Testing for Information Asymmetry in the Mortgage Servicing Market

Helmi Jedidi
Georges Dionne

March 2019

CIRRELT-2019-13

Testing for Information Asymmetry in the Mortgage Servicing Market

Helmi Jedidi, George Dionne*

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Department of Finance, HEC Montréal, and Canada Research Chain in Risk Management, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

Abstract. Our main objective is to test for evidence of information asymmetry in the mortgage servicing market. Does the sale of mortgage servicing rights (MSR) by the initial lender to a second servicing institution unveil any residual asymmetric information? We analyze the originator's selling choice of MSR using a large sample of U.S. mortgages that were privately securitized during the period of January 2000 to December 2013 (more than 5 million observations). Our econometric methodology is mainly non-parametric and the main test for the presence of information asymmetry is driven by kernel density estimation techniques (Su and Spindler, 2013). We also employ the non-parametric testing procedure of Chiappori and Salanié (2000). For robustness, we present parametric tests to corroborate our results after controlling for observable risk characteristics, for econometric misspecification error, and for endogeneity issues using instrumental variables. Our empirical results provide strong support for the presence of second-stage asymmetric information in the mortgage servicing market.

Keywords: Mortgage servicing market, mortgage servicing right, information asymmetry test, MSR-purchaser, parametric model, non-parametric model, kernel estimation, instrumental variable.

---

* Corresponding author: Georges.Dionne@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2019

I. INTRODUCTION

Rapid development of financial markets and advances in structured finance have enabled lenders to overcome the traditional lending scheme by removing mortgages they originate from their balance-sheets before the scheduled maturity through securitization. Securitization enables mortgage originators to sell mortgage-related cash flows to third-party investors in the form of liquid interest-bearing securities traded on financial markets (commonly known as mortgage-backed securities, MBSs). The two main advantages of securitization are to improve liquidity by converting long-term illiquid mortgages into highly tradable securities, and reduce regulatory capital requirements. The process of securitization involves numerous entities such as the special purpose vehicle (SPV), underwriter, credit enhancement entity, credit rating agency, and, more importantly for our research, the mortgage servicer.

Once the securitization process is achieved and the underlying MBSs are sold to investors, an important player intervenes, the mortgage servicer, who ensures the ongoing management and upkeep of interest payments. In general, the main task of a mortgage servicer is collecting principal and interest payments from borrowers and passing the proceeds on to the underlying MBS investors in the secondary market. These cash flows are passive claims linked to the pool of mortgages packaged by the SPV and held by MBS investors. Typically, the mortgage originator can act as the servicer of the deal by guaranteeing the connection of cash-flow streams between borrowers and MBS-investors. However, originators are also able to further reduce the borrower's default risk by selling the underlying mortgage servicing rights (MSRs) to a third party, hereafter referred to the MSR-purchaser or new servicer. In such case, the new servicer replaces the originator in ensuring ongoing mortgage management; borrowers become directly linked to the new servicer, to which they make monthly debt payments.

In case of borrower delinquency, the servicing cost of mortgages increases significantly as the servicer incurs additional costs related to managing these loans, which can significantly reduce the profitability of the servicing activity. For instance, the mortgage servicer is required to deploy additional resources to investigate and collect delinquent payments, to perform loss mitigation activities, or to manage a foreclosure process. The servicer is also required to advance payments to investors, insurers, and tax authorities, and may be required to pay third-party fees related to foreclosure proceedings. Mortgage servicers could incur additional significant costs related to unreimbursed foreclosure costs and real-estate owned losses. For these reasons, servicing inferior-quality mortgages could hinder the performance of mortgage servicing.

The main objective of this study is to test for evidence of information asymmetry in the mortgage servicing market. We answer the following main question: Does selling the mortgage servicing rights unveil any residual asymmetric information? In a typical principal-agent relationship, we hypothesize that the mortgage originator (the agent) possesses an informational advantage over the MSR-purchaser (the principal) in the market for mortgage servicing rights. This privileged information about both loan risk characteristics and borrower credit quality is collected at the time of the original underwriting, and the originator could have inducements to adversely exploit this information asymmetry.

Although a large body of theoretical and empirical literature has examined asymmetric information through the securitization process (see Ambrose et al. (2005), Keys et al. (2010, 2012), Agarwal et al. (2012), Malekan et al. (2014), Albertazzi et al. (2014) and Elul (2016), among many others), we are the first to investigate this second-stage asymmetric information. The above-mentioned studies focus on information asymmetry between lenders and investors at the first stage of securitization. The main research question for most studies that test for asymmetric information through the securitization process is investigating the originators' decision to securitize a given loan. For instance, most studies compare the ex-ante risk characteristics as well as the ex-post default likelihood of mortgages that the originator chooses to securitize *versus* those kept on its balance-sheet. In this study, we primarily focus on mortgages that have already been securitized. We consequently dig deeper in the data as we scrutinize these securitized mortgages to test for second-stage information asymmetry.

To empirically test for evidence of asymmetric information in the market for mortgage servicing rights, we analyze the originator's selling choice of MSR using a large sample of U.S. mortgages that were issued and privately securitized during the period of January 2000 to December 2013 (90 million loan-month observations). In the first step, we contrast the ex-ante risk profile of mortgages for which the originator chooses to sell the underlying servicing rights to a third party with those for which it chooses to hold and service. In the second step, we compare the ex-post default risk of these observably similar mortgages. Our econometric methodology is merely non-parametric in the sense that we do not make any restrictive assumptions about either the conditional distribution of the originator's MSR-selling decision or the functional form of the relationship between the decision to switch the mortgage servicer and the mortgage default risk. The main advantage of this methodology is that inferences about the distribution are made purely from the data, and the density estimation is thus more data-driven than it would be if the density function were constrained to fall in a given parametric family. Our methodology is inspired by the non-parametric tests of asymmetric information proposed by Su and Spindler (2013). The test is mainly driven by kernel density estimation techniques. We also employ the non-parametric testing procedure of Chiappori and Salanié (2000). To verify robustness, we present a battery of parametric tests to corroborate our results after controlling for observable risk characteristics, for econometric misspecification error, and for endogeneity issues using the instrumental variable estimation procedure.

Our empirical results provide strong support for the presence of second-stage asymmetric information in the mortgage servicing market. We obtain a significant positive association between lenders' decision to switch the servicer of the deal and the probability of mortgage default. For instance, our results show that the higher the likelihood of switching the mortgage servicer, the higher the probability that the borrower defaults.

Our evidence suggests that originating lenders are indeed taking advantage of privileged information about both loan risk characteristics and borrower credit quality they obtain at the time of the original underwriting. In this context, it is clear that asymmetric information influences the decision of mortgage originators to keep servicing mortgages they originate or to sell the underlying servicing rights to a third party. Two explanations based on contract theory remain possible. First, the originator could retain superior-quality loans with a low probability of default on its servicing portfolio and adversely sells lemons with high default risk; an outcome related to adverse selection. Alternatively, the transfer of mortgage servicing rights could reduce the originator's effort to screen applicants and monitor borrowers as soon as the underlying servicing rights have been planned to be sold to another servicer; an outcome related to moral hazard.

The remainder of this paper proceeds as follows. Section II introduces the servicing activity and briefly describes the income stream of mortgage servicers. We also present and discuss the main risks that mortgage servicers encounter. We introduce the non-parametric kernel density estimation techniques in section III and present the proposed non-parametric information asymmetry test in section IV. Section V describes the data as well as the variables used in our study. Section VI reports the main empirical results of the non-parametric testing procedure. For robustness, we also report the results of commonly used parametric tests. Section VII concludes the paper.

II. OVERVIEW OF THE MORTGAGE SERVICING TASK AND VALUE

II.1 Representation of mortgage servicing process

Figure 1 shows the various contracted parties involved in the mortgage lending process along with the generated cash flows in every step.

A typical mortgage lending process starts with a borrower applying for a mortgage in order to buy a property or to refinance an existing mortgage to take advantage of lower interest payments. The originator is the financial institution that makes the mortgage lending transaction with the borrower. Usually, the mortgage originator is a commercial bank, a credit union, or a non-depository retail lender. Whatever the case, the mortgage originator administers the complete loan-granting process. Based on its information set, the originating lender expends effort to assess the borrower's reliability and creditworthiness. Eventually, if the borrower meets the lending requirements, the mortgage application is approved and funds are released as represented by cash flow ***1*** in Figure 1.

Obviously, the borrower is required to repay the loan principal and interest as scheduled. The debt payments in absence of securitization are represented by cash flow ***2***.

In this traditional lending scheme the originator bears all risks directly associated with its lending activity, mainly the borrower's default risk as long as the mortgage amount appears on its balance-sheet. Following the development of financial markets and advances in structured finance, this is no longer the only potential relationship. Nowadays, many newly originated mortgages are removed from the originator's balance-sheet and sold in the secondary financial market in the form of mortgage-backed securities (MBSs) through securitization. This activity is defined as the process whereby illiquid loans extended to borrowers are converted into liquid securities traded on financial markets. This process is summarized in steps ***3*** to ***6***.

[Figure 1 about here]

In the first step, the originating institution transfers the mortgage to a special purpose vehicle (hereafter referred to as SPV) defined as a legally separate entity created to handle the securitization process. The mortgage transfer is marked by cash flow ***3***. The SPV packages the illiquid mortgages and transforms them into liquid securities. This process of handling securitization involves external parties such as the underwriter that assists with the sale, the credit enhancement agency, and the credit rating agency that rates the interest-bearing securities. Once the tradable MBS are created and rated, the SPV sells them to investors, as depicted in cash flows ***4*** to ***5***. Finally, the SPV uses the proceeds of the MBSs sale to pay back the entity that originated the underlying debt, as illustrated by cash flow ***6***.

Once the securitization process is finalized and the underlying MBSs are sold to investors, the mortgage servicer ensures ongoing management and the upkeep of the payments. The main task of a mortgage servicer is collecting principal and interest payments from the borrower (cash flow ***7***) and passing the proceeds along to the underlying MBS investors in the secondary market (cash flow ***8***). These cash flows are claims that are linked to the pool of loans packaged by the SPV and that are held by MBS investors in the secondary market. They are passive in the sense that the underwriting decision has already been made. Thus, as the borrower makes interest and principal payments, the servicer of the deal ensures that the cash flows are paid back to investors in accordance with the terms laid out in the securities prospectus.

The mortgage originator can act as the servicer of the deal by guaranteeing the connection of cash-flow streams between the borrower and the MBS-investors, or it can sell the Mortgage Servicing Rights (MSR) to a third party involved in this process, hereafter referred to as the new servicer or the MSR purchaser.

Should the mortgage originator choose to sell the underlying MSR of the mortgage to a new servicer, the sale of mortgage servicing rights and the corresponding cash proceeds are indicated by cash flows ***9*** and ***10***, respectively. In this case, the buyer of the mortgage servicing rights replaces the original servicer of the deal and ensures the ongoing mortgage management. Therefore, borrowers become directly linked to the new servicer to whom they continue making monthly debt payments (cash flow ***11***) that the former passes along to the MBS investors in the secondary market, as indicated by cash flow ***12***. Customarily, in return for these services, the new mortgage servicer is paid a monthly fee generally specified as a fixed percentage of the declining unpaid balance of the underlying mortgage loan. The servicer is also entitled to collect other fees such as float income, late payment fees, and other ancillary income. All these income streams are represented by cash flow ***13***. Finally, if a delinquent borrower defaults on a loan and stops making monthly payments due to financial distress, the mortgage servicer is required to advance funds to MBS-investors in the secondary market in keeping with the terms and conditions of the loan servicing contract, as indicated by cash flow ***14***.

At this point, it is crucial to note that neither the new servicer nor the investors in the secondary market observe all the background information on the borrower's application and report, of which the mortgage seller is hypothesized to take advantage. As stated by Keys et al. (2010), only the hard information about the borrower (*e.g.* FICO score) and the contractual terms (*e.g.* loan balance, initial interest rate, initial term) are observed by investors that buy these loans as part of a securitized pool. The rest of soft information (*e.g.* measures of future income and employment stability for the borrower, how many years of documentation were provided by the borrower, marital status and the joint income status, etc.) are kept private by the originating lender. We suppose that the soft information is also kept private by the originating lender when selling the MSR to a second servicing institution.

In this environment, information asymmetry theory suggests that if one assumes that the originating lender is better informed about the borrower's credit quality than are the purchasers of the securitized debt and the new servicer, then the originating lender may have incentives to exploit this informational advantage and pass on lemons to the new servicer of the deal and retain higher-quality loans in its servicing portfolios; the outcome of an adverse selection problem (JFE). However, the information asymmetry can also decrease the originating lenders' incentives to spend efforts to screen applicants and monitor borrowers for the mortgages to be securitized and sold to MBS investors. The underlying servicing rights of these loans will be sold to the new servicer; an outcome of a moral hazard problem (Keys). In this research, we do not separate the two information problems because we do not have access to a dynamic relationship between the services, as in Abbring et al. (2003) and Dionne et al. (2013) for repeated insurance contracting. Our data does not permit to separate the two information problems as in Dionne et al. (2013) or Keys et al. (2011, 2012).

II.2 Cash flows and risks of the servicer[1]

The most important source of revenue for a mortgage servicer is the servicing fee, generally specified as a fixed percentage of the declining unpaid balance of the underlying mortgage. The servicing fee currently collected by servicers is typically earned per active-loan per month. Therefore, servicers do not collect servicing fee revenue for non-performing loans for which borrowers are delinquent. The current compensation structure was established in the 1980s, in conjunction with the boom in the mortgage securitization market, and has not been changed since then. Another potential source of valuable income for a mortgage servicer is the interest earned on principal and interest, and tax and

[1] The contents of this section are based on the works of FitzGerald (2016), Hernandez et al. (2015), and Federal Reserve Board (2016).

insurance escrows collected and held by the servicer before distribution. The value of this income to the servicer largely depends on the opportunity costs of funds, which in turn depends on the current short-term interest rate. Finally, the servicer may also collect ancillary fees in the form of late fees, transfer fees, loan modification fees, and various other miscellaneous fees.

The mortgage servicer incurs a variety of expenses associated with the servicing activity, which include the direct cost-to-service as well as delinquency and foreclosure costs. The former expenses consist of the basic costs of operating a business of servicing. In the case of delinquency or default, the cost-to-service of a given mortgage increases significantly as the servicer incurs additional costs related to managing these loans, which can significantly reduce the profitability of the servicing activity. For instance, if a mortgage is delinquent, the mortgage servicer will be required to deploy additional resources to investigate and collect delinquent payments, to perform loss mitigation activities, or to manage the foreclosure process. Most importantly, the servicer may be required to advance payments to investors, insurers, and tax authorities, and may be also required to pay third-party fees related to foreclosure proceedings. Lastly, a servicer will incur additional significant costs related to unreimbursed foreclosure costs and real-estate owned losses.

There are three main risks associated with the mortgage servicing activity: prepayment risk, default risk, and operational risk. The prepayment risk is defined as the possibility of an early unscheduled full repayment of the loan. The default risk is defined as the hazard that a borrower will be unable to honor the required principal and/or interest payments on the mortgage agreement in a timely manner. Typically, the default risk is closely related to the quality of mortgage underwriting, as well as to macroeconomic conditions and to house market conditions.

