Liar's Loan?

Effects of Origination Channel and Information Falsification on Mortgage Delinquency[1]

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

This Draft: April 2011

[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 Vyacheslav Fos, Chris Mayer, Atif Mian, Daniel Paravisini, Tomasz Piskorski, David Scharfstein, Amit Seru, Bob Van Order, and seminar/conference participants at the Association for Public Policy Analysis and Management Annual Research Conference of 2009, Columbia, Georgia State, Kansas City Federal Reserve Bank, the NBER 2009 Summer Institute, the FDIC 2009 Mortgage Symposium, The Philadelphia Fed's Conference on Recent Developments in Consumer Credit and Payments of 2009, the Sixth Annual Credit Risk Conference by Moody's & NYU Stern School of Business of 2010, and the 2010 AEA have contributed to this draft. The authors also thank Erica Blom, Vyacheslav Fos, Andres Liberman, Sunyoung Park, and Mike Tannenbaum for excellent research assistance.
[2] Corresponding author. Columbia Business School, Finance and Economics Division, Tel: 212 854 9002, Email: wj2006@columbia.edu.
[3] Indiana University, School of Public and Environmental Affairs, Tel: 812 855 5971, Email: ashlyn@indiana.edu.
[4] Yale University, Department of Economics, Tel: 203 436 3994, Email: edward.vytlacil@yale.edu.

Liar's Loan?

Effects of Origination Channel and Information Falsification on Mortgage Delinquency

ABSTRACT

This paper presents a comprehensive analysis of mortgage delinquency between 2004 and 2008 using a unique loan-level dataset from a major national mortgage bank. Our analysis highlights two major problems underlying the mortgage crisis: a heavy reliance on mortgage brokers who tend to originate lower quality loans, and a high prevalence of low-documentation loans—known in the industry as "liars' loans"—which results in information falsification by borrowers. While up to three-quarters of the difference in delinquency rates between bank and broker channels can be attributed to observable loan and borrower characteristics, all of the delinquency difference between full- and low-documentation mortgages is due to unobservable heterogeneity. We provide evidence that this unobserved heterogeneity results from income falsification among low-documentation loans. Further, we find little evidence that the loan pricing compensates for the high delinquency rates among the brokered and low-documentation loans.

The recent crisis in the housing and mortgage debt market has drawn considerable attention from regulators and market participants. A decade-long boom in the housing market and related financial sectors was followed in 2007 by a market bust with falling house prices and a rapid increase in mortgage defaults and foreclosures. Those caught in the crisis included large financial institutions that experienced sharp expansion in, and profited from, their exposure to mortgage loans. The crisis that started in the mortgage market quickly spread to other financial markets and throughout the economy.

We use the experience of a major national mortgage bank to uncover the determinants of the mortgage crisis and the evolution of the crisis at a micro level. The particular bank provides an ideal context for the study, presenting a representative and yet amplified version of the boom-and-bust cycle experienced by the national mortgage sector in the last decade. First, the bank was among the nation's top ten mortgage lenders in 2006 and was one of the fastest growing players in the mortgage market; a majority of the loans they issued were for low- and no-documentation mortgage products (nicknamed "liars' loans"). Second, the bank suffered some of the largest losses in the industry since the 2007 crisis; by 2009, loans issued by the bank since the beginning of 2004 reached a delinquency rate of 26%. Finally, the borrowers and properties underlying the bank's loans during our sample period are fairly represented across all 50 states. Therefore, lessons from this particular bank have general implications for the national mortgage market.

Our proprietary data set contains the most detailed and disaggregated information used thus far in the mortgage loan literature. The data set includes all 721,767 loans that the bank originated between January 2004 and February 2008. For each of these loans, we observe all information collected by the bank at origination, as

well as monthly performance data through January 2009. Our data set includes not only information about the loan (pricing, loan product, and other contractual terms) and the property (address, appraisal value, owner occupancy status, etc.), but also about borrower demographic characteristics (race, age, gender, etc.) and economic conditions (income, cash reserves, employment status, etc.). Finally, we are able to use the property address information to match loans to community attributes (e.g., demographics and employment opportunities) at a narrow geographic level.

We divide our sample into four distinct subsamples by a two-way sorting. The first sorting variable is the loan origination channel: whether a loan is originated directly by the bank, or by a third party originator or broker. The second sorting variable is the loan documentation level: whether a loan is originated with full documentation of the borrower's economic conditions, or with various reduced levels of documentation (including no documentation). Throughout the paper we refer to the four subsamples (with the initial letters capitalized) as: Bank/Full-Doc; Bank/Low-Doc; Broker/Full-Doc; and Broker/Low-Doc. The Bank (Broker) subsamples include both the Bank/Full-Doc and Bank/Low-Doc (Broker/Full-Doc and Broker/Low-Doc) subsamples, and the Full-Doc (Low-Doc) subsamples are defined analogously.

Our empirical analysis uncovers two major problems in mortgage lending that constitute the fundamental causes of high loan delinquency rates, and by extension, the mortgage crisis. The first is a heavy reliance on mortgage brokers, driven by the credit expansionary policies pursued by many large lending institutions. We find that loans in the Broker subsamples have delinquency probabilities that are 10-14 percentage points higher than those for the Bank subsamples, reflecting the incentive structure in broker commissions which compensates brokers based on origination volume rather than loan performance. A decomposition exercise attributes up to three-quarters of the Bank-Broker delinquency gap to differences in observable borrower characteristics. Hence, the higher delinquency rates among brokered loans are largely explained by brokers serving an observably worse quality clientele as compared to the bank (as measured by credit score, loan-to-value ratio, income, etc.).

