probability of securitization for prime loans. The probability of default, on the other hand, yields a negative and significant impact on the securitization decision for prime loans, confirming the results reported in Ambrose et al. (2005). Similarly, for subprime loans, while the probability of default does not have a significant impact on the securitization decision, as in Agarwal et al. (2012), the expected loss yields a negative and significant result for two of the three subsamples. These comparisons establish that the two measures of default risk indeed produce very different results.

In Table 8, we extend the analysis of Ambrose et al. (2005) and Agarwal et al. (2012) to test the impact of the probability of default on a lender's securitization decision using the 620 FICO score as a threshold. A comparison of Panel A and Panel B of Table 8 uncovers an interesting result. It shows that when we focus on whether a loan is easier or harder to securitize using the FICO score cutoff line, rather than whether a loan is a prime loan or not, the probability of default yields opposite and significant estimates for better quality loans (FICO greater than 620 versus prime loans). Using expected loan loss, on the contrary, yields consistently negative coefficient estimates for the different classifications of better quality or easier-to-securitized loans, as shown in Table 7.

Expected loan loss (E(L)) and probability of default (P(D=1) both serve as measures of loan quality and are closely related to each other. Then, why do results vary when using alternative measures of loan quality? To illustrate the relationship between these two measures, we write out the following equation:

$$E(L) = E(L \mid D = 1) * P\ (D = 1) + E\ (L \mid D = 0) * P\ (D = 0)$$

Assume that $E\ (L \mid D = 0) = 0$, the above equation simplifies to:

$$E(L) = E(L \mid D=1) * P\ (D=1)$$

This equation reveals a positive correlation between expected loss and the probability of default, through the magnitude of E(L|D=1). If all defaulted loans incur the same loan loss rate (i.e., E(L|D =1) = constant), sorting loans by either probability of default or expected loan loss would yield the same results. The variation in results arises from the fact that not all defaults lead to the same loan loss rate. This understanding helps to interpret the difference in the results from different measures. For example, to reconcile the differences in results for prime loans reported in Panel A Tables 7 versus 8, all else equal, loans with higher E(L|D=1) but smaller P(D=1) need to be more actively securitized. Conversely, those with smaller E(L|D=1) and higher P(D=1) should be less securitized. Furthermore, since loans with higher E(L|D=1) are more actively securitized, assuming all else is the same, the disparity in results between the alternative measures is expected to be more pronounced among loans with higher E(L|D=1) than among loans with lower E(L|D=1).

To examine the relationship between alternative measures, we include both expected loss and probability of default into the securitization decision equation, as reported in Table 9. Expected loan losses (Expct Loss) are out-of-sample estimations derived using the coefficient estimates from the corresponding regressions in the first stage regressions with the same model specifications in Table 4. Loan loss rates are tracked 36 and 48 months after origination. Default probabilities (Default Prob) and prepayment probabilities (Prepay Prob) are out-of-sample estimations that represent cumulative default/prepayment probabilities within 12 or 24 months after origination. We include prepayment probability as a robustness check since prepayment risk could also affect lenders' securitization decisions (Agarwal, et al., 2012). These probabilities are calculated using the coefficient estimates from the corresponding first stage multinomial logit regressions (competing risk models) with the same set of control variables as the loan loss regressions in Table 4.

To further differentiate E(L|D=1) and explore whether the discrepancy between the two results from alternative measures of loan quality is more pronounced in the subsample with higher E(L|D=1), we introduce an interaction term, Expct Loss * LTV1PCT where LTV1PCT equals one if LTV is in the top one percent of the mortgage pricing brackets and zero otherwise. For example, for loans with LTV in the LTV pricing bracket 80.01% to 85%, LTV1PCT equals one if LTV is between 84% and 85% and zero otherwise. As detailed in the next section, loans with LTV in the top one percent of the mortgage pricing brackets are considered to have a similar default probability as other loans in the same pricing brackets but are likely to incur a higher loss rate due to higher appraisal inflation at loan origination (Calem, et. al., forthcoming and Diop, et. al., 2024). Therefore, the interaction term is included to investigate whether the disagreement between the two measures is more pronounced among loans with potentially higher ex ante loan loss rates.

Across various specifications, the findings in Table 9 consistently indicate that expected loan loss provides additional information beyond default probability. Moreover, as the previous analysis indicates, the disagreement in results is more pronounced among loans with higher ex ante expected loan loss.

3.3 Why Adverse Selection in Prime Loans and Lack of Adverse Selection in Subprime Loans

Our finding that lenders sell lower quality prime loans to MBS investors indicates adverse selection in prime loan securitization. However, we find no evidence that lenders sell lower quality subprime loans to MBS investors. This section offers some potential reasons why lenders choose to sell lower quality prime loans to the private securitization market but do not seem to be cherry-picking when it comes to subprime loans.

3.3.1 Why Adverse Selection in Prime Loans

Adverse selection occurs when asymmetric information is exploited. In our analysis, we infer lenders' perceived loan quality from expected loan loss, and expected loan loss is estimated from "observable" loan characteristics. However, asymmetric information can exist between lenders and investors even with regard to "observables." One reason is that prime loan investors are not likely to scrutinize loan characteristics of each individual loan in a mortgage pool. Because of the complex nature of mortgage-backed securities, investors, especially those unsophisticated prime loan investors (e.g., pension funds), typically rely on credit ratings to infer the riskiness of the securities and to price the securities correspondingly. However, during the pre-crisis period, rating agencies failed to rate many MBSs properly (Benmelech and Dlugosz, 2010; Ashcraft et al., 2011; Becker and Milbourn, 2011; and Griffin, 2019). As an example, more than fifty percent of the structured finance issues received the highest credit rating before the subprime crisis, which is obviously hard to justify by the later deterioration in MBS performances. Since rating agencies failed to distinguish between lower quality prime loans and higher quality prime loans, investors were not able to price in the differences in loan qualities among prime loans. This created an opportunity for lenders to sell prime loans with lower quality but still fit into a triple-A MBS deal, and keep the better quality ones in their books. In addition, the perceived default rate for loans originated in the years 2005 and 2006 was low even for lower quality prime loans, so lenders' reputational concerns were not likely to play a major role in the prime loan securitization decisions. One possible outcome of this result is that it induced lenders to loosen lending standards at origination and sell more low quality prime loans. This would lead to a decline in the overall quality of prime loans. Thus, lenders' loosened lending standards due to securitization could play a role in the poor performance of prime loans during the mortgage crisis. As our data set does not have the secondary market mortgage pricing information, we are not able to investigate how secondary mortgage pricing interacts with a lender's securitization decision. However, even if the secondary mortgage market pricing somewhat reflects the default risk of prime mortgages, if lenders are able to sell lower observable quality loans to investors, their motivation to lessen the lending standard and originate lower quality loans is still likely to increase, which may lead to a deteriorating quality of prime mortgage loans. When this happens on a large scale, it could contribute to the formation of a financial crisis.