Consequently, the default risk is known to have a significant effect on the profitability of mortgage servicing activity: if a borrower’s ability to make monthly payments is impaired, the mortgage servicer’s income stream vanishes. Undoubtedly, the cost of financing the advance could be exorbitant if the number of delinquent mortgages in the servicer's portfolio surges. The risks due to the operational side of the servicing business are of an entirely different nature than the above-mentioned risks associated with prepayment and default rates. For example, in servicing the deal there is the possibility that the initial mortgage was made based on fraudulent information. In this study, we focus on loan default risks.

III. KERNEL ESTIMATION FRAMEWORK

III.1 Motivation for non-parametric model

Consider any continuous random variable, X^c, that has a probability density function f^c. Suppose that we have a random sample of *n* observed data points, $\{X_i\}_{i=1}^{n}$, that we assume to be drawn from an unknown probability density function. The main goal of the density estimation framework is the construction of an estimate of the density function from a given dataset, which we do not possess.

The non-parametric approach is widely known as distribution-free because we do not assume any specific distributional form for the data. As a result, inferences about the distribution are made purely from the data. Although we will be assuming that the distribution has a deterministic probability density f^c, the estimation of f^c will be entirely data-driven in the sense that the data will be allowed to speak for themselves, more than would be the case if f^c were constrained to fall in a given parametric family. Obviously, the primary advantage of the non-parametric approach is its robustness; it can be applied in a broader range of situations even where the parametric conditions of validity are not met. A second advantage of the non-parametric approach is that it can be applied using small sizes of data points. For instance, using parametric methods could deliver misleading results if coupled with a very small sample of data that does not meet the sample size guidelines and for which one might not be able to properly ascertain the distribution of the data. Another notable advantage of the non-parametric approach is its ability to handle various data types (*e.g*. continuous, ordinal, and ranked data) even if measured imprecisely or if the data comprise outliers, anomalies widely recognized to seriously affect the routine of parametric tests. Below we describe the most important non-parametric method of estimating density functions, namely *kernel density estimation*.

III.2 Multivariate kernel density estimation with mixed data types

Consider again the randomly drawn sample of a continuous random variable X^c composed of *n* data points, $\{X_i^c\}_{i=1}^n$. Technically, a kernel is defined as a weighting function that weights the observations X_i^c in the sample based on their distance from a specific value *x*, usually referred to as the *smoothing point*, within a fixed range known as the *bandwidth*, *h*. The weights given by the kernel function to the observations in the sample are known as the *local weights*. The kernel density estimator is basically calculated as the sample average of the local weights that are given by the kernel function for all the data points in the sample.

The multivariate kernel density function for a given data type is estimated by using the product of the univariate kernel functions. Therefore, for *q* continuous variables, the estimator of the multivariate density function takes the following form:

$$\hat{f}(x) = \hat{f}(x_1, x_2, \dots, x_q) = \frac{1}{n(\hat{h}_1 \dots \hat{h}_q)} \sum_{i=1}^{n} \prod_{s=1}^{q} K\left(\frac{X_{i,s} - x_s}{\hat{h}_s}\right) \qquad \mathbf{(1)}$$

where $K(\cdot)$ denotes a kernel weighting function.

Similarly, the estimator of the multivariate density function of *p* discrete variables is represented as follows:

$$\hat{f}(x) = \hat{f}(x_1, x_2, \dots, x_p) = \frac{1}{n} \sum_{i=1}^{n} \prod_{r=1}^{p} l(X_{i,r}, x_r, \hat{\gamma}_r) \qquad \mathbf{(2)}$$

where $l(\cdot)$ is a weighting kernel function that depends on the estimated bandwidth $(\hat{\gamma}_r)$.

An estimation framework that involves a mixture of continuous and discrete variables using the non-parametric kernel density estimation technique is widely known as *mixed data types kernel estimation framework*. Examples of works that have contributed to the development of the non-parametric estimator are presented by Ahmad and Cerrito (1994), Racine (2008), and Li and Racine (2007, 2008).

Formally, for a multivariate density function including *q* continuous variables and *p* discrete variables, the general form of the kernel density estimator is merely the product of the univariate kernel functions of the mixed-type variables in the model. The general form could be represented as follows:

$$\hat{f}(x) = \hat{f}(x_1, \ldots, x_q, x_1, \ldots, x_p) = \frac{1}{n(\hat{h}_1 \ldots \hat{h}_q)} \sum_{i=1}^{n} \prod_{s=1}^{q} K\left(\frac{X_{i,s} - x_s}{\hat{h}_s}\right) . \prod_{r=1}^{p} l(X_{i,r}, x_r, \hat{\gamma}_r) \quad \textbf{(3)}$$

In practice, the mixed data-type kernel estimation framework enlarges the applications of the non-parametric estimation techniques. For instance, most of the topics that researchers investigate involve a mixture of discrete and continuous variables. Moreover, the mixed kernel allows to have a non-parametric counterpart for the discrete choice models like probit, logit, multinomial logit, or ordered logit.

III.3 Bandwidth selection for kernel density estimators

It is well known that the performance of kernel density estimators depends crucially on the value of the smoothing parameter or the bandwidth (denoted *h* for the continuous variable kernel and γ for the discrete variable kernel).

The optimum value for the bandwidth is the value minimizing the *integrated mean square error*, or IMSE, simply defined as a measure of the discrepancy between the estimated density $\hat{f}_h$ and the true density *f*. The IMSE can be expressed as follows:

$$IMSE_{\hat{f}_h(x)} = E\left\{\int \left(\hat{f}_h(y) - f(y)\right)^2 dy\right\}$$

$$= \int Biais(\hat{f}_h(y))^2 dy + \int Var(\hat{f}_h(y)) dy \qquad \textbf{(4)}$$

The optimal bandwidth is a function of the second derivative of the true density, which is unknown in the model. The bandwidth approximation methods use some underlying assumptions about the true density, which are useful in an application with a large number of variables or large sample size.

III.4 Multivariate conditional kernel density estimation

The core of our information asymmetry test is the estimation of the conditional density function. Let *y* be the vector of the values of a mixed-type random variable, $y = \left\{y_1^c, \dots, y_{q_y}^c; y_1^d, \dots, y_{p_y}^d\right\}$, and *x* be the vector of the values taken by another random variable with mixed data type, $x = \left\{x_1^c, \dots, x_{q_x}^c; x_1^d \dots, x_{p_x}^d\right\}$. Then, the conditional kernel density estimator for random variable *y*, given values in *x*, $\hat{f}_{Y|X}(y|x)$ takes the following form:

$$\hat{f}_{Y|X}(y|x) = \frac{\frac{1}{n\left(\hat{h}_1 \dots \hat{h}_q\right)} \sum_{i=1}^{n} \prod_{s=1}^{q} K\left(\frac{Z_{i,s}^c - z_s^c}{\hat{h}_s}\right) . \prod_{r=1}^{p} l\left(Z_{i,r}^d, z_r^d, \hat{\gamma}_r\right)}{\frac{1}{n\left(\hat{h}_1 \dots \hat{h}_q\right)} \sum_{i=1}^{n} \prod_{s=1}^{q_x} K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right) . \prod_{r=1}^{p_x} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right)} \qquad (\mathbf{5})$$

where *z(.)* denotes the variables *y(.)* and *x(.)* in the joint density function, as $z^c = \left\{y_1^c, y_2^c, \dots, y_{q_y}^c, x_1^c, x_2^c, \dots, x_{q_x}^c\right\}$ and $z^d = \left\{y_1^d, y_2^d, \dots, y_{p_y}^d, x_1^d, x_2^d, \dots, x_{p_x}^d\right\}$.

For the purpose of our testing procedure, we use the Nadaraya-Watson (Nadaraya, 1965; Watson, 1964) kernel regression to estimate the conditional distribution function. The general expression takes the following form:

$$\hat{F}_{Y|X}(y^c|x) = \frac{\frac{1}{n\left(\hat{h}_1 \dots \hat{h}_q\right)} \sum_{i=1}^{n} . I(Y_i^c \leq y^c) . \prod_{s=1}^{q} K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right) . \prod_{r=1}^{p} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right)}{\frac{1}{n\left(\hat{h}_1 \dots \hat{h}_q\right)} \sum_{i=1}^{n} \prod_{s=1}^{q} K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right) . \prod_{r=1}^{p} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right)} \quad (\mathbf{6})$$

where $I(Y_i^c \leq y^c)$ denotes an indicator function.

The last two equations represent the core of the non-parametric test described by Su and Spindler (2013). In the next section, we provide a detailed description of the testing procedure and the hypotheses to be tested as well as their intuition.

IV. NON-PARAMETRIC INFORMATION ASYMMETRY TEST

We hypothesize that mortgage originators are better informed than both new servicers and investors in the secondary market because they possess privileged information about loan risk characteristics and borrower credit quality obtained at the time of original underwriting. We also posit that this informational advantage influences the behavior of mortgage originators either by selling the underlying MSRs of lemons to new servicers or by reducing their efforts to screen and monitor applicants when they anticipate transferring mortgage servicing.[2]

First, we contrast the ex-ante risk profile of mortgages for which the lender has ceased servicing and sold the underlying servicing rights to a new servicer with the loans that the lender choses to retain be held in its servicing portfolio. Second, we compare the ex-post default likelihood of these observably similar mortgages. In particular, we verify whether the mortgages in our sample experience a higher default rate if the originator decides to sell the underlying MSRs.

Formally, let *Y* denote the dependent variable under study, *X* the set of independent variables, and *Z* the decision variable. In our context of switching the servicer of the deal of securitized mortgages, *Y* refers to the default event on a given mortgage, *X* a vector of exogenous variables that encompass both loan risk characteristics and borrower credit quality observed by the originator, and *Z* stands for the originator's decision to sell the mortgage servicing rights rather than keep servicing the mortgage. According to the theory of asymmetric information, the decision variable, *Z*, should provide no additional information if and only if the prediction of *F(Y)* given *X* and *Z* jointly coincides with its prediction given *X* alone (Dionne et al. 2001). Formally, we can write the following expression in terms of conditional probability functions:

$$F(Y/X,Z) = F(Y/X) \text{ or}$$

[2] There is a complementary behavior that may explain our results. According to Levitin and Twomey (2011), there may be a principal-agent conflict between third-party mortgage servicers and MBS investors. Since a new servicer has no interest in the loan performance, his decision to foreclose or renegotiate a loan is mainly related to its own benefit and cost payoff, which seems to favor foreclosure.

$$Pr(Default_i/Charac_i, Switch_i) = Pr(Default_i/Charac_i) \qquad \textbf{(7)}$$

where *Pr(Default$_i$)* refers to the probability of default on mortgage *i*. *Charac$_i$* refers to the set of observable risk characteristics for loan *i* (*e.g.* borrower FICO score, loan amount, interest rate, Loan-To-Value ratio, payment type, ...). Finally, *Switch$_i$* denotes a dummy variable that equals 1 if the originator of loan *i* chooses to cease servicing the deal and sells the servicing right to another mortgage servicer and 0 if it chooses to preserve the servicing task of the mortgage *i*.

In other words, equation (7) means that the original lender's decision to switch the servicer or to continue servicing the deal do not convey any additional information in predicting the probability of default on the mortgage, as long as all loan and borrower observable risk characteristics are taken into account.

Given a set of randomly drawn observations $\{Y_i, Z_i, X_i^c, X_i^d\}_{i=1}^n$, the non-parametric test is mainly based on comparing the following two conditional CDF estimates: $F(y/x^c, x^d, z)$ and $F(y/x^c, x^d)$. The conditional CDF, $F(\cdot \mid \cdot)$, is estimated using the local constant Nadaraya-Watson method, as represented in equation (6), augmented by our decision variable *Z* as follows.

$$\hat{F}_{Y|X,Z}(y|x^c, x^d, z) = \frac{\frac{1}{n(\hat{h}_1 \dots \hat{h}_q)} \sum_{i=1}^n . I(Y_i \leq z). \prod_{s=1}^q K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right). \prod_{r=1}^p l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right). I(Z_i \leq z)}{\frac{1}{n(\hat{h}_1 \dots \hat{h}_q)} \sum_{i=1}^n \prod_{s=1}^q K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right). \prod_{r=1}^p l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right). I(Z_i \leq z)} \qquad \textbf{(8)}$$

We then measure the variation in $F(y/x^c, x^d, z)$ across different values of *z* and different observations as follows:

$$D_n = \sum_{i=1}^n \left[\hat{F}_{Y|X,Z}\left(y_i \middle| x_i^c, x_i^d, z_i = 1\right) - \hat{F}_{Y|X,Z}\left(y_i \middle| x_i^c, x_i^d, z_i = 0\right)\right]^2 . a(x_i^c) \qquad \textbf{(9)}$$

where a(·) is a uniformly bounded nonnegative weight function with compact support X^c that lies within the support of X_i^c.This serves to perform trimming in areas of sparse support. It can be expressed as follows:

$$a(x_i^c) = \prod_{j=1}^{p_c} I\left(q_j(0.025) \leq X_{ij}^c \leq q_j(0.975)\right) \qquad (\mathbf{10})$$

where $\mathrm{q_j}(\alpha)$ denotes the α^{th} sample quantile of $\mathrm{X_{ij}^c}$ and p_c is the total number of continuous variables.

We can compute the test statistic D_n as described in equation (9). The test statistic could be viewed as the difference between the expected probability of default depending on whether the originator switches servicers or not. Su and Spindler (2013) show that D_n is asymptotically normally distributed under the null hypothesis of independence. The authors also demonstrate that the test statistic, after being appropriately recentered and scaled, is asymptotically distributed as $N(0,1)$ under the null hypothesis. We can implement a bootstrap procedure to obtain the corresponding test *p*-values.

V. DATA, VARIABLES, AND SUMMARY STATISTICS

V.1 Data source and sample construction

To empirically test for asymmetric information in the mortgage servicing market, we use a large data set provided by *MBSData, LLC*. The data comprise U.S. mortgages that were securitized through the non-agency channel. Mortgages securitized through the private-label channel have fundamental risk characteristics that make them riskier-than-average. In general, mortgages securitized through this channel do not conform to the prudent lending guidelines set by the government-sponsored enterprises *Freddie Mac*, *Fannie Mae* and *Ginnie Mae*. For example, most of the mortgages do not meet the GSE requirements in terms of loan size (*e.g.* jumbo loans with original loan amount exceeding the conforming loan limits), documentation (*e.g.* loans with no or low level of documentation) and, loan-to-value ratios (*e.g.* LTV ratio above 80%).

Owing to the lack of a government guarantee, holding these private-label securities carries a significantly higher risk than carrying the agency counterparts. For instance, without government back-up, private-label mortgage originators rely on both credit rating agencies and credit enhancements to attract MBS investors and convince them that the underlying mortgage is safe.

The non-agency market had witnessed tremendous growth during the pre-crisis period. For instance, the outstanding quantity of non-agency mortgages grew from roughly $600 billion at the end of 2003 to $2.2 trillion at its peak in 2007, according to JP Morgan (2010). This tremendous growth in the non-agency market is widely recognized by both researchers and practitioners as being one of the main triggers of the financial crisis.

Our dataset consists of mortgages issued between January 2000 and December 2013. The initial sample consists of more than 25 million mortgages that were originated throughout the U.S. The mortgages are granted by diverse types of lenders ranging from top investment banks to regional small retailers. The yearly distribution of loan origination follows a pattern similar to that observed in the entire U.S. mortgage market.