The Low-Doc subsample also exhibits worse performance than the Full-Doc subsample, and the difference in delinquency is 5-8 percentage points. However, the same decomposition method reveals that unobserved heterogeneity explains nearly 100% of the delinquency difference. In contrast to the Broker channel, the Low-Doc channel does not necessarily compromise lending standards along the observed metrics; rather, low-documentation mortgages suffer from adverse selection along unobservables. We argue that the unobserved heterogeneity in loan performance arises because low-documentation borrowers' reported information (notably income) is verified less carefully, and borrowers are screened less diligently along hard-to-quantify measures (such as other major expenditures). This finding highlights a major agency problem between lenders and borrowers, wherein borrowers may hide or even falsify unfavorable information when lax screening and verification permits. Further, these agency problems are exacerbated in the Broker channel, resulting in the highest delinquency rates among Broker/Low-Doc mortgages. We provide evidence of borrower information falsification at both individual variable and aggregate levels. First, we find that the in-sample goodness-of-fit

(out-of-sample predictive power) of our delinquency prediction model is about 1.5 (2.5) times as high for the Full-Doc subsamples as for the Low-Doc subsamples. These differences suggest that borrower information collected for low-documentation loans is of lower quality--either in terms of inaccurately recorded data, omitted data, or intentionally falsified information--which compromises the ability of such information to predict delinquency. Second, reported income exhibits a perverse relation to delinquency probabilities among low-documentation loans, especially in the Broker/Low-Doc subsample where brokers both apply looser lending standards and are less diligent in verifying borrower information. The most plausible explanation for this observed pattern is information falsification. Through further analysis, we conservatively estimate that the median magnitude of income exaggeration is about 20% among low-documentation borrowers.

Finally, we find little evidence that the bank's interest rate scheme adequately priced for the different delinquency rates across loan types. While we find that low-documentation loans do indeed command a modest interest rate premium of 16-24 basis points, there is virtually no rate premium for broker-originated loans. These findings may be explained by a number of factors, including weakened incentives for screening due to high securitization rates during our sample period; fair lending regulations preventing banks from charging a premium for broker-originated mortgages; and the lack of evidence until mid-2007 (near the end of our sample period) of delinquency differences across origination channels and documentation types.

Our paper builds on a fast-growing literature on the mortgage crisis,[5] and most closely relates to a few recent empirical papers exploring the causes of the mortgage crisis using large sample micro-level archival data. Mian and Sufi (2008) identify the effects of the increase in the supply of mortgage credit on fueling the housing bubble between 2001 and 2005. Demyanyk and Van Hemert (2008) and Keys, Mukherjee, Seru, and Vig (2008) both use securitized loan data from LoanPerformance. Demyanyk and Van Hemert (2008) focus on the deterioration in loan quality between 2001 and 2006, while Keys, et al. (2008) focus on how securitization weakens the incentive of lenders to screen loan applicants. Commercial or government agency loan data sets typically used in the mortgage loan literature do not often contain borrower demographic characteristics, detailed loan contractual terms, or location (address) information, and usually only include securitized loans. Some earlier papers (e.g., Munnell, Tootell, Browne, and McEneaney (1996)) obtain demographic information from government data sources such as those reported for compliance with the Home Mortgage Disclosure Act (HMDA). However, loan performance and detailed location information are absent from these data sources, as are certain central economic variables such as the borrowers' credit scores and the loan-to-value ratio.

The contribution of this paper can be summarized as follows. First, our unique data set allows us to present the most comprehensive and updated predictive model of delinquency in the literature. Because we observe all loan and borrower attributes collected by the bank at origination, we are able to decompose delinquency differences into loan and borrower characteristics observed by the bank, versus those attributable to

[5] An incomplete list includes Chomsisengphet and Pennington-Cross (2006), Dell'Ariccia, Igan, and Laeven (2008), Mayer, Pense, and Sherlund (2008), and Ben-David (2008).

unobserved heterogeneity. Such decomposition provides us with an accurate calibration of the information possessed by the bank, which is essential for analyzing moral hazard and adverse selection problems in the loan market.

Second, the composition of loans in this data set reflects the mix of borrowers and loan products originated nationally both before and during the mortgage crisis. Our sample includes both prime and subprime loans, full- and low-documentation loans, loans kept on the bank's balance sheet and loans sold to the secondary mortgage market. As such, we are able to obtain separate analyses for different loan types partitioned by origination channel and documentation status, and to attribute delinquency and pricing to loan types with minimal omitted variable bias (in terms of the bank's information set). Moreover, with loan performance information updated through early 2009, we are able to capture the full effect of the crisis on the mortgage market.

Finally, we examine the extent to which mortgage pricing reflected market participants' recognition of the default risk associated with broker-originated and low-documentation loans. Our access to the bank's full information set on borrower characteristics allows us to conduct this analysis with minimal risk of an omitted variable bias.

The rest of the paper is organized as follows. The next section provides a description of the data. Section II contains a comprehensive analysis of predictive models of loan delinquency. Section III models borrowers' choices of loan origination channel and documentation level, and decomposes the cross-subsample differences in delinquency rates into two components: one reflecting observable lending standards and another reflecting unobservable borrower heterogeneity. Section IV documents and quantifies borrower information falsification among low-documentation loans. Section V discusses the extent to which mortgage interest rates reflected the incentive conflicts presented in the analysis. Finally, Section VI concludes.

I. Data and Sample Overview

A. Data Sources and Description

Our proprietary data set contains 721,767 loans funded by the bank between January 2004 and February 2008. The data set contains all information obtained 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. Loan contract information includes the loan terms (such as loan amount, loan-to-value (LTV) ratio, interest rate, and prepayment penalty), product category (such as whether the interest rate is fixed or adjustable), loan purpose (such as home purchase or refinance), origination channel (broker versus bank-originated), and documentation requirements.

