Securitization and Loan Performance: A Contrast of Ex Ante and Ex Post Relations in the Mortgage Market[1]

Wei Jiang[2] Ashlyn Nelson[3] Edward Vytlacil[4]

First Draft: June 2009; This draft: July 2010

[1]The authors thank a major national mortgage bank for providing the data and assistance in data processing and the National Science Foundation (NSF Grant #SES-0851428) for financial support. Comments and suggestions from Chris Mayer, Tomasz Piskorski, David Scharfstein, Amit Seru and seminar/conference participants at Columbia, Georgia State, George Mason, Fudan, Peking University, Tsinghua, SHUFE, and the NBER Summer Institute have contributed to this draft. The authors also thank Greg Klochkoff, An Li, and Slava Fos for excellent research assistance.

[2]Corresponding author. The Graduate School of Business, Columbia University, Tel: 212 854 9002, Email: wj2006@columbia.edu.

[3]Indiana University, School of Public and Environmental Affairs, Tel: 408 674 3053, Email: ashlyn@indiana.edu.

[4]Yale University, Department of Economics, Tel: 203 436 3994, Email: edward.vytlacil@yale.edu.

Securitization and Loan Performance:
A Contrast of Ex Ante and Ex Post Relations in the Mortgage Market

ABSTRACT

This study presents an intriguing contrast of the ex ante and ex post relations between mortgage securitization and loan performance. While the paper supports prior research in that the bank applies lower screening efforts on loans that have higher ex ante probability of being securitized, it further shows that loans remaining on the bank's balance sheet are, ex post, of worse quality than sold loans. Most of the differences can be explained away by secondary market investors' information advantage over the originating bank due to the time lag between loan origination and loan sale. While many blame the presence of the secondary market for the emergence of "liars' loans," we find that ironically these loans hurt the originating bank more than it did the secondary market.

1 Introduction

Traditional banks were lenders that held loans until they matured or were paid off. These loans were funded by direct obligations of the bank, principally by deposits and sometimes by debt. Such a model no longer describes modern banks or other financial intermediaries that increasingly combine assets into pools, which are split into shares through securitization, and sold to investors who share the risk and reward of the performance of those assets. Loan sale (securitization) has the benefits of reducing the impact of bank-specific or local funding shocks on credit supply and reducing the cost of funding by enhancing a bank's liquidity (Loutskina and Strahan 2007). On the other hand, loan sale inevitably gives rise to agency problems when the lender does not bear the full consequences of its actions that affect loan performance. Such agency problems include weakened incentives for monitoring by the original lender (Gorton and Pennacchi 1995, Drucker and Puri 2009) and inefficiency in both the ex ante contracting with multiple creditors (Gilson, John, and Lang 1990, Bolton and Scharfstein 1996) and ex post renegotiation (Sufi 2006).

Loan securitization rose to the headlines during the financial crisis that started in 2007. Many consider loan sale to be a major cause for the loosening lending standards that led to the mortgage crisis (Keys, Mukherjee, Seru, and Vig 2010b) and the difficulty in renegotiation that continues to aggravate the crisis (Piskorski, Seru, and Vig 2009). There are two potential effects of loan sales on loan quality for the mortgage market. At the macro level, the rapid expansion of loan originations – especially among low-documentation loans originated through the broker channel (Jiang, Nelson, and Vytlacil 2009) – would not have been possible if the loans could not be sold but instead had to remain on the originating bank's balance sheet. The originating bank would not have had sufficient liquidity to rapidly expand its loan offerings without the existence of a secondary market for loan sales. At the micro level, a loan of dubious quality is more likely to be originated if it is expected to be sold (Keys, Mukherjee, Seru, and Vig 2010b).

This paper explicitly distinguishes between the ex ante versus ex post relationships between loan performance and loan sale. The ex ante relationship is the relationship as perceived by the bank when deciding whether to originate the loan. The ex ante relationship is thus one between the probability that the loan will eventually become delinquent and the probability that the loan will be sold, given the information known by the bank at the time of loan origination. The ex post relationship applies at the time the loan is offered for sale, after it has been originated. The ex post relationship is thus one between the probability that the loan will eventually perform poorly and the sale status of the loan, conditional on the loan having already been originated and given the information known to market participants at the time of loan sale. As we discuss further below, there is often a several month gap in time between loan origination and when the loan is offered for sale.

Using a unique data set from a leading national mortgage bank, we analyze both the ex ante and ex post relationships between loan performance and loan sales, and we find an intriguing contrast between the two. While we confirm the finding of prior research that loans with higher ex ante probability of being sold entail *higher* delinquency rates, we further show that, conditional on the information known by the bank at the time of loan origination, loans actually sold by the bank have a *lower* delinquency rate than the loans retained. The implications of such a contrast is two-fold. On one hand, the prospect of unloading delinquency risk to investors in the secondary market weakens the lending bank's incentive to carefully screen borrowers and offer them appropriate contracts, which leads to an average deterioration in loan quality compared to the state of no securitization. On the other hand, the agency problem on the bank's part ironically hurts the lending bank more than the secondary market investors because the adverse selection works against the bank once the loans are originated: Investors are able to select relatively higher quality loans for purchase by exploiting information revealed between the time of loan origination and the time of loan sale that is predictive of loan performance, including specific information on individual loans (such as the current payment status), borrower status (such as changes in credit score and debt balances), general information about loan products (such as the performance of particular categories of loans), and about the neighborhoods in which the mortgaged properties reside (such as the change in housing prices or unemployment rates).

The idea can be graphically illustrated as follows. Suppose there exists a state variable that does not affect loan performance but has a direct impact on the probability of loan sale. Fig. 1(a) plots such a situation where each dot represents a loan observation and where loans are partitioned into high and low sale probability groups by the state variable. In the absence of any moral hazard, i.e., if the prospect of loan sale does not affect the bank's screening efforts, the delinquency probability of loans with higher loan sale propensity (in the right pane) should be about the same as that of loans with lower sale propensity (in the left pane). Once moral hazard is introduced, the delinquency probability of the easier-to-sell group (as classified by the state variable) becomes stochastically higher than the hard-to-sell group as Fig. 1(b) indicates. Finally, after the loans are originated and are offered for sale, investors are able to pick the relatively better loans from each group (though they tend to buy more loans from the easier-to-sell group), as shown in Fig. 1(c). As a result, loans retained by the bank are, ex post, worse on average despite the higher delinquency rates among loans with ex ante higher propensity for sale.

[Insert Figure 1 here.]

Our paper is related to an expanding list of recent empirical papers that analyze the relationship between securitization and loan performance and the role that loan securitization played in the financial crisis (Keys, Mukherjee, Seru, and Vig 2010b, Mian and Sufi 2009, Johnson, Mayer, and Faltin-Traeger 2009, Bubb and Kaufman 2007, Piskorski, Seru, and Vig 2009, Benmelech, Dlugosz, and Ivashina 2007, Elul 2009, Adelino, Gerardi, and Willen 2009). Our paper differs from the prior and concurrent research on this topic in two main aspects. First, we present an integrated analysis relating both the ex ante prospects of loan sale and the ex post actual securitization to loan performance. The opposite effects from the two stages help reconcile the mixed evidence documented in the prior research,[1] and explain the irony that while many blame the lending banks for unloading low quality loans to investors through securitization, these same banks also suffered the heaviest losses among all financial institutions during the crisis.

Second, we observe all loan and borrower attributes collected by the bank at the time of loan origination, including data on loan contract terms, property characteristics, and borrower attributes. In comparison, most research in the literature is based on commercial or government agency loan databases which usually do not include borrower demographic characteristics or detailed loan contractual terms, and sometimes include only particular types of loans (such as subprime loans or securitized loans). The comprehensive information available to our research not only provides us a better understanding of the determinants of loan delinquency and loan sale, but also provides us an accurate calibration of the information possessed by the bank in making loan origination decisions, information that is essential for our analyses of the moral hazard and adverse selection problems in the loan market.

The prior work that is the most closely related to ours is Keys, Mukherjee, Seru, and Vig (2010b) , which focuses on the ex ante relation between securitization and loan performance using a dataset comprised exclusively of securitized loans. Our work complements theirs by using a comprehensive dataset of all loans originated by the bank (including both securitized and unsold loans) to highlight how the direction of the effect of securitization on loan performance switches in ex ante and ex post settings. In our ex ante analysis, we apply the methodology of Keys, Mukherjee, Seru, and Vig (2010b) to identify the impact of securitization on loan performance by exploiting the discontinuity in the probability of securitization at certain credit score cutoffs. We also extend their analysis by incorporating other covariates (borrower and loan characteristics) in the analysis, allowing the relationship between these variables and credit scores to be discontinuous at the same thresholds at which the probability of securitization is expected to be a discontinuous function of credit score.

[1]For example, Elul (2009) and Keys, Mukherjee, Seru, and Vig (2010b) found a negative relation between securization and loan performance. Work by Benmelech, Dlugosz, and Ivashina (2007) and Bubb and Kaufman (2007) , on the other hand, support no difference in performance due to securization.

Such an extension is made possible by both the partial linear model we adopt and our access to the exhaustive set of variables that the bank observes at the time of loan origination. While we confirm Keys, Mukherjee, Seru, and Vig's (2010b) main result that the discontinuity in the ex ante ease of securitization around certain credit score threshold values is associated with a discontinuity in ex post loan performance, we further show that the same threshold values are also associated with jumps in other covariates which also impact loan performance.

Importantly, controlling for a comprehensive set of covariates reveals new and different aspects of weakened screening efforts by the lending bank due to the ease of securitization. More specifically, the weakened screening for borrowers just above the 620 credit score threshold (commonly considered to be the cut-off between "poor" and "OK" credit quality) mostly relates to hard-to-observe/quantify attributes that affect loan quality, and therefore, the jump in the delinquency rates is preserved or even strengthened when the jumps in other covariates are controlled for. On the other hand, when a borrower's credit score surpasses the threshold value of 660 (the score that separates "OK" from "Good" credit quality), the bank is more likely to be looser on lending standards based on observables. As a result, the jump in the delinquency rates is reduced to insignificance after controlling for the effect of the other covariates.

The rest of the paper is organized as follows. Section 2 provides a data description and an overview of the institutional background of mortgage securitization. Section 3 models the determinants of loan sale. Section 4 and Section 5 analyze the ex ante and ex post relations between loan sale and loan performance, respectively. Finally, Section 6 concludes.

2 Data Overview and Institutional Background

2.1 Sample description

Our proprietary data set contains all $721,767$ residential mortgage loans (includes prime, Alt-A, and subprime mortgages) funded by a leading national mortgage bank between January 2004 and February 2008. The data set contains all information recorded by the bank at loan origination, including the loan contract terms, property data, and borrower financial and demographic data, as well as monthly performance data updated through January 2009. Panel A of Table 1 details the definitions of all major variables used in the paper, and Panel B reports their summary statistics.

[Insert Table 1 here.]

More specifically, loan contractual information includes all variables used to characterize the mortgage, including: product category (fixed versus adjustable rate, payment structure, prepayment penalty structure, etc.), loan purpose (home purchase versus refinance, primary versus secondary lien), loan size (loan size, loan-to-value ratio (LTV) and combined loan-to-value ratio (CLTV)), interest rates (at initiation and at the sample close), origination channel (broker versus bank origination), documentation requirement (full-documentation versus various levels of reduced documentation), and securitization status (whether the loan remained in the bank's portfolio or was sold on the secondary market to government-sponsored agencies, investment banks, or into private label securities).

In addition, each loan is linked to monthly performance data updated through January 2009, including unpaid loan balance and whether the loan status is current or delinquent. For delinquent loans, the monthly performance data specify the number of days past due (30, 60, 90, or more than 120 days), whether the loan is in a state of short sale, whether foreclosure proceedings were initiated, and whether the property is Real Estate Owned (REO) by the bank following foreclosure proceedings.

Our sample loans form 320 product categories spanned by details on interest rates, benchmark rates, rate adjustment periods, and payment options, etc. For our research purpose we classify them using three dummy variables. The first dummy variable is ARM, which takes the value of one if the loan is an adjustable rate mortgage. The second dummy variable is $OptionARM$, which takes the value of one if the loan is an option ARM (nicknamed "pick-a-payment") product that offers the borrower multiple options during the initial period, including a specified minimum payment, an interest-only payment, or a fully amortizing payment for 15 or 30 years. Option ARMs are often offered with a very low teaser rate (often as low as 1%), and most of the borrowers ended up choosing payment levels below full amortization. The third dummy, IO, classifies loans starting with interest-only payments, and these loans could have either fixed or adjustable rates. To create mutually exclusive categories, we exclude interest-only products from the $OptionARM$ category, and exclude both interest-only and option ARM products from the ARM category. Such a classification results in 11.4%, 16.4%, and 34.7% of our sample having ARM, $OptionARM$, and IO values of one.

Products that take the value one on any of these three dummy variables tend to enable borrowers to qualify for a larger loan than their income and credit condition would allow under the traditional fixed rate terms. Both option ARM and interest-only products can generate "negative amortization" which increases the chance of negative equity, especially in a stagnant or declining housing market. Not surprisingly, all such products are associated with higher delinquency rates.

Borrower economic data includes all financial and credit information collected during loan underwriting: Borrower income, assets and cash reserves; expenditures and debts; number of borrowers on the contract; employment characteristics (tenure in current job and whether the borrower is self-employed); credit score; and past bankruptcies and foreclosures. Borrower demographic data include variables collected under the Home Mortgage Disclosure Act (HMDA): race (white, black, Asian, and others), ethnicity (Hispanic versus non-Hispanic), gender, and age.