The *MBSData*, *LLC* database consists of two main datasets. A static file reporting detailed information collected at the time of loan origination and a dynamic file reporting monthly-updated information of the loan. The first static file provides detailed information on the mortgage, the borrower, as well as the house securing the loan. For instance, it reports the borrower's FICO credit score and its Debt-To-Income (DTI) ratio as measures of creditworthiness and indebtedness, respectively. The dataset also reports detailed loan-level information such as loan amount, Loan-To-Value (LTV) ratio, loan purpose, payment type, initial interest rate, private insurance coverage, and prepayment penalty. The information regarding the property backing the loan includes house value, address, state and zip code. For the originator identity, the database reports both the lender's name and type, the original servicer of the deal as well as the most recently updated servicer name. All this information, contained in a static file, is recorded at the time of original underwriting.

The second dataset consists of historical files that include data that have been collected over the loan lifetime on a monthly basis. The key variables recorded in the monthly remittance files are: current loan balance, current interest rate, scheduled principal and interest, next due date, and more importantly, a monthly delinquency code indicator compiled according to the methodologies of both the Office of Thrift Supervision (OTS) and the Mortgage Banker Association (MBA). The delinquency codes include: current, paid-off, +30, +60, or +90 days delinquent, in foreclosure, in bankruptcy, or real-estate-owned (REO).

The dataset also provides information on losses and loan modification. Loss files mainly report loan-level loss amount, loss severity, recovery amount, loan liquidation proceeds, and current value at liquidation. Loan modification datasets report the modification type, modified loan amount, pre- and post-modification interest rates, term modification, deferred payment period schedules and the modification effective date.

While constructing our sample, we impose several restrictions in order to create a homogenous loan sample. For instance, we focus on mortgages in a first-lien position on the property securing the mortgage and exclude second mortgages and home equity lines of credit (HELOCs). Our choice is primarily motivated by the fact that first-lien mortgages have priority over all other subsequent claims (*i.e.* second-lien or junior) on a property in the event of borrower default. We restrict our attention to single-family owner-occupied homes and exclude multifamily and/or non-owner occupied properties. We also exclude loans whose main purpose is designated as home improvement, and retain loans with the main purpose identified as a house purchase or refinancing an existing mortgage (both cash-out and no cash-out). We also exclude planned unit developments (PUDs) and mobile homes. All these restrictions result in a final sample including 5,591,353 distinct mortgages originated between January 2000 and December 2013 and tracked until December 2015 on a monthly basis.

V.2 Variables and hypotheses

The main variable of interest in our empirical analysis is the mortgage servicer switching indicator denoted, as *Switch_Servicer*. This variable is a dummy indicator taking the value of one if the originating lender decides to sell the mortgage servicing right to another servicer and zero if the lender decides to continue servicing the loan it originates. The second most important variable of interest is the *Default* dummy variable, which denotes whether a given mortgage becomes 90+ days delinquent or is in default (*i.e*. when a loan is first reported as the borrower having missed three or more consecutive monthly payments), as in Agarwal et al. (2012) and Keys et al. (2010) used 60+ days delinquent. We will also report results with 60+ days delinquent. As discussed in the literature, this definition of default is considered to be a relatively “early” definition, compared with foreclosure or bankruptcy, which usually occur several months later. See for example Ambrose et al. (2005), Krainer and Laderman (2009), and Elul (2011), among others, who investigate the originator’s decision to securitize. In line with the literature, we adopt the standard 90+ definition of default to avoid the ambiguity of differences in state laws governing foreclosure, which are widely recognized as having an effect on the length of time it takes to conclude a foreclosure.

The set of covariates includes several explanatory variables that measure the risk characteristics of the borrower and the mortgage, all recorded at the time of origination. All variables are presented in Table A1 of the Appendix. The first variable we consider is the borrower’s FICO score. In general, the FICO score measures individuals’ creditworthiness by taking into account their payment history, length of credit history, current level of indebtedness, and types of credit used. The score ranges between 300 and 850. Typically, a FICO score above 660 is indicative of a good credit history. We expect that the originator will keep servicing mortgages with high FICO scores.

The second independent variable is the Loan-To-Value ratio, *LTV*, calculated by lenders as the percentage of the first-lien mortgage to the total appraised value of the purchased property. The LTV ratio is one of the key risk factors used by U.S. lenders when qualifying borrowers for a mortgage. In the United States, mortgagors with LTV ratios higher than 80% are required to buy private mortgage insurance to protect the lender from the default risk, which increases the cost of borrowing. The LTV ratio also measures the equity stake of borrowers in a given property. The higher the LTV ratio, the lower the down-payment, so the lower the borrower’s equity stake in that house. Therefore, because a higher LTV ratio mirrors a risky mortgage, where the borrower holds a lower equity stake in a given house, we expect the lender’s decision to switch the servicer of the deal will be positively correlated with the LTV ratio.

Another key explanatory variable in our analysis is *No/Low documentation*, a dummy variable indicating whether the lender has collected the required level of documentation on the borrower. As discussed above, the borrower is asked to fill out a credit application and provide a number of statements and proofs of employment status and income when applying for a loan. Based on this documentation, the lender expends effort to assess the borrower's creditworthiness. Therefore, a no/low-documentation loan is a loan for which the lender has not gathered a sufficient level of information on the borrower's income. In terms of default risk, there is no reason to presume that no/low-documentation loans will default more frequently than full-documentation mortgages, because it is not a direct measure of the credit risk of a given loan.

The next independent variable is the ARM indicator; ARM stands for Adjustable-Rate Mortgages (commonly referred to as variable-rate mortgages). The ARM variable indicates whether the interest rate paid on the outstanding balance of a given mortgage varies according to a specific benchmark. Usually, the initial interest rate is fixed for a period of time, after which it is reset periodically, often every month. The interest rate paid by the borrower is usually based on a benchmark plus an additional spread, called the ARM margin. In terms of risk, ARM-type mortgages transfer part of the interest-rate risk from the lender to the borrower. Indeed, these mortgages are generally used when interest rates fluctuate and are difficult to predict (which make fixed-rate mortgages, FRMs, difficult to obtain). In terms of servicing choice, a positive statistical relationship is expected between terms of servicing choice and interest rate.

We also include a conforming indicator as an explanatory variable that denotes loans with characteristics that obey the GSEs' (Fannie Mae and Freddie Mac) lending guidelines. The *GSE_conforming* dummy variable indicates whether the mortgage was eligible to be sold to the GSEs at origination. Following the GSEs' recommendations[3], we classify a mortgage as conforming if the borrower's FICO score is above 660 and the loan amount

[3] For details about the GSE classification, please refer to the Federal Reserve Bank of St. Louis website. The document "What Is Subprime Lending?" can be viewed at: https://files.stlouisfed.org/files/htdocs/publications/es/07/ES0713.pdf. For additional details on the lending guidance, please see: www.federalreserve.gov/boarddocs/press/bcreg/2007/20070302/default.htm.

is below the conforming loan limit in place at time of origination and the LTV is either less than 80 percent or the loan has private mortgage insurance if the LTV is greater than 80 percent. Given that conforming loans meet the GSE lending standards, we expect a negative correlation with the default event. Indeed, falling within the GSE prudence guidelines should significantly reduce the probability of default. Regarding the choice of switching the servicer, we presume that both signs are plausible. On the one hand, being GSE-conforming increases the ease of finding a buyer of the underlying MSRs. For instance, because these loans are originated following the GSE standards, it would be easier to find buyers of the securitized pool of loans in the market. Thus, a positive sign is expected. On the other hand, being GSE-conforming increases the probability that the lender will be paid back as scheduled. Lenders may therefore keep these good-quality loans on their balance-sheets because the risk of default on these loans is significantly low. Therefore, the sign of the conforming coefficient is an empirical matter.

V.3 Descriptive statistics

We start the empirical analysis by providing summary statistics of some of the key variables used in our analysis. Because we are focusing on the non-agency market, we pay special attention to the role of credit scores, loan-to-value ratios, amount of documentation collected by the lender, and some interest rate features. Table 1 reports descriptive statistics for the sample of mortgages during the entire study period from January 2000 to December 2013. It also reports summary statistics segmented by origination year. Table 2 breaks down the sample by payment type (FRM vs. ARM), loan type (Prime vs. Subprime), before/after financial crisis, default status and switching servicer status.

The first two columns of Table 1 provide a comprehensive picture of the evolution of the non-agency segment of the mortgage market over the 14-year study period. At first glance, the table shows that mortgage origination has witnessed two major trends explained by the financial crisis. First, the market expanded rapidly from 2000 to 2006 and reached its highest level just before the financial crisis. Afterwards, mortgage origination plunged dramatically. During and after the financial crisis the market also sustained a dramatic drop; origination of new mortgages during 2008-2009 did not even sum up to one billion. After the financial crisis (2010 and beyond), origination increased slightly but remained far from its level before the financial crisis.

The third column of Table 1 displays the average FICO credit score in the sample. Unsurprisingly, the credit score, on average, is 4 points lower than the 660 threshold. The next column also shows that less than half of the sample (48%) is composed of loans granted for borrowers with credit scores higher than 660. The evolution of the FICO credit score over the years is interesting. For one, the credit quality of borrowers was below the 660 threshold before the financial crisis (655) but above it afterward (671). For instance, the credit score averaged 615 and 644 in years 2000 and 2002. However, after the crisis, credit quality improved significantly; the average FICO score is consistently higher than 770 in the 2010-2013 period.

[Figure 2 about here]

Figures 2 and 3 further examine the evolution of borrowers' credit quality; they depict the evolution of FICO scores by payment type (ARMs *versus* FRMs) and by loan type (Prime *versus* Subprime). As shown in Figure 2, ARM borrowers have lower credit scores than FRM borrowers, on average. For example, in 2002, the average FICO score for ARMs and FRMs are 619 and 672, respectively. This trend is almost verified for the period before the financial crisis, after which the difference in credit scores is reduced to 10 points. Table 2 shows that the ARM-FRM FICO score differential over the study period is about 34 points, statistically significant at the 1% level. Figure 3 suggests that, unsurprisingly, the average credit score for subprime loans is significantly lower than for prime loans. For illustration, in 2002 the average FICO score for subprime loans is almost 120 points lower than for prime borrowers (616 versus 735). Table 2 indicates that over the study period the average FICO scores for prime and subprime borrowers are 731 and 635, respectively. The difference of 96 FICO points is statistically significant at the 5% level. After the financial crisis, the average credit score tended to improve each year, mainly due to the drop in subprime lending. As column 4 of Table 1 indicates, almost all loans originated after the financial crisis have a credit score higher than 660.

[Figure 3 about here]

Regarding the loan-to-value (LTV) ratio of sampled mortgages, columns 5 and 6 of Table 1 show that the average LTV ratio in the sample is 77% and 60% of loans in the sample have an LTV ratio higher than 80%. Regarding the evolution of the LTV ratio over the years, the LTV ratio plunged significantly soon after the financial crisis. For instance, column 6 of Table 1 shows that more than 60% of loans have an LTV ratio higher than 80% throughout the pre-crisis period. However, this proportion drops to almost 20% in the 2010-2013 post-crisis period. We further split our sample according to payment type (ARMs *versus* FRMs) and loan type (Prime *versus* Subprime). Table 2 shows the results over the entire studied period.

We also investigate the lender's effort to gather all documentation required at the date of original underwriting. The statistics show that lenders did not gather sufficient documentation on applicants in almost half of the cases (47% of the time, lenders granted funding to borrowers but gathered little or no documentation on borrowers' income and employment status). Yearly statistics show that this practice of granting funding without the required documentation increased steadily in the early 2000s. For illustration, the proportion of loans granted with no/low documentation increased from an initial level of 34% in 2000 to 51% in 2005 and 52% in 2006. This practice peaked in early 2007, when almost 60% of loans were granted without gathering sufficient information. This could be viewed is an additional evidence that lenders in the subprime market did not make an adequate effort to gather the required level of information on borrowers' income and employment status before the financial crisis. In contrast, the proportion of loans with no/low documentation fell to around 2% and 3% in 2010 and 2012. As shown in Figure 4, the high proportion of no/low documentation is mainly driven by the practice in the subprime segment; this proportion peaked at 70% from 2005 through 2007.

[Figure 4 about here]

In general, the lending strategy appears to radically change after the financial crisis. This shift in lending strategy entailed (*i*) increasing loans granted for borrowers with good credit quality, (*ii*) reducing loans with a small down payment (LTV ratio higher than 80%), and (*iii*) reducing the proportion of loans granted with insufficient documentation. These changes in underwriting patterns are consistent with lenders looking for new ways to limit risk exposure after the financial crisis.

We also examine the proportion of loans that conform to the GSE prudent lending guidelines. Although the proportion was quite low before the financial crisis (only 17% of the originated mortgages were conforming), the results show that this proportion drops to zero in the post-crisis period (see Table 1).

To motivate our empirical analysis, we further contrast the ex-ante risk characteristics of mortgages for which the originator chooses to sell the underlying servicing rights to another servicer with mortgages that it chooses to continue to service. Overall, we note that for 54.7 percent of the sampled mortgages (3,060,083 mortgages) the originator chooses to switch the servicer of the deal. For the remaining loans (45% of the sample), the originator decides to keep servicing the mortgages and to hold them in its servicing portfolio until maturity. Table 2 shows that the average servicing fee is 44 bp, which does not change very much before and after the crisis. On average lenders in the sample tend to charge significantly higher fees than the average servicing fees applied by the GSEs and the FHA/VA, at 25bp and 19bp respectively.

Regarding the borrower's credit quality, the results show that lenders tend to keep servicing loans granted to borrowers with superior credit quality. For illustration, the average credit score for loans held in the originator's servicing portfolio is 661 while the average credit score for loans for which the lender decides to switch servicing is 654, namely 3 basis points below the sample average. The two-sample mean difference (untabulated) is 6.39 points, statistically significant at the 1% level. Table 2 also shows that the fraction of loans granted for borrowers with FICO scores higher than the 660 threshold is significantly larger for loans held in portfolio (51% for non-switch versus 46% for switch).

These results indicate that lenders switch servicing of the deal for loans that are more risky, and keep servicing mortgages that are less risky. For instance, the pool of loans for which the servicer has changed is characterized by higher loan-to-value ratios and slightly higher debt-to-income ratios. Regarding the subprime loan type, the primary statistics are not informative in that the propensity to switch the servicer of the deal is 52% for prime loans and slightly higher, at 56%, for subprime loans. The results also suggest that 15% of loans for which the servicer is switched follow the GSEs' prudent lending guidelines, whereas this percentage increases to 20% for loans held in the originator's servicing portfolio. The proportion of loans that conform to the GSE lending guidelines at origination represents only 17% of the sample.

To summarize, based on the observable risk characteristics of originated mortgages, these preliminary results are consistent with the evidence of lenders selling MSR rights for low-quality loans to other servicers and retaining high-quality mortgages in their own servicing portfolios.

To better understand the originators' motive to switch the servicing of the deal, we further break down the mortgage sample by default status. The statistics show that, not surprisingly, lower FICO scores, higher LTV ratios, higher debt-to-income ratios, and higher interest rates are the risk characteristics that are more likely to be associated with the default outcome. For instance, 55% of loans that never entered delinquency are granted to borrowers with FICO scores above the 660 threshold. In addition, 72% of loans identified as being in default exhibit an LTV ratio higher than 80%. Not surprisingly, following the GSE guidelines significantly reduces the observed default frequency in that only 10% of defaulting loans follow the GSE prudent lending guidelines.