Property data used in our analysis includes the property address, whether the property will be owner-occupied as a primary residence or used as an investment property/second home, and home appraisal value. Borrower data includes protected class demographic variables collected under the Home Mortgage Disclosure Act (HMDA) such as race, ethnicity, gender, and age, as well as all financial and credit information collected at

origination: income, cash reserves, expenditures, additional debts, bankruptcy and/or foreclosure status at loan origination, credit score,[6] employment status, employment tenure, self-employment status, and whether there are multiple borrowers (usually used as a proxy for marital status). Finally, we have monthly performance data for each loan through January 2009, including the monthly unpaid balance and the loan status: whether the loan payments are current or delinquent, the number of days delinquent, and whether the property is in a state of short sale (i.e., the sale of a home at a loss, in which the lender agrees to avoid foreclosure by accepting the sale proceeds in forgiveness of the outstanding loan balance) or foreclosure.

We use the recorded property addresses to match approximately three-quarters of the loans to community attributes—such as mean demographic characteristics and economic conditions—obtained at narrow levels of geography.[7] Using ArcGIS geocoding software and Decennial Census geographic boundary files, we match each property address to its census tract, zip code, metropolitan statistical area (MSA), and county. We obtain the following information at the census tract level from the Decennial Census and the Bureau of Labor Statistics: population count, median age of the residents, percent of residents who are black or Hispanic, and unemployment rate. In addition, we obtain zip-code level average household income information from the Internal Revenue Service's Individual Master File system. Finally, we obtain state-level housing price changes before and after loan origination using state-level housing price indices from the Federal Housing Finance Agency (FHFA).

B. Sample Overview

During the sample period, the bank experienced substantial changes in the composition of its loans and borrowers, as did the national mortgage market. Figure 1 reveals several salient patterns. First, the bank experienced a rapid increase in loan production during the mortgage boom, followed by a sharp decline during the housing bust; new loan originations increased from about 20,000 in the first half of 2004 to a peak of over 154,000 in the second half of 2006, followed by a precipitous decline starting in the second half of 2007.

[Insert Figure 1 here.]

Figure 1 also shows that the rapid expansion in loan production was driven almost exclusively by increased loan originations via the broker channel, and in particular by the expansion of low-documentation loans via the broker channel. Broker-originated loans represented 73% of all loan originations in the first half of 2004, increasing to 94% by the second half of 2006. While broker-originated low-documentation loans accounted for 39% of originations in early 2004, they comprised 75% of originations by late 2006.

Cumulative delinquency rates increased progressively and substantially over the time period in our sample (shown in Figure 2). At 18 months after origination only 6.7% of loans originated in the first half of 2004 were ever more than 60 days delinquent, as compared to 23.9% of loans originated in the second half of 2007.

[6] The credit score used by the bank is the median score obtained from the three major credit-reporting bureaus--Equifax, Experian, and TransUnion--and is numerically comparable and analytically equivalent to the FICO score by the Fair Isaac Corporation used in many other studies.

[7] Approximately one-quarter of the property addresses were unmatched, mostly due to variations in address recording (such as non-standard abbreviations), and in some cases, recording errors.

Demyanyk and Van Hemert (2008) document a similarly deteriorating trend for subprime loans from 2001-2006 using the LoanPerformance database.

[Insert Figure 2 here.]

We define all variables used in this paper in the Appendix I, and report their mean and standard deviation values by origination year in Table 1.

[Insert Table 1 here.]

The time trends in the key determinants of delinquency mostly reflect changes in housing prices, the loosening of lending standards during the boom period (2005 - 2006), and the subsequent tightening of loan underwriting guidelines by the bank starting in 2007. Mean loan-to-value ratios decreased from above 70% in 2004-2005 to 67% in 2006 before climbing to 77% in early 2008. Average borrower credit scores and job tenure (a proxy for job stability) also exhibited a U-shaped trend during the sample period. At the same time, average reported monthly income levels increased from $6,670 in 2004 to $7,490 in 2007, before trending downward. The growth in borrowers' incomes through early 2007 may have resulted from the booming economy, or from borrower income falsification on low-documentation loans.

The housing boom welcomed many first-time homebuyers to the mortgage market. In 2004, only 9.7% of borrowers in the sample were first-time homebuyers, a figure that climbed to 17.5% by 2006 before falling to 15.5% by 2008. During the sample period, black and Hispanic borrowers gained a significantly higher share of new loan originations, representing 5.7% and 9.6% of the borrower population in 2004, and 10.3% and 23.3% by the end of the sample period. At the same time, the national mortgage market experienced a similar increase in the percentage of first-time homebuyers and the expansion of credit to minority households who were disproportionately first-time homebuyers. According to national HMDA data on home purchase loans,[8] 6.6% (10.8%) of borrowers were black (Hispanic) in 2004; the numbers increased to 8.7% (14.4%) in 2006.

C. Sample Representativeness

Because our analyses rely on information from a single bank, it is natural to ask how representative this sample is and to what extent our results can be generalized. The properties in our sample are fairly represented across all 50 states, and their geographic distribution is roughly proportional to population density.