Lenders can also enjoy cherry-picking opportunities by utilizing their superior knowledge and expertise in mortgage lending.[25] We next provide direct evidence of one such opportunity. It is well known that mortgage pricing and mortgage insurance premiums are a step function of the LTV ratio. For example, loans with an LTV ratio between 80.01% and 85% have the same interest rate and the same mortgage insurance premium, ceteris paribus. In other words, the credit risk is priced the same for all the LTV ratios in the 80.01%-85% bracket. However, loans with an LTV ratio closer to the upper limit of each bracket can be significantly riskier and carry higher loan loss than loans with a lower LTV ratio in the same pricing bracket (Calem, et. al., forthcoming and Diop, et. al., 2024). It is reasonable to assume that lenders understand this and may correspondingly cherry-pick loans with LTVs in the lower spectrum of each LTV pricing bracket to retain in their portfolios.

We use Fannie Mae's Loan-Level Price Adjustment Matrix (LLPA) for LTV pricing brackets and report the proportion of loans within the top one percent of the upper limit of each LTV pricing bracket in Table 10. For example, for the LTV pricing bracket of 80.01%-85%, loans within the top one percent of the upper limit have LTV ratios between 84% and 85%. Since LLPA is used for conforming loan pricing, we separate jumbo and conforming loan results. Table 10 shows that prime sold loans have a higher proportion of loans in the riskier upper one percent of each LTV pricing bracket than prime portfolio loans. The difference is more striking for conforming loans. Subprime loans seem to have the opposite pattern where sold subprime loans have a lower proportion of loans with an LTV ratio in the riskier top one percent of the LTV pricing bracket than subprime portfolio loans.

[25] For example, Adelino et al. (2019) find that mortgage performance improves with time to sale in seasoned privately securitized mortgage market. This is because sellers of high-quality mortgages have a lower cost of waiting to sell as they face lower probabilities of default, thus they can use time to sell as a signal of the quality of their mortgages.

Table 11 conducts Logit regressions to investigate whether securitized loans are associated with a different probability of having a mortgage with an LTV ratio in the upper one percent of LTV pricing brackets than otherwise similar portfolio loans. Loan-level controls are the same as in Table 4, other than the LTV ratio. Panel B adds in LTV pricing bracket dummies to have a within LTV pricing brackets comparison. The results consistently show that sold prime loans are significantly more likely to have loans with LTV ratios in the upper one percent of the corresponding LTV pricing bracket. Subprime sold loans do not show significant results. Since loans with LTVs in the upper one percent of the LTV pricing bracket have higher credit risk than otherwise similar loans in the same LTV pricing bracket, but have the same mortgage rates, the results provide supporting evidence of adverse selection in prime loans.

3.3.2 Why Lack of Adverse Selection in Subprime Loans

Interestingly, for subprime loans, we do not find any evidence of adverse selection. In fact, in some subsamples, lenders choose to sell better quality subprime loans and keep lower quality ones on their balance sheets. In general, subprime loans have lower credit quality and the default rate is relatively high. In addition, the Servicing and Pooling Agreement typically includes clauses that require lenders to repurchase loans that default soon after securitization. Since loans with extremely low quality, such as the lower quality subprime loans, are likely to default early, lenders may choose to sell relatively better observable quality subprime loans to investors. Compared with prime loans where about 37 percent (14097/(14097+24218)=36.79%) are kept in banks' balance sheets, only about 9 percent

(1864/(18389+1864)=9.2%) of subprime mortgages are kept as portfolio loans. Since lenders seem to securitize as many subprime loans as possible, another likely explanation is that the remaining 9 percent of subprime mortgages might be loans with very low quality that fail to meet the securitization parameters and are difficult to securitize. This explanation seems plausible since subprime mortgage-backed securities are generally purchased by more sophisticated buyers (e.g., hedge funds) who scrutinize the detailed loan characteristics disclosed by the sellers in their purchase decisions.[26]

Another possible explanation for the lack of adverse selection in subprime loans is related to mortgage servicing efficiency between portfolio loans and securitized loans. The valuation of mortgages could be different between investors and lenders due to their different abilities in loss mitigation. Previous studies document that delinquent securitized mortgages are more likely to enter into foreclosure while delinquent portfolio mortgages are more likely to have soft resolutions such as renegotiation and short sale (Agarwal et al., 2011; Piskorski, Seru, and Vig, 2010; Kruger, 2016). In Table 12, we report the percentage of loans that ended in a foreclosure sale one or two years after the first default. Panel A reports the results for loans defaulted within one year of origination and Panel B reports the results for loans defaulted within two years of origination. The pattern is clear and consistent with previous literature; once in default, sold loans are more likely to end up in a foreclosure sale than portfolio loans. The difference in servicing efficiency of defaulted loans could lead to different valuations of mortgages by the lenders and investors. This is especially true for

[26] Other financial institutional details such the prevalence of bank versus nonbanking lending in the prime versus subprime markets might also play a role in explaining different outcomes for prime and subprime loans. However, in exchange for making this data available for academic research, financial institution information has been excluded from the data set.

subprime loans, where loss mitigation effort is more likely. On the contrary, for prime loans, mortgage servicing efficiency is less likely to play a major role in valuation since the probability of a credit event is lower. When the loss mitigation efficiency differential leads investors' valuation of subprime loans to fall significantly below that of lenders, it benefits the lenders to keep the lower quality and undervalued subprime mortgages in their portfolios.

4 Robustness Checks

This section conducts additional robustness checks for our main results.

4.1 Mortgage Servicing

Agarwal et al. (2011) document that delinquent securitized mortgages are more likely to enter foreclosure while delinquent portfolio mortgages are more likely to have soft resolutions such as renegotiation and short sale. Different loan terminations could lead to different loan losses. The servicing treatment as a function of securitization status might lead to biased coefficient estimates in the first stage regression and the bias might convey securitization status information. To address this concern, we conduct robustness checks by using portfolio loans only or sold loans only in the first stage regressions. When using the portfolio loans only coefficient estimates from the first stage regression, the expected loss of the holdout sample is calculated as if all loans were treated as portfolio loans. Similarly, when using the sold loans only estimation from the first stage regression, the expected loss of the holdout sample is calculated as if all loans were treated as sold loans. This alleviates any concerns about variation in servicing treatments of portfolio versus sold loans with respect to expected loan loss. Table 13 reports the regression results. Except for Panel A using portfolio loans only, and Panel B using sold loans only for the first stage regressions, other model specifications are the same as in Tables 4 and 5. We report the results of the second stage regressions for prime loans and subprime loans. Loan performances are tracked for 36, 48, and 60 months after origination. The results are consistent with the previous findings: lenders sell lower quality prime loans but not subprime loans. This verifies that our results are robust to potential differences in servicing treatments due to securitization status.

4.2 Using Imputed Loss to Address Censoring Issue

To address the censoring issue of loan loss, we track loan performance until June 2016 (the end of our data) and calculate the imputed losses. The imputed losses are calculated as a linear interpolation between the first default and liquidation. We then use the imputed losses in the first stage regressions. Table A.4. Panel A in the appendix reports the second stage regression results where the first stage regression is a Tobit model. Since Tobit model could be sensitive to specifications, as an additional robustness check, Table A.4. Panel B reports the second stage regression results where we use OLS regression in the first stage. Other than that, the model specifications are the same as in Tables 4 and 5. For prime loans, the results across different specifications consistently show that higher loan losses are associated with a higher probability of securitization. For subprime loans, all the coefficients are statistically insignificant.