Contrasting the distribution of loans that were chosen for servicer switch with the default outcome yields additional interesting findings. For instance, almost 62% of loans for which the servicer of the deal has been switched are reported as being in a default status, compared with 26% of loans held in the servicer's portfolio.[4] When comparing the default propensities between the switch and non-switch groups, the results show that 50% of loans defaulting have the servicer switched, compared with 18% of loans in the non-

[4] The high default rate of 37% should be interpreted with caution because it is sample-specific and does not necessarily represent the default rate in the overall mortgage market. Notably, we are using a database that focuses primarily on mortgages securitized through the private-label channel, which are widely recognized to be more risky than loans sold to Government-Sponsored Enterprises, GSE-labelled (60 days).

default category. In general, these preliminary results suggest a positive association between the originator's decision to switch servicers and the default outcome.

Overall, the univariate analysis shows that the mortgages for which the servicer has been switched are generally of low credit quality and are commonly associated with a higher default likelihood. These primary results give us the first insight into the possible presence of asymmetric information because there is a clear association between the originator's decision to switch the servicer of the deal and the likelihood that the borrower defaults. In the next section, we further examine these patterns in more detail in a multivariate framework.

VI. EMPIRICAL RESULTS

VI.1 Non-parametric methods

The two non-parametric testing approaches that we consider have been used to test for asymmetric information in the automobile insurance market. The first approach is the non-parametric testing procedure proposed by Chiappori and Salanié (2000). Their framework is mainly based on a sequence of Pearson's χ^2-test of independence. The second approach is proposed by Su and Spindler (2013). It is mainly driven by the Kernel Density Estimation (KDE) technique, described in Section III.

The Chiappori and Salanié (2000) method

In our context, the main objective is to examine the relationship between the originator's decision to switch the servicer of the deal and the default event. Therefore, the null hypothesis of Pearson's χ^2-test of independence is that there is no significant relationship between the decision to switch servicers and the likelihood that the borrower defaults.

To apply the methodology, we need to consider only binary variables (*i.e*. discrete variables with only two categories). Therefore, throughout this analysis we convert two continuous variables, FICO score and LTV ratio, into binary variables: *FICO660* and *LTV80*. The first variable denotes borrowers with a FICO score higher than 660, and the second variable denotes borrowers with an LTV ratio higher than 80%. The final set of explanatory variables that we consider in our analysis are *FICO660*, *LTV80*, *ARM*, *No/Low documentation*, *Balloon*, *GSE conforming*, *Subprime*, and *Prepayment Penalty*. For robustness, we consider various configurations of the variables.[5] The upper part of Table 3 displays the different configurations that we use to define the set of control variables.

We first choose a set of m exogenous control variables. These variables are binary, so we construct $M = 2^m$ cells in which all mortgages have the same values for the selected control variables. For example, for m=3 including *FICO660*, *LTV80*, and *ARM* as control variables, the first cell (0,0,0) comprises all mortgages granted to borrowers with FICO scores lower than 660, have LTV ratios lower than 80%, and are of the FRM payment type. Next, we draw a 2-by-2 contingency table for the two variables of interest (default and switch) to illustrate the occurrence of the default event depending on the servicer's switch decision, and conduct the Pearson's χ^2-test of independence in each cell. We obtain M Pearson's test statistics at the end of this procedure. Under the null hypothesis of independence, each test statistic is distributed asymptotically as $\chi^2_{(1)}$.

We use three methods to test conditional independence: the Kolmogorov-Smirnoff non-parametric test and the $\chi^2_{(1)}$ test that counts the number of rejections of the null hypothesis in each individual cell. Third, we sum all χ^2 test statistics within the M cells. The sum, denoted S, is asymptotically distributed $\chi^2_{(M)}$ under the null hypothesis of independence.

Table 3 displays the results of Chiappori and Salanié's (2000) testing procedure. The table reports the number of control variables included in each configuration and the total number of cells. We first examine the p-values of the Kolmogorov-Smirnoff (KS) one-sample test. Clearly, using all possible combinations, we unequivocally reject the null hypothesis at the 1% significance level. Using the second method, the rejection rate of the

[5] We do not include all of these variables simultaneously because some of them, for example, *GSE conforming* and *Subprime*, are a function of the others.

null hypothesis of independence in individual cells is high for all configurations. For instance, almost all test statistics calculated within the individual cells exceed the $\chi^2_{(1)}$ critical value of 3.84 (at a 5% significance level). The highest rejection rate is reached with configuration I, which includes 4 control variables *FICO660*, *LTV80*, *ARM*, and *NoLow_doc*. The latter method confirms these findings in that the aggregate test statistics according to all configurations are above the critical values of the $\chi^2_{(M)}$ theoretical distribution.

The Su and Spindler (2013) method

We begin by documenting how well the kernel density estimation fits the data. As mentioned above, the main advantage of the non-parametric approach is that it does not restrict the distribution of the data or the functional form of the density. Therefore, all inferences in our non-parametric framework are purely data-driven in the sense that the data will be allowed to speak for themselves. Figures 5 and 6 display histograms for our two continuous variables: the borrower's FICO score and the LTV ratio. For better visualization, both histograms are augmented with curves of the kernel (non-parametric) and normal (parametric) density functions. Clearly, we can see that, especially for the LTV ratio (see Figure 6), the non-parametric kernel density function (KDE) has a much better fit to the actual data than the parametric *pdf* does. For illustration, the histogram of the LTV ratio suggests that loans with LTV ratios falling in the 75-80% range are over-represented in the sample. The normal density curve underestimates that proportion by 5.5%, whereas the KDE presents accurate estimates.

[Figure 5 about here]

Figure 7 highlights the key role of the smoothing parameter (or bandwidth) in the estimation and displays the sensitivity of the fit of the kernel density estimation technique to the data. In particular, the figure displays the KDE fitting for three different values of the bandwidth: a very high bandwidth, an optimal bandwidth, and a very low bandwidth. We now describe how we obtained the optimal bandwidth. The figure shows that failing to select the optimal bandwidth could be costly because it could result in over-fitting or under-fitting the data. In fact, the bandwidth, as a smoothing parameter, controls the size of the neighborhood around a given point of estimation *x*. We use the Maximum Likelihood Cross-Validation (MLCV) method to estimate the bandwidth from the sample by optimizing the loss objective function on the true density. The estimation results show that the optimal bandwidth values are 3.357 for the FICO score and 0.716 for the LTV based on the MLCV method. These values of optimal bandwidths suggest a significant kernel density estimate because the bandwidths are higher than zero. We also include additional discrete (binary) control variables such as indicator variables for the ARM payment type, Balloon loan type, No/Low documentation, Subprime loan type and/or GSE conforming loan. For all discrete variables, the optimal bandwidth values are within the [0,1] interval, which, according to Li and Racine (2007, 2008) and Racine (2008), means that these variables are relevant to the model.[6]

[Figure 6 about here]

Studies using the non-parametric framework commonly employ graphical representations to display the results where a continuous variable typically serves as a support. In our context of mortgage servicing and loan default, we use the borrower's FICO score as a support to visualize our results. Our choice is motivated by the fact that this variable represents a direct measure of the credit quality that the originator could use to assess the likelihood of borrower's default, and then to decide whether or not to sell the underlying servicing rights after securitization. Thus, this particular continuous variable could be directly linked to both the propensity of borrower default and the originator's decision to switch the servicer of the deal. Consequently, all frequencies and probabilities are plotted below with respect to the borrower's FICO score, which seems to highlight our evidence.

[Figure 7 about here]

[6] Li and Racine (2007, 2008) and Racine (2008) assert that the CV methods produce high bandwidth values for the irrelevant continuous variables and bandwidths close to 1 for irrelevant discrete variables. Interested readers could refer to the three contributions for additional details on bandwidth selection methods.

We present the results of the non-parametric kernel density estimation approach in Figure 8, which displays the conditional probability of mortgage default. Conditional means that the probability of mortgage default is conditional on observed risk characteristics on both the borrower and the granted loan. Obviously, the set of conditioning information is collected and recorded at the time of original underwriting. For comparison purposes, Figure 8 also displays the fitted values of a linear (parametric) model. This model suggests a statistically significant negative slope for the FICO score-mortgage default linear relationship. The kernel density estimation corroborates this finding and suggests that the relationship could be non-monotonic in some parts of the data.

Now we move to the core of the asymmetric information test. Figure 9 displays the estimated probability of default conditional on all observed risk characteristics and, most importantly, on the originator's switching decision. We observe distinct default estimates based on the decision to switch the servicer of the deal. In particular, the two plots labelled "Switched" and "not Switched" display the estimated probability of default over a set of different FICO score values conditional on the originator's decision to switch the servicer of the deal or not, respectively. More formally, the two plots represent $\hat{f}\left(y_i \middle| x_i^c, x_i^d, z_i = 1\right)$ and $\hat{f}\left(y_i \middle| x_i^c, x_i^d, z_i = 0\right)$, respectively.

[Figure 8 about here]

At first glance, both plots show that the conditional probability of default decreases as the borrower's credit quality improves. However, the plots point to a significant difference in the probability of default if we take into account the originator's decision to switch the servicer. For illustration, the estimated probability of default for loans granted to borrowers with an average FICO score of 450 is about 75%. If we consider the case where the originator chooses to switch the servicer of the deal and sells the underlying MSR, the probability that a mortgagor enters delinquency and defaults increases to 80% (The probability becomes higher if the FICO score is lower than 450). However, mortgages that the originator chooses to keep in its servicing portfolio have an expected probability of default of 70%, resulting in an almost 10% drop in the expected probability of default.

Note that borrowers of these mortgages under consideration share many characteristics because they belong to the same FICO score cohort. The only variable that makes the difference here is the originator's decision to switch the servicer of the deal. Figure 9 also shows that this pattern is valid not only for low-quality borrowers but also for those with superior credit quality. Although the average default likelihood drops significantly by almost 70% (see the discussion above) if we consider high-quality borrowers, the default likelihood drops much more if the originator chooses to keep the loan in its servicing portfolio. For illustration, if we consider loans granted to borrowers with FICO scores higher than 750, the estimated conditional probability of default is about 19% if the originator sells the underlying MSRs, and is almost zero if it chooses to keep the loan in its servicing portfolio.

These results are in line with those found in the previous section using Chiappori and Salanié's (2000) method. For instance, the results suggest a positive association between the probability of default and the originator's decision to switch the servicer of the deal. Our results show that, for a given pool of observably similar mortgages (*i.e*. mortgages that have similar risk characteristics and that are granted to borrowers with very similar credit scores), the mortgage originator's decision to sell the underlying MSRs to a new servicer single-handedly increases the expected probability of default of a given loan.

We use the bootstrap method to obtain the bootstrap *p*-values and conclude our asymmetric information test. The bootstrap method consists of three main steps and can be summarized as follows. First, for every bootstrap iteration $b= 1\dots B$, we create a bootstrap resample of the data with replacement, denoted as $(X_i^{c^b}, X_i^{d^b}, Y_i^b, Z_i^b)$ where the superscript b denotes the b^{th} resampled data set. Next, for every resample b, we estimate the conditional probability of default given all observed characteristics as well as the originator's switching decision, and calculate the corresponding test statistic. After repeating these steps for all B iterations, we obtain a total of B bootstrap test statistics. Lastly, the bootstrap *p*-value is simply calculated as the frequency of the event where the bootstrap statistic is higher than or equal to the original test statistic.

The set of explanatory variables that we consider is *FICO*, *LTV80*, *ARM*, *No/Low documentation*, *Balloon*, *GSE conforming*, *Subprime*, and *Prepayment Penalty*. For robustness, we try several inclusion combinations for the control variables, as we did for the Chiappori and Salanié (2000) model (upper panel of Table 3). The total number of bootstrap replications is set to B= 500. In all possible configurations not reported here (but available), we find bootstrap p-values lower than the 5% standard significance level. Clearly, low p-values enable us to reject conditional independence. In other words, low bootstrap p-values suggest that in all cases we can refute the null hypothesis of absence of asymmetric information at the 5% level. This means that the likelihood of mortgage default and the decision to switch the servicer of the deal are closely linked. Indeed, a significant relationship exists between these two variables.

The failure to reject the null hypothesis of absence of asymmetric information can be interpreted as follows: The original lender's decision to switch the servicer or to continue servicing a given deal conveys an important piece of information in predicting the probability of default on that mortgage, even after taking into account all loan and borrower risk characteristics. Apparently, the originator's decision to switch the servicer of the deal plays a key role in predicting the probability of default. For instance, we have seen that, conditional on all observed risk characteristics, the default likelihood of a given mortgage increases by 10%, on average, if the originator has sold the underlying MSRs and switched the servicer of the deal. Hence, the presence of asymmetric information in the mortgage servicing market has an evident impact on the probability of mortgage default.

VI.2 Robustness checks: results of the parametric methods

In this section, we provide additional support for our evidence of the presence of information asymmetry in the mortgage servicing market based on common parametric models. First, we use a probit model to investigate the determinants of the default likelihood (see Table 4). We then use a variety of model specifications to account for potential endogeneity issues, for econometric misspecification, and for simultaneity (see

Table 5). In particular, we use the two-stage instrumental variable probit model in order to account for potential endogeneity issue. We also use the two-step estimation procedure proposed by Dionne, La Haye, and Bergerès (2015) in order to account for econometric misspecification error and correct the linear-imposed relationship. Additionally, we use simultaneous probit regressions (Bivariate-Probit) that jointly model both the decision to switch the servicer of the deal and the event of mortgage default in a system of simultaneous equations (Chiappori and Salanié, 2000).[7]

Table 4 displays the estimation results for the standard probit model where the dependent variable is the mortgage default binary variable. The table reports various inclusion configurations for the set of control variables. We control for *(i)* fundamental borrower and loan risk characteristics, *(ii)* general economic conditions, *(iii)* housing market conditions, *(iv)* bond market conditions, and *(v)* state legal structure.

As shown in Table 4, all explanatory variables display the expected sign (column I). For instance, borrowers with good credit scores (high FICO scores) who can afford larger down payments (low LTV ratios) experience a lower probability of mortgage default. Likewise, following GSE prudent lending guidelines and collecting a sufficient amount of required documentation significantly reduce the likelihood of mortgage default. Conversely, having an ARM or Balloon payment structure significantly increases the risk of mortgage default in that their coefficients exhibit a statistically significant positive sign.

The parametric test consists primarily of investigating the statistical significance of the link between the decision to switch the servicer of the deal and the likelihood of mortgage default. Nevertheless, the main challenge for our empirical test is the endogeneity issue (Dionne et al., 2009, 2015).

The first stage of Table 5 shows the results of the two-stage instrumental variable probit model where the dependent variable is the default likelihood, and where Income and

[7] Other recent parametric applications have been developed by Adams et al. (2009) and Crawford et al. (2018). These authors factored in the market conditions of the lending market to develop their tests; we do not do so in this research since we do not have access to the necessary information.

Divorce are used as instrumental variables.[8] Not surprisingly, the income growth rate is negatively correlated with the default likelihood; the coefficient is statistically significant at the 1% level. In contrast, the divorce rate is positively related to the default likelihood, suggesting that marital breakdown represents a significant factor in determining mortgage default. All other coefficients have the expected sign. The second-stage regression, which includes modeling the originator's decision to switch the servicer of the deal as a dependent variable, shows that this variable is positively correlated with the default event, even after controlling for endogeneity.