Property data include the exact property address, purchase price, property type (e.g., single or multifamily dwelling or condominium), and owner-occupancy status (whether the property is the borrower's primary residence, a second home, or investment property). Using geocoding software, we are able to link approximately three-quarters of the loans to their census tract, zip code, metropolitan statistical area (MSA), and county. We then link census tract level data demographic data (such as population count, median age, the racial/ethnic composition of residents) and economic data (such as the unemployment rate) from the Decennial Census and the Bureau of Labor Statistics. We also use the Internal Revenue Service's Individual Master File system to link zip code-level average household income information to each property address. Finally, we match property addresses to the MSA level (or county or state when MSA information is unavailable) to obtain housing price changes from a combination of three major indices: the Case-Shiller Index, the First American Index, and the Office of Federal Housing Enterprise Oversight (OFHEO) Index.

Our sample properties are present in all 50 states, and their distribution is roughly proportional to population density. Due to the business model of our sample bank that outsources origination and distribution, our sample has a significantly higher representation of broker originated loans, low-documentation loans, and loans that are securitized. The average loan size and the fraction of the borrowers who are Hispanic in our sample are higher than the corresponding national figures. On the other hand, the distribution of several key loan characteristics in our sample is comparable to the distribution in the general market, including the loan-to-value ratio, the representation of subprime loans, and credit score. Table 2 outlines the comparison between our sample and the general market, and we refer the readers to Jiang, Nelson, and Vytlacil (2009) for a more detailed description of the sample data and its representativeness.

[Insert Table 2 here.]

Though not completely representative, this particular bank's experience presents an amplified version of the boom-bust cycle that the national mortgage industry experienced from 2004 to 2008. The bank in our analysis enjoyed a 50% annual growth rate in terms of total loan values during the boom (from 2004 to 2006), followed by a 28% delinquency rate by early 2009. Both numbers are much higher than the national average. The bank's rapid growth in loan originations and the subsequent poor performance of those loans are both associated with the securitization of a high fraction of the bank's loans. As a result, this particular bank offers unique insight into the impact of securitization on mortgage performance.

2.2 About loan sale/securitization

The sample bank adopted an "outsource origination to distribution" business model in which the vast majority (89%) of loans originated during our sample period were pooled into mortgage-backed securities (MBSs) to be sold in the secondary mortgage market. Buyers included the Government-Sponsored Enterprises ("GSE agencies," including Freddie Mac, Fannie Mae, Ginnie Mae), investment banks, institutional investors (such as hedge funds), and private investors. Banks benefit from selling mortgages to the secondary market because such sales provide the bank with liquidity to originate new loans and collect additional origination fees, and because banks are able to off-load relatively concentrated sources of cash flow risk (interest rate, default, and prepayment risk) to the more diversified investors who purchase these loans as part of their portfolio strategies.

The bank's decision to sell a mortgage into an MBS or retain it in portfolio is primarily driven by the bank's ongoing capital requirements and by comparing the margin earned on the loan under different sell/hold scenarios. Conventional mortgages – underwritten and approved in accordance with agency guidelines, such as Fannie Mae's Desktop Underwriter and Freddie Mac's Loan Prospector – are often priced for agency sale immediately after origination. The MBS market works differently for non-agency, private investors. To price MBSs for non-conforming loans, private investors typically requested a 5-10% random sample of mortgages, along with all relevant origination data (excluding HMDA demographic data); some investors also requested an "adverse sample" of the lowest quality loans in the pool to assess the risk of buying the pool. Due to this procedural difference, mortgages sold to private investors typically remained on the bank's balance sheet for a longer period of time. In addition, investors could request a put-back option in their contract, allowing them to put back early-delinquent loans (usually loans becoming delinquent within 90-120 days past origination) to the lender. We refer the readers to Ashcraft and Schuermann (2006) for a detailed overview on the mortgage securitization process.

The sample bank has two distinctive institutional features related to securitization. First, the sample bank retained servicing rights for the almost all of its loans, allowing the bank to strip off a portion of the note rate as a servicing fee. Second, the bank was among the first movers in the Option/ARM/IO product space and generated many such loans during the sample period. These loans exhibited unusually high delinquency rates during the recession (when borrowers could not catch up with the balloon-nature payments) and housing market downtown (when the negative amortization led to negative equity).

3 Determinants of Loan Sales: Preliminary Analysis

As a first step, we analyze the determinants of loan sales by relating loan, borrower, and neighborhood characteristics to the outcome of loan sale by estimating the following probit model:

$$\begin{aligned} LoanSale_i^* &= \beta_1 LoanChar_i + \beta_2 BorrowerChar_i + \beta_3 Neighborhood_i + \lambda_t + \varepsilon_i, \qquad (1) \\ LoanSale_i &= (LoanSale_i^* > 0). \end{aligned}$$

In (1), $LoanSale_i^*$ is the latent propensity for a loan to be sold, and $LoanSale_i$ is the actual binary outcome. $LoanChar_i$, $BorrowerChar_i$, and $Neighborhood_i$ represent vectors of loan characteristics, borrower characteristics, and neighborhood characteristics associated with the mortgaged property's address; λ_t is a vector of year dummy variables; and ε_i is the disturbance term.

Given that most loans were originated with the expectation of being sold, and a great majority (89%) of the loans were sold on the secondary market, the relation between loan sales and the covariates mostly reflects the preferences of the secondary market investors. We do not include

HMDA protected class information (notably race/ethnicity, age, and gender) as covariates determining loan sale since the bank is not allowed to reveal such information to investors. Results are reported in Table 3.

[Insert Table 3 here.]

Column 1 of Table 3 reports that results that only include loan-level variables as covariates, while Column 2 additionally includes neighborhood level variables. The coefficients on the loan-level covariates are very similar across the two specifications. The table reveals the following patterns. First, investors in general avoid characteristics that are associated with a significantly higher probability that the loan will become delinquent (Jiang, Nelson, and Vytlacil 2009), such as loans with high combined leverage ($CLTV$), large loan amounts ($Loan$), second lien loans ($SecondLien$), loans with single borrowers ($OneBorrower$, a proxy for unmarried or single-earning households), low income where the income could be stated rather than verified in low-documentation loans ($Income$), low credit scores ($CreditScore$), loans for non owner-occupied properties ($OwnerOccup$) and loans originated by self-employed borrowers ($SelfEmploy$). Investors also prefer loans obtained by first time home buyers ($FirstOwner$) who are on average less likely to become delinquent. Of course, the characteristics that indicate higher risk should naturally be priced into the loan interest rate; investor taste for lower-delinquency loans may reflect investors' beliefs that the loan pricing does not adequately compensate for risk, or that the price-risk trade-off is too high for high risk loans.

Second, investors prefer loans that have lower prepayment and interest rate risk. We observe from Table 3 that investors prefer fixed-rate loans (ARM), refinance loans rather than home purchase loans ($Refinance$) and loans with a hard prepayment penalty ($HardPenalty$) – a payment penalty for refinancing the mortgage or selling the home within a particular time period, typically 1-3 years – which lower the investor's risk of losing the loan due to a future refinance or home sale. Investors' preference for loans with low cash reserves ($CashResv$) is consistent with the low prepayment risk hypothesis, but may also result from the fact that the bank often decided to retain loans for which the borrower had an established relationship with the bank (e.g., had deposit or credit card accounts with the bank), where such accounts tend to have relatively high cash reserves. On the other hand, option ARM ($OptionARM$) loans are more likely to be sold than fixed-rate ones, and the interest-only (IO) loans do not make a significant difference.

Third, loans originated through the broker channel ($Broker$) are slightly more likely to be sold, especially when the neighborhood characteristics are controlled for. Although Jiang, Nelson, and Vytlacil (2009) find that these loans are more likely to become delinquent, the differences in loan performance across origination channels did not materialize until approximately 2006. Hence investors might have been largely unaware of this issue for most of our sample period. As such, during the early part of the sample period, investors differentiated borrower and loan quality based on their reported attributes, rather than origination channel. Our discussions with bank officers also reveal that agency (GSE) investors did not request or use origination channel information at any point in the loan purchase decision; private investors did not request or use such information until 2006. Indeed, if we put the interactive term $Broker \cdot Year$ in the regression (not tabulated), the coefficient would be significantly negative at the 1% level. Further, fair lending guidelines prohibited banks from using a "two-door" policy in which banks provided different pricing for bank- versus broker-originated loans, so the delinquency risk associated with broker-originated mortgages would not have been captured in the loan interest rate. Finally, borrowers who had an established relationship with the bank are more likely to seek a mortgage directly from the bank, and such loans are more likely to be retained.

We also find strong evidence of investor taste for low-documentation loans ($LowDoc$) and loans with missing income information ($IncomeMiss$, a feature associated with low-documentation loans). Rapid expansion in low-documentation mortgages was driven by both strong borrower and investor demand for such products, fueled by a widespread perception that these loans cut down transaction cost and had favorable risk-reward trade-offs. Lower dimensions of information relevant for pricing also facilitate the packaging of mortgages in relatively homogenous pools. Moreover, borrowers paid pricing premiums for originating low-documentation loans, despite having similar credit and reported economic conditions as full-documentation borrowers.

Of course, the assumption that low-documentation borrowers would honor their payments at rates similar to full-documentation borrowers rested on the assumptions that adverse selection into low-documentation loans and information falsification by low-documentation borrowers were both relatively limited, assumptions which proved to be incorrect (Jiang, Nelson, and Vytlacil 2009). Yet, investor learning about the quality of low-documentation loans might only have occurred toward the later half of our sample period because it was not until 2007 that low documentation loans began to exhibit delinquency rates that were considerably higher than those of their full-documentation counterparts. Similarly, the interactive term $LowDocr \cdot Year$ (not tabulated) is significantly negative at the 1% level, suggesting investor learning over time.[2]

Fourth, examining the coefficients on the neighborhood level covariates reported in Column 2, investors seem to prefer buying loans backing properties in neighborhoods with low black ($PctBlack$) and Hispanic ($PctHispanic$) representation. Presumably these two groups of borrowers have higher delinquency rates on average, conditional on other observable characteristics (Jiang, Nelson, and Vytlacil 2009). While investors do not observe the racial and ethnic identities of the individual borrowers, they can infer the neighborhood demographics from the property addresses. Moreover, loans originated in areas that experienced higher recent housing price appreciation ($HPI6mBefore$) are less popular with investors. While appearing counter-intuitive, this relation is actually consistent with investors' preferences discussed above. Loans are more likely to be refinanced in the near future if the underlying property is in an area that has recently experienced rapid housing price appreciation. Moreover, such loans also suffer from a higher delinquency rate, a relation that we will analyze in more detail in the next section.

Finally, the year dummy variables indicate that loans were more likely to be sold in 2005 and 2006 (as compared to the base year, 2004), during a period of industry-wide expansion in the secondary mortgage market. The trend reversed itself in 2007 as several major banks failed and the secondary market evaporated following the colossal losses of the Bear Sterns hedge funds and the collapse of Lehman Brothers.

4 Prospects of Loan Sales and Delinquency Propensity: Ex Ante

4.1 Overview of Discontinuity if Loan Sale and Delinquency

We first analyze the ex ante relationship between the prospect of loan sale and the propensity for the loan to become delinquent. Based on economic theory and prior research, we expect this relationship to be positive because a higher ex ante probability for the bank to off-load the loan

[2] Demyanyk and Van Hermert (2009) argue that securitizers were partially aware of the deteriorating quality of loans during the years leading to the crisis. However, the problems were much masked by high house price appreciation.

from its own balance sheet to the secondary market weakens the bank's incentive to carefully screen loan applications, leading to higher delinquency rates. The identification of the causal effect is, however, non-trivial because of the potentially non-random selection of loans into securitization. Ideally, the analysis would rely on the existence of instrumental variables that affect the prospect of loan sale but are not correlated with loan quality (which affects delinquency probability) other than through the process of securitization.

There are no obvious candidates for such instruments. We therefore adopt a refined version of the approach of Keys, Mukherjee, Seru, and Vig (2010b) that achieves identification through discontinuities in the probability of loan sale as a function of credit score. We map the terminologies of regression discontinuity following Imbens and Lemieux (2008) to the variables in our context as follows: The "assignment variable" (also called the "treatment variable") is a dummy variable for whether the bank exercises weak screening effort ($WeakScreen$); the "forcing variable" (also called "running variable") is credit score ($CreditScore$); the "outcome variable" is delinquency ($Delinquency$); and the covariates include all loan and borrower characteristics in our sample. Finally, we apply a "sharp regression discontinuity" model in that the assignment variable switches from zero to one with probability one when the forcing variable gets across the threshold value. Our goal is to identify a jump in the outcome variable at the threshold value of the forcing variable, conditional on all the covariates (including the their possible jumps at the same threshold values of the forcing variable).

The validity of such a method requires that the credit quality of potential borrowers varies continuously with credit score, while there exist certain thresholds at which the bank's screening effort drops discretely because the ease of securitization as a function of credit score jumps at the threshold values. The combination of the continuity of potential borrower quality as a function of credit score and the discontinuities in the screening effort due to the ease of securitization allows the identification of a causal effect of the prospects of securitization on loan performance. Our method differs from Keys, Mukherjee, Seru, and Vig (2010b) in that we control for all observable loan characteristics (allowing for possible jumps in their conditional distributions at the same threshold values) while exploiting the discontinuities in probability of loan sale.