4.3 Future Market Conditions

This paper uses expected loan loss as a proxy of lenders' expectation of loan quality at origination. We estimate the loan loss equation by including loan and borrower characteristics available to lenders at origination as explanatory variables. Our analyses do not assume that lenders have ex ante knowledge of future market conditions or future policy changes, as those variables are excluded from the loan loss regressions as explanatory variables. Since future market conditions can affect realized loan loss (the dependent variable in the loan loss regression), those after-origination variables would be captured by the error term of the first stage loan loss regression. When those after-origination variables are correlated with securitization status, this might lead to biased coefficient estimates of the second stage securitization decision regression. However, as discussed in Section 4.1, using sold or portfolio loans only to estimate the loan loss equation ensures that future market conditions as omitted variables are not correlated with securitization status and thus should not bias the coefficient estimates in the second stage securitization decision regression. In other words, the robustness checks for mortgage servicing as reported in Table 13 also work as robustness checks for future market conditions.

Nonetheless, as another robustness check, we calculate the imputed loan loss as of the end of year 2007 before the major downturn of the housing market and use that as the loan loss measure in the first stage regression. This setting represents a relatively "normal" ex-ante market condition from lenders' perspective. Table A.5 reports the second stage regression results. The first stage regression uses portfolio loans only. As additional robustness checks to address the potential incidental parameter problem with nonlinear models (Lancaster, 2000; Angrist and Pischke, 2008), we report three sets of results in Table A.5: Tobit for the first stage and Logit for the main regression, OLS for the first stage and Logit for the main regression, and OLS for the first stage and OLS for the main regression. The results remain consistent with our previous findings.

4.4 Unobservable Loan Characteristics

Since our empirical model includes most of the commonly used credit risk characteristics in the first stage regression, the impact from unobservable loan characteristics on securitization should be relatively small. Tables 7 and 8 focus on full documentation loans to reduce the impact of unobservable. Nonetheless, if lenders use some loan characteristics not observed in the McDash data to predict loan losses, and if those unobservable loan characteristics are correlated with securitization status, this might incur bias in the second stage estimation. The direction of the bias from unobservable is determined by the product of the correlation between the omitted variable and loan loss, and the correlation between the omitted variable and securitization, CORR=Corr(omitted, loss)*Corr(omitted, securitization). If the impact of expected loss on securitization has the same sign from both unobservable and observables, then our conclusion of adverse selection based on observables will remain unchanged.

While we are not able to observe unobservable variables directly, we infer the likely sign of CORR through the reduced form regression. Table A.6 in the appendix reports the results of the reduced form regressions. Other than adding securitization status as an additional independent variable, other specifications are the same as the previous first stage regressions in Table 4. Since the model controls for observable loan characteristics, the securitization variable captures unobservable loan characteristics and/or servicing treatment associated with sold loans. The positive sign of securitization in Table A.6 indicates that unobservable loan characteristics/servicing treatment associated with securitization status are also associated with larger loan losses. In other words, the omitted variables (as a whole) move in the same direction with both securitization and loan loss (a positive CORR).[27] A positive CORR means that the expected loan loss from unobservable is associated with a higher probability of securitization. For prime loans, the previous analyses show that expected loss from observables increases the chance of securitization. Since expected loss from observables and unobservables have the same sign of impact on securitization, the conclusion of adverse selection drawn from observable loan characteristics is unlikely to reverse due to bias from omitted variables. For subprime loans, the omitted variables are likely to have a small impact as the reduced form regressions are not consistently significant.

4.5 Jumbo Loans

Non-jumbo loans can be sold to GSEs or the private securitization market. Jumbo loans, however, can only be sold to the private MBS pools. Focusing on jumbo loans helps eliminate any potential complications that might arise from the securitization channel. As an additional robustness check, Table A.7 in the appendix includes expected prepayment risk as an additional explanatory variable for the securitization decision equation. Table A.7 shows that jumbo loan results (with prepayment risk as an additional explanatory variable) remain consistent with our main findings.

[27] In case of multiple omitted variables, an aggregate variable or a linear combination of individual variables, moves in the same direction with both securitization and loss.

The various robustness checks presented above confirm our result that lenders sell lower quality prime loans into securitization, which could lead to a decrease in the overall prime market loan quality. Since prime mortgages constitute a large proportion of the overall mortgage market, our findings shed light on one potential channel through which loan quality deteriorated during the pre-crisis years. The results also highlight the importance of investigating prime and subprime mortgage markets separately.

5 Conclusion

Due to data constraints, earlier studies of adverse selection in securitization have used default probability as a proxy for loan quality. In this paper, we utilize a unique data set that allows us to adopt a more comprehensive proxy for loan quality, unconditional loan losses. Though the probability of default and loan losses both measure credit risk, they could capture mortgage performance quite differently. Loans with a higher probability of default may not experience larger losses at liquidation and vice versa. In addition, a significant portion of defaults were later self-cured by the borrowers and incurred no loan losses. Indeed, we find a low correlation between default and loss rate in our sample.

Our analysis of prime loans shows that higher expected loan losses are associated with a higher probability of securitization. Lenders sell prime loans with lower quality and keep higher quality loans on their books. We do not find consistent evidence of adverse selection for subprime loans. These results contradict earlier studies that use the probability of default as a proxy for default losses. The disparity in results between the alternative measures is more pronounced among loans with higher expected loss given default. These differences between our results and earlier results show that the choice of proper loan quality measure is critical as using expected loan loss alters some of the results in the literature in a significant way.

One implication of our prime loan results is that private securitization led to lower loan origination quality prior to the recent financial crisis through adverse selection. This is important as prime loans constitute a large proportion of the overall mortgage market. It is possible that some of the deterioration in observable loan quality of prime loans might be captured in the pricing of mortgages and mortgage-backed securities. However, even if the secondary mortgage market pricing properly reflects some of the credit risks of prime mortgages, the ability of lenders to sell lower observable quality loans to investors will give them incentives to lessen the lending standards and originate lower quality loans that they would not have otherwise. This adverse effect on loan origination quality is a crucial concern for the financial stability of the markets and for policymakers, regardless of whether the deterioration in loan quality is fully priced or not.

Non-agency issuance of residential mortgage-backed securities suffered a dramatic decline following the financial crisis. Although the private securitization market has been recovering slowly, the overwhelming majority of new mortgages are issued with government backing and the market share of GSEs is much greater now than it was before the financial crisis. In a July 2017 speech, the Fed chair Gerome Powell described the status quo as "unacceptable" and referred to the current situation as "may feel comfortable, but unsustainable," as it "leaves us with both potential taxpayer liability and systemic risk."[28] Given the political climate, we are likely to see major changes in the role of the GSEs and an increasing role of private securitization in residential mortgage markets in the coming years. Thus, a good understanding of the role of private securitization in the last financial crisis is likely to become even more critical.

[28] https://www.federalreserve.gov/newsevents/speech/powell20170706a.htm

References

Adelino, M., Gerardi, K., & Hartman-Glaser, B. (2019). Are lemons sold first? Dynamic signaling in the mortgage market. *Journal of Financial Economics*, *132*(1), 1-25.

Adelino, M., K. Gerardi, and P. S. Willen (2013). Why don't lenders renegotiate more home mortgages? Re-defaults, self-cures and securitization. *Journal of Monetary Economics 60*(7), 835–853.

Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphet, and D. Evanoff (2011). The role of securitization in mortgage renegotiation. *Journal of Financial Economics* 102, 559-578.

Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphet, T. Piskorski, and A. Seru (2017). Policy intervention in debt renegotiation: evidence from the Home Affordable Modifcation Program. *Journal of Political Economy* 125, 654-712.