The large mortgage bank under analysis operated under an "outsource origination to distribution" business model wherein nearly 90% of loans were broker-originated and 72% of loans were originated by non-exclusive brokers. These figures are considerably higher than those for mortgage banks with more traditional models.[9] In addition, more than 85% of our sample loans were sold to the secondary market, a considerably higher proportion than the 60% figure reported in Rosen (2007) for the 2005-2006 period but comparable to the

[8] Source of information: http://www.ffiec.gov/hmdaadwebreport/NatAggWelcome.aspx.
[9] For example, a 2007 Wall Street Journal article estimates that brokers originate around 60% of all home loans. See "Mortgage Brokers: Friends or Foes?" by James Hagerty, *The Wall Street Journal*, May 30, 2007.

national securitization rate of 75-91% reported in "Inside Mortgage Finance" during the same period for subprime and non-conforming loans.[10]

We further compare our 2004-2008 sample average statistics to those covered by McDash Analytics, the most comprehensive commercial database on mortgage performance.[11] Our sample averages exhibit a comparable LTV, loan amounts that are 15% higher on average, and slightly lower borrower credit scores (about 5-8 points lower).[12] Finally, low-documentation loans represent 70% of the loans in our sample due to the lender's specialization in low-documentation products, but represent just 20% of all loans in the McDash database.

Last, subprime loans—which constitute 14-15% of our sample—are not over-represented.[13] Nationally, 18-21% of loans originated during 2004-2006 were subprime.[14] Our sample affords analyses on the full spectrum of the market, thereby complementing prior research focusing on the subprime sector (e.g., Keys, et al. (2008) and Demyanyk and van Hemert (2007)) and highlighting the widespread crisis beyond the subprime sector.

In summary, the bank in our analysis pursued an aggressive expansion strategy relying heavily on broker originations and low-documentation loans in particular. The strategy allowed the bank to grow at an annualized rate of over 50% from 2004 to 2006. Such a business model is typical among the major players that enjoyed the fastest growth during the housing market boom and incurred the heaviest losses during the downturn. By January 2009, the delinquency rate among the bank's outstanding loans approached 26%; while this figure is significantly higher than the industry average of 10.4%, the delinquency rate among subprime loans is comparable to the industry subprime average of 39%.[15]

Overall, this particular bank experienced a representative and yet amplified version of the boom-bust cycle experienced in the mortgage industry, thereby providing unique insights into the major problems underlying the mortgage crisis. To avoid generalizing on empirical relations that emerge from the bank's particular loan composition, we conduct our analyses on subsamples partitioned by loan type (origination channel and documentation level) rather than on the pooled sample.

II. Prediction of Loan Delinquency: Model Specification

A. General Framework

[10] Source of information: http://www.imfpubs.com/data/mortgage_securitization_rates.html.

[11] The comparison dataset is used in recent studies including Pikorski, Seru, and Vig (2009). We thank Amit Seru for providing the summary statistics for this data set.

[12] Part of the difference can be attributed to the fact that McDash over-represents prime loans. McDash covers about 60% of the entire mortgage market but only 30-40% of subprime originations.

[13] Despite its wide usage, there is actually no official definition of "subprime loans," which are loosely defined as loans to borrowers who might have difficulty repaying due to their poor credit, lack of credit history, low income, or high leverage. Our sample bank considers credit scores below 620 to be subprime, but with exceptions made in cases of mitigating financial circumstances. Given that we use full credit score information in our analysis, we do not flag subprime loans separately in our regressions.

[14] Source of information: *The State of the Nation's Housing, 2008* by the Joint Center for Housing Studies at Harvard University. Webpage: http://www.jchs.harvard.edu/publications/markets/son2008/son2008.pdf. The report mostly relies on the credit score cutoff at 640 for subprime classification.

[15] Source of information: Loan Processing Services (LPS). Webpage: http://www.lpsvcs.com/NewsRoom/IndustryData/Pages/default.aspx.

Delinquency prediction is one of the most important questions in the mortgage literature. We maintain the standard definition of delinquency as the borrower being at least 60 days behind in payment, or in a more serious condition of default (such as short sale or foreclosure). Our main model uses probit regressions complemented by several sensitivity checks.

All analyses throughout the paper, unless otherwise stated, control for loan origination year fixed effects and report standard errors that are robust to heteroskedasticity and within-cluster correlation of observations at the MSA level[16] to account for common shocks to real estate markets in the same MSA. The effective number of observations for the purpose of computing standard errors of estimated parameters is on the order of the number of clusters, which is 983 in the full sample. Finally, we use the 5% level as the criterion for statistical significance.

B. Main Analysis: Probit Model

The probit regression specification is as follows:

$$\begin{aligned} &Delinquency_i^* = X_i\beta + \varepsilon_i; \\ &Delinquency_i = 1 \text{ if } Delinquency_i^* \geq 0; = 0 \text{ otherwise.} \end{aligned} \tag{1}$$

In equation (1), $Delinquency_i^*$ is the underlying propensity of delinquency, and $Delinquency_i$ is an indicator variable for actual delinquency, defined as a loan being in a delinquent state (at least 60 days behind payment or worse) by the end of our sample period. In our sample, 25.6% of the loans are delinquent: 11.0% of the loans are 60 or more days behind in payments, 4.7% are in a state of short sale, and 9.9% are in a state of foreclosure.

We consider two sets of covariates *X*. Our short regressions include covariates that are largely exogenous in the sense that they are unlikely to be affected by the loan origination channel or documentation level. These covariates include whether the property is owner-occupied (*OwnerOccupied*); whether there is only one borrower on the loan application (*OneBorrower*); borrower income (*Income*), cash reserves (*CashResv*), credit score (*CreditScore*), gender (*Female*), ethnicity (*Hispanic, Black,* and *Asian*), age (*Age*), tenure (*Tenure*), self-employed status (*SelfEmploy*), and whether the borrower is a first-time homeowner (*FirstTimeOwner*); and housing price changes at the state level during the six months before and after loan origination (*HPI6MBefore* and *HPI6MAfter*).[17]

The longer regressions additionally include loan contract terms[18] and product categories: loan-to-value ratio (*LTV*); additional leverage on the same property (*AddLTV*); loan size (*Loan*); second-lien status (*SecondLien*); refinance status (*Refinance*); and variables indicating whether the mortgage interest terms have

[16] For observations where an address cannot be matched to any MSA, we form the clusters at the state level.