Based on institutional features of the mortgage market, we conjecture that the conditional probability of loan sale conditional on credit score should be discontinuous at two credit-score thresholds, at 620 and 660. A score of 620 is considered to be the cut-off between "Poor" and "OK" credit quality, and 660 separates "OK" from "Good." The mortgage market and banking regulatory agencies generally considers 620 (660) as the bottom (top) cut-of of prime (sub-prime) loans. We now verify this conjecture. To this end, we estimate the following probit model:

$$LoanSale_i^* = \sum_{j=2}^{18} \gamma_j (CreditScore_i \in CreditScoreRange_j) + \beta X_i + \lambda_t + \varepsilon_i, \tag{2}$$

$$LoanSale_i = (LoanSale_i^* > 0).$$

In (2), $CreditScoreRange_j$ is a dummy variable for the credit score to fall within one of the 17 ranges with an even width of 20 points, i.e., from $[\cdot, 499]$, $[500, 519]$, $[520, 539]$,..., to $[780, 799]$, and $[800, \cdot]$, with the first range serving as the omitted category in the regression. X_i is a vector of variables that includes all regressors appearing in the first column of Table 3 except $CreditScore$.

From (2) we obtain $\widehat{P}_j$, $j = 2, ..., 18$, the average partial effect of each credit score range corresponding to each $\widehat{\gamma}_j$ coefficient. The set of $\widehat{P}_j$ reflect the incremental probability of loan sale in each credit score category relative to the omitted category ($j = 1$). To back out the estimated probability of loan sale for all categories, conditional on the X covariates, we simply rescale $\widehat{P}_j$ so that the differences in the probabilities across different categories are maintained, while the expected number of loan sale computed from the probabilties is equal to the actual number in the sample. More specifically, we back out $\widehat{P}_1$ according to:

$$\sum_{i=1}^{n} LoanSale_i = \widehat{P}_1 \sum_{i=1}^{n} (CreditScore_i \in CreditScoreRange_1) + \sum_{j=2}^{18} \widehat{P}_j \sum_{i=1}^{n} (CreditScore_i \in CreditScoreRange_j),$$

and then add $\widehat{P}_0$ to all $\widehat{P}_j$ ($j \neq 1$) values. Fig. 2 plots the estimated probability of loan sales conditional on covariates against credit score ranges.

[Insert Figure 2 here.]

A prominent feature arising from Fig. 2 is that the estimated probability of loan sale is non-monotone in credit score. More specifically, the probability is roughly stable for credit scores up to 619. It rises steeply at the $[620, 639]$ range by about 2.7 percentage points, and then ascends substantially again at the $[660, 679]$ range by about 1.9 percentage points. The first jump confirms the premise for the identification strategy used in Keys, Mukherjee, Seru, and Vig (2010b) . The probability stays roughly steady at the $[660, 699]$ range, and then drops continuously as the score approaches the maximum level.

The two breaking points, 620 and 660, confirm our conjecture based on the the institutional features of the mortgage market. The empirical results imply that the most marketable loans are those with intermediate quality, possibly reflecting the secondary market's aversion to loans with high levels of information asymmetry (below 660, especially below 620) and to loans with relatively low yields (above 700).

It is worth noting that the increase in the conditional rate of loan sales (among all originated loans) may underestimate the unconditional rate of loan sale (among all potential loans) if the bank issues more loans to borrowers with credit scores above 620 and 660 precisely because of the perceived ease of securitization. Keys, Mukherjee, Seru, and Vig (2010a) provided a simple analysis on the issue. A significant increase in the number of securitized loans just above the threshold values is indication of a jump in the ease of securitization, provided that the distribution of credit score in the population (not in the sample of approved borrowers) does not exhibit similar jumps at those threshold values. Fig. 3 displays the histograms of sold loans with respect to credit score for the full sample and a more detailed view for loans by borrowers with credit score between 600 and 700. It shows that 620 and 660 are indeed the cut-off values that see the steepest jump in the frequency of sold loans. Given the continuous nature of credit scores, such jumps are unlikely to be present among all potential borrowers unless there is prevalent and accurate manipulations of credit score, an issue we will discuss in Section 4.2.2.

[Insert Figure 3 here.]

To the extent that banks have ex ante knowledge about the discrete jump in the ease of loan sale at the breaking points, the monitoring incentives should be significantly weakened for loans just above the thresholds compared to those just below. As a result, Keys, Mukherjee, Seru, and Vig (2010b) argue that the loan delinquency rate should show a local perverse change as the credit score improves across the breaking points. As a first step to analyze this hypothesis, we replicate Keys, Mukherjee, Seru, and Vig's (2010b) results by estimating the following univariate relation:

$$Delinquency_j = f(CreditScore_j) + \varepsilon_j,\ j = 500, ...800, \tag{3}$$

with $E(\varepsilon_j|CreditScore_j) = 0$, so that

$$\Pr[Delinquency_j = 1 \mid CreditScore_j] = f(CreditScore_j).$$

In (3), our full sample of 719,695 observations with credit score and loan performance information available are collapsed into group data where each group is indexed by j, a value of credit score between 500 and 800. Observations with credit scores below 500 (above 800) are combined with the first (last) group. The $f(.)$ function in (3) is estimated with a cubic polynomial function,[3] and the estimation is conducted separately on the three regions based on credit scores: $[500, 619]$, $[620, 659]$, and $[660, 800]$. This estimation procedure constraints $f(CreditScore)$ to be continuous within regions while allowed to be discontinuous at the 620 and 660 threshholds. We plot the actual delinquency rates ($\%Delinquency_j$), the expected delinquency rates ($\widehat{f}(CreditScore_j)$), and the 95% confidence interval associated with the estimates for the expected rates, in Fig. 4.

[Insert Figure 4 here.]

We now examine the evidence for a jump at the 620 and 660 thresholds using different degrees of smoothing. There are enough borrowers with credit scores exactly on either side of the thresholds that we can test for a jump without smoothing. The delinquency rate for the 529 loans with credit score exactly equal to 619 is 41.2%, compared to 46.8% for the 6,213 loans with credit scores equal to 620. This difference, at 5.6 percentage points, is both economically and statistically significant (t-statistic = 2.50). The delinquency rates at credit scores of 659 (3,788 loans) and 660 (5,592 loans) are 36.2% and 39.1%, respectively. The difference of 2.9 percentage points is again economically and statistically significant (t-statistic = 2.83). We can also compare delinquency rates averaging over a range of credit scores on either side of the cutoffs, improving precision at the cost of some degree of bias. Comparing the average delinquency rates in the two intervals on either side of the 620 cut-off, $[615, 619]$ vs. $[620, 624]$, the pattern is much weakened and the difference

[3] The $f(.)$ function can also be estimated using the nonparametric kernel method. We opt for the parametric method for two reasons. First, the resulting graph from nonparametric estimation is very similar to the polynomial function but, as expected, with wider standard errors bounds. Second, the asymptotic properties of nonparametric estimation in this case is non-standard because there is a fixed number of groups (i.e., the number of different credit score values). By estimating the function at the credit score group level, the aymptotics rely on the degeneracy of ε_j due to the increasing of observations within each group.

is no longer statistically significant. The average delinquency rate for the former is 43.2%, and that for the latter is 43.6%. The average rates for the range $[655, 659]$ and $[660, 664]$ are 36.4% and 38.5%, respectively. This difference remains statistically significant at the 5% level. The expected delinquency rates, estimated using the polynomial, however, only see significant (at the 5% level) jumps at the 660, but not at the 620 credit score value. The lack of significance at the latter breaking point is due to the fact that the discrete jump from 619 to 620 is more blurred if delinquency rates in the narrow range below 619 and that above 620 are considered.

4.2 Conditions for Identification through Discontinuity

4.2.1 Accommodating Jumps in Covariates: the Partial Linear Model

The first challenge to the identification method based on the discontinuity of an outcome as a function of one predictive variable without controlling for other covariates is the validity of the following key assumption: the conditional distribution of none of the other covariates that has predictive power for delinquency has a discrete jump at the same critical threshold levels of credit score. To assess this premise, we perform a two-sample mean difference t-test for the following null hypothesis for each of the covariates used in this study, testing a separate null for each of the two cutoffs:

$$\begin{aligned} H_{0j}: &\quad E(X_j|CreditScore \in [618, 619]) = E(X_j|CreditScore \in [620, 621]), \\ H'_{0j}: &\quad E(X_j|CreditScore \in [658, 659]) = E(X_j|CreditScore \in [660, 661]). \end{aligned} \tag{4}$$

Each null hypothesis is that the population mean of the particular covariate conditional on the credit scores being just below 620 (or 660) is equal to population mean conditional on the credit score being just above 620 (660). A rejection of a null indicates a jump in the conditional distribution of that covariate, and challenge the premise that the jump in the loan sale probability is the sole reason for the jump in delinquency. Results are reported in Table 4. In addition to the comparison of covariate means across the threshold values, the last column of Table 4 shows whether the individual covariates have significantly positive ("+"), significantly negative ("−"), or non-significant (at the 5% level, "*n.s.*") effects on delinquency based on a delinquency probit analysis using the same set of covariates (also see (Jiang, Nelson, and Vytlacil 2009)).

[Insert Table 4 here.]

It seems that the conditional means of most covariates have jumps at the same breaking points. While the jumps in most cases are associated with increased delinquency risk (such as ARM, $OptionARM$, IO, $Broker$, $HardPenalty$, $Hispanic$, $Loan$, $LowDoc$, and $OneBorrower$), some have the opposite effects (notably $Refinance$ and $InitialRate$). Therefore the identifying assumption that excludes jumps in conditional distributions all covariates except $CreditScore$ is questionable, which calls for a formal method that incorporates the effects on delinquency from other covariates, allowing for possible jumps in their conditional distributions at the same breaking points for $CreditScore$.

The method we use is a variant of a partially linear regression model (see, e.g., (Robinson 1988)). More specifically, the estimation entails two steps. In the first step, we group all observations by their credit scores. That is, an observation i with credit score equal to j belongs to group j. And we consider 301 groups where the credit score ranges from 500 to 800. About 2% of the observations fall out of this range, and we combine observations with credit scores below 500 (above 800) to the bin of $j = 500$ ($j = 800$). Suppose our delinquency prediction follows the linear probability specification:

$$Delinquency_i^j = f(CreditScore_i^j) + \gamma X_i^j + \varepsilon_i^j. \tag{5}$$

Subtracting from each variable its group mean with the same credit score on both sides of the above equation yields:

$$\begin{aligned} Delinquency_i^j - \overline{Delinquency}^j &= \left\{ f\left(CreditScore_i^j\right) - f\left(\overline{CreditScore}^j\right)\right\} + \gamma\left(X_i^j - \overline{X}^j\right) + \left(\varepsilon_i^j - \overline{\varepsilon}^j\right) \\ &= \gamma\left(X_i^j - \overline{X}^j\right) + \left(\varepsilon_i^j - \overline{\varepsilon}^j\right) \end{aligned} \tag{6}$$

Therefore, in the above regression, the variable $CreditScore$ is differenced out. The γ coefficients on other covariates could be consistently estimated by regressing the within-group difference in the dependent variable ($Delinquency_i^j - \overline{Delinquency}^j$) on the differences in all covariates except $CreditScore$, $\left(X_i^j - \overline{X}^j\right)$. This is equivalent to a linear regression of $Delinquency_i$ on X_i and a set of dummy variables set to one if the i-th observation's credit score is equal to all possible credit score values between 500 and 800.

In the second step, we plug the estimates $\widehat{\gamma}$ from (6) into (5) to form the function that attributes the residual delinquency rate to the credit score:

$$\widetilde{Delinquency}_i^j \equiv Delinquency_i^j - \widehat{\gamma} X_i^j = \widehat{f}(CreditScore_i^j) + \widehat{\varepsilon}_i^j. \tag{7}$$

In (7), $\widehat{f}(CreditScore_i^j)$ is equivalent to the estimated coefficients on the dummy variable $(CreditScore_i = CreditScore^j)$.

We conduct the two step estimation (equations (5) to (7) separately for three segments of data based on the following credit score ranges: $[500, 619]$, $[620, 659]$, and $[660, 800]$. For graphical illustration, we plot a smoothed version of $\widehat{f}(CreditScore_i^j)$ for each segement using the cubic polynomial function.[4] Results are plotted in Fig. 5.

[Insert Figure 5 here.]

After controlling for the effects of covariates (which are allowed to have jumps in their conditional distributions at the break points and to have different coefficients in different segements), we learn from Fig. 5 that the jump in delinquency rate at the 620 threshold is much strengthened. The point estimate of the jump is 4.5 percentage points, significant at the 5% level. On the other hand, the point estimate of the jump at the 660 threshold is weakened to 68 basis points, and is no longer significant at the conventional levels.

Both Fig. 4 and Fig. 5 and the associated analyses are informative about the effect of the jump in the ease of loan sale at credit score values of 620 and 660 on the loan screening incentive of the bank. However, the two figures also convey somewhat different aspects of the weakened screening efforts by the bank. The weakened screening for borrowers just above the 620 credit score mostly relates to attributes affecting loan quality that are hard to observe/quantify; therefore, the jump in delinquency rates is preserved when the jumps in other covariates are controlled for. On the other hand, when a borrower's credit score surpasses the threshold value of 660, the bank is more likely to be looser on lending standards based on observables. As a result, the jump in delinquency rates is reduced to insignificance after controlling for the covariate effects. Overall results support

[4] Alternatively, we could use a nonparametric method, such as a kernel function, to present the smoothed function $\widehat{f}(CreditScore_i^j)$. We find that the cubic polynomial function fits the data almost just as well while providing sharper confidence intervals.

the hypothesis regarding the positive relation between delinquency rates and the ex ante prospect of loan sale.