Agarwal, S., Y. Chang, and A. Yavas (2012). Adverse selection in mortgage securitization. *Journal of Financial Economics 105*(3), 640–660.

Albertazzi, U., G. Eramo, L. Gambacorta, and C. Salleo (2015). Asymmetric information in securitization: An empirical assessment. *Journal of Monetary Economics 71*, 33–49.

Ambrose, B. W., J. Conklin, and J. Yoshida (2016). Credit rationing, income exaggeration, and adverse selection in the mortgage market. *The Journal of Finance 71*(6), 2637–2686.

Ambrose, B. W., M. LaCour-Little, and A. B. Sanders (2005). Does regulatory capital arbitrage, reputation, or asymmetric information drive securitization? *Journal of Financial Services Research 28*(1-3), 113–133.

Ambrose, B. W., A. B. Sanders, and A. Yavas (2016). Servicers and Mortgage-Backed Securities Default: Theory and Evidence. *Real Estate Economics*, 44: 462–489.

An, X, D. Deng and S.A. Gabriel (2011). Asymmetric information, adverse selection, and the pricing of CMBS**.** *Journal of Financial Economics* 100, 304-325

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics*. Princeton university press.

Ashcraft, A., P. Goldsmith-Pinkham, and J. Vickery (2011). MBS ratings and the mortgage credit boom. FRB of New York Staff Report 449. Federal Crisis Inquiry.

Barnichon, R., C. Matthes, and A. Ziegenbein (2018). The Financial Crisis at 10: Will We Ever Recover? FRBSF Economic Letter 2018-19.

Becker, B. and T. Milbourn (2011). How did increased competition affect credit ratings? *Journal of Financial Economics* 101(3):493-514.

Benmelech, E. and J. Dlugosz (2010). The credit rating crisis. *NBER Macroeconomics Annual 24*(1), 161-208.

Black, L., Krainer, J., & Nichols, J. (2020). Safe Collateral, Arm's-Length Credit: Evidence from the Commercial Real Estate Market. *Review of financial Studies.*

Bubb, R. and A. Kaufman (2014). S ecuritization and moral hazard: Evidence from credit scoring cutoff rules. *Journal of Monetary Economics 63*, 1-18.

Calem, P., C. Henderson, and J. Liles (2011). "Cherry Picking in Subprime Mortgage Securitizations: Which Subprime Mortgages Were Sold by Depository Institutions Prior to the Crisis of 2007?" *Journal of Housing Economics*, 20(2): 120-140.

Calem, P. S., Lambie-Hanson, L., & Nakamura, L. I. (forthcoming). Appraising home purchase appraisals, *Real Estate Economics*.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics 123*(1), 177–218.

DeMarzo, P. M. (2005). The pooling and tranching of securities: A model of informed intermediation. *The Review of Financial Studies*, *18*(1), 1-35.

Demyanyk, Y. and O. Van Hemert (2011). Understanding the subprime mortgage crisis. *Review of financial Studies 24*(6), 1848–1880.

DiNardo, J. and D. S. Lee (2004). Economic impacts of new unionization on private sector employers: 1984–2001. *The Quarterly Journal of Economics 119* (4), 1383–1441.

Diop, M., Yavas, A. and Zhu, S., 2024. Appraisal inflation and private mortgage securitization. Available at SSRN 3657595.

Downing, C., D. Jaffee and N. Wallace (2009). Is the market for mortgage-backed securities a market for lemons? *Review of Financial Studies*, 22, 2457-2494.

Drechsler, I., A. Savov and P. Schnabl (2019). How Monetary Policy Shaped the Housing Boom, NBER Working Paper, #25649.

Elul, R. (2016). Securitization and mortgage default. *Journal of Financial Services Research*, 49 (2-3), 281-309.

The Financial Crisis Inquiry Report: Final Report of the National Commission on the

Causes of the Financial and Economic Crisis in the United States (2011). *Washington, DC: U.S. Government Printing Office. January 2011*. ISBN 978-0-16-087727-8.

Griffin, J. M. (2019). Ten Years of Evidence: Was Fraud a Force in the Financial Crisis? (February 18). Available at SSRN: https://ssrn.com/abstract=3320979 or http://dx.doi.org/10.2139/ssrn.3320979

Government Accountability Office (2013). Financial Regulatory Reform: Financial Crisis Losses and Potential Impacts of the Dodd-Frank Act. Report to Congressional Requesters

Hwang, M., Miller, C., & Van Order, R. A. (2016). Effects of FHA loan limit increases by ESA 2008: Housing demand and adverse selection. *Available at SSRN 2789620*

Inside Mortgage Finance (2007). The 2007 Mortgage Market Statistical Annual.

Jiang, W., A. A. Nelson, and E. Vytlacil (2013). Securitization and loan performance: Ex ante and ex post relations in the mortgage market. *The Review of Financial Studies 27*(2), 454–483.

Kahn, C.M., and A. Yavas (1994). The economic role of foreclosures. *Journal of Real Estate Finance and Economics.* 8, 35–51.

Keys, B., T. Mukherjee, A. Seru, and V. Vig (2010a). Did securitization lead to lax screening? E vidence from subprime loans. *Quarterly Journal of Economics 125*, 307–362.

Keys, B., T. Mukherjee, A. Seru, and V. Vig (2010b). 620 FICO Take 2: Securitization and Screening. Working Paper.

Keys, B., A. Seru, and V. Vig (2012). Lender screening and the tole of securitization: Evidence from prime and subprime loans. *The Review of Financial Studies 25*, 2071–2108.

Krainer, J. and E. Laderman (2014). Mortgage loan securitization and relative loan performance. *Journal of Financial Services Research 45*(1), 39–66.

Kruger, S. A. (2016). The effect of mortgage securitization on foreclosure and modifications,working paper.

LaCour-Little, M., & Zhang, Y. (2014). Default probability and loss given default for home equity loans. *Office of the Comptroller of the Currency Economics Working Paper, 1*.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics, 95*(2), 391-413.

Mian, A. and A. Sufi (2009). The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *The Quarterly Journal of Economics 124*(4), 1449–1496.

Mian, A., & Sufi, A. (2019). Credit Supply and Housing Speculation. NBER Working Paper No. 24823

Parlour, C. A. and G. Plantin (2008). Loan sales and relationship banking. *The Journal of Finance 63* (3), 1291–1314.

Pennington-Cross, A. (2003). Subprime & prime mortgages: Loss distributions.

Piskorski, T., Seru, A., Vig, V. (2010). Securitization and distressed loan renegotiation: evidence from the subprime mortgage crisis. *Journal of Financial Economics* 97, 369-397.

Posner, E. A. and L. Zingales (2009). A loan modification approach to the housing crisis. *American Law and Economics Review,* 11, 575-607.

Purnanandam, A. (2011). Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies 24* (6), 1881–1915.

Qi, M. and X. Yang (2009). Loss given default of high loan-to-value residential mortgages. *Journal of Banking & Finance 33* (5), 788–799.

Yavas, A. and Zhu, S. (2023). Misreporting of Second Liens in Portfolio Mortgages and Privately Securitized Mortgages, *Real Estate Economics*, forthcoming

Zhang, Y., Ji, L., & Liu, F. (2010). *Local housing market cycle and loss given default: Evidence from sub-prime residential mortgages*. International Monetary Fund.