The last two columns in Table 5 confirm the above findings of a positive association between the two variables using the parametric model of Dionne, La Haye, and Bergerès (2015) and Chiappori and Salanié (2000). Moreover, the results show that the estimated correlation coefficient is about 0.60 and is statistically significant at the 1% level. All other explanatory variables remain statistically significant and keep the expected sign.

In the Appendix, we present the results of Table 4 and Table 5 with a different definition of default variable (60+ days) and a different time period of 2001-2006 and our results are robust to these alternatives observed in the literature.

VII. CONCLUSION

In this paper, we analyze the servicing switching decision in the securitization market. Our main objective is to verify whether information asymmetry between servicers affects loan default. Specifically, we investigated whether a first-level service decision to sell the mortgage servicing rights to a second-level service reveals any residual information asymmetry in the mortgage servicing market.

Our empirical results reveal interesting and important conclusions related to the US mortgage servicing market. We observe that information asymmetry between servicers

[8] We provide tests of the validity of these two instruments. It is clear that aggregate ratio of Income and Divorce could affect the loan default probability. They should not significantly affect the servicer's decision to switch the loan servicing. Usual test with linear probability models rejects the Wu-Hausman test as well the weak instruments test. Results are available from the authors.

influences the loan default probability significantly. The mortgage originator uses its private information advantage to sell more risky loans to the MSR-purchaser.

This result has important consequences for the securitization market. Recent regulation has introduced a retention provision for banks that use securitization. Since December 2014, securitizers must keep an economic interest (retention) in the credit risk of the securitized assets (Morgan Lewis, 2018). Only the original creditor must keep the economic interest, which means that the risk retention cannot be allocated to a subsequent purchaser. It would be interesting to investigate how this new rule may have affected the type of information asymmetry effect that we have measured.

REFERENCES

Abbring, J., P.A. Chiappori, and J. Pinquet (2003). "Moral Hazard and Dynamic Insurance Data." *Journal of the European Economic Association*, *1*, pages 767-820.

Adams, W., L. Einav, and J. Levin (2009) "Liquidity Constraints and Imperfect Information in Subprime Lending." *American Economic Review*, *99*, pages 49-84.

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

Ahmad, I. and P. Cerrito (1994) "Nonparametric estimation of joint discrete-continuous probability densities with applications." *Journal of Statistical Planning and Inference*, *41*, pages 349-364.

Aitchison, J. and C.G.G. Aitken (1976) "Multivariate Binary Discrimination by the Kernel Method." *Biometrika, 63,* pages 413-420.

Akerlof, G.B. (1970) "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, *89*, pages 488-500.

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

Aldrich, S.P.B., W.R. Greenberg, and B.S. Payner (2000) "A Capital Markets View of Mortgage Servicing Rights." *Journal of Fixed Income, 11*, pages 37-54.

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*, pages 113-133.

Buttimer, R.J. and C.C. Lin (2005) "Valuing U.S. and Canadian Mortgage Servicing Rights." *Journal of Housing Economics, 14*, pages 194-211.

Chiappori, P.A. and B. Salanié (2000) "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, *108*, pages 56-78.

Chiappori, P.A. and B. Salanié (2013) "Asymmetric Information in Insurance Markets: Predictions and Tests." In: Georges Dionne (Ed.) *Handbook of Insurance*, Springer, pages 397-422.

Crawford, G.S., N. Pavanini, and F. Schivardi (2018) "Asymmetric Information and Imperfect Competition in Lending Markets." *American Economic Review*, *108*, pages 1659-1701.

Deheuvels, P. (1977) "Estimation non paramétrique de la densité par histogrammes généralisés." *Revue de statistique appliquée*, *25*, pages 5-42.

Dionne, G., C. Gouriéroux, and C. Vanasse (2001) "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment." *Journal of Political Economy*, *109*, pages 444-453.

Dionne, G., M. La Haye, and A.S. Bergerès, (2015) “Does Asymmetric Information Affect the Premium in Mergers and Acquisitions?” *Canadian Journal of Economics*, *48*, pages 819-852.

Dionne, G., P. St-Amour, and D. Vencatachellum (2009) “Asymmetric Information and Adverse Selection in Mauritian Slave Auctions.” *Review of Economic Studies*, *76*, pages 1269-1295.

Dionne, G., P.C. Michaud, and M. Dahchour, (2013) “Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France,” *Journal of the European Economic Association*, *11*, pages 897-917.

Elul, R. (2016) “Securitization and Mortgage Default.” *Journal of Financial Services Research*, *49*, pages 281-309.

Federal Reserve Board (2016) “Report to the Congress on the Effect of Capital Rules on Mortgage Servicing Assets.” Washington, USA.

FitzGerald, G. (2016) “The Servicing Industry in 2016: A Changing Landscape.” *Servicing Management*, 3 pages. http://www.bkfs.com/Data/TLArticles/Feb.%202016_George%20 FitzGerald_Servicing%20Executive.pdf.

Hernandez, R. et al. (2015) “The Changing Dynamics of the Mortgage Servicing Landscape.” Report of the Mortgage Bankers Association in cooperation with PwC,

JP Morgan (2010) “Non-Agency Mortgage-Backed Securities, Managing Opportunities, and Risks.

Keys, B.J., A. Seru, and V. Vig (2012) “Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets,” *The Review of Financial Studies*, *25*, pages 2071-2108.

Keys, B.J., T. Mukherjee, A. Seru, and V. Vig (2010) “Did Securitization Lead to Lax Screening? Evidence from Subprime Loans,” *Quarterly Journal of Economics*, *125*, pages 307-362.

Levitin, A.J. and T. Twomey (2011). “Mortgage Servicing.” *Yale Journal on Regulation*, *28*, pages 1-90.

Li, Q. and J. Racine (2007) “Nonparametric Econometrics: Theory and Practice.” *Princeton University Press*.

Li, Q. and J. Racine (2008) “Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data.” *Journal of Business and Economic Statistics*, *4*, pages 423-434.

Lin, C.C. and L.C. Ho (2005) “Valuing Individual Mortgage Servicing Contracts: A Comparison between Fixed-Rate Mortgages and Adjustable-Rate Mortgages.” *Review of Pacific Basin and Financial Markets and Policies*, *8*, pages 131-146.

Lin, C.C., T.H. Chu, and L. Prather (2006) "Valuation of Mortgage Servicing Rights with Foreclosure Delay and Forbearance Allowed." *Review of Quantitative Finance and Accounting*, *26*, pages 41-54.

Malekan, S. and G. Dionne (2014) "Securitization and Optimal Retention under Moral Hazard." *Journal of Mathematical Economics*, *55*, pages 74-85.

McConnell, J. J. (1976) "Valuation of a Mortgage Company's Servicing Portfolio." *Journal of Financial and Quantitative Analysis, 11*, pages 433-53.

Morgan Lewis, 2018 "Guide to the Credit Risk Retention Rules for Securitization." https://www.morganlewis.com/~/media/files/handouts/final_risk_retention_guide_handout.ashx.

Nadaraya, É.A. (1965) "On non-parametric estimates of density functions and regression curves." *Theory of Probability & Its Applications*, *10*, pages 186–190.

Pagan, A. R. and A. Ullah (1999) "Nonparametric econometrics." *Themes in modern econometrics*, Cambridge University Press, Cambridge, New York.

Parzen, E. (1962) "On Estimation of a Probability Density Function and the Mode." *The Annals of Mathematical Statistics*, *33*, pages 1065-1076.

Racine, J. (2008). "*Nonparametric econometrics: a primer*." Vol. 3, Now Pub.

Rosenblatt, M. (1956) "Remarks on some nonparametric estimates of a density function." *The Annals of Mathematical Statistics*, pages 832-837.

Silverman, B. (1986) "Density estimation for statistics and data analysis." Vol. 26, Chapman & Hall/CRC.

Su, L. and M. Spindler (2013) "Nonparametric Testing for Asymmetric Information." *Journal of Business and Economic Statistics*, *31*, pages 208-225.

Van Drunen, L. D. and J. J. McConnell (1988) "Valuing Mortgage Loan Servicing." *Journal of Real Estate Finance and Economics*, *1*, pages 5-22.

Watson, G.S. (1964) "Smooth regression analysis." *Sankhyā: The Indian Journal of Statistics, Series A*, *26*, pages 359-372.

Woodroofe, M. (1970) "On choosing a delta-sequence." *The Annals of Mathematical Statistics*, *41*, pages 1665-1671.

Table 1. Summary statistics by origination year

The table reports summary statistics for the sample of U.S. mortgages originated over the period from January 2000 to December 2013. The mortgages have been securitized through the non-agency channel. The first row reports statistics over the 2000-2013 study period for the total sample of 5,591,353 distinct mortgages while the next rows report statistics by origination year. The first two columns *Volume (in %)* and *Volume (in $B)* refer to the total origination volume expressed in percentage of the total sample and in US$ billions, respectively. *FICO score* abbreviates the borrower's Fair Isaac Corporation score attributed at origination. *FICO.660* denotes the fraction of loans granted to borrowers with FICO scores higher than 660. *LTV* abbreviates the initial loan-to-value ratio. *LTV.80* denotes the fraction of loans with LTV ratios higher than 80%. *DTI* stands for the debt-to-income ratio. *No/Low doc.* indicates whether the originator collected either no or low documentation. *Interest rate* is the coupon rate applied at origination. *Balloon* denotes balloon payment mortgages. *ARM* and *ARM margin* denote adjustable-rate mortgages and the corresponding margin. *GSE conf.* denotes the fraction of loans that conform to the Government-Sponsored Enterprises' prudent lending guidelines. *Prep. Penalty* measures the fraction of loans with prepayment penalties.

Origination year	***Volume (in %)***	***Volume (in $B)***	***FICO score***	***FICO. 660***	***LTV ratio***	***LTV. 80***	***DTI***	***No/Low doc.***	***Interest rate***	***Balloon***	***ARM***	***ARM margin***	***GSE conf.***	***Prep. Penalty***
All period	100.0	1509.1	657.12	0.48	76.93	0.60	38.65	0.47	6.97	0.06	0.63	5.00	0.17	0.49
2000	1.05	8.87	615.49	0.31	78.20	0.62	38.65	0.34	10.08	0.07	0.34	6.13	0.17	0.41
2001	2.47	32.07	648.33	0.47	76.87	0.56	37.74	0.29	8.56	0.03	0.36	6.09	0.23	0.33
2002	5.74	69.08	644.97	0.42	77.47	0.58	37.84	0.33	7.92	0.02	0.54	5.92	0.21	0.38
2003	11.46	170.89	670.12	0.56	75.14	0.51	36.95	0.38	6.60	0.01	0.49	5.18	0.25	0.31
2004	16.93	232.68	657.75	0.49	77.60	0.60	36.81	0.44	6.30	0.00	0.70	4.75	0.19	0.52
2005	27.28	411.36	658.81	0.49	76.97	0.62	38.33	0.51	6.51	0.02	0.69	4.91	0.16	0.53
2006	27.11	422.92	650.45	0.44	77.44	0.63	39.90	0.52	7.44	0.15	0.66	5.07	0.12	0.57
2007	7.79	153.39	668.92	0.56	75.92	0.56	39.17	0.57	7.32	0.12	0.52	4.50	0.15	0.47
2008	0.02	0.60	717.06	0.80	73.25	0.44	36.59	0.40	7.16	0.03	0.48	3.13	0.03	0.17
2009	0.00	0.21	774.60	1.00	53.11	0.06	36.00	0.30	4.79	0.00	0.83	2.04	0.03	0.00
2010	0.01	0.43	772.33	1.00	61.78	0.17	32.48	0.02	4.93	0.00	0.08	1.64	0.01	0.21
2011	0.02	1.17	770.62	1.00	66.60	0.23	32.98	0.17	4.72	0.00	0.06	1.83	0.01	0.16
2012	0.06	2.79	773.08	1.00	66.42	0.20	34.00	0.03	4.06	0.00	0.02	2.25	0.00	0.13
2013	0.06	2.67	771.14	1.00	66.24	0.19	30.80	0.00	3.91	0.00	0.01	2.53	0.00	0.01

Table 2. Summary statistics by loan type and status

The table reports summary statistics for the sample of 5,591,353 U.S. mortgages originated over the period from January 2000 to December 2013. The mortgages have been securitized through the non-agency channel. The table breaks down the sample by payment type (FRM *vs.* ARM), loan type (Prime *vs.* Subprime), financial crisis era (Before *vs.* After), default status, and servicer switch status. *FICO score* abbreviates the borrower's Fair Isaac Corporation score at origination. *FICO.660* denotes the fraction of loans granted to borrowers with a FICO score higher than 660. *LTV* abbreviates the initial loan-to-value ratio. *LTV.80* denotes the fraction of loans with LTV ratios greater than 80%. *DTI* stands for the debt-to-income ratio. *No/Low doc.* indicates whether the originator collected either no or low documentation. *Interest rate* is the coupon rate applied at origination. *Balloon* denotes balloon payment mortgages. *ARM* and *ARM margin* denote adjustable-rate mortgages and the corresponding margin. *Subprime* and *Prime* are sub-prime loan classifiers. *GSE conf.* denotes the fraction of loans conforming to the GSEs' lending guidelines. *Prep. Penalty* indicates the fraction of mortgages with prepayment penalty. *Service fee* is the mortgage servicer fee expressed in percentage of the remaining balance. *Switch servicer* indicates the fraction of mortgages for which the originator switched the servicer of the deal. *Default* denotes the fraction of mortgages in default. *Age at default* is the average age of defaulting mortgages. *Default 12, Default 18,* and *Default 24,* refer to the fraction of loans defaulting within 12, 18, and 24 months since origination, respectively.