[17] Strictly speaking, housing price changes post origination (*HPI6MAfter*) are not known at loan origination, and can only serve as a predictive variable after a six month lead. For this reason, this variable is not included in some of the later analyses that rely strictly on information obtained at origination. In addition to controlling for housing prices, we conduct a further sensitivity analysis by excluding loans originated in California and Florida. Results are qualitatively indistinguishable from those of the full sample, indicating that they are not driven by a handful of "hot markets."

[18] Loan maturity is not included in the list of regressors due to a lack of variation; 30-year loans comprise 93% of our sample (the majority of the remainder are 15-year and 40-year loans).

adjustable rate (*ARM*), option ARM (*OptionARM),* or interest-only *(IO*, which may have either fixed or adjustable rates) features. Option ARM mortgages—nicknamed "pick-a-payment" mortgages—offer the borrower multiple payment options for a short time following origination, usually with low initial "teaser" rates, and most borrowers with these loan products choose payment levels below full amortization. To create mutually exclusive categories we exclude interest-only products from the OptionARM category and exclude both 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 equal to one. We do not include interest rates as a regressor in our main delinquency analysis because of two major complications. First, interest rates are partly set to price in delinquency propensity. Second, neither the initial nor the current interest rates in our data set are comparable across loans due to the presence of adjustable rate and variable payment products that re-set interest rate terms at different stages during the life of the loan. In Section V, we analyze interest rates in detail by examining specific subsamples in which the rate information is comparable across observations.

For both the short and long regressions, we conduct the analysis separately for each of the four subsamples and report the results in Table 2. We report the estimated coefficients of the probit model ($\hat{\beta}$) and t-statistics based on standard errors robust to clustering at the MSA level. We also report the statistics $\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$ at the bottom of each column, where $\phi(\bullet)$ is the standard normal probability density function, such that the empirical analog to the average partial effects (APE, or $E\left(\partial \Pr(Delinquency_i = 1 \mid X_i) / \partial X_i\right)$) can be calculated as $\hat{\beta}\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$.

[Insert Table 2 here.]

Most of the coefficients in Table 2 are intuitive. We defer discussion of these results along with additional analyses to Sections III and IV.

C. Sensitivity Analyses

We conduct three sensitivity analyses to supplement the previous section. First, we employ a hazard model to analyze the per-period "failure rate" (where a failure event is delinquency). Though closely related, the probit and hazard analyses differ in two aspects. First, the probit analysis does not distinguish loans that become delinquent at different points in time. In contrast, the hazard analysis captures the information of "survival time," i.e., the time from origination to delinquency. Second, the probit model considers delinquency the only "risk" (i.e., a loan is terminated before maturity and is not right-censored), while a hazard model can potentially incorporate prepayment as a competing risk to delinquency (Deng, Quigley, and Van Order (2000)). Given that a high proportion (22.3%) of loans in our sample were prepaid before maturity, we adopt the model developed by Fine and Gray (1999) to incorporate competing risks into a Cox (1972) proportional hazard model. The model acknowledges that the competing failure events (prepayment) censor the event of main interest (delinquency).

Results from the hazard analysis are presented in Table IA.1 of the Internet Appendix, where the estimates represent the semi-elasticity of hazard rate of delinquency with respect to the covariates (that is, $d \ln h / dX$). Though these coefficients are not directly comparable to those in the probit analysis for reasons stated above, their signs and statistical significance are highly consistent.

The only notable difference is that the coefficient on *HPI6MBefore* is positive (negative) in the probit (hazard) analysis. This contrast suggests that higher recent past housing price run-ups are associated with higher eventual delinquency rates, but early delinquency becomes less likely. Presumably, properties in areas with high recent past housing price appreciation played a larger role in the housing bubble, had more appraisal inflation, and were purchased by borrowers who hurried to buy without careful calculation. On the other hand, given the persistence in housing price appreciation, these borrowers were less likely to enter get into delinquency quickly.

This relation is confirmed by our second sensitivity analysis, where we conduct the probit analysis separately for early (within 12 months of loan origination) delinquency and later delinquencies. We report results in Table IA.2 of the Internet Appendix. Early and later delinquencies may be caused by quite different reasons. While later delinquencies are plausibly due to unforeseen negative economic shocks, the early delinquencies are more likely to be driven by adverse factors that already existed at loan origination (such as income falsification, excess leverage, and appraisal exaggeration). Indeed, high *HPI6MBefore* is negatively associated with early delinquency, but contributes to high eventual delinquency.

There are a few additional differences between early and later delinquencies that are worth highlighting. First, refinance loans have higher eventual delinquency rates but somewhat lower delinquency rates in the short run. Second, option ARM loans are associated with high delinquency rates only after 12 months post-origination. The mostly insignificant effect of option ARMs on near-term delinquency is plausibly due to artificially lower teaser rates in the initial period following loan origination.

In our final sensitivity analysis, we replace the dependent variable—delinquency at the end of the sample period—with an indicator variable for loans that were ever delinquent at some point in time (*EverDelinquent*). In our sample, 33.1% of the loans are ever delinquent; among these loans, 33.3% are still behind the payment schedule at the end of the sample period, 44.2% are in a state of short-sale or foreclosure, and the remaining 22.5% are considered "self-cured" (the borrower either resumed normal payments, or paid off the loan in full). We report results in Table IA.3 of the Internet Appendix. Again, results are very similar to those in Table 2.