Zhu, S. and Pace, R.K., (2021). Statutory right of redemption and its influence on defaulted mortgage outcomes. *Journal of Banking & Finance, 124,* p.106058.

Table 1: Summary Statistics of Loan Characteristics at Origination

Variable	Portfolio Loans		Securitized Loans	
	Mean	Std Dev	Mean	Std Dev
FICO (in 100s)	7.094	0.631	6.898	0.671
LTV	0.807	0.134	0.786	0.098
Second Lien	0.176	0.381	0.222	0.416
Full Doc	0.712	0.453	0.575	0.494
Low Doc	0.288	0.453	0.425	0.494
DTI	0.339	0.133	0.372	0.124
Jumbo	0.284	0.451	0.299	0.458
Owner Occupied	0.813	0.390	0.800	0.400
Term30	0.872	0.334	0.887	0.316
ARM	0.582	0.493	0.726	0.446

Table 2: Pearson Correlation Coefficients between Default and Loan Loss Rate

Track Time	Loss36m	Loss48m	Loss60m
Default12m	0.321	0.257	0.229
Default24m	0.505	0.481	0.449
Default Prob12m	0.284	0.292	0.284
Default Prob24m	0.340	0.369	0.370

Notes: This table reports the Pearson correlation coefficients between default/default probability and loan loss rate. Loan losses are tracked 36 months, 48 months and 60 months after origination (Loss36m-36 months after origination, Loss48m-48 months after origination, and Loss60m-60 months after origination). Loan loss rates are calculated as $\frac{outstandingloanbalance - liquidationprice}{outstandingloanbalance}$. The outstanding loan balance is the unpaid loan balance at the time of default. Liquidation price is the final property sale price. Defaults are tracked one year and two years after origination (Default12m-12 months after origination, and Default24-24 months after origination). The default probability within one and two years after origination (Default Prob12m and Default Prob24m) are estimated from the Logit regression of default.

Table 3: Loan Loss Rates

Variable	Portfolio Loans Mean	Portfolio Loans Std Dev	Securitized Loans Mean	Securitized Loans Std Dev
Full Sample				
Loss36m	0.007	0.054	0.029	0.112
Loss48m	0.016	0.087	0.053	0.150
Loss60m	0.025	0.108	0.069	0.169
Prime Sample				
Loss36m	0.004	0.044	0.008	0.061
Loss48m	0.011	0.070	0.022	0.098
Loss60m	0.019	0.095	0.036	0.124
Subprime Sample				
Loss36m	0.025	0.101	0.056	0.152
Loss48m	0.060	0.157	0.094	0.192
Loss60m	0.072	0.172	0.112	0.207
FICO $\geq$ 620				
Loss36m	0.006	0.052	0.027	0.108
Loss48m	0.015	0.083	0.050	0.147
Loss60m	0.023	0.104	0.067	0.167
FICO $<$ 620				
Loss36m	0.012	0.074	0.041	0.131
Loss48m	0.032	0.121	0.065	0.166
Loss60m	0.043	0.141	0.079	0.180

Notes: This table reports the means and the standard deviations of the unconditional loan loss rates for the full sample and the various sub samples. Loan loss rate is calculated as $\frac{outstandingloanbalance - liquidationprice}{outstandingloanbalance}$. The outstanding loan balance is the unpaid loan balance at the time of default. Liquidation price is the final property sale price. Loan losses are tracked 36 months, 48 months and 60 months after origination (Loss36m-36 months after origination, Loss48m-48 months after origination, and Loss60m-60 months after origination).

Table 4: First Stage Tobit Regression of Loan Loss

	Prime Loans			Subprime Loans		
Variable	(1) 36m	(2) 48m	(3) 60m	(4) 36m	(5) 48m	(6) 60m
Intercept	−1.0698***	−0.6036***	−0.6224***	−0.9154***	−0.8261***	−0.7138***
	(0.2760)	(0.1700)	(0.1250)	(0.1312)	(0.0994)	(0.0864)
FICO	−0.3276***	−0.2992***	−0.2418***	−0.1222***	−0.0705***	−0.0488***
	(0.0159)	(0.0105)	(0.0080)	(0.0078)	(0.0060)	(0.0055)
LTV	1.6300***	1.4503***	1.3672***	1.0818***	0.8781***	0.8074***
	(0.1044)	(0.0653)	(0.0489)	(0.0556)	(0.0430)	(0.0388)
Second Lien	0.2203***	0.1796***	0.1750***	0.0385***	0.1121***	0.1294***
	(0.0163)	(0.0111)	(0.0088)	(0.0112)	(0.0089)	(0.0081)
DTI	0.2497***	0.2004***	0.2491***	0.3187***	0.3493***	0.3010***
	(0.0597)	(0.0380)	(0.0295)	(0.0396)	(0.0310)	(0.0283)
Low Doc	0.1906***	0.1926***	0.1507***	0.1470***	0.1291***	0.1190***
	(0.0163)	(0.0109)	(0.0084)	(0.0087)	(0.0069)	(0.0063)
ARM	0.4806***	0.4187***	0.3142***	0.2174***	0.1512***	0.0931***
	(0.0237)	(0.0148)	(0.0106)	(0.0177)	(0.0131)	(0.0114)
Jumbo	−0.2550***	−0.2183***	−0.1957***	0.0038	−0.0181**	−0.0353***
	(0.0175)	(0.0117)	(0.0091)	(0.0101)	(0.0081)	(0.0076)
Owner Occupied	0.0104	−0.0362***	−0.0450***	−0.2130***	−0.2419***	−0.2309***
	(0.0189)	(0.0122)	(0.0095)	(0.0090)	(0.0072)	(0.0067)
Term30	0.1104***	−0.0151	−0.0334**	−0.0879***	−0.1096***	−0.0911***
	(0.0251)	(0.0166)	(0.0136)	(0.0093)	(0.0074)	(0.0068)
Sigma	0.6974***	0.6608***	0.6252***	0.5437***	0.4992***	0.4821***
	(0.0134)	(0.0081)	(0.0059)	(0.0050)	(0.0036)	(0.0031)
State FE	Y	Y	Y	Y	Y	Y
Orig Quarter FE	Y	Y	Y	Y	Y	Y
R2	0.2881	0.2643	0.2441	0.1994	0.2027	0.1991
N	115069	115069	115069	60410	60410	60410

Notes: This table reports the coefficient estimates and the standard errors of the Tobit regressions (the first stage regression in Equation (1)). We report the results for the prime and the subprime loans separately. The dependent variable is loan loss rate. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 5: Lender's Securitization Decision - Second Stage Logit Regression Using Expected Loan Loss as the Proxy for Loan Quality

Panel A	Prime Loans			Subprime Loans		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	21.4790***	10.7628***	6.9773***	−0.1647	−0.7854	−0.6517
	(2.1759)	(1.1416)	(1.0720)	(0.9957)	(0.5928)	(0.5723)
R2	0.0478	0.0479	0.0459	0.1126	0.1129	0.1128
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864
Panel B	**FICO ≥ 620**			**FICO < 620**		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	15.7343***	9.2162***	7.3123***	4.3937*	0.6842	−0.0965
	(2.5471)	(1.0798)	(0.4739)	(2.3277)	(1.3786)	(1.0835)
R2	0.0481	0.0481	0.0469	0.1876	0.1860	0.1859
N Securitized	35848	35848	35848	6759	6759	6759
N Portfolio	14544	14544	14544	1417	1417	1417