4.2.2 Manipulation of Credit Score and Sufficient Condition for Identification

The second major challenge to the identification through discontinuity is the possible manipulation of the forcing variable (i.e., credit score), especially around the threshold values (620 and 660 in our context). If loans are easier to obtain or could be obtained at more favorable terms just above the 620 or 660 scores than below them, potential borrowers might have an incentive to manipulate their credit scores in order to stand a better chance qualifying for loans. Such manipulation could span a full spectrum of behavior from legitimate credit management, to dubious credit "repair," and all the way to outright fraud. Though Fair Isaac strives to maintain the integrity of credit score by keeping secret its decision model and by constantly updating its model to accommodate and preempt strategic behaviors, we cannot rule out the possibility of manipulation; rather, we discuss the implications of manipulation.

First, we would like to point out that the assumption to rule out individual influence over their scores is fundamentally untestable (Lee 2008). MaCrary (2000) proposes a test based on the smoothness of the density of the forcing variable of all participants, which has been applied in some recent empirical papers on the mortgage market (e.g., Bubb and Kaufman 2007) . Though suggestive, MaCrary (2000) acknowledges that the smooth density across the threshold values is neither necessary nor sufficient for the presence of individuals' influence over their scores. Even if we had the credit score of all loan applicants – and not just that of all approved borrowers – a jump in the density at 620 or 660 does not necessarily contradict the absence of manipulation. Such a jump could occur if a potential applicant is more likely to actually apply for a loan if her credit score is 620 than if it were 619 knowing that the former is viewed significantly more favorably than the latter. Unfortunately, a large sample of credit score values that is representative of the population is not available.

Second, instead of contributing to the debate on the presence or absence of credit score manipulation, we take the view that, in our context, identification through discontinuity is valid even with the presence of score manipulation. Use the 620 threshold as an example and let 620^- and 620^+ denote credit scores immediately below and at or above 620. We posit the following as the necessary and sufficient condition our identification: Provided that the bank exercises the same due diligence in screening two otherwise identical loans except one has credit score 620^{-} and the other 620^{+}, the delinquency outcome as a function of credit score should be smooth at 620. Based on this principle, one of the following two conditions, both of which allow manipulation, would be sufficient for identification.

The first condition requires that the propensity to manipulation, conditional on credit score and other observable characteristics, is not positively correlated with the propensity of delinquency. Intuitively, it assumes that among borrowers with credit scores just below 620, the worse (i.e., high delinquency propensity) type along unobserved dimensions is no more likely to manipulate than the better type. While the assumption appears questionable for fraud-type manipulation – people who commit fraud presumably are the "worse" type as far as loan performance is concerned – it is plausible for legitimate credit management: successful management should be correlated with informedness and self-control, both are indications of the better, or at least no-worse type.

Suppose the first condition fails. The second, and weaker condition is that there is some randomness in the outcome of manipulation conditional on borrower characteristics and actions. This is the condition discussed by Lee (2008) . More formally, suppose borrowers who have pre-manipulation scores in the 620^{-} region aim to reach a score $v \geq 620$ through strategic behavior. In the absence of perfect control of their credit score, the outcome will be a random variable with a smooth (but not necessarily symmetric) density function around v. In this case, the manipulators (who are of the worse type if the first condition fails) do not form a discrete mass at 620 due to the randomness in the outcome of manipulation. As such, within the "manipulation range" the negative relation between delinquency propensity and credit score will be weakened or even perverted; nevertheless, there should be no discrete jump in delinquency at 620. In other words, identification of a discrete jump in delinquency at 620 is sufficient to conclude that the bank exercises different levels of screening care for loans at 620^{-} and those at 620^{+}.

We believe that the second condition is highly plausible. In the absence of a straight-forward and publicly known formula in calculating credit score and with the uncertainty in individuals' influence over the outcomes of their actions, it is unlikely that an individual is able to manage her score exactly to her target of 620 (or 660). Moreover, the bank in our sample pulled all three credit scores reported at the three different credit bureaus (TransUnion, Experian and Equifax) and used the median of the three credit scores to price the loan. Therefore, even if the borrowers knew what they could do to improve their score generally, it is impossible for them to influence the decisions at all three credit bureaus in such a way that their median score hits the 620 or 660 target without error. In summary, we conclude that the only plausible explanation for the discrete jumps shown in Fig. 4 and Fig. 5 is a weakened screening effort by the bank at credit score just above the 620 and 660 levels.

5 Actual Loan Sale and Delinquency: Ex post

5.1 Modeling Selection Effects

Results from the previous section suggest that the higher ex ante probability of loan sale is associated with the origination of more delinquency-prone loans. This relation does not, however, necessarily imply the same ex post allocation of loans between the bank (retained loans) and the secondary investors (sold loans). Availability of information on both securitized and non-securitized loans allows us to perform the ex post analysis on the relation between loan sales and loan quality.

Panel B of Table 1 shows that a simple two-sample comparison results in sold loans having a lower delinquency rate (28.2%) than that of the retained loans (32.9%), and the difference of 4.7 percentage points is statistically significant at the 1% level. Such a simple statistic is suggestive of a counter-intuitive result of worse quality loans being retained in the bank's portfolio. Moreover, the difference is significantly in favor of the sold loans for all the subsamples except the Bank/Full-Doc subsample, and more so in the Low-Doc subsamples than the Full-Doc ones. Such differences are suggestive of the hypothesis that the adverse selection works again the bank once loans are originated, except in the small sample of full-documentation loans issued directly by the bank where presumable the bank is more likely to possess soft information about loan quality. However, the simple statistics presented in Table 1 Panel B is by no means conclusive given the significant differences in loan attributes between sold and retained loans, as shown in Table 4. A formal analysis that examines the selection effect underlying the ex post relation between actual loan sale and delinquency is therefore necessary.

To assess the performance of sold loans relative to those retained, one potential approach would be to include the dummy variable *LoanSold* as a regressor in a delinquency regression and to make inference about the treatment effect of loan sale on delinquency based on the magnitude and statistical significance of the coefficient. We refrain from conducting such an analysis since it does not answer our research question. Such an analysis would answer a treatment effect question: if two ex ante identical loans – identical along both observable and unobservable dimensions – were assigned to different sold and retained status, how would their delinquency propensity differ ex post? We argue that there is no conventional "treatment effect" of loan sales in our context: not only all loans, both sold and retained, are serviced by the bank, but also post-origination monitoring has little effect on delinquency.[5] It is the screening at loan origination that has the critical impact on loan quality. As a result, the difference in the delinquency outcome that is correlated with loan sales should be predominantly attributed to the "selection effect;" that is, the bank, or the investors, or both are offering or picking loans based on the information they possess at loan origination and purchase, and such information is correlated with delinquency propensities.

We estimate the following bivariate probit model that controls for the information set of the bank at the time of loan origination, and that of the investors at the time of loan sale:

$$
\begin{aligned}
Delinquency_i^* &= X_i\beta + \varepsilon_i, \\
Delinquency_i &= (Delinquency_i^* > 0), \\
LoanSold_i^* &= Z_i\gamma + \eta_i, \\
LoanSold_i &= (LoanSold_i^* > 0).
\end{aligned}
\tag{8}
$$

In (8), $Delinquency_i^*$ ($Delinquency_i$) and $LoanSold_i^*$ ($LoanSold_i$) represent the latent propensity of loan delinquency and loan sale. X is the full vector of observable characteristics that predict delinquency (as analyzed in Jiang, Nelson, and Vytlacil 2009) that are known by the bank at the time of loan origination. The Z vector overlaps with, but is not identical to X (details to follow), and represents secondary market investors' information set at the time of loan sale. We impose that (ϵ_i, η_i) is distributed bivariate normal.

The bivariate probit system in (8) differs from a collection of two independent probit equations in that the two residuals in the system are allowed to be correlated, that is, $\rho = corr(\varepsilon_i, \eta_i) \neq 0$. It

[5] The same could not be said about foreclosure. Piskorski, Seru, and Vig (2009) argue that banks are less diligent in renegotiating with the delinquent borrowers if the loans are securitized, which results in more foreclosure among the securitized vs. retiained loans *conditional on* delinquency. We are able to replicate their result using our sample.

is important to note that the correlation between ϵ_i and η_i is identified semiparametrically using that Z_i contains elements not contained in X_i, so that while we will exploit the joint normality assumption in estimation the identification of ρ is not exclusively driven by the joint normality assumption.

Given the joint normality assumption, system (8) can be estimated using the full-information maximum likelihood (FIML) method. Define:

$$q_{i1} = 2Delinquency_i^* - 1,$$
$$q_{i2} = 2LoanSold_i^* - 1.$$

Then the integrated log-likelihood function becomes:

$$\ln L = \sum_i \ln \Phi(w_{i1}, w_{i2}, \rho^*),$$

where $\Phi(\cdot,\cdot,\cdot)$ is the cdf of the standard bivariate normal distribution, and

$$w_{i1} = q_{i1}(X_i\beta);\ w_{i2} = q_{i2}(Z_i\gamma);\ \rho^* = q_{i1}q_{i2}\rho.$$

The sign of $\rho = corr(\varepsilon_i, \eta_i)$ relates to three hypotheses on the relative informational advantage of the bank vis-à-vis the investors in the secondary market.

Hypothesis One: The bank does not possess any soft information (beyond the observable X variables), or it does not use such information in deciding whether to sell a particular loan to the secondary market. The investors do not have additional information, either. Under this hypothesis, $\rho = 0$, that is, a loan is randomly sold conditional on observables.

Hypothesis Two: The bank possesses soft information (beyond the observable X variables) that is predictive of delinquency, and it uses such information in deciding whether to sell a particular loan to the secondary market. The investors do not have additional information. Under this hypothesis, $\rho > 0$, that is, worse loans tend to be sold conditional on observables.

Hypothesis Three: The bank tries to sell as many loans as possible. Investors in the secondary market possess better information about the loans' prospects at the time of the loan sale than the bank did at loan origination. Under this hypothesis, $\rho < 0$, that is, better loans tend to be sold conditional on observables.

Prior to loan sale, investors are able to access the full set of hard information collected by the bank at the time of loan origination, except for HMDA-reported protected class data (e.g. borrower race, gender, ethnicity, and age). The bank is legally prohibited from releasing the HMDA data to investors, though the law also does not allow the bank to use HMDA information in loan approval and pricing decisions. Thus, it is important to emphasize the potential information advantage of investors (including government agencies, investment banks, and hedge funds) relative to the bank comes mainly from the time lag between loan origination and loan sale, typically in months with variations related to general market conditions.

While most loans that meet criteria for sale to government-sponsored enterprises (Fannie Mae, Freddie Mac, and Ginnie Mae) are sold immediately or within 30 days, loan sales to non-government agency investors typically take longer; on average, loan sales are completed 60 days after origination, although longer time horizons (up to 120 days after origination) are not unusual. After 120 days on the bank's balance sheet, the probability of loan sale drops dramatically. In addition, investor contracts often contain "put" options, which allow investors to return loans to the bank if they become delinquent soon after sale. Hence a delay in loan sales means loans that become delinquent within a few months of origination will not be sold.

During the interval between loan origination and sale, investors may gain additional information that is crucial for delinquency predictions, such as changes in real estate values in the region where the loan was originated, changes in the default rate for particular loan types, or changes in the borrower's credit score or debt load (this information is obtained by pulling an updated credit report). Such updated information may prove powerful in a quickly changing market environment, like the one experienced since the second half of 2007. These combined forces could turn the table against the bank if the bank's information advantage at the loan origination stage was not strong enough. On the net, the corresponding ρ could turn negative.

5.2 Negative Ex Post Relation between Loan Sale and Delinquency

Table 5 reports the estimates of the system (8), with the correlation coefficients ρ and the associated Wald test statistics highlighted in the bottom of the table. In this table, we take the full set of individual loan and borrower variables as our X variables, and we exclude personal demographic variables (gender, age, ethnicity, and race) from the Z variables (consistent with the practice in Table 3). The estimated ρ coefficients are negative in all subsamples, and are highly significant in all subsamples except for the Bank/Full-Doc subsample in which the estimated ρ is still significant at the 10% level. These results indicate that Hypothesis Three above is the dominant force in our data. Moreover, the estimated ρ in the Broker subsamples ($-.110$ and $-.102$) are more than twice as high in magnitude as in the Bank subsamples ($-.037$ and $-.047$). Such a contrast is intuitive given that the bank would not have much useful soft information on the quality of loans originated by third party brokers. As a result, the information advantage that investors have over the bank would be much stronger for such loans. On the other hand, the difference in the estimated ρs between Full-Doc and Low-Doc subsamples is much smaller; presumably, both the bank and the investors have less accurate information on Low-Doc loans, so neither party appears to have a particular information advantage.

[Insert Table 5 here.]