III. Loan Types and Attribution of Differences in Delinquency

Table 2 reveals that broker-originated loans exhibit much higher delinquency rates than bank-originated ones; the difference is greater than 10 percentage points. We find similar delinquency differences based on loan documentation level; delinquency rates for low-documentation mortgages are 5-8 percentage points higher than for full-documentation ones. This section discusses the differences in loan performance across loan types along two dimensions. First, we examine which covariates determine a borrower's choice of loan type. Second, we

decompose the differential delinquency rates across loan types into differences due to observable versus unobservable characteristics.

A. Choice of Loan Origination Channel and Documentation Level

We start with a probit analysis where the dependent variables indicate the loan type (e.g., *Broker* or *Low-Doc*). Results are presented in Table 3. The first three columns use only loan and borrower characteristics as regressors, while the next three add neighborhood characteristics to the list of covariates. The sample size for the latter is about 25% smaller due to the additional data requirement.

[Insert Table 3 here.]

The following variables predict a higher likelihood that a borrower will obtain a loan from a broker rather than from the bank: high debt level, original purchase (as opposed to refinance), first lien, first-time owner, owner-occupied, low income, low credit score, female borrower, minority borrower, young borrower, short employment tenure, and self-employed. All non-white borrowers favor the Broker channel in comparison to white borrowers. Most of these characteristics (except perhaps the first-lien and self-employed variables) are associated, on average, with lower financial sophistication, less experience with mortgages, and lower credit quality.

Theoretically, a borrower living in any location can apply for a loan directly from the bank. In regions where the bank does not have branch operations, the loan application can be completed via phone or internet. The sorting of less financially sophisticated borrowers into the broker channel is compatible with two (non-mutually exclusive) explanations. On one hand, borrowers may select brokers because they believe brokers possess better knowledge about the products offered by different institutions, can help shop for competitive pricing, and can provide more personalized experiences and hand-holding throughout the loan origination process. On the other hand, borrowers may lack knowledge about alternative origination channels, or be unaware that they can approach the lender directly.

This particular lender did not have an established history as a brick-and-mortar depository institution, though they did expand branch operations in some regions during the last decade. As a result, and as indicated by the empirical results in Table 3, the bank relied on brokers for the majority of its origination volume, especially as it expanded rapidly into underserved communities. The bank relied on brokers both to increase origination volumes in the absence of high visibility as a depository institution, and to earn credit under the Community Reinvestment Act (CRA), a federal law regulating banks to ensure they meet the credit needs of low- and moderate-income households in the communities in which they hold a charter.

The variables that predict choosing a low-documentation loan have the following contrasts with those that predict choosing a broker. First, borrowers with low loan-to-value (LTV) ratios but high loan size are more likely to choose low documentation. Second, first-time owners and those purchasing owner-occupied properties are less likely to choose low documentation. Third, borrowers with high income and credit scores tend to choose low

documentation, and age is not correlated with documentation level. Finally, black borrowers do not appear disproportionately in low documentation loans, while Hispanic and Asian borrowers do. To summarize, low documentation loans do not necessarily attract less-experienced borrowers. The most prominent summarizing feature of these borrowers seems to be that they are "good on paper." That is, borrowers who have favorable "hard" information (i.e., information that is quantifiable and could potentially be verified, such as LTV, prior mortgage experience, high income, and high credit score) choose low documentation.

Prior research has shown that lending practices and borrower characteristics are correlated with neighborhood characteristics (e.g., Calem, Gillen, and Wachter (2004), Nelson (2009)). Columns 4-6 of Table 3 report the relation between neighborhood characteristics and the respective likelihoods that a borrower will select the broker channel or apply for a low-documentation loan. The model's regressors include average per capita income (*Avgincome)* at the zip code level, and also include the following regressors at the census tract level: Log population size (*Population*)[19], percentage of residents who are black (*Pctblack*) and Hispanic (*Pcthisp*), median age (*Medage*), unemployment rate (*Unemprate*), and the change in housing prices during the six-month period before loan origination at the state level (*HPI6MBefore*).

Brokers seem to predominate in neighborhoods with low minority representation and young residents. The combination of results from earlier columns indicates that minority households in non-minority neighborhoods are the prime clients of mortgage brokers. Low-documentation loans, on the other hand, are significantly more popular in minority neighborhoods and in booming neighborhoods (with low unemployment rates, high recent past housing price appreciation, and young populations).

B. Decomposition of Pairwise Subsample Differences in Delinquency

The analyses in this section attempt to decompose the difference in loan performance between different loan types into two components: one that can be predicted based on borrower and loan characteristics that are observable to the lender at loan origination, and another that can only be attributed to unobservables (information that is likely unknown to the bank at origination). Such a dichotomy has implications for understanding why delinquency rates vary across subsamples.[20]

We apply a non-linear version of the Blinder-Oaxaca (1973) decomposition to the probit model to separate the effects of observable qualities from the effects of unobserved heterogeneities. Let $D = 0, 1$ be the indicator variable for the two subsamples for comparison, and let Y be the indicator variable for loan delinquency. Specifically, we will compare loans from the Bank ($D = 0$) and Broker ($D = 1$) channels, controlling for the documentation level, and loans issued as Full-Doc ($D = 0$) and Low-Doc ($D = 1$), controlling for the origination

[19] The average and median population size of a census tract is between 5,000 and 6,000 residents.
[20] While an earlier study by Alexander, Grimshaw, McQueen, and Slade (2002) also documents higher delinquency rates among brokered loans, the study does not contain the level of borrower detail used in this study and hence cannot decompose the difference into differences due to characteristics that are observable versus unobservable to the bank.

channel. For all subsamples, we obtain coefficient estimates (β_0 and β_1 , corresponding to the D = 0 and D = 1 subsamples) from the probit model, reported in Table 2.