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The dependent variable is the securitization status. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the corresponding regressions in Table 4. Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. Panel A reports the results of the prime versus subprime mortgages. Panel B reports the results of the high versus low credit score mortgages. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Standard errors are clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 6: Lender's Securitization Decision - Sub Sample Robustness Checks

Panel A	Prime		Subprime	
	OrigY 2005	OrigY 2006	OrigY 2005	OrigY 2006
		36m		
Expct Loss	63.0666***	15.0447***	−1.2331	−0.3816
	(10.1559)	(1.5104)	(1.6791)	(0.9311)
N Securitized	15811	8407	10739	7650
N Portfolio	6980	7117	1689	175
		48m		
Expct Loss	20.5925***	7.2770***	−1.0984	−0.7036
	(2.2382)	(0.9829)	(0.8384)	(0.6116)
N Securitized	15811	8407	10739	7650
N Portfolio	6980	7117	1689	175
		60m		
Expct Loss	11.6933***	4.4801***	−0.8210	−0.7940
	(1.2095)	(0.9074)	(0.7861)	(0.5724)
N Securitized	15811	8407	10739	7650
N Portfolio	6980	7117	1689	175

Panel B	FICO≥620		FICO<620	
	OrigY 2005	OrigY 2006	OrigY 2005	OrigY 2006
		36m		
Expct Loss	15.7010***	12.3605***	4.1559	−5.8311***
	(4.1512)	(1.3714)	(3.6308)	(2.1692)
N Securitized	22155	13414	4260	2499
N Portfolio	7766	6716	869	548
		48m		
Expct Loss	8.6093***	7.6013***	0.2078	−4.0449***
	(1.7069)	(0.5660)	(1.8786)	(1.2491)
N Securitized	22155	13414	4260	2499
N Portfolio	7766	6716	869	548
		60m		
Expct Loss	7.4030***	5.9742***	−0.2156	−3.4277***
	(0.9062)	(0.4928)	(1.5936)	(1.0914)
N Securitized	22155	13414	4260	2499
N Portfolio	7766	6716	869	548

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision) for the various sub samples. The dependent variable is the securitization status. Expected loan losses are the out-of-sample estimation using the coefficient estimates from the corresponding Tobit regressions (first stage regression). Other independent variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Standard errors are clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7: Lender's Securitization Decision - Using Expected Loan Loss as the Proxy for Loan Quality, Full Documentation Sample

Panel A	Prime Loans			Subprime Loans		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	12.0311***	5.3739***	2.6938*	−3.0011	−3.3327**	−2.7721*
	(2.7012)	(1.7633)	(1.4060)	(2.7587)	(1.5952)	(1.4556)
R2	0.0792	0.0787	0.0779	0.1089	0.1111	0.1109
N Securitized	13647	13647	13647	11036	11036	11036
N Portfolio	10487	10487	10487	863	863	863

Panel B	FICO ≥ 620			FICO < 620		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	19.0701***	9.9204***	6.2132***	3.6080	−0.1985	−0.8352
	(5.8306)	(2.5063)	(1.1894)	(2.8085)	(1.5729)	(1.2138)
R2	0.0679	0.0657	0.0619	0.2136	0.2127	0.2129
N Securitized	18682	18682	18682	6001	6001	6001
N Portfolio	10108	10108	10108	1242	1242	1242

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). Only loans with full documentation status are included in the analysis. The dependent variable is the securitization status. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the corresponding regressions in the corresponding first stage regressions. Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. Panel A reports the results of the prime versus subprime mortgages. Panel B reports the results of the high versus low credit score mortgages. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Standard errors are clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 8: Lender's Securitization Decision - Using Default Probability as the Proxy for Loan Quality, Full Documentation Sample

	Prime Loans		Subprime Loans	
	(1)	(2)	(3)	(4)
	12m	24m	12m	24m
Default Prob	-6.7280^{***}	-1.7466^{*}	0.6224	-0.5640
	(2.3125)	(0.9942)	(1.3720)	(0.9862)
R2	0.0797	0.0784	0.1085	0.1088
N Securitized	13647	13647	11036	11036
N Portfolio	10487	10487	863	863
Panel B	FICO $\geq$ 620		FICO $<$ 620	
	(1)	(2)	(3)	(4)
	12m	24m	12m	24m
Default Prob	11.1675^{***}	4.2068^{***}	1.4502	0.4705
	(1.3772)	(0.5694)	(1.2416)	(0.7121)
R2	0.0739	0.0717	0.2135	0.2132
N Securitized	18682	18682	6001	6001
N Portfolio	10108	10108	1242	1242

Notes: This table uses default probability as the proxy for loan quality and reports the coefficient estimates and the standard errors of the Logit regressions of the securitization decisions (the second stage regression). Only loans with full documentation status are included in the analysis. The dependent variable is the securitization status. Default probabilities are the out-of-sample estimations representing the cumulative default probabilities within 12 or 24 months after origination. The default probability is calculated using the coefficient estimates from the corresponding first stage default regressions. Other independent variables in the second stage regressions include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. The first stage default regressions use the multinomial logistic model (competing risk model) and have the same set of control variables as the loan loss regressions as in Table 4. Standard errors are clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 9: Lender's Securitization Decision - Expected Loan Loss vs Expected Default Probability

	D:12m L:36m		D:24m L:36m		D:24m L:48m	
	(1)	(2)	(3)	(4)	(5)	(6)
Expct Loss	22.3100***	20.4289***	32.5953***	31.2002***	23.9476***	23.1818***
	(5.5892)	(5.5841)	(9.2188)	(9.4943)	(4.8545)	(4.8245)
Expct Loss*LTV1PCT	12.0742**	12.6355**	11.9541**	12.3202**	7.5772**	7.7377**
	(5.7804)	(5.6609)	(5.7868)	(5.7377)	(3.3795)	(3.3025)
Default Prob	−15.1042***	−12.6445***	−6.7698***	−6.0591***	−10.0309***	−9.2553***
	(4.2036)	(4.0944)	(2.2834)	(2.3161)	(2.6676)	(2.7051)
Prepay Prob		6.4138***		2.2976***		2.2959***
		(1.6068)		(0.8843)		(0.8841)
R2	0.0893	0.1043	0.0905	0.0969	0.0959	0.1022

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The analysis includes only prime loans with full documentation status. The dependent variable is the securitization status. Expected loan losses (Expct Loss) are out-of-sample estimations derived using the coefficient estimates from the corresponding regressions in the first stage regressions with the same model specifications in Table 4. Loan loss rates are tracked 36 months and 48 months after origination. Default probabilities (Default Prob) and prepayment probabilities (Prepay Prob) are out-of-sample estimations that represent cumulative default/prepayment probabilities within 12 or 24 months after origination. These probabilities are calculated using the coefficient estimates from the corresponding first stage multinomial logit regressions (competing risk models) with the same set of control variables as the loan loss regressions in Table 4. LTV1PCT equals one if LTV is in the top one percent of the mortgage pricing bracket, and equals zero otherwise. For example, for loans with LTV in the LTV pricing bracket 80.01% to 85%, LTV1PCT equals 1 if LTV is between 84% and 85% and equals 0 otherwise. Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. Standard errors are clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 10: Proportion of Loans in the Top One Percent of Each Mortgage Pricing LTV Bracket

	Prime Loans		Subprime Loans	
Sample	Portfolio	Securitized	Portfolio	Securitized
Overall	0.461	0.588	0.717	0.664
NonJumbo	0.399	0.560	0.693	0.642
Jumbo	0.605	0.626	0.826	0.801

Notes: This table reports the proportion of loans that fall in the top one percent of each LTV bracket used in Fannie Mae Loan-Level Price Adjustment(LLPA) Matrix. For example, for the LTV bracket of 80.01% to 85%, loans in the top one percent of the bracket have LTV ratios between 84% and 85%.