What do the ρ values reported in Table 5 tell us about the sensitivity of delinquency propensity to the propensity of loan sales? Recall that the standard deviations of the residual terms in the bivariate probit model (equation (8)) are normalized to one. Hence,

$$\varepsilon = \rho\eta + \varepsilon', \;\; \varepsilon' \sim N(0, 1-\rho^2). \tag{9}$$

That is, ρ can be interpreted as the increase in delinquency propensity due to a one standard deviation increase in the shocks to the propensity of loan sale. If we substitute (9) into the first equation of (8) and treat η as an auxiliary regressor, we can derive the average partial effect (APE) of η as follows:

$$APE = E\left[\frac{\partial}{\partial\eta}\Pr(Delinquency_i = 1|X_i, \eta_i)\right] = E\left[\frac{\partial}{\partial\eta}\Phi\left(\frac{X_i\beta + \rho\eta_i}{\sqrt{(1-\rho^2)}}\right)\right],$$

with the empirical analogue of

$$\widehat{APE} = \frac{\widehat{\rho}}{\sqrt{(1-\hat{\rho}^2)}}\frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{X_i\hat{\beta} + \hat{\rho}\hat{\eta}_i}{\sqrt{(1-\hat{\rho}^2)}}\right),$$

where $\phi()$ represents the probability density function of the standard normal distribution. The estimated APE is reported at the bottom of Table 5. For Bank loans, the delinquency rate decreases $0.8 - 1.3$ percentage points for a one standard deviation increase in the shocks to the propensity of loan sale. The sensitivity is much higher at $2.5 - 2.8$ percentage points for Broker loans.

The overall negative values of the ρ coefficient suggest that the bank's information advantage at the loan origination stage, beyond the recorded information (most of which is accessible to the investors when the loan is up for sale), is limited. This could be due to the automated system that is heavily used by the bank in loan approval where the objectives of the system was to streamline the process as well as to maintain "objectivity" (so as to ensure compliance with fair lending requirements). At the same time, such a system inevitably weakens the effective use of soft information in the process, especially among loans originated by third parties.

5.3 Explaining the Negative Ex Post Relation

What explains this negative relation between ex post loan sales and delinquency? The easiest explanation is the presence of quickly delinquent loans. If a loan goes bad shortly after origination and before a buyer comes along, it will remain with the originating bank. Such a selection effect does not require any information advantage of the potential buyers other than requesting loans with good current standing. In addition, some loan sale contracts allow investors to force the bank to re-assume loans that go bad shortly after loan sale. To assess the explanatory power of quick delinquency, we add $EarlyDelinq$ (a dummy variable for the delinquency of a loan within six months of loan origination) into the Z regressor vector in (8). The new variable does not enter the X vector because the information of quick delinquency is not available at loan origination. Early delinquent loans account for $2.6 - 3.3\%$ of the Bank subsamples, and $5.3 - 6.4\%$ of the Broker subsamples. Results are reported in Table 6 Panel A. For the economy of space, we only tabulate the bottom panel of the table that shows the test results for $H_0 : \rho = 0$. The coefficients on the covariates are qualitatively similar to those in Table 5.

[Insert Table 6 here.]

Not surprisingly, early delinquency significantly (at the 1% level in all columns) reduces the probability of loan sales with the average partial effect amounting to $6 - 7$ percentage points. Interestingly, though quickly delinquent loans account for a small percentage of the loan samples, their inclusion substantially changes the relative information advantage between the originating bank and the secondary market. The ρ values for the Bank subsamples are no longer statistically significant (and the APE values are economically insignificant as well). That is, the selection effect does not favor either of the two sides (originator versus investor) once the quick delinquency information is incorporated by the secondary loan market. The Broker subsamples, on the other hand, continue to exhibit a significant negative relation between delinquency and loan sale, but the average partial effect, now at $0.6 - 1.0$ percentage points, is less than half the magnitude of that using the full sample.

A second explanation attributes the negative relation to the "music chair" effect associated with the credit crisis. That is, when the subprime crisis suddenly hit the market in late 2007, the secondary market disappeared quickly. When the "music" of credit boom came to an abrupt halt, the originating bank was forced to retain loans that would have been sold under normal circumstances. At the same time, the overall delinquency rate increased, leading to higher delinquency rates among retained loans. Panel B of Table 6 entertains this possibility by restricting the sample to loans originated no later than June 2007, when two Bear Sterns hedge funds revealed colossal losses due to exposure to subprime securities, marking the beginning of the mortgage crisis.

Loans originated prior to June 2007 should have been sold without general issue in the secondary market. Results show that the Bank subsamples exhibit insignificant relations between delinquency and loan sale; but the negative relation in the Broker subsample is even stronger as shown in Table 5. Therefore, the market-wide crisis cannot explain the relatively poor performance of Broker loans that stay on the bank's balance sheet in comparison to those loans that were sold. In other words, the negative ex post relationship between loan sale and delinquency is not specific to the market during the crisis.

Last, the time lag between loan origination and loan sale affords investors additional information that helps predict delinquency. Given the anonymity of borrower identities, such additional information may come from changes in borrower creditworthiness or changes in the housing market. We are unable to examine whether post-origination changes in borrower credit score are related to delinquency, as we do not possess updated borrower credit scores in our data set (or the borrower permissions required to obtain them). However, we are able to examine changes in local housing markets; presumably, loans from regions that have experienced housing price appreciation (depreciation) since loan origination are less (more) vulnerable to delinquency, a relation that is confirmed by our data.

To assess the importance of this potential explanation, we match property addresses to three major housing price indices. The first index is the Case Shiller index at the MSA-monthly level; the second is the First American index at the county-quarterly level; and the third is the Office of Federal Housing Enterprise Oversight (OFHEO) index at the county (or in some cases, state)-quarterly level. The Case Shiller index is considered to be of higher quality than the First American index, which in turn is considered to be better than the OFHEO index in terms of coverage, reporting frequency, and refined locality. For each property address, we use the best available index to construct the local housing price change (in percentage) during the six months before loan origination ($HPI6mBefore$) and six months after the date ($HPI6mAfter$). We add $HPI6mBefore$ to both equations in (8) (because this information is available to both the originating bank and investors), and $HPI6mAfter$ only to the second equation (because this information was not available to the bank at loan origination). We lose about one quarter of the observations due to this additional data requirement. Panel C of Table 6 reports the results from this specification.

Housing price changes both before and after loan origination play an important role. $HPI6mBefore$ is significantly associated with higher delinquency rates. Presumably, properties in areas with high housing price appreciation played a larger role in the housing bubble, had more appraisal inflation, and the borrowers were in a greater hurry to buy without careful calculation. In average partial effects terms, a 10 percentage point increase in $HPI6mBefore$ is associated with a $2.2 - 4.7$ percentage point increase in delinquency rates across the four loan types. Secondary market investors seem to be aware of this relation and avoid buying loans from areas with high-flying recent housing price indices. For every 10 percentage point increase in $HPI6mBefore$, the probability of loan sale drops by $3.3 - 4.8$ percentage points on average.

The investors exhibit similar sophisticated selection behavior with regard to housing price changes post loan origination: they avoid buying loans from areas that have experienced negative housing price changes since loan origination. A 10 percentage point decrease in $HPI6mAfter$ is associated with a $4.6 - 5.1$ percentage point decrease in the probability of loan sale, and the effect is very close across the four loan types.

After the information about local housing prices is incorporated, the adverse selection effect is completely accounted for: all four $\widehat{\rho}$s are indistinguishable from zero, both statistically and economically. Results are qualitatively similar if we use a three-month window for the change in local housing price indices. Panel C of Table 6 confirms that investors used local housing market information in loan purchase decisions. Such a strategy by investors renders the bank vulnerable in a declining housing market.

In sum, we find support for Hypothesis Three which prescribes a negative ρ value to the joint system of delinquency and loan sale. That is, loans with higher propensity to delinquency are, ex post, less likely to be sold to investors. Moreover, we find that about half of the adverse selection effect that works against the bank can be explained away by about five percent of loans that went delinquent within six months of origination and many of which the bank was forced to retain. Among loans that survive the first six months, the remaining adverse selection effect can be explained by the fact that investors have access to post-origination local housing market information, and use such information in loan purchase decisions.

6 Conclusion

The stark contrast between the ex ante and ex post relation between loan quality and loan sale has several interesting implications. First, it challenges the bank's incentive and ability to collect meaningful "soft" information about borrower quality in a time of rapid growth supported mostly by the sector of broker-originated, low documentation loans. This dark side of securitization has prompted theoretical work on the optimal contract of securitization with moral hazard (Barney, Piskorski, and Tchistyi 2007). Second, once a loan is originated, investors' information advantage over the bank gains over time, and they indeed use such information strategically against the bank. The agency problem on the bank's part due to the presence of the secondary market ended up, ironically, hurting the bank more than it did the secondary market due to the investors' ability to select higher quality loans for purchase by exploiting information revealed between the time of loan origination and the time of loan sale. Finally, our paper presents an interesting empirical case to the research on contract theory on the intriguing relation between ex ante incentives and ex post allocation.

References

Adelino, Manuel, Kristopher Gerardi, and Paul Willen, 2009, Why don't lenders renegotiate more home mortgages? Redefaults, self-cures, and securitization, NBER Working Paper 15159.

Ashcraft, Adam, and Til Schuermann, 2006, Understanding the securitization of subprime mortgage credit, *Foundations and Trends in Finance* 2, 191–309.

Barney, Hartman-Glaser, Tomasz Piskorski, and Alexei Tchistyi, 2007, Optimal securitiztion with moral hazard, Columbia Business School working paper.

Benmelech, Effi, Jennifer Dlugosz, and Victoria Ivashina, 2007, Securitization without adverse selection: The case of CLOs, Harvard University working paper.

Bolton, Patrick, and David Scharfstein, 1996, Optimal debt structure and the number of creditors, *Journal of Political Economy* 104, 1–25.

Bubb, Ryan, and Alex Kaufman, 2007, Securitization and moral hazard: Evidence from a lender cut-off rule, Public Policy Discussion Paper from Federal Reserve Bank of Boston.

Demyanyk, Yulia, and Otto Van Hermert, 2009, Understanding the subprime mortgage crisis, *Review of Financial Studies* forthcoming.

Drucker, Steve, and Manju Puri, 2009, On loan sales, loan contracting and lending relationship, *Review of Financial Studies* 22:7, 2835–2872.

Elul, Ronel, 2009, Securitization and mortgage default: Reputation vs. adverse selection, Federal Reserve Bank of Philadelphia working paper.

Gilson, Stuart, Kose John, and Larry Lang, 1990, Troubled debt restructurings: An empirical study of private reorganization of firms in default, *Journal of Financial Economics* 27, 315–353.

Gorton, Gary, and George Pennacchi, 1995, Banks and loan sales: Marketing nonmarketable assets, *Journal of Monetary Economics* 35, 389–411.

Imbens, Guido, and Thomas Lemieux, 2008, Regression discontinuity designs: A guide to practice, *Journal of Econometrics* 142, 615–635.

Jiang, Wei, Ashlyn Nelson, and Edward Vytlacil, 2009, Liar's loan?–Effects of origination channel and information falsification on mortgage loan delinquency, Working Paper, Columbia Business School.

Johnson, Kathleen, Christopher Mayer, and Oliver Faltin-Traeger, 2009, Partipants in the securitization process and their impact on performance, Working Paper, Columbia Business School.

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010a, 620 FICO, take II: Securitization and screening in the subprime mortgage market, Working Paper, University of Chicago.

——— , 2010b, Did securitization lead to lax screening? Evidence from subprime loans, *Quarterly Journal of Economics* 125:1, 307–362.

Lee, David, 2008, Randomized experiments from non-random selection in U.S. house elections, *Journal of Econometrics* 142, 675–697.

Loutskina, Elena, and Philip Strahan, 2007, Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage acceptance rates, working Paper, Boston College.

MaCrary, Justin, 2000, Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics* 142.

Mian, Atif, and Amir Sufi, 2009, The consequences of mortgage credit expansion: Evidence from the 2007 mortgage default crisis, *Quarterly Journal of Economics* 124, 1449–1496.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig, 2009, Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis, working paper, Columbia Business School.

Robinson, Peter, 1988, Root-n-consistent semiparametric regression, *Econometrica* 56, 931–954.

Sufi, Amir, 2006, Information asymmetry and financing arrangements: Evidence from syndicated loans, *Journal of Finance* 62, 629–668.