The difference in the delinquency rates between two subsamples can be expressed as:

$$\begin{aligned} &E(Y \mid D=1) - E(Y \mid D=0) \\ &= \left\{ E\left[\Phi(X\beta^0) \mid D=1\right] - E\left[\Phi(X\beta^0) \mid D=0\right] \right\} + \left\{ E\left[\Phi(X\beta^1) - \Phi(X\beta^0) \mid D=1\right] \right\}, \end{aligned} \tag{2}$$

or as:

$$\begin{aligned} &E(Y \mid D=1) - E(Y \mid D=0) \\ &= \left\{ E\left[\Phi(X\beta^1) \mid D=1\right] - E\left[\Phi(X\beta^1) \mid D=0\right] \right\} + \left\{ E\left[\Phi(X\beta^1) - \Phi(X\beta^0) \mid D=0\right] \right\}. \end{aligned} \tag{3}$$

Equations (2) and (3) are numerically different (because they use different "base" samples) but employ the same logic. There is no *a priori* reason to favor one choice over the other. For the economy of space, we present results using (2) where the subsample with higher delinquency rates serves as the base sample for covariate weights.

The left sides of the equations are the difference in the expected value of the outcome variable (delinquency) between the two subsamples. The right sides of the equations feature a sum of two terms. In labor economics, the first term is called the "endowment effect"; that is, the difference in the outcome due to different distributions of the covariates (the X variables) in the two subsamples, using the same set of coefficients for both subsamples. The second term is called the "coefficient effect" (in a production function, the coefficients are also referred to as "returns to factors") and estimates the hypothetical difference in delinquency if the two subsamples had identical covariate distributions but the coefficients remained different. The coefficient effect encompasses two possibilities: a differential sensitivity of the outcome to the covariates in the underlying model, or the effects of missing variables that spill over to the remaining covariates. Both possibilities reflect unobserved heterogeneity.

Given that all loans--regardless of origination channel or documentation level–are serviced by the same bank, there is no difference in ex post "treatment" in the sense of account monitoring and payment collection. Nevertheless, we cannot rule out a "treatment effect" in the sense that otherwise similar borrowers choosing different loan channels or documentation levels could also be exposed to different endogenous conditions (such as contract terms and degree of information falsification) that affect loan performance.

For this reason, we present two versions of the decomposition. The first version includes in the regressions only covariates that are largely exogenous, corresponding to the short regressions in Table 2 (columns (1) to (4)). These variables are unlikely to be affected by the broker channel or documentation level.[21] In the second version, we additionally include loan contract terms and product categories, corresponding to the long regressions in Table 2 (columns (5) to (8)) which could be endogenously chosen and potentially affected by the

[21] Two covariates, *Income* and *OwnerOccupied,* may not satisfy the strict exogeneity condition due to potential falsification by low-documentation borrowers, especially those who with broker-originated loans. For this reason, we present a variation of Table 4 Panel A that excludes these two variables. The results, presented in Table IA.4 in the Internet Appendix, are qualitatively similar to those in Table 4 Panel A.

Broker and *LowDoc* variables. The total difference (the left side of equation (2)) is reported in the bottom row, and is, by construction, 100% of the difference. The t-statistics are based on standard errors obtained through block bootstrap and clustered at the MSA level.

[Insert Table 4 here]

Panel A of Table 4 reveals that almost 100% of the 5−8 percentage-point difference in delinquency rates between Full-Doc and Low-Doc loans (conditional on origination channel) should be attributed to the "coefficient effect." The estimated "endowment effect" is small and is not statistically significant, regardless of whether we use all covariates known at loan origination in the prediction, or only a subset of relatively exogenous controls. We thus conclude that Low-Doc loans are just as "good on paper" as Full-Doc loans, but encompass more adverse selection along unobserved dimensions.

The comparison between Bank and Broker loans conditional on documentation level offers a different picture, as shown by Table 4 Panel B. Here, the endowment effect accounts for three-quarters of the 10 percentage-point total difference in delinquency rates between Bank and Broker loans among Full-Doc loans when all observable covariates are included, and approximately half of the 14 percentage-point difference in delinquency rates among Low-Doc loans. When controlling for the exogenous subset of covariates only, the share of the endowment effect is lower but still over half. Put differently, if the bank and its brokers had loaned to borrowers of the same *observable* quality, between half and three-quarters of the difference in the delinquency rate would have disappeared. The bank should have been aware of the observable differences in credit quality between borrowers from the two channels, yet relied on the broker channel for boosting volume and earning credit under the Community Reinvestment Act. The bank may also have underestimated the difference because evidence of higher delinquency from the broker channels did not surface until 2007 when the housing market softened.

Moreover, one-quarter (half) of the delinquency difference between Bank and Broker loans among Full-Doc (Low-Doc) loans is due to unobserved heterogeneity. Because unobserved heterogeneity constitutes a larger portion of the delinquency difference between Bank and Broker low-documentation loans, we have suggestive evidence that the degree of information falsification is worse among broker-originated mortgages.

It is worth noting that the above decomposition does not afford a clear conclusion regarding whether the coefficient effects are due to selection on unobservables that are not affected by brokers, or to broker endorsement of behaviors that are unobserved by the bank (including information falsification). To some extent, the message from Table 4 becomes stronger: If brokers originate loans on behalf of borrowers with worse measured characteristics while also facilitating falsification of those characteristics, then our results for broker-issued mortgages with worse *measured* characteristics form a lower bound on the same results with worse *true* characteristics.