	All	***Payment type***		***Loan type***		***Financial crisis***		***Default***		***Switch Servicer***	
		FRM	***ARM***	***Prime***	***Subprime***	***Before***	***After***	***No***	***Yes***	***No***	***Yes***
FICO score	657.12	678.00	644.84	730.93	634.87	655.92	671.02	669.62	635.77	660.62	654.23
FICO.660	0.48	0.61	0.41	1.00	0.33	0.48	0.57	0.55	0.37	0.51	0.46
LTV	76.93	73.89	78.73	63.48	80.99	77.04	75.72	74.86	80.48	76.49	77.30
LTV.80	0.60	0.48	0.67	0.00	0.78	0.60	0.55	0.53	0.72	0.58	0.61
DTI	38.65	37.64	39.08	35.71	39.10	38.60	39.09	37.63	39.91	38.02	38.95
No/Low doc.	0.47	0.46	0.48	0.63	0.43	0.46	0.55	0.45	0.50	0.49	0.45
Interest rate	6.97	7.10	6.89	5.57	7.39	6.94	7.26	6.71	7.41	6.86	7.05
Balloon	0.06	0.04	0.08	0.01	0.08	0.06	0.12	0.03	0.11	0.04	0.08
ARM	0.63	0.00	1.00	0.45	0.68	0.64	0.52	0.59	0.70	0.59	0.66
ARM margin	5.00	.	5.00	2.86	5.42	5.03	4.50	4.80	5.27	4.93	5.04
Subprime	0.77	0.66	0.83	0.00	1.00	0.77	0.72	0.70	0.89	0.75	0.78
Prime	0.23	0.34	0.17	1.00	0.00	0.23	0.28	0.30	0.11	0.25	0.22
GSE Conf.	0.17	0.25	0.12	0.56	0.05	0.17	0.15	0.21	0.10	0.19	0.15
Prep. Penalty	0.49	0.34	0.58	0.24	0.57	0.50	0.46	0.42	0.63	0.49	0.50
Purchase	0.37	0.30	0.42	0.22	0.42	0.38	0.30	0.36	0.40	0.36	0.39
Refin. cash-out	0.47	0.49	0.45	0.46	0.47	0.46	0.51	0.46	0.47	0.49	0.45
Refin. no cash-out	0.16	0.21	0.13	0.31	0.11	0.15	0.19	0.18	0.12	0.15	0.16
Service fee	0.44	0.38	0.47	0.33	0.47	0.44	0.39	0.42	0.46	0.41	0.46
Switch servicer	0.55	0.50	0.58	0.52	0.56	0.56	0.44	0.18	0.50	0.00	1.00
Default	0.37	0.30	0.41	0.18	0.43	0.35	0.54	0.00	1.00	0.26	0.62
Age at default	36.64	45.25	32.98	47.72	35.21	37.41	30.81	.	36.64	38.07	35.47
Default 12	0.11	0.06	0.12	0.03	0.12	0.10	0.12	.	0.11	0.09	0.12
Default 18	0.23	0.15	0.26	0.08	0.24	0.22	0.26	.	0.23	0.20	0.24
Default 24	0.35	0.24	0.40	0.15	0.38	0.34	0.44	.	0.35	0.32	0.37

Table 3. Results of the Chiappori and Salanié non-parametric test

The table reports the results of the Chiappori and Salanié (2000) non-parametric testing methodology. The overall sample includes 5,591,353 U.S. mortgages originated over the period from January 2000 to December 2013. The mortgages have been securitized through the non-agency channel. The upper panel of the table reports 10 different configurations of the control variables. The table displays the number of variables included in each configuration as well as the resulting number of cells. *KS p-value* is the *p*-value of the Kolmogorov-Smirnov non-parametric test. $\chi^2_{(1)}$ *crit. value* is the theoretical value of the χ^2 distribution at the 5% significance level. *Rejection rate* provides the frequency of rejection of the null hypothesis of independence among all individual cells. *S value* is the sum of individual test statistics among all cells.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***	***X***	***XI***
FICO.660	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LTV.80	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ARM	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No/Low doc.	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balloon	-	-	Yes	-	-	-	Yes	Yes	Yes	Yes	Yes
GSE Conf.	-	-	-	Yes	-	-	Yes	-	-	Yes	-
Subprime	-	-	-	-	Yes	-	-	Yes	-	-	Yes
Prep. penalty	-	-	-	-	-	Yes	-	-	Yes	Yes	Yes
# variables	3	4	5	5	5	5	6	6	6	7	7
# cells (M)	8	16	32	32	32	32	64	64	64	128	128
Method 1:											
KS p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Method 2:											
$\chi^2_{(1)}$ *crit. value*	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84
Rejection rate	0.75	1.00	0.81	0.92	1.00	0.91	0.75	0.81	0.83	0.73	0.83
Method 3:											
$\chi^2_{(M)}$ *crit. value*	15.51	26.30	46.19	46.19	46.19	46.19	84.82	84.82	84.82	124.34	124.34
S value	6388.6	4491.4	6840.3	5577.2	4491.4	9638.9	7628.9	6840.3	11089.8	11230.5	11089.8

Table 4. Results of the Probit model

The table reports estimation results of the parametric Probit regressions. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. The dependent variable, *Default*, is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +90 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo* R^2 is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***
LTV ratio	0.0169***	0.0172***	0.0170***	0.0171***	0.0169***	0.0175***	0.0170***	0.0172***	0.0179***
ARM	0.0980***	0.1324***	0.1290***	0.1064***	0.0866***	0.0755***	0.0940***	0.0911***	0.1206***
Balloon	0.6336***	0.5681***	0.5770***	0.5887***	0.6384***	0.6373***	0.6344***	0.6264***	0.4146***
No/Low doc.	0.3726***	0.3742***	0.3741***	0.3707***	0.3673***	0.3602***	0.3721***	0.3690***	0.3396***
GSE Conf.	-0.1939***	-0.1914***	-0.1895***	-0.1920***	-0.1905***	-0.1959***	-0.1918***	-0.1910***	-0.1567***
B. Economic general conditions									
GDP growth		-14.808***							-1.9725***
C. Housing market conditions									
HPI growth			-3.4660***						-7.6275***
D. Bond market conditions									
σ interest				0.4669***					1.0679***
Credit spread					0.3561***				1.8900***
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0464***		-0.0421***
SRR								-0.0868***	-0.0853***
Intercept	0.2878***	0.6870***	0.5697***	-0.1014***	0.6244***	-0.1253***	0.3277***	0.3385***	1.9433***
Pseudo R^2	8.40	9.10	8.82	9.04	8.53	9.39	8.43	8.46	11.60
Log-likelihood	-3.37e+06	-3.35e+06	-3.36e+06	-3.35e+06	-3.37e+06	-3.34e+06	-3.37e+06	-3.37e+06	-3.25e+06
Wald p-value									0.00
LR p-value									0.00

Table 5. Results of the Two-stage and Bivariate Probit models

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S. *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e*. is labelled as +90 days delinquent). *Switch serv.* denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption, and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0007***		-0.0002***				
Divorce	0.2896***		0.2104***				
Pr(Default=1)		0.5334***					
Ê(Default)				0.4871***	0.1683***		
Default					0.3188***		
FICO score	-0.0035***		-0.0011***			-0.0035***	-0.0001***
LTV ratio	0.0180***	0.0029***	0.0051***	0.0004***	0.0004***	0.0180***	0.0030***
ARM	0.1212***	-0.1867***	0.0430***	-0.0795***	-0.0795***	0.1184***	-0.1707***
Balloon	0.4129***	-0.0225***	0.1596***	-0.0525***	-0.0525***	0.4085***	0.0582***
No/Low doc.	0.3395***	0.1579***	0.1062***	0.0437***	0.0437***	0.3416***	0.1699***
GSE Conf.	-0.1537***	0.0777***	-0.0432***	0.0699***	0.0699***	-0.1524***	0.0020
GDP growth	-4.9640***	4.4784***	-1.8759***	3.5329***	3.5329***	-1.9603***	-0.5005***
HPI growth	-7.5731***	-5.8147***	-2.5375***	-0.8476***	-0.8476***	-7.5398***	-7.7918***
σ interest	0.9305***	0.6387***	0.2736***	0.0887***	0.0887***	1.0688***	0.8380***
Credit spread	1.9934***	1.1789***	0.6298***	0.0044***	0.0044***	1.8957***	1.9713***
Judicial	-0.0425***	0.0129***	-0.0129***	0.0066***	0.0066***	-0.0426***	0.0018
SRR	-0.0844***	0.0321***	-0.0267***	0.0145***	0.0145***	-0.0851***	0.0321***
$\boldsymbol{R^2}$	11.7	3.8	13.8	31.2	38.2		
$\boldsymbol{\rho}$						0.5965***	

Figure 1. Lending and securitization processes

Legend:

1. Original loan.
2. Payments before securitization.
3. Loan sale (securitization).
4. Issue of mortgage-backed securities.
5. Subscriptions to MBSs.
6. Proceeds of loan sale.
7. Payments after securitization.
8. Servicing principal and interest before servicer switch.
9. Sale of mortgage servicing rights.
10. Proceeds of MSRs sale.
11. Payments after securitization and servicer switch.
12. Servicing principal and interest after servicer switch.
13. Servicing fees.
14. Advances in case of borrower default.