Figure 1. Graphic Illustration:
Ex Ante and Ex Post Relations between Loan Sale and Delinquency

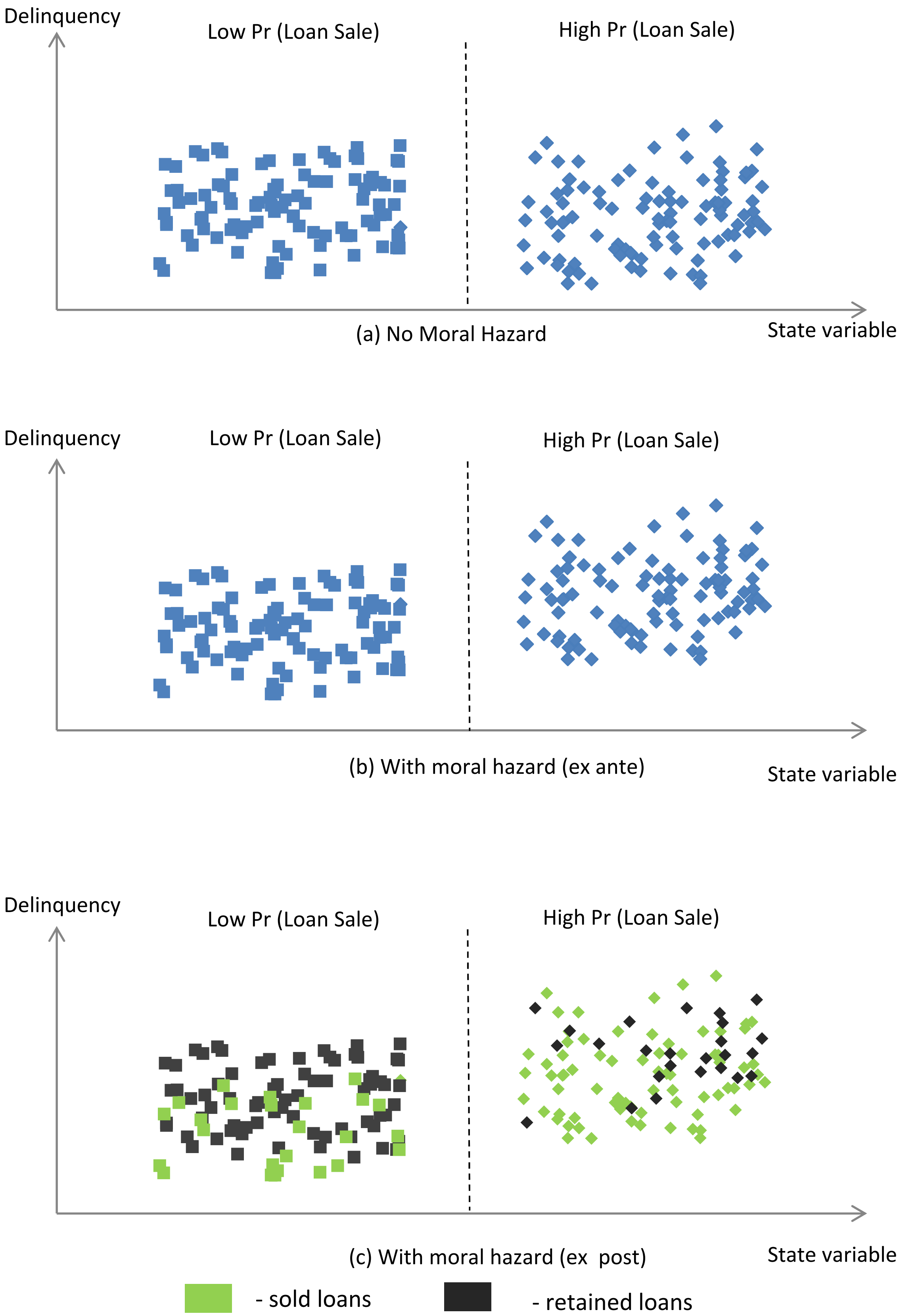


Figure 2. Credit Score and Probability of Loan Sale

Each dot in this figure represents the estimated probability of loan sale for loans with credit scores falling into the individual ranges with an even width of 20 points, conditional on all other regressors that appear in Table 4.

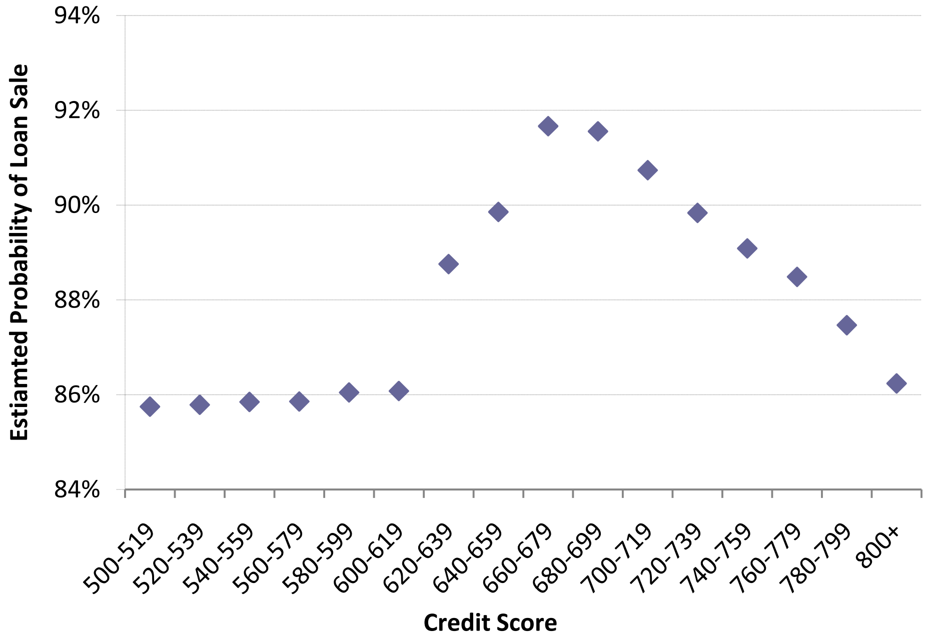
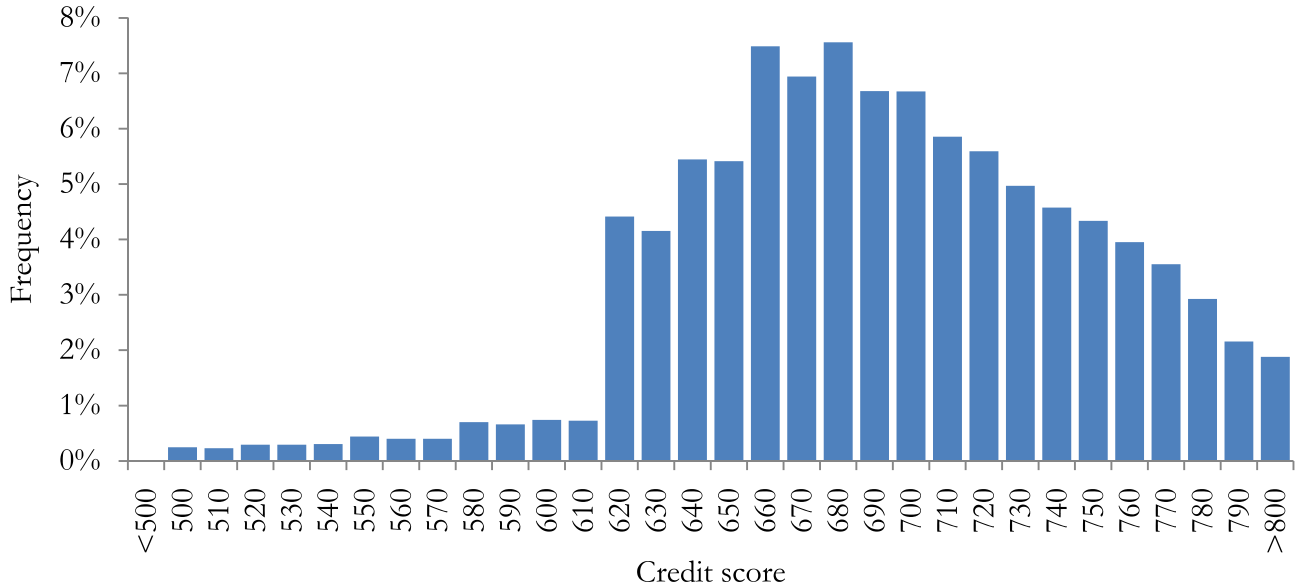
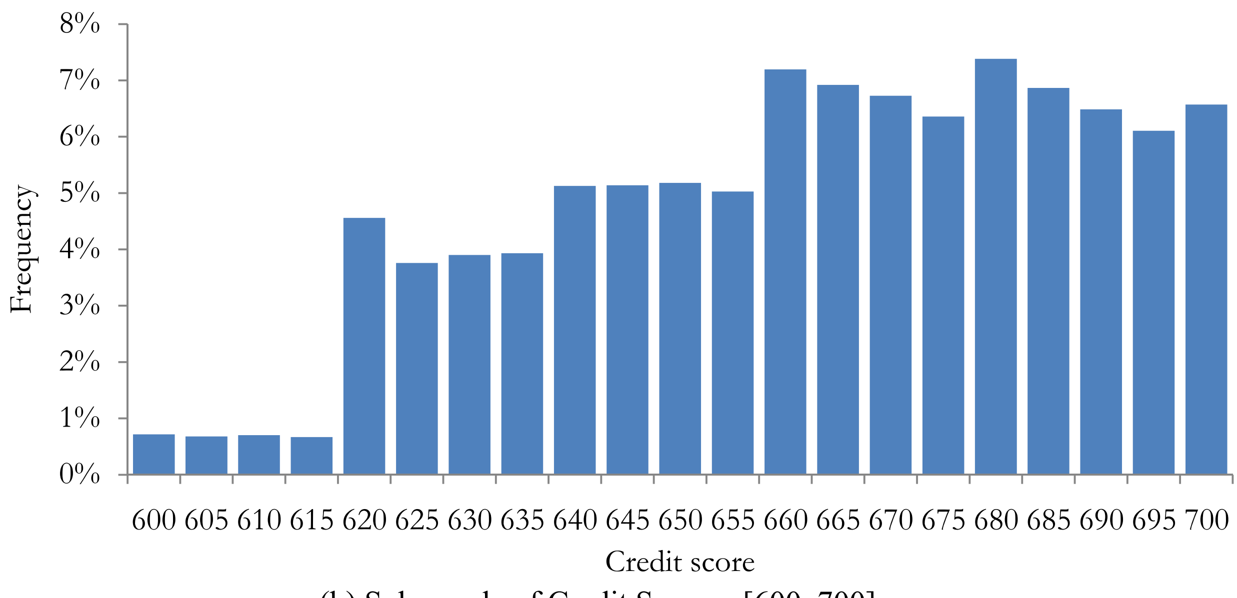


Figure 3. Histograms of Sold Loans vs. Credit Score

Figure (a) plots the histogram of sold loans vs. credit score of the full sample, and figure (b) plots the same histogram for loans with borrowers whose credit score falls between 600 and 700. The horizontal axis marks the lower bound of each bin.



(a) Full Sample



(b) Subsample of Credit Score = [600, 700]

Figure 4. Actual and Estimated Delinquency Probability vs. Credit Score: Univariate Analysis

The scattered dots represent the average actual delinquency rates of loans with credit scores equal to the individual values ranging from 500 to 800. Observations with creditor score below 500 (above 800) are combined with the first (last) group. The solid lines represent the expected delinquency rate from equation (3) using cubic polynomials. The estimation is conducted separately on the three regions based on credit scores: [500,619], [620,659], and [660,800]. Finally, the dashed lines are the 95% confidence interval associated with the estimates for the expected delinquency rates.

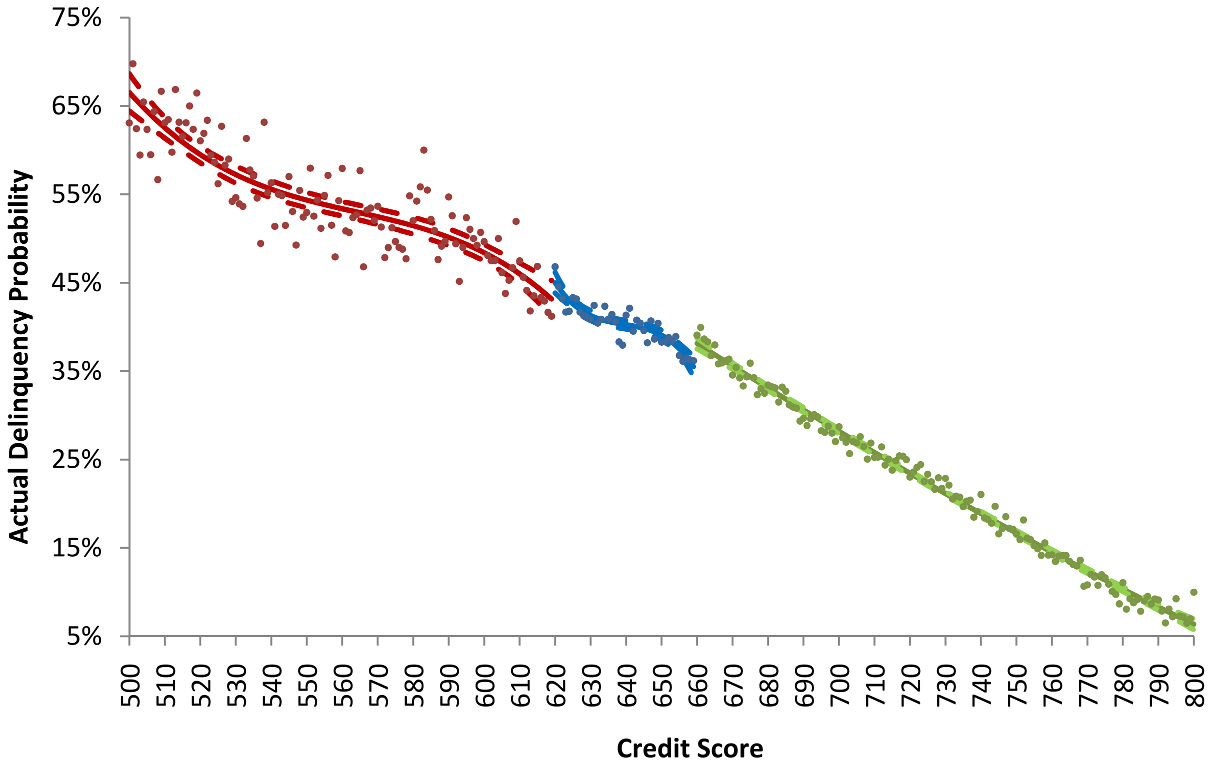


Figure 5. Residual Delinquency Probability vs. Credit Score: Multivariate Analysis

This figure plots the residual delinquency probability, after filtered out the effects of all covariates except credit scores, versus the credit score using a partial linear model as specified in equations (5) to (7). The scattered dots represent the average residual delinquency rates of loans with credit scores equal to the individual values ranging from 500 to 800. Observations with creditor score below 500 (above 800) are combined with the first (last) group. The solid lines represent the expected residual delinquency rates using cubic polynomials. The estimation is conducted separately on the three regions based on credit scores: [500,619], [620,659], and [660,800]. Finally, the dashed lines are the 95% confidence interval associated with the estimates for the expected residual delinquency rates.