To summarize, the Low-Doc channel does not necessarily compromise lending standards along verifiable metrics (such as LTV and credit score), but suffers from less careful verification—and potential falsification—of

some reported information (such as income and owner-occupancy status), or less diligent screening of borrowers along hard-to-quantify measures (such as other major expenditures).[22] In comparison, the Broker channel—which also lacked incentives for careful screening—served a borrower pool that was of significantly worse quality, even by observable, quantifiable, and potentially verifiable standards. Nevertheless unobserved heterogeneity also plays a role in causing the difference. Section IV provides supporting evidence that unobserved heterogeneity in the Low-Doc channel results in part from information falsification, and the effect is also most evident among broker-issued low-documentation mortgages.

C. Differences within the Broker Channel

The broker channel can be further divided into pure (non-correspondent) brokers and correspondent brokers. Pure brokers act as matchmakers and submit loan applications to a variety of banks for competitive pricing. In contrast, the correspondent brokers in our sample have long-term, established, and near-exclusive relationships with the bank for at least one product type (such as prime loans) and abide by the bank's particular underwriting guidelines in exchange for expedited loan processing. Thus, in comparison to non-correspondent brokers, the incentives of correspondent brokers are more closely aligned with those of the bank. To examine the difference between the two broker types, Table 5 replicates the long regressions in Table 2 for correspondent and non-correspondent brokers separately, interacted with the Full-Doc/Low-Doc sorting.

[Insert Table 5 here.]

A comparison of Table 5 to Table 2 confirms our conjecture. The patterns revealed in the Correspondent subsamples are always between those of the Bank subsamples and those of the Non-Correspondent subsamples, and tend to be closer to the former. In addition, relationships between loan performance and individual covariates are more similar when comparing the Bank and Correspondent subsamples than when comparing the Correspondent and Non-Correspondent subsamples.

Because results for the correspondent channel fall between those for the bank and non-correspondent broker channels, we supplement the channel choice prediction with an ordered probit analysis, where the high, middle, and low outcome values are assigned to the non-correspondent broker, correspondent broker, and bank channels, respectively. Results reported in column 7 of Table 3 confirm that non-correspondent brokers, more than the correspondents, issued mortgages to borrowers (as measured by high leverage, first-time home buying status, low credit scores, and minority status) and in neighborhoods (as measured by lower average income and age, and lower housing prices) with lower average credit quality.

[22] Borrowers may choose a low-documentation mortgage for a number of reasons, including because their earned income is not reported on a standard W-2 (e.g.., due to self-employment or cash compensation); because they would like to hide income that was not reported to the IRS; because they would like to hide adverse financial and credit information (e.g., prior bankruptcies or financial liabilities from divorce or legal issues); or because they would like to exaggerate their income in order to qualify for a larger loan amount.

IV. Liar's Loan: Model Predictive Power and Information Falsification

We have previously found that delinquency differences between Full-Doc and Low-Doc loans result fully from unobserved heterogeneity, and that the proportion of the delinquency difference between Bank and Broker loans attributable to unobserved heterogeneity is higher in the Low-Doc channel. In this section, we provide suggestive evidence that this unobserved heterogeneity results from income falsification among low-documentation loans--the "liar's loan" problem.

"Liar's loan" encompasses various forms of borrower information falsification, possibly at the encouragement of brokers who have stronger incentives to close deals than to screen applicants. Such falsification appears primarily among low- or no-documentation loans, where much of the recorded information is self-reported without strict verification. Anecdotal evidence[23] suggests that the following falsifications are among the most common: exaggerating income or assets; hiding other major expenditures; and claiming that properties purchased for investment/speculation purposes will be owner-occupied as primary residences.

Despite ample anecdotes, there are no formal empirical analyses of borrower information falsification and its impact on loan performance. Our paper fills this void by presenting two pieces of analysis. First, we use model predictive power as an aggregate measure of the quality of information recorded at loan origination. Second, we offer evidence of the falsification of individual variables by exploring how their relationship to loan performance differs between the Full-Doc and Low-Doc subsamples.

A. Model Predictive Power across Different Loan Types

Inaccurately recorded loan and borrower characteristics--whether due to unintentional mistakes or intentional falsification--will attenuate the empirical relationship between these variables and loan performance, thereby compromising the model's fit and predictive power. Because the bank services and maintains records for all loans in our sample, there is no obvious reason to believe that data recording quality should vary systematically across the subsamples after loan origination.

In Table 2, we observe that the goodness-of-fit (i.e., the in-sample model predictive power) is indeed substantially different across the four subsamples. More specifically, the two Full-Doc subsamples have much higher pseudo R-squared statistics[24] (22.4% and 18.4% for the Bank and Broker subsamples) than the two Low-Doc subsamples (14.1% and 14.9%), indicating higher quality explanatory variables in the Full-Doc subsamples.

Though the pseudo R-squared is a popular goodness-of-fit measure for non-linear models, it suffers from two major drawbacks. First, its interpretation is not as intuitive as that of the R-squared metric for linear models in term of the percent of variation explained. Second, the in-sample goodness-of-fit should not be equated with

[23] See, for example, "My Personal Credit Crisis" by Edmund Andrews, which appeared in the *New York Times* on May 17, 2009. The author provides a detailed description of his personal experience in qualifying for a loan far beyond his financial means by hiding, forging, and strategically managing information with the help of his mortgage broker.

[24] We adopt the standard pseudo R-squared: $(1 - \ln L / \ln L_0)$, where lnL is the maximized log likelihood value using all covariates, and lnL_0 is the maximized log likelihood value of the same model, but with a constant as the sole regressor.

model predictive power. When economic agents (the bank or its brokers) make decisions, their predictions are based on information revealed at the time, without knowledge of the full sample. Therefore