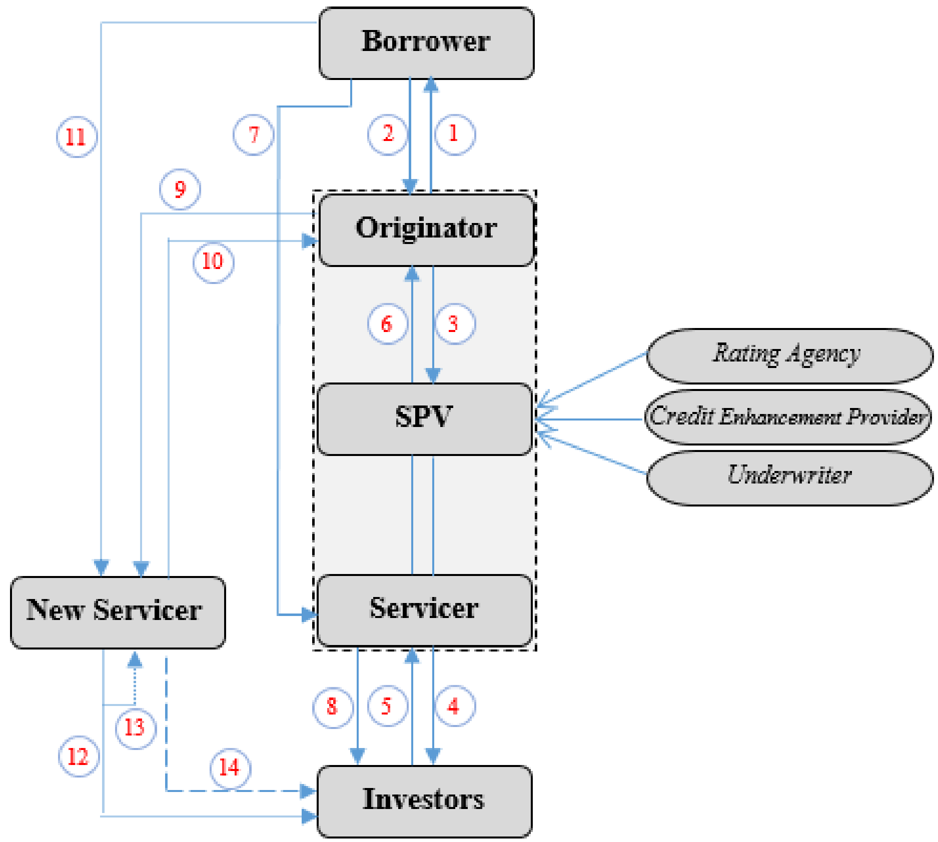


Figure 2. FICO scores at origination by payment type

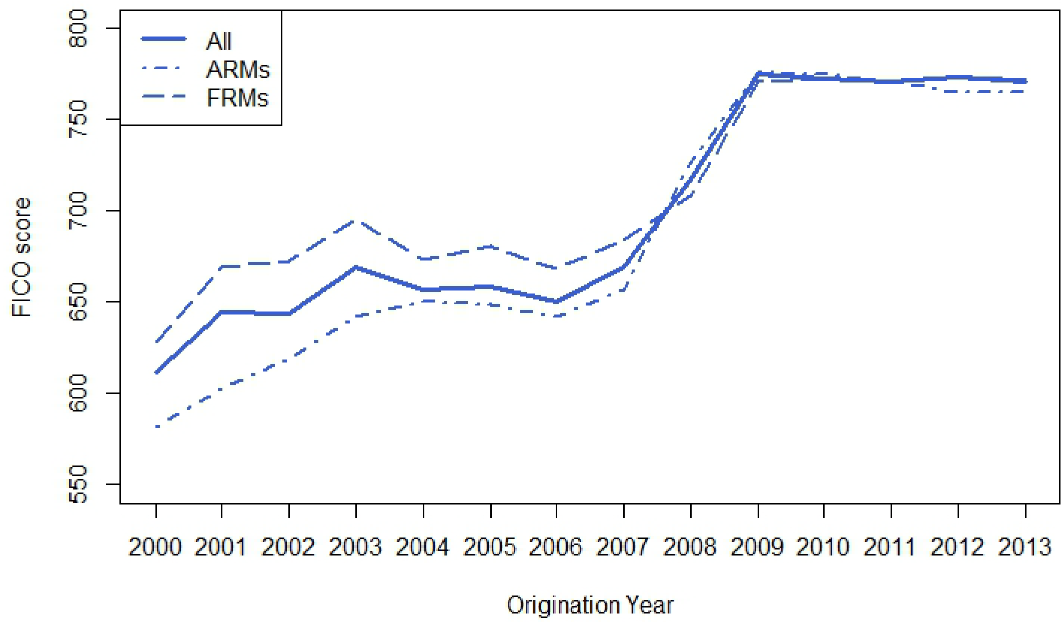


Figure 3. FICO scores at origination by loan type

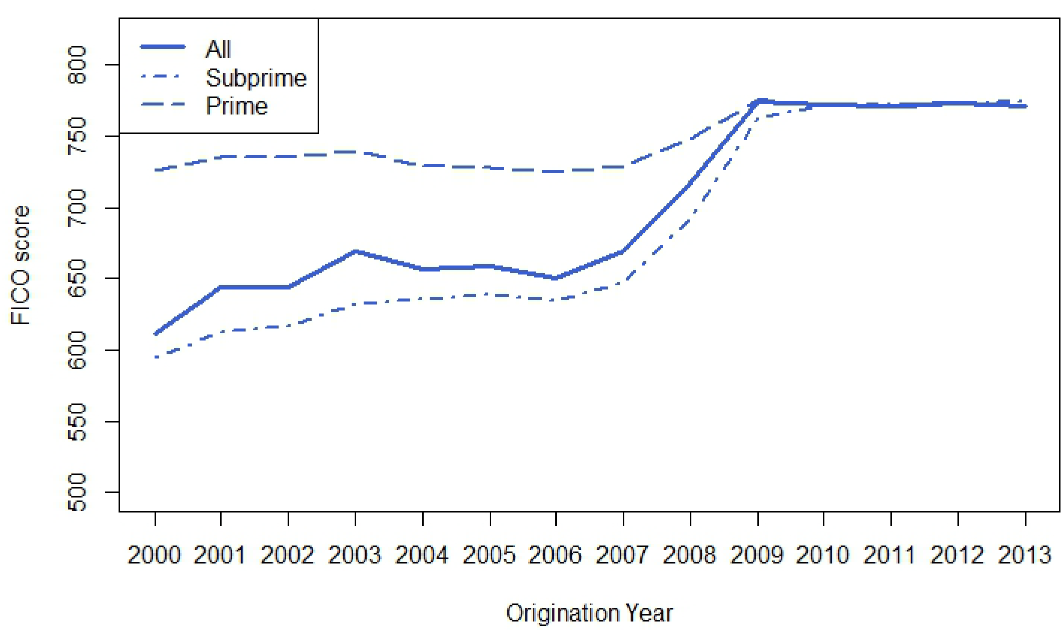


Figure 4. No/Low documentation at origination by payment type

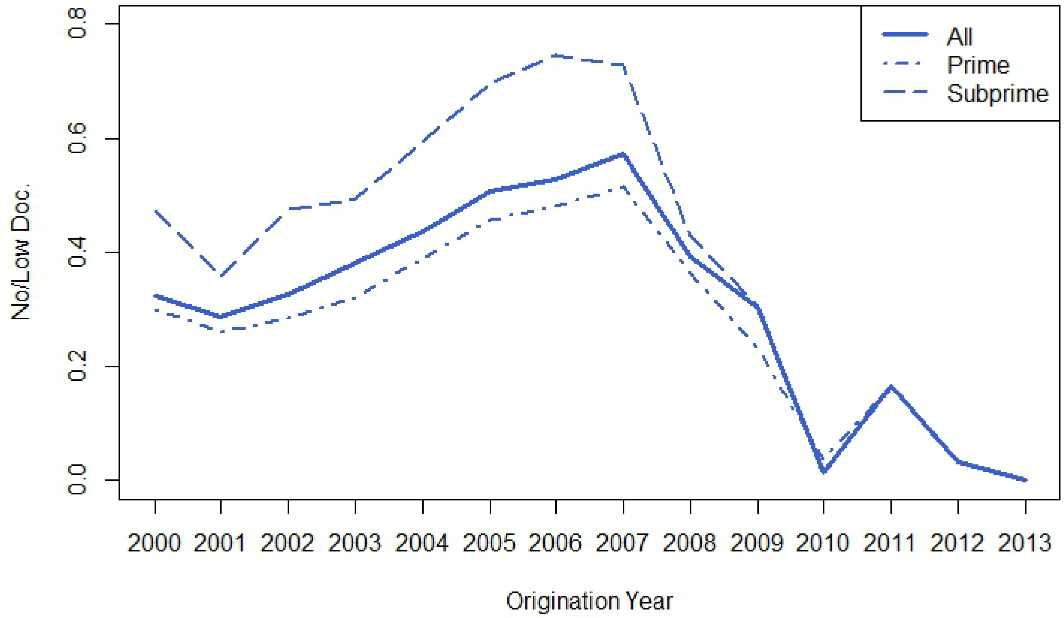


Figure 5. Kernel density fitting of the FICO score

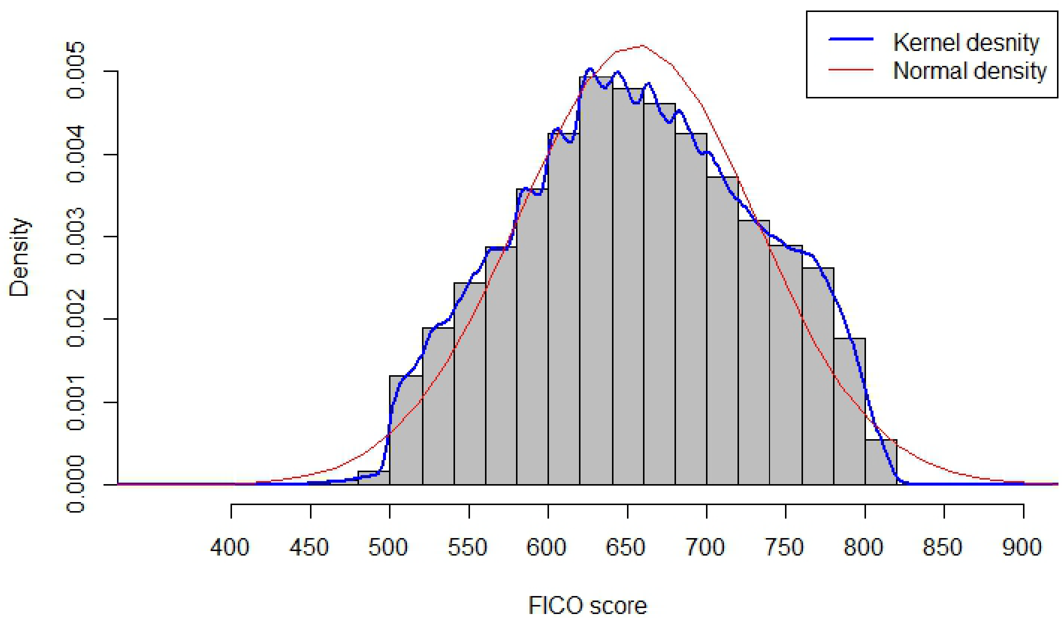


Figure 6. Kernel density fitting of the LTV ratio

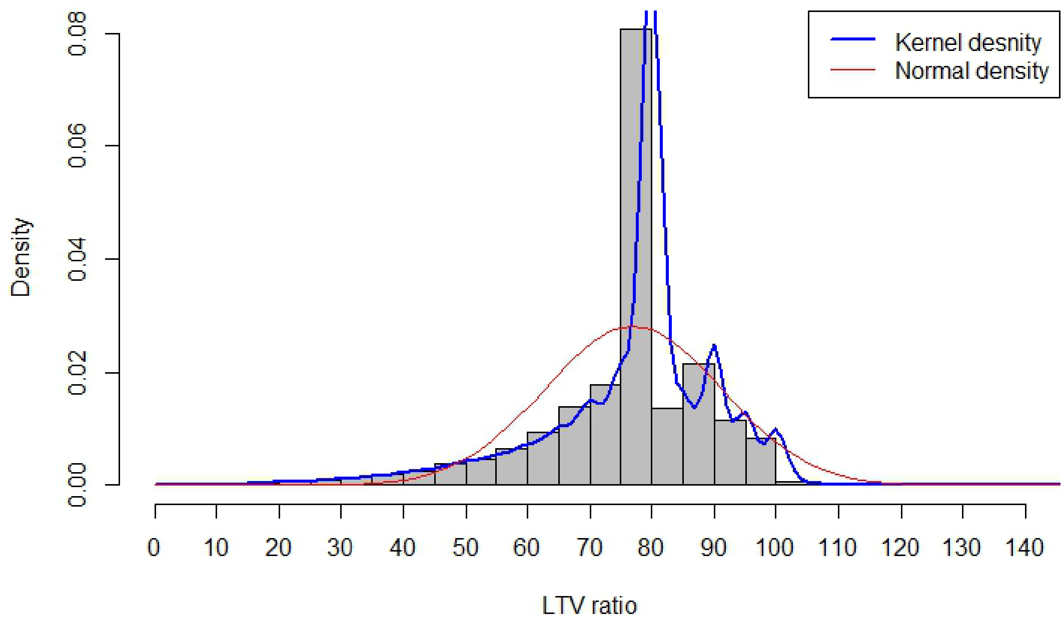


Figure 7. Fitting of the KDE with multiple bandwidths

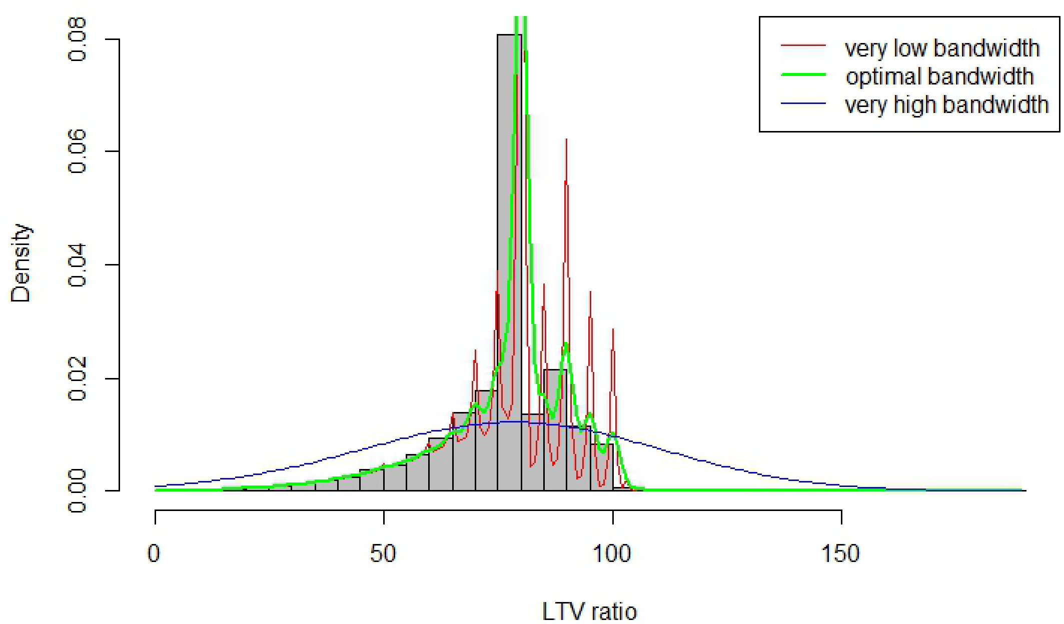


Figure 8. Credit quality vs conditional probability of default

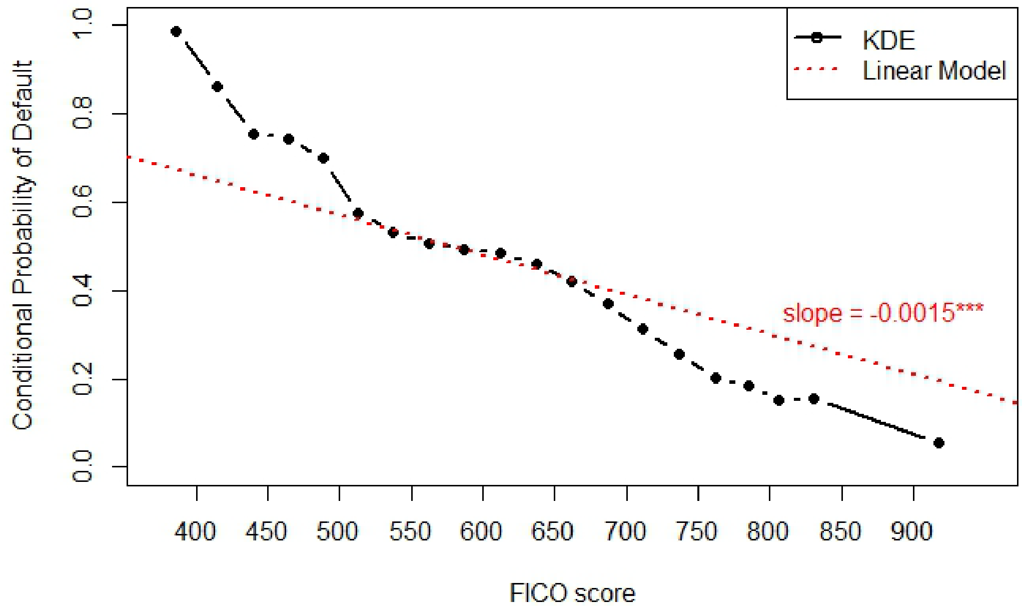
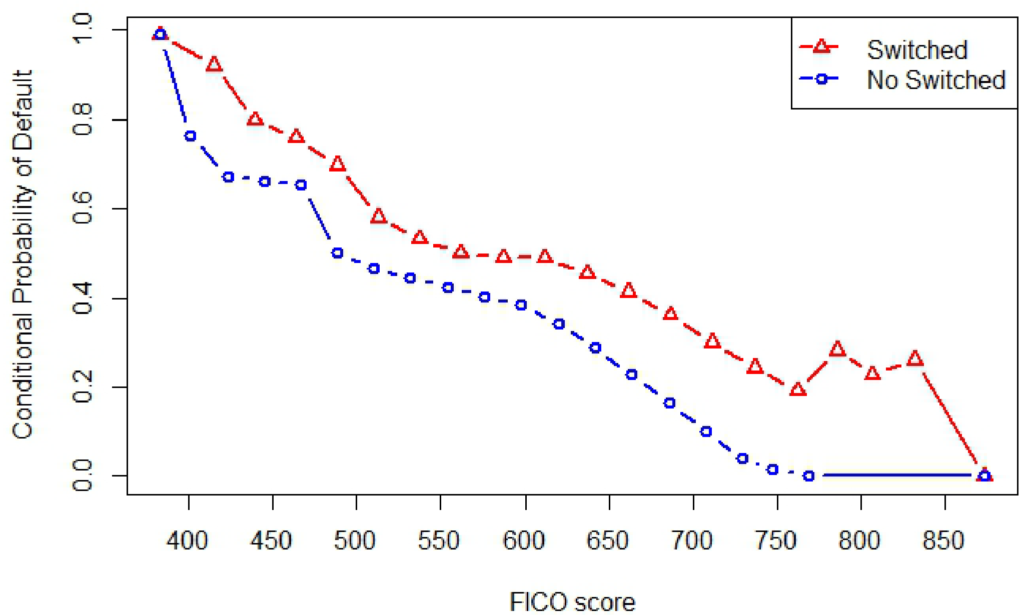


Figure 9. Credit quality vs conditional probability of default



APPENDIX

Table A1. Variable definition and source

Name	***Type***	***Description***	***Source***
Switch Servicer	Binary	Denotes the decision of the originating lender to sell or to retain the mortgage servicing right of a given loan. Takes the value of 1 if the originator decides to sell the underlying MSR and 0 if the he retains the MSR and continues servicing the loan.	MBSData
Default	Binary	Denotes mortgage default. Takes the value of 1 if the borrower of a given mortgage misses three or more consecutive monthly payments (*i.e.* when the mortgage status is first labeled as 90+ days delinquent).	MBSData
FICO score	Continuous	The borrower's FICO score created and calculated by the Fair Isaac Corporation. It measures the credit quality of borrowers by taking into account individual's payment history, length of credit history, current level of indebtedness, and types of credit used by the borrower.	MBSData
FICO660	Binary	Takes the value of 1 if the borrower's FICO score is above 660 and 0 otherwise. In general, a FICO score above 660 indicates that the individual has a good credit history.	MBSData
LTV	Continuous	The Loan-To-Value ratio calculated as the percentage of the first-lien mortgage to the total value of the property. It is one of the key risk factors used by U.S. lenders when qualifying borrowers for a mortgage. A high LTV ratio mirrors a loan with low down payment for which the borrower has little equity stake in the property.	MBSData
LTV80	Binary	Takes the value of 1 if the LTV ratio is equal or higher than 80%.	MBSData
DTI	Continuous	The Debt-To-Income ratio calculated as the fraction of monthly mortgage payments to the borrower's monthly income. DTI measures the borrower's ability to honor periodic debt payments as it compares debt payments to the borrower's income.	MBSData
No/Low doc.	Binary	Takes the value of 1 if the documentation level is labelled "*missing*" or "*low*", and 0 otherwise. No- or Low-documentation mortgages designate loans for which the lender did not gathered a sufficient level of information on the borrower's reliability and credit worthiness.	MBSData
ln Amount	Continuous	The natural logarithm of the initial balance of the mortgage. Does not include neither interest nor taxes nor fees.	MBSData
Interest	Continuous	The interest rate initially applied at the time of original underwriting. Higher interest rates usually reflect loans granted for borrowers with inferior credit quality, which increase their monthly debt payments.	MBSData
ARM	Binary	Takes the value of 1 if the loan type is Adjustable-Rate Mortgage and 0 if Fixed-Rate Mortgage. *ARM* indicates whether the interest rate of a given mortgage is fluctuation over time based on a benchmark index plus an additional spread, called an ARM margin.	MBSData
ARM margin	Continuous	A fixed component added to the interest rate for ARM mortgages. The margin is constant throughout the lifetime of the mortgage while the benchmark index fluctuates over time according to general market conditions.	MBSData