Table 11: Securitization and LTV in the Top One Percent of the Mortgage Pricing LTV Brackets

Panel A	Prime Loans			Subprime Loans		
	Overall	Jumbo=0	Jumbo=1	Overall	Jumbo=0	Jumbo=1
LTV1PCT	0.1745***	0.2324***	0.0881***	0.0213	0.0286	0.0416
	(0.0119)	(0.0150)	(0.0202)	(0.0295)	(0.0298)	(0.0785)
LTV Brackets	N	N	N	N	N	N
Controls	Y	Y	Y	Y	Y	Y
R2	0.0510	0.0688	0.0603	0.0436	0.0295	0.0178
Panel B	**Prime Loan**			**Subprime Loan**		
	Overall	Jumbo=0	Jumbo=1	Overall	Jumbo=0	Jumbo=1
LTV1PCT	0.1202***	0.1598***	0.0629***	0.0213	0.0280	−0.0350
	(0.0133)	(0.0169)	(0.0222)	(0.0295)	(0.0315)	(0.0833)
LTV Brackets	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
R2	0.0863	0.1180	0.0662	0.0436	0.0479	0.0234

Notes: This table investigates securitization status and LTV being in the top one percent of the mortgage pricing brackets. The dependent variable is securitization status. LTV1PCT equals one if LTV is in the top one percent of the mortgage pricing bracket, and equals zero otherwise. For example, for loans with LTV in the LTV pricing bracket 80.01% to 85%, LTV1PCT equals 1 if LTV is between 84% and 85% and equals 0 otherwise. We report the Logit regression coefficient estimates and standard errors. Control variables in Panel A include the same loan characteristics as in Table 4, other than LTV ratio. Control variables in Panel B include the same loan characteristics as in Table 4, and dummy variables for each LTV pricing bracket. *· $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 12: Proportion of Defaulted Loans that Ended Up in a Foreclosure Sale

Sample	Prime Loans: Portfolio	Prime Loans: Securitized	Subprime Loans: Portfolio	Subprime Loans: Securitized
Panel A: Loans Defaulted within One Year of Origination				
Foreclosure Sale in 12M	0.173	0.287	0.267	0.4370
Foreclosure Sale in 24M	0.291	0.447	0.436	0.6052
Panel B: Loans Defaulted within Two Years of Origination				
Foreclosure Sale in 12M	0.133	0.325	0.319	0.4483
Foreclosure Sale in 24M	0.243	0.504	0.492	0.6122

Notes: This table reports the proportion of defaulted loans that ended up in a foreclosure sale within one year or two years after first default. Panel A includes loans defaulted within one year of origination. Panel B includes loans defaulted within two years of origination.

Table 13: Lender's Securitization Decision - Robustness Checks Using Portfolio Loans Only or Sold Loans Only for the First Stage Regression

Panel A	Expected Loss Based on the First Stage Estimation Using Portfolio Loans Only					
	Prime Loans			Subprime Loans		
	(1) 36m	(2) 48m	(3) 60m	(4) 36m	(5) 48m	(6) 60m
Expct Loss	7.3584***	4.5460***	2.4797**	−0.3423	−0.8598	−0.7195
	(2.4312)	(1.2408)	(1.0699)	(0.9466)	(0.5785)	(0.5600)
R2	0.0374	0.0385	0.0367	0.1126	0.1129	0.1129
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864

Panel B	Expected Loss Based on the First Stage Estimation Using Securitized Loans Only					
	Prime Loans			Subprime Loans		
	(1) 36m	(2) 48m	(3) 60m	(4) 36m	(5) 48m	(6) 60m
Expct Loss	19.0315***	9.9019***	5.8983***	−1.4745	−0.9809	−0.8130
	(2.9154)	(2.0150)	(1.4614)	(1.5073)	(0.6276)	(0.5288)
R2	0.0403	0.0389	0.0381	0.1128	0.1405	0.1344
N Securitized	24218	24218	24218	18389	17622	17893
N Portfolio	14097	14097	14097	1864	1812	1818

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The dependent variable is the securitization status. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the first stage loan loss equation. Panel A reports the results using the expected loss calculated from the first stage regression using portfolio loans only. Panel B reports the results using the expected loss based on the first stage regression using securitized loans only. Using portfolio loans only or sold loans only for the first stage estimation ensures that securitization does not enter into the expected loss through potential different servicing treatments between securitized loans versus portfolio loans. Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Standard errors are clustered by state. * p<0.10, ** p<0.05, *** p<0.01.

A Appendix

Table A.1: Lender's Securitization Decision - Robustness Checks Using Estimation Sample in the Same Year as the Holdout Sample

Panel A	Prime Loans			Subprime Loans		
	(1) 36m	(2) 48m	(3) 60m	(4) 36m	(5) 48m	(6) 60m
Expct Loss	20.2725***	10.6539***	6.8310***	0.8919	−0.6218	−0.5457
	(1.9840)	(1.1112)	(1.0546)	(1.0119)	(0.6460)	(0.6175)
R2	0.0478	0.0485	0.0464	0.1127	0.1146	0.1128
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864
Panel B	**FICO ≥ 620**			**FICO < 620**		
	(1) 36m	(2) 48m	(3) 60m	(4) 36m	(5) 48m	(6) 60m
Expct Loss	15.2245***	8.8767***	7.1222***	5.1798**	1.0626	0.4697
	(2.6690)	(1.0903)	(0.4837)	(2.1946)	(1.1984)	(1.0455)
R2	0.0482	0.0477	0.0470	0.1888	0.1862	0.1972
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The dependent variable is the securitization status. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the corresponding regressions in Table 4. The estimation sample and the holdout sample are in the same loan origination year (e.g., loans originated in year 2005 in the estimation sample are used to predict loan losses for loans originated in year 2005 in the holdout sample). to Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Panel A reports bootstrapped standard errors clustered by state. Panel B reports bootstrapped standard errors clustered by loan origination quarter.* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.2: Lender's Securitization Decision - Robustness Checks Using Bootstrapped Standard Errors

Panel A	Prime Loans			Subprime Loans		
	(1) 36m	(2) 48m	(3) 60m	(4) 36m	(5) 48m	(6) 60m
Expct Loss	21.4789***	10.7627***	6.9028***	−0.1620	−0.7881	−0.6517
	(3.7238)	(1.8007)	(1.8799)	(2.7460)	(1.2470)	(1.1760)
R2	0.0478	0.0479	0.0459	0.1126	0.1129	0.1128
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864
Panel B	**Prime Loans**			**Subprime Loans**		
	(1) 36m	(2) 48m	(3) 60m	(4) 36m	(5) 48m	(6) 60m
Expct Loss	21.4789***	10.7627***	6.9028***	−0.1620	−0.7881	−0.6517
	(5.0350)	(2.4009)	(1.3689)	(2.2994)	(1.1229)	(0.9535)
R2	0.0478	0.0479	0.0459	0.1126	0.1129	0.1128
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The dependent variable is the securitization status. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the corresponding regressions in Table 4. Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Panel A reports bootstrapped standard errors clustered by state. Panel B reports bootstrapped standard errors clustered by loan origination quarter.* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.3: Lender's Securitization Decision - Robustness Checks for Lender Effects, Housing Price Dynamics, and FICO Thresholds