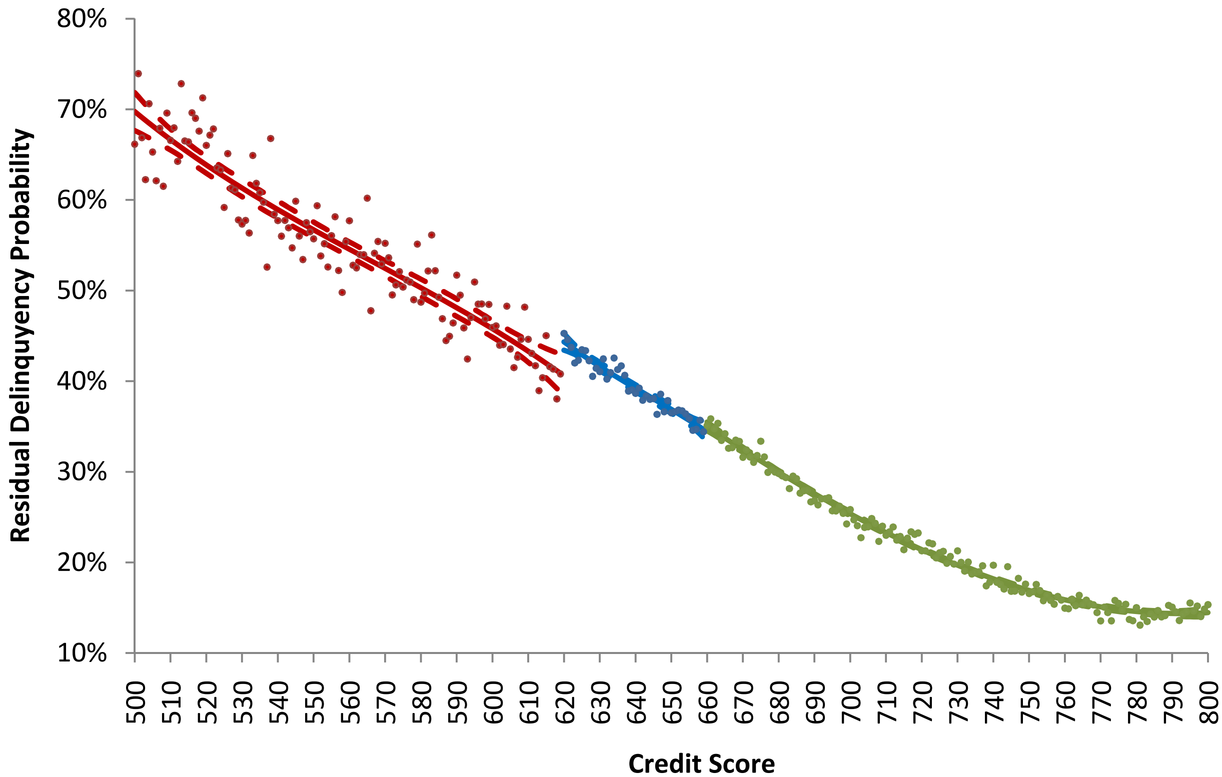


Table 1. Variable Definitions and Summary Statistics

Panel A provides definitions of the main variables. Panel B reports their summary statistics, including mean, variance, and values at the 25th, 50th (median), and the 75th percentiles. Panel C reports the difference in delinquency rates between retained and sold loans by origination channels and documentation status.

Panel A: Variable Definitions

	Definition
Age	Age of the borrower
ARM	Dummy variable = 1 if the loan is adjustable rate, but not Option/ARM or IO
Asian	Dummy variable = 1 if the borrower is Asian
AvgIncome	Average income per capita of the census tract where the property is located
Black	Dummy variable = 1 if the borrower is black
Broker	Dummy variable = 1 if the loan originated through the broker channel
CashResv	Cash reserves, in multiples of monthly mortgage payments
CLTV	Combined loan-to-value ratio
CreditScore	Median of the borrower's TransUnion, Experian, and Equifax credit scores.
Delinq	Dummy variable = 1 if borrower is ever delinquent, defined as at least 60 days behind in payment
EarlyDelinq	Dummy variable = 1 if borrower is delinquent within 6 months of loan origination
Female	Dummy variable = 1 if the borrower is female
FirstTimeOwner	Dummy variable = 1 if the borrower is a first-time mortgage borrower
HardPenalty	Dummy variable = 1 if there is hard prepayment penalty in the loan contract
Hispanic	Dummy variable = 1 if the borrower is Hispanic
HPI6mAfter	Change in housing price index during the 6 months after to loan origination
HPI6mBefore	Change in housing price index during the 6 months prior loan origination
Income	Monthly income of the borrower in $1,000
IncomeMiss	Dummy variable = 1 if the income information is missing
InitialRate	Initial interest rate on the mortgage
IO	Dummy variable = 1 if the loan is interest only
Loan	Total loan amount
LoanSold	Dummy variable = 1 if the loan was sold
LowDoc	Dummy variable =1 if low documentation loan
LTV	Loan-to-value ratio
OneBorrower	Dummy variable = 1 if there is only one borrower on the mortgage
OptionARM	Dummy variable = 1 if the loan is option/ARM but not IO
OwnerOccupied	Dummy variable = 1 if the property is the owner's primary residence
PctBlack/PctHisp	Proportion of black/Hispanic households in the census tract where the property is located
Population	Population size of the census tract where the property is located
Refinance	Dummy variable = 1 if the mortgage is for refinancing
Secondlien	Dummy variable = 1 if the mortgage is a second-lien
SelfEmploy	Dummy variable = 1 if the borrower is self-employed
Tenure	Number of months that the borrower has been employed in the current job
TenureMiss	Dummy variable = 1 if the tenure information is missing
Unemprate	Unemployment rate in the census tract where the property is located

Panel B: Summary Statistics

Variable	Obs	Mean	Std. Dev.	25%	median	75%
1. Loan information						
ARM	721767	0.114	0.318	0	0	0
Broker	721767	0.904	0.294	1	1	1
CLTV	721744	0.811	0.170	0.722	0.800	0.950
FirstOwner	699682	0.154	0.361	0	0	0
HardPenalty	721767	0.101	0.301	0	0	0
InitialRate	721767	0.0636	0.0258	0.0588	0.06625	0.075
IO	721767	0.347	0.476	0	0	1
Loan	721764	268003	198557	132000	227000	356000
LowDoc	721767	0.710	0.454	0	1	1
OneBorrower	721767	0.679	0.467	0	1	1
OptionARM	721767	0.164	0.370	0	0	0
OwnerOccup	721767	0.849	0.358	1	1	1
Refinance	721767	0.581	0.493	0	1	1
SecondLien	721767	0.096	0.295	0	0	0
2. Borrower Demorgraphics						
Age	678084	43.74	12.57	34	43	52
Asian	721767	0.052	0.223	0	0	1
Black	721767	0.082	0.274	0	0	1
Female	721767	0.335	0.472	0	0	1
Hispanic	721767	0.188	0.391	0	0	1
3. Borrower economic conditions						
CashResv	721767	12.54	32.32	0.0	2	10
CreditScore	719974	695.92	56.19	660	694	736
Income	568957	9.45	94.00	4.81	7.00	10.42
IncomeMiss	721767	0.212	0.167	0	0	0
SelfEmploy	694917	0.200	0.400	0	0	1
Tenure	540955	85.73	90.40	24	58	120
TenureMiss	721767	0.251	0.433	0	0	1
4. Neiborhood information						
AvgIncome	539817	30.192	22.727	18.805	25.033	33.627
HPI6mAfter	538779	-0.011	0.071	-0.049	-0.009	0.027
HPI6mBefore	538807	-0.018	0.066	-0.053	-0.014	0.024
PctBlack	539815	0.122	0.208	0.012	0.035	0.119
PckHispanic	539815	0.185	0.217	0.033	0.094	0.252
5. Loan performance						
Delinq	721767	0.287	0.453	0	0	1
EarlyDelinq	721767	0.066	0.248	0	0	0
LoanSold	721767	0.890	0.313	0	1	1

Panel C: Difference in Delinquency Rates between Sold and Retained Loans

	(1) Bank/Full-Doc	(2) Bank/Low-Doc	(3) Broker/ Full-Doc	(4) Broker/ Low-Doc	Full Sample
Sold	13.7%	18.1%	23.2%	31.7%	28.2%
Retained	10.2%	24.5%	27.2%	39.0%	32.9%
Difference	3.4%	-6.5%	-3.9%	-7.3%	-4.7%
All Loans	13.3%	19.0%	23.8%	32.4%	28.7%
Difference/All	26.0%	-33.9%	-16.6%	-22.6%	-16.4%
(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 2. Sample Representativeness—Comparison with the National Market

This table compared the key summary statistics of our sample to those of the national market.

	Our sample	National market
% loans originated by brokers	90%	60%(1)
% loans securitized	85%	60%-80% for all; (2) 75-91% for subprime and Alt-A loans(3)
% low-doc	70%	25%(4)
% subprime	15%	18-21%(4)
LTV	About the same(4)	
Loan amount	Our sample is about 15% higher(4)	
Credit score	Our sample is about 5-8 points lower(4)	
Demographics	Our sample has higher representation of Hispanic borrowers(5)	
Annual growth 2004-2006	> 50%	30-40%(6)
% Delinquency (early 2009)	26%	11% for all, 39% for subprime(7)

(1) Source: "Mortgage Brokers: Friends or Foes?" by James Hagerty, *The Wall Street Journal*, May 30, 2007.
(2) Source: Rosen, Richard, 2007, The Role of Securitization in Mortgage Lending, *Chicago Fed Letter,* No. 244.
(3) Source: http://www.imfpubs.com/data/mortgage_securitization_rates.htm.
(4) Source: McDash Analytics.
(5) Source: National HMDA data, http://www.ffiec.gov/hmdaadwebreport/NatAggWelcome.aspx
(6) Source: Dell'Ariccia, Giovanni, Deniz Igan and Luc Laeven, 2008, Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market, IMF Working Paper..
(7) Source: Loan Processing Services (LPS), http://www.lpsvcs.com/NewsRoom/IndustryData/Pages/default.aspx.

Table 3. Prediction of Loan Sale

The dependent variable is the dummy variable for loan sale (*LoanSale*), and the estimation method is probit. The definitions of all variables are given in Table 1 Panel A. Reported are the coefficients (coef), t-statistics (t-stat) that adjust for clustering at the MSA level, and the change in marginal probability for a one unit change in the regressors (dPr/dX). At the bottom of the table, we report the sample frequency of delinquency, the pseudo R-squared, the number of observations and the number of clusters (at the MSA level). Columns (1) includes as regressors only loan-level variables while columns (2) also includes neighborhood level variables. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(continued)	(1)	(2)
CLTV	-0.2133***	-0.2336***	CreditScore/100	0.0678***	0.0782***
	[-6.22]	[-4.63]		[7.63]	[8.36]
	-2.87%	-3.02%		0.91%	1.01%
Loan	-0.2204***	-0.1732***	Tenure	-0.0164***	-0.0155***
	[-21.57]	[-13.75]		[-4.39]	[-2.66]
	-2.97%	-2.24%		-0.22%	-0.20%
SecondLien	-0.5969***	-0.5399***	TenureMiss	-0.3286***	-0.3645***
	[-18.64]	[-13.01]		[-16.21]	[-12.01]
	-11.21%	-9.49%		-4.93%	-5.34%
Refi	0.0083	0.0511***	SelfEmploy	-0.0507***	-0.0606***
	[0.77]	[3.98]		[-5.77]	[-5.73]
	0.11%	0.67%		-0.70%	-0.81%
ARM	-0.1014***	-0.1304***	PctBlack		-0.1040***
	[-4.49]	[-4.34]			[-7.36]
	-1.45%	-1.81%			-1.34%
OptionARM	0.1136***	0.0894***	PctHisp		-0.1576***
	[5.66]	[3.31]			[-6.98]
	1.45%	1.10%			-2.04%
IO	-0.0272	-0.0234	AvgIncome		-0.0002
	[-1.15]	[-0.95]			[-1.38]
	-0.37%	-0.30%			0.00%
HardPenalty	0.1164***	0.0620*	HPI6mBefore		-4.4607***
	[5.80]	[1.73]			[-4.63]
	1.46%	0.77%			-57.66%
FirstOwner	0.0193**	0.0386***	y2005	0.9049***	0.7764***
	[2.28]	[2.75]		[37.81]	[23.77]
	0.26%	0.49%		8.46%	7.21%
OwnerOccup	0.1015***	0.0654***	y2006	1.0300***	1.1573***
	[4.57]	[3.11]		[40.18]	[20.76]
	1.44%	0.88%		11.93%	12.59%
OneBorrower	-0.0952***	-0.0856***	y2007	-0.0541**	0.2280**
	[-12.46]	[-9.87]		[-2.55]	[2.11]
	-1.25%	-1.08%		-0.74%	2.80%

Income	0.0431***	0.0325***	LowDoc	0.2055***	0.2153***
	[7.08]	[3.31]		[18.41]	[15.17]
	0.58%	0.42%		2.95%	2.98%
IncomeMiss	0.2270***	0.2108***	Broker	0.0136	0.0366*
	[12.45]	[7.78]		[0.98]	[1.75]
	2.77%	2.48%		0.19%	0.48%
CashResv	-0.0435***	-0.0386***	Constant	3.2940***	2.5445***
	[-9.80]	[-7.34]		[26.56]	[14.59]
	-0.59%	-0.50%			
Observations	683591	390359			
R-squared	0.141	0.170			

Table 4. Testing Covariates Equality across Credit Score Breaking Points

This table compares all covariates except credit score in loan delinquency prediction (as shown in Table 2) for loans with credit score in the range of [618,619] ([658,659]) versus those in the range of [620,621] ([660,661]). We report the mean differences and the associated t-statistics based on standard errors clustered at the MSA level. The final column of the table shows the effects of the individual covariates on delinquency probabilities where "+" ("-") indicates a significantly positive (negative) effect and "n.s." stands for "not significant" at the 5% level.