Balloon	Binary	Takes the value of 1 if the mortgage has a balloon payment structure, 0 otherwise. Balloon mortgagors make only interest payments during the lifetime of the loan. At the term end, the borrower repays the entire principal at once.	MBSData
GSE conforming	Binary	Takes the value of 1 if the lender follows the GSEs' lending guidelines and 0 otherwise. Following the GSEs' recommendations, we classify a mortgage as conforming if the borrower's FICO score is above 660 and the loan amount was below the conforming loan limit in place at time of origination and the LTV is either less than 80% or the loan has private mortgage insurance in the case that the LTV ratio is above 80%. Since conforming loans meet the GSE lending standards, the conforming dummy variable indicates whether the mortgage was eligible to be sold to the GSEs at origination.	MBSData
Subprime	Binary	Denotes subprime mortgages. A mortgage is labelled "*Subprime*" at origination if the borrower's FICO score is lower than 580 or the LTV ratio is higher than 90%.	MBSData
Prime	Binary	Denotes prime mortgages. A mortgage is considered as "*Prime*" if the borrower's FICO score is higher than 660 and the LTV ratio is lower than 80%.	MBSData
Prep. Penalty	Binary	Equals to 1 if the mortgage contract includes a prepayment penalty clause, and 0 otherwise. Accordingly, the borrower will pay a penalty if he chooses to pre-pay the loan within a certain time period. The penalty is based on the remaining mortgage balance and the number of months worth of interest.	MBSData
Purchase	Binary	Takes the value of 1 if the loan purpose is labeled "*Purchase*" a property, and 0 otherwise.	MBSData
Refin. cash-out	Binary	Equals to 1 if the loan is granted for the purpose to refinance an existing loan with "*cash-out*". A cash-out refinance mortgage is a new loan in which the amount is greater than the existing mortgage amount, which will be refinanced. Since the borrower refinances for more than the amount owed, he/she takes the difference in cash.	MBSData
Refin. no cash-out	Binary	Equals to 1 if the loan is granted for the purpose to refinance an existing loan with "*no-cash-out*". A no-cash-out refinance mortgage is a new loan in which the amount is equal or lower than the existing mortgage amount. The main purpose of such loans is usually to lower the interest rate charge on the loan.	MBSData
Service fee	Continuous	The servicing fee that the servicer of the deal charges as a compensation for costs he bears. It is expressed as a fixed percentage of the declining balance of the mortgage.	MBSData
Age at default	Continuous	The age-at-default is measured as the total number of months since origination when the default is first recorded.	MBSData
Default N	Binary	Denoting the fraction of mortgages that default within *N* months since origination.	MBSData
Income	Continuous	The annual growth rate of personal income, which is defined as an individual's total earnings from wages, investment interest, and other sources. The seasonally unadjusted U.S. real disposable (after deducting tax) personal income data is retrieved from the US. Bureau of Economic Analysis' web site.	bea.gov
Divorce	Continuous	The annual divorce rate calculated as the ratio of the number of marriages contracted and ended in divorce and the numbers of all marriages contracted in the same year. The divorce rate is commonly used as an indicator of social stress in the society. The seasonally unadjusted divorce rate is retrieved from the U.S. Census Bureau' web site.	census.gov

GDP growth	Continuous	The annual growth rate of the U.S. Real Gross Domestic Product. The real GDP is collected from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org
HPI growth	Continuous	The annual growth rate of the House Price Index for the U.S. We use the seasonally unadjusted purchase-only HPI index retrieved from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org
σ interest	Continuous	The interest rate volatility calculated as the volatility on the 1-Year Treasury Constant Maturity Rate over the 24 months before origination. The monthly seasonally unadjusted treasury rate is collected from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org
Credit spread	Continuous	The yield spread between *AAA* and *Baa* bond indexes. It is calculated as the interest rate difference between Moody's Aaa and Baa Corporate Bond Yields. Both variables are seasonally unadjusted recorded on a monthly basis and retrieved from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org
Judicial	Binary	Takes the value of 1 if the state laws require judicial procedures to foreclose on a mortgage, and 0 if not. The variable is compiled based on information from the National Center for State Courts' web site.	ncsc.org
SRR	Binary	Stands for Statutory Right of Redemption and takes the value of 1 if the state has statutory redemption laws. The variable is compiled based on information from the National Center for State Courts' web site.	ncsc.org

Table A2. Results of the Probit model using +60 days default definition

The table reports estimation results of the parametric Probit regressions. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. The dependent variable, *Default*, is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +60 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo R^2* is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0036^{***}	-0.0036^{***}	-0.0036^{***}	-0.0035^{***}	-0.0036^{***}	-0.0036^{***}	-0.0036^{***}	-0.0036^{***}	-0.0037^{***}
LTV ratio	0.0164^{***}	0.0166^{***}	0.0165^{***}	0.0165^{***}	0.0164^{***}	0.0168^{***}	0.0165^{***}	0.0166^{***}	0.0173^{***}
ARM	0.0773^{***}	0.1117^{***}	0.1087^{***}	0.0858^{***}	0.0660^{***}	0.0587^{***}	0.0735^{***}	0.0713^{***}	0.1010^{***}
Balloon	0.6303^{***}	0.5647^{***}	0.5731^{***}	0.5852^{***}	0.6350^{***}	0.6368^{***}	0.6310^{***}	0.6240^{***}	0.4114^{***}
No/Low doc.	0.3740^{***}	0.3758^{***}	0.3758^{***}	0.3721^{***}	0.3687^{***}	0.3631^{***}	0.3736^{***}	0.3709^{***}	0.3417^{***}
GSE Conf.	-0.1844^{***}	-0.1817^{***}	-0.1798^{***}	-0.1825^{***}	-0.1810^{***}	-0.1871^{***}	-0.1823^{***}	-0.1819^{***}	-0.1475^{***}
B. Economic general conditions									
GDP growth		-14.788^{***}							-1.8866^{***}
C. Housing market conditions									
HPI growth			-3.5125^{***}						-7.6803^{***}
D. Bond market conditions									
σ interest				0.4624^{***}					1.0581^{***}
Credit spread					0.3491^{***}				1.8732^{***}
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0447^{***}		-0.0412^{***}
SRR								-0.0752^{***}	-0.0737^{***}
Intercept	0.5435^{***}	0.9450^{***}	0.8307^{***}	0.1598^{***}	0.8735^{***}	0.1316^{***}	0.5821^{***}	0.5876^{***}	2.1942^{***}
Pseudo R^2	8.50	9.19	8.92	9.12	8.62	9.43	8.52	8.54	11.60
Log-likelihood	$-3.41e^{+06}$	$-3.39e^{+06}$	$-3.40e^{+06}$	$-3.39e^{+06}$	$-3.41e^{+06}$	$-3.38e^{+06}$	$-3.42e^{+06}$	$-3.42e^{+06}$	$-3.30e^{+06}$
Wald p-value									0.00
LR p-value									0.00

Table A3. Results of the Probit model using 2001-2006 period

The table reports estimation results of the parametric Probit regressions. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. The dependent variable, *Default*, is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +90 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo* R^2 is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0034***	-0.0033***	-0.0034***	-0.0033***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0035***
LTV ratio	0.0169***	0.0169***	0.0168***	0.0172***	0.0169***	0.0170***	0.0170***	0.0170***	0.0178***
ARM	0.0978***	0.1157***	0.1028***	0.1191***	0.0820***	0.0824***	0.0942***	0.0930***	0.0885***
Balloon	0.6464***	0.6156***	0.6361***	0.5640***	0.6522***	0.6571***	0.6469***	0.6412***	0.4378***
No/Low doc.	0.3444***	0.3476***	0.3455***	0.3398***	0.3375***	0.3379***	0.3440***	0.3417***	0.3064***
GSE Conf.	-0.1921***	-0.1946***	-0.1928***	-0.1834***	-0.1869***	-0.1979***	-0.1902***	-0.1901***	-0.1474***
B. Economic general conditions									
GDP growth		-8.4588***							11.941***
C. Housing market conditions									
HPI growth			-0.7645***						-6.5539***
D. Bond market conditions									
σ interest				0.6250***					1.4068***
Credit spread					0.4258***				1.8676***
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0413***		-0.0402***
SRR								-0.0621***	-0.0545***
Intercept	0.2735***	0.4876***	0.3351***	-0.2557***	0.6864***	-0.1608***	0.3083***	0.3090***	1.1864***
Pseudo R^2	8.22	8.41	8.24	9.37	8.42	9.21	8.24	8.25	11.50
Log-likelihood	-3.03e^{+06}	-3.03e^{+06}	-3.03e^{+06}	-2.99e^{+06}	-3.03e^{+06}	-3.00e^{+06}	-3.03e^{+06}	-3.03e^{+06}	-2.92e^{+06}
Wald p-value									0.00
LR p-value									0.00

Table A4. Results of the Probit model using 2001-2006 period and +60 days default definition

The table reports estimation results of the parametric Probit regressions. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. The dependent variable, *Default,* is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +60 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo* R^2 is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0036***	-0.0036***	-0.0036***	-0.0035***	-0.0036***	-0.0036***	-0.0036***	-0.0036***	-0.0037***
LTV ratio	0.0163***	0.0164***	0.0163***	0.0166***	0.0163***	0.0163***	0.0164***	0.0165***	0.0172***
ARM	0.0775***	0.0954***	0.0826***	0.0986***	0.0619***	0.0660***	0.0740***	0.0735***	0.0694***
Balloon	0.6409***	0.6099***	0.6303***	0.5582***	0.6465***	0.6543***	0.6414***	0.6367***	0.4336***
No/Low doc.	0.3452***	0.3486***	0.3465***	0.3407***	0.3383***	0.3402***	0.3448***	0.3431***	0.3080***
GSE Conf.	-0.1820***	-0.1845***	-0.1827***	-0.1733***	-0.1769***	-0.1885***	-0.1803***	-0.1803***	-0.1382***
B. Economic general conditions									
GDP growth		-8.4972***							11.542***
C. Housing market conditions									
HPI growth			-0.7942***						-6.5161***
D. Bond market conditions									
σ interest				0.6198***					1.3822***
Credit spread					0.4142***				1.8360***
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0391***		-0.0388***
SRR								-0.0502***	-0.0428***
Intercept	0.5317***	0.7481***	0.5958***	0.0110	0.9336***	0.1011***	0.5649***	0.5605***	1.4420***
Pseudo R^2	8.31	8.49	8.32	9.43	8.49	9.25	8.32	8.33	11.51
Log-likelihood	-3.08e^{+06}	-3.07e^{+06}	-3.08e^{+06}	-3.04e^{+06}	-3.07e^{+06}	-3.04e^{+06}	-3.08e^{+06}	-3.08e^{+06}	-2.97e^{+06}
Wald p-value									0.00
LR p-value									0.00

Table A5. Results of the Two-stage and Bivariate Probit models using +60 days default definition

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S. *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e*. is labelled as +60 days delinquent). *Switch serv*. denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption, and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0006***		-0.0002***				
Divorce	0.3028***		0.2069***				
Pr(Default=1)		0.5197***					
Ê(Default)				0.4443***	0.1183***		
Default					0.3260***		
FICO score	-0.0037***		-0.0012***			-0.0037***	-0.0001***
LTV ratio	0.0174***	0.0030***	0.0051***	0.0007***	0.0007***	0.0174***	0.0030***
ARM	0.1018***	-0.1840***	0.0371***	-0.0749***	-0.0749***	0.1005***	-0.1712***
Balloon	0.4097***	-0.0194***	0.1557***	-0.0435***	-0.0435***	0.4053***	0.0582***
No/Low doc.	0.3416***	0.1590***	0.1087***	0.0472***	0.0472***	0.3431***	0.1696***
GSE Conf.	-0.1446***	0.0785***	-0.0426***	0.0676***	0.0676***	-0.1434***	0.0015
GDP growth	-4.7726***	4.3899***	-1.8273***	3.5229***	3.5229***	-1.8682***	-0.5094***
HPI growth	-7.6137***	-5.8278***	-2.5897***	-0.9331***	-0.9331***	-7.6081***	-7.7881***
σ interest	0.9214***	0.6408***	0.2792***	0.1016***	0.1016***	1.0572***	0.8377***
Credit spread	1.9721***	1.1882***	0.6366***	0.0236***	0.0236***	1.8768***	1.9701***
Judicial	-0.0416***	0.0122***	-0.0130***	0.0062***	0.0062***	-0.0410***	0.0019
SRR	-0.0729***	0.0294***	-0.0233***	0.0118***	0.0118***	-0.0743***	0.0326***
R^2	11.7	3.8	14.1	31.2	38.6		
ρ						0.6190***	

Table A6. Results of the Two-stage and Bivariate Probit models using 2001-2006 period

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e*. is labelled as +90 days delinquent). *Switch serv*. denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption, and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0012***		-0.0004***				
Divorce	3.8303***		1.3157***				
Pr(Default=1)		0.6350***					
Ê(Default)				0.3639***	0.0550***		
Default					0.3089***		
FICO score	-0.0035***		-0.0011***			-0.0035***	-0.0001***
LTV ratio	0.0181***	0.0004***	0.0050***	0.0002***	0.0002***	0.0179***	0.0031***
ARM	0.0926***	-0.2652***	0.0328***	-0.0912***	-0.0912***	0.0866***	-0.2397***
Balloon	0.4051***	-0.0681***	0.1604***	-0.0403***	-0.0403***	0.4340***	0.0663***
No/Low doc.	0.2985***	0.0895***	0.0904***	0.0296***	0.0296***	0.3079***	0.1282***
GSE Conf.	-0.1384***	0.0823***	-0.0376***	0.0550***	0.0550***	-0.1427***	0.0196***
GDP growth	7.5664***	16.7512***	1.8333***	7.3254***	7.3254***	11.7905***	18.9797***
HPI growth	-3.3490***	-5.6139***	-1.1534***	-0.8432***	-0.8432***	-6.4116***	-6.9225***
σ interest	0.5626***	1.0320***	0.1439***	0.2578***	0.2578***	1.3962***	1.3139***
Credit spread	1.7920***	1.6191***	0.5576***	0.1905***	0.1905***	1.8542***	1.9820***
Judicial	-0.0420***	0.0109***	-0.0127***	0.0062***	0.0062***	-0.0410***	0.0088***
SRR	-0.0525***	0.0467***	-0.0162***	0.0161***	0.0161***	-0.0545***	0.0553***
R^2	12.1	3.0	14.0	30.4	37.3		
ρ						0.6004***	

Table A7. Results of the Two-stage and Bivariate Probit models using 2001-2006 period and +60 days default definition

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e*. is labelled as +60 days delinquent). *Switch serv*. denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption, and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0012***		-0.0004***				
Divorce	3.7299***		1.3072***				
Pr(Default=1)		0.5789***					
Ê(Default)				0.3292***	0.0136***		
Default					0.3156***		
FICO score	-0.0038***		-0.0012***			-0.0037***	-0.0001***
LTV ratio	0.0174***	0.0001**	0.0049***	0.0004***	0.0004***	0.0172***	0.0031***
ARM	0.0734***	-0.2599***	0.0269***	-0.0881***	-0.0881***	0.0690***	-0.2402***
Balloon	0.4015***	-0.0562***	0.1566***	-0.0329***	-0.0329***	0.4295***	0.0663***
No/Low doc.	0.3003***	0.0931***	0.0929***	0.0319***	0.0319***	0.3090***	0.1280***
GSE Conf.	-0.1294***	0.0791***	-0.0370***	0.0531***	0.0531***	-0.1337***	0.0191***
GDP growth	7.3051***	16.9002***	1.8576***	7.4274***	7.4274***	11.4314***	18.9506***
HPI growth	-3.4066***	-5.7247***	-1.1831***	-0.9145***	-0.9145***	-6.4057***	-6.9189***
σ interest	0.5607***	1.0537***	0.1501***	0.2713***	0.2713***	1.3720***	1.3126***
Credit spread	1.7675***	1.6508***	0.5631***	0.2088***	0.2088***	1.8242***	1.9798***
Judicial	-0.0406***	0.0102***	-0.0125***	0.0058***	0.0058***	-0.0388***	0.0089***
SRR	-0.0409***	0.0438***	-0.0125***	0.0144***	0.0144***	-0.0435***	0.0558***
R^2	12.0	2.9	14.2	30.4	37.7		
ρ						0.6230***	