Panel A	Prime Loans			Subprime Loans		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	20.2876***	11.7866***	7.2738***	−2.3896	−2.6780**	−2.4804**
	(3.7347)	(1.7602)	(1.5486)	(1.6883)	(1.0903)	(1.2235)
R2	0.0807	0.0841	0.0793	0.1172	0.1219	0.1216
N Securitized	6164	6173	6174	4807	4807	4807
N Portfolio	4219	4221	4221	335	335	335
Panel B	**Prime Loans**			**Subprime Loans**		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	20.3595***	11.6463***	7.1075***	−0.0254	−1.4820	−1.3014
	(3.1300)	(2.8218)	(1.3885)	(2.2134)	(1.4268)	(1.4868)
R2	0.0884	0.0921	0.0873	0.3815	0.3831	0.3828
N Securitized	6168	6169	6169	3049	3049	3049
N Portfolio	4139	4139	4139	243	243	243

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The dependent variable is the securitization status. In estimating the out-of-sample expected loan losses (Expct Loss), the first stage regressions in Panel A incorporate lender fixed effects alongside the independent variables outlined in Table 4. For the first stage regressions in Panel B, both lender fixed effects and housing price dynamics in preceding years (house price appreciation one year and three years prior to loan origination at MSA-level) are included, in addition to the independent variables specified in Table 4. The second stage regressions include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility as control variables. In Panel B, for subprime loans, an additional FICO threshold dummy variable is introduced, taking the value of 1 if a borrower's credit score is greater than or equal to 620 or 580 for corresponding low or full documentation loans. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Standard errors are bootstrapped and clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.4: Lender's Securitization Decision - Robustness Checks Using Imputed Loan Loss in the First Stage Regression

	Expected Loss Based on Imputed Loss in the First Stage Estimation - Tobit Model					
Panel A	Prime Loans			Subprime Loans		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	2.5710***	2.4711**	2.4241**	−0.0226	0.0047	0.0261
	(0.9822)	(0.9656)	(0.9577)	(0.7269)	(0.7204)	(0.7208)
R2	0.0383	0.0382	0.0381	0.1126	0.1126	0.1126
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864

	Expected Loss Based on Imputed Loss in the First Stage Estimation - OLS Model					
Panel B	Prime Loans			Subprime Loans		
	(1)	(2)	(3)	(4)	(5)	(6)
	36m	48m	60m	36m	48m	60m
Expct Loss	2.1993**	2.0673**	2.0091**	−0.3144	−0.1996	−0.1546
	(1.0289)	(1.0076)	(1.0006)	(0.7756)	(0.7768)	(0.7772)
R2	0.0364	0.0362	0.0361	0.1126	0.1126	0.1126
N Securitized	24218	24218	24218	18389	18389	18389
N Portfolio	14097	14097	14097	1864	1864	1864

Notes: This table reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The first stage loan loss estimation uses the imputed loan loss instead of realized loan losses. Panel A uses Tobit model in the first stage estimation and Panel B uses OLS model in the first stage estimation. The dependent variable is the securitization status. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the first stage loan loss equations. Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility.
Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. Standard errors are clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.5: Lender's Securitization Decision - Robustness Checks for Rational Expectations

	1:Tobit 2:Logit		1:OLS 2:Logit		1:OLS 2:OLS	
	Prime	Subprime	Prime	Subprime	Prime	Subprime
Expct Loss	3.8981**	−0.9153	8.5046***	−1.7533**	1.9165***	−0.0898*
	(1.8914)	(0.5843)	(2.3252)	(0.6816)	(0.5175)	(0.0540)
R2	0.0361	0.1352	0.0378	0.1145	0.0365	0.1923
N Securitized	24218	18389	24218	18389	24218	18389
N Portfolio	14097	1864	14097	1864	14097	1864

Notes: This table conducts robustness checks for the rational expectation assumption. We report the coefficient estimates and the standard errors of the second stage regression (lender's securitization decision). The first stage loan loss estimation uses the imputed loan loss by the end of year 2007 instead of loan losses tracking n months after origination. The first stage regression uses portfolio loans only. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the first stage loan loss equations. Other control variables include jumbo loan dummy, ARM, yield spread, credit spread, yield curve, and interest rate volatility. We report three sets of results: Tobit (first stage) and Logit (main regression), OLS (first stage) and Logit (main regression), and OLS (first stage) and OLS (main regression). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.6: Reduced Form Regression of Loan Loss

	Prime Loans			Subprime Loans		
	36m	48m	60m	36m	48m	60m
Securitization	0.0382**	0.0791***	0.1261***	-0.0123	0.0142	0.0293**
	(0.0182)	(0.0128)	(0.0101)	(0.0155)	(0.0133)	(0.0123)
Control	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Orig Quarter FE	Y	Y	Y	Y	Y	Y
N	217141	217141	217141	109900	109900	109900
R2	0.2283	0.2126	0.2047	0.1688	0.1743	0.1705

Notes: This table reports the coefficient estimates and the standard errors of the reduced form regression of loan loss. Other than including securitization status as an additional independent variable, other model specifications are the same as the first stage regression in Table 4. We report the results for the prime and subprime loans separately. The dependent variable is loan loss rate. Loan loss rates are tracked 36 months, 48 months, and 60 months after origination. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A.7: Jumbo Loan Results with Prepayment Risk as an additional Risk Factor

	Prime Loans			Subprime Loans		
	36m	48m	60m	36m	48m	60m
Expct Loss	22.7453***	11.2233***	9.1719***	0.1529	−3.1188***	−3.3039**
	(2.3756)	(1.2087)	(1.2979)	(0.9941)	(1.0078)	(1.2960)
Expct Prepay	0.0301	0.2496	0.6621	−1.0338**	−2.0004***	−2.1069***
	(0.4167)	(0.4137)	(0.5167)	(0.4061)	(0.5001)	(0.6106)
R2	0.0446	0.0445	0.0450	0.1127	0.1139	0.1137
N Securitized	10115	10115	10115	2579	2579	2579
N Portfolio	4123	4123	4123	330	330	330

Notes: This table reproduces Table 5 Panel A by focusing on jumbo loans and including predicted prepayment risk as an independent variable. It reports the coefficient estimates and the standard errors of the Logit regressions (the second stage regression in Equation (4) and (5) of the securitization decision). The dependent variable is the securitization status. Expected loan losses (Expct Loss) are the out-of-sample estimations using the coefficient estimates from the corresponding regressions in Table 4. Expected prepayment (Expct Prepay) is the out-of-sample estimation using the otherwise same model specifications as in Table 4 (the only two differences are that the dependent variable is prepayment status within n months after origination and the model is Logit.) Other control variables include ARM, yield spread, credit spread, yield curve, and interest rate volatility. Loan loss rates/prepayment are tracked 36 months, 48 months, and 60 months after origination. Standard errors are clustered by state. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.