Credit score ranges	[618,619] vs. [620,621]		[658,659] vs. [660,661]		Effect on Delinquency
	Mean difference	t-stat	Mean difference	t-stat	
ARM	-0.060	-4.08	-0.017	-3.93	+
Age	-0.103	-3.38	-0.030	-3.08	n.s.
Asian	0.006	1.09	-0.001	-0.25	n.s.
Black	-0.051	-3.30	-0.009	-1.67	+
Broker	0.059	6.24	0.021	5.47	+
CashResv	0.406	7.60	-0.100	-4.92	-
CLTV	0.020	1.81	-0.001	-0.39	+
Female	-0.048	-2.98	-0.011	-1.45	n.s.
FirstOwner	0.132	5.04	0.018	2.59	-
HardPenalty	0.058	7.33	-0.003	-0.65	+
Hispanic	0.091	4.27	0.014	2.52	+
Income	-0.190	-6.27	0.014	0.65	n.s.
Incomemiss	0.189	11.62	0.011	1.41	+
InitialRate	-0.596	-5.13	-0.163	-2.84	+
IO	0.137	11.86	0.017	2.52	+
Loan	0.196	6.09	0.018	1.53	+
LowDoc	0.283	19.69	0.036	5.63	+
OneBorrower	0.113	6.60	0.000	-0.04	+
OptionARM	0.074	19.31	0.022	3.95	+
OwnerOccup	-0.089	-2.41	0.005	1.03	-
Refinance	-0.158	-6.51	-0.028	-3.65	+
SecondLien	-0.037	-2.85	0.021	3.46	+
SelfEmploy	0.041	3.20	0.001	0.10	+
Tenure	-0.164	-2.14	-0.102	-3.19	-

Table 5. Loan Sale and Delinquency

This table reports estimates of the bivariate probit model equation (9). The bottom of the table reports the estimated ρ coefficients, the correlation between the residuals from the *Delinq* equation and that from the *LoanSale* equation. The Wald statistic tests the null hypothesis that the two equations are uncorrelated. The average partial effects (APE) are the effects on the delinquency probability for one-standard deviation increase in the shocks to the propensity of loan sales. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

	(1) Bank/Full-Doc		(2) Bank/Low-Doc		(3) Broker/Full-Doc		(4) Broker/Low-Doc	
dependent variable	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold
CLTV	1.5512***	-0.5159***	2.2527***	-0.1364	1.9041***	-0.0992	2.9657***	-0.1296***
	[14.95]	[-6.32]	[17.97]	[-1.23]	[18.80]	[-1.61]	[21.16]	[-3.75]
Loan	0.1036***	-0.2288***	0.1574***	-0.3115***	0.2056***	-0.2581***	0.2142***	-0.1881***
	[3.84]	[-7.18]	[6.49]	[-15.39]	[8.99]	[-14.25]	[8.39]	[-17.74]
SecondLien	0.1678***	-0.7619***	0.2190***	-0.7725***	0.3435***	-0.4529***	0.3810***	-0.7050***
	[2.63]	[-8.79]	[4.36]	[-12.96]	[8.98]	[-11.66]	[8.57]	[-25.14]
Refi	-0.0207	0.2206***	0.0090	0.1497***	-0.0460**	-0.0212	0.0850***	-0.0263**
	[-0.52]	[5.89]	[0.34]	[3.88]	[-2.01]	[-1.25]	[4.84]	[-2.38]
ARM	0.2191***	0.2514***	0.1426***	-0.0151	0.2025***	0.1555***	0.1801***	-0.3075***
	[6.87]	[3.83]	[6.03]	[-0.34]	[12.11]	[5.10]	[11.43]	[-14.57]
OptionARM	0.1842***	0.3669***	0.3407***	-0.0611	0.2356***	0.4015***	0.2877***	0.0036
	[3.23]	[3.27]	[9.77]	[-0.70]	[6.28]	[8.39]	[11.11]	[0.16]
IO	0.1791***	-0.0719	0.1962***	-0.0584	0.1198***	0.1504***	0.2011***	-0.1546***
	[6.02]	[-1.60]	[10.87]	[-1.57]	[6.46]	[4.53]	[13.07]	[-8.83]
HardPenalty	0.0568	0.3314***	-0.0406	0.1532***	-0.0348*	0.3518***	0.0508***	0.0510**
	[1.07]	[4.18]	[-0.98]	[2.83]	[-1.67]	[10.96]	[3.86]	[2.07]
FirstOwner	-0.1456***	-0.2404***	-0.0360	-0.2750***	-0.0034	-0.0424***	-0.0490***	0.0469***
	[-3.30]	[-3.96]	[-0.59]	[-4.69]	[-0.22]	[-3.04]	[-3.66]	[4.53]
OwnerOccup	-0.2152***	0.1748***	-0.2605***	-0.1661***	-0.3585***	0.2115***	-0.2868***	0.1042***
	[-4.71]	[4.12]	[-7.60]	[-4.65]	[-13.79]	[10.66]	[-11.34]	[3.93]

OneBorrower	0.2571***	-0.1492***	0.3449***	-0.0775***	0.2947***	-0.0888***	0.2962***	-0.1269***
	[12.18]	[-6.33]	[15.55]	[-3.50]	[19.29]	[-6.67]	[16.83]	[-17.14]
Income	-0.1153***	-0.0107	0.0047	0.1283***	-0.0731***	0.0758***	0.0439***	0.0154**
	[-7.65]	[-0.37]	[0.26]	[6.60]	[-5.41]	[5.88]	[5.42]	[2.04]
IncomeMiss	-0.0559	0.1480	-0.0414	0.4663***	-0.1703***	0.1206**	0.1780***	0.1268***
	[-0.47]	[1.27]	[-0.85]	[8.50]	[-3.19]	[1.96]	[8.07]	[5.41]
CashResv	-0.0447***	-0.0424***	-0.0195***	-0.0692***	-0.0879***	-0.0408***	-0.0687***	-0.0363***
	[-5.33]	[-3.92]	[-2.73]	[-8.43]	[-17.65]	[-6.16]	[-16.66]	[-8.76]
CreditScore/100	-0.0083***	-0.0002	-0.0075***	-0.0013***	-0.0083***	0.0016***	-0.0070***	0.0000
	[-47.79]	[-1.23]	[-34.18]	[-7.38]	[-52.58]	[12.82]	[-73.51]	[0.35]
Female	-0.0408		-0.0160		-0.0051		0.0020	
	[-1.63]		[-0.86]		[-0.38]		[0.24]	
Hispanic	0.2695***		0.2165***		0.3867***		0.2720***	
	[5.46]		[3.71]		[7.89]		[10.77]	
Black	0.1278***		0.1635***		0.1687***		0.1207***	
	[2.74]		[2.93]		[5.19]		[4.60]	
Asian	-0.0661		-0.0523		0.0228		0.0338	
	[-0.66]		[-1.04]		[0.76]		[1.21]	
Age	-0.0874***		0.0135		-0.0172		0.0057	
	[-3.54]		[0.73]		[-1.46]		[0.62]	
Tenure	-0.0156*	-0.0193**	-0.0410***	0.0069	-0.0091	-0.0376***	-0.0337***	-0.0008
	[-1.72]	[-2.01]	[-4.73]	[0.78]	[-1.44]	[-6.53]	[-6.80]	[-0.19]
TenureMiss	-0.0585	-0.5695***	-0.1408***	-0.1129**	-0.2324***	-0.2877***	-0.2322***	-0.2001***
	[-0.93]	[-9.09]	[-3.10]	[-2.06]	[-7.33]	[-10.98]	[-10.11]	[-7.59]
SelfEmploy	-0.0026	-0.1553***	0.0650***	-0.0923***	0.0595***	-0.0886***	0.0132	-0.0394***
	[-0.05]	[-3.01]	[3.41]	[-3.59]	[2.83]	[-4.91]	[1.25]	[-4.73]
y2005	-0.0087	1.2571***	0.0898**	0.8599***	-0.0187	1.2079***	0.0947***	0.8997***
	[-0.24]	[17.97]	[2.39]	[12.62]	[-0.72]	[44.81]	[4.03]	[24.69]
y2006	-0.0034	1.2397***	0.1057***	0.9327***	0.0196	1.2794***	0.2251***	0.9995***
	[-0.09]	[24.00]	[2.63]	[19.94]	[0.45]	[38.32]	[5.60]	[27.76]
y2007	-0.1863***	-0.3984***	0.0573	-0.3101***	-0.0687	0.0664**	0.1467***	-0.0237

	[-3.82]	[-11.22]	[1.11]	[-10.19]	[-1.37]	[2.28]	[3.38]	[-0.72]
y2008	-0.2723***		-0.0982		-0.2157***		-0.0522	
	[-3.62]		[-1.28]		[-3.93]		[-0.86]	
Constant	2.6592***	4.3900***	0.5485*	5.9776***	1.0887***	2.7239***	-1.0835***	3.6419***
Observations	31,408		35,553		166,402		425,181	
ρ	-0.037	-1.68	-0.047	-3.11	-0.110	-10.35	-0.102	18.31
APE	-0.80%		-1.30%		-2.68%		-3.00%	
Wald test of ρ = 0:								
chi2(1) and p-val	2.83	0.09	9.67	0.00	107.10	0.00	335.26	0.00

Table 6. Loan Sale and Delinquency: Additional Analyses

This table repeats the analysis in Table 5 but with additional covariates or confines the analysis to various subsamples. Panel A adds the dummy variable *EarlyDelinq* to the *LoanSale* equation. Panel B confines the sample loans to those originated before July 2007. Panel C further adds *HPI6mAfter* to the specification in Panel A. For the economy of space, only the coefficients on the new variables and the summary ρ coefficient and the Wald-test are reported. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

Panel A: Early Delinquency Considered

	(1) Bank/Full-Doc		(2) Bank/Low-Doc		(3) Broker/Full-Doc		(4) Broker/Low-Doc	
dependent variable	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold
EarlyDelinq		-0.2388***		-0.3636***		-0.2763***		-0.4660***
		[-3.03]		[-6.17]		[-11.95]		[-47.57]
Observations	31,408		35,553		166,402		425,181	
ρ	0.0093	[0.32]	-0.0009	[-0.05]	-0.0416***	[-3.61]	-0.0212***	[-3.26]
APE	0.20%		-0.02%		-1.01%		-0.62%	
Wald test of ρ = 0:								
chi2(1) and p-val	0.10	0.75	0.00	0.96	13.01	0.00	10.63	0.00

Panel B: Pre-July 2007

	(1) Bank/Full-Doc		(2) Bank/Low-Doc		(3) Broker/Full-Doc		(4) Broker/Low-Doc	
dependent variable	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold
EarlyDelinq		-0.5786***		-0.2996***		-0.4713***		-0.4974***
		[-5.37]		[-3.52]		[-16.65]		[-33.85]
Observations	25,426		29,958		132,160		377,381	
ρ	0.0106	[0.29]	-0.0151	[-0.66]	-0.0751***	[-5.48]	-0.0619***	[-7.68]
APE	0.24%		-0.42%		-1.82%		-1.84%	
Wald test of ρ = 0:								
chi2(1) and p-val	0.08	0.77	0.43	0.51	30.06	0.00	59.01	0.00

Panel C: Information from the Time Lag between Loan Origination and Loan Sale

	(1) Bank/Full-Doc		(2) Bank/Low-Doc		(3) Broker/Full-Doc		(4) Broker/Low-Doc	
dependent variable	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold	Delinq	LoanSold
HPI6mBefore	1.0268***	-3.4537***	1.4751***	-3.0795***	1.5314***	-3.6647***	1.6376***	-2.5432***
	[3.09]	[-6.59]	[4.04]	[-4.65]	[6.02]	[-6.95]	[5.99]	[-4.62]
EarlyDelinq		-0.1725**		-0.2796***		-0.2612***		-0.4623***
		[-2.11]		[-3.75]		[-10.41]		[-32.72]
HPI6mAfter		3.9323***		3.7111***		3.5420***		3.7576***
		[10.35]		[9.33]		[12.29]		[9.29]
Observations	22522	22522	24252	24252	124573	124573	319226	319226
ρ	0.0114	[0.37]	0.0214	[1.00]	-0.0034	[-0.25]	0.0012	[0.14]
APE	0.24%		0.57%		-0.08%		0.03%	
Wald test of $\rho = 0$:								
chi2(1) and p-val	0.14	0.71	1.01	0.32	0.06	0.80	0.02	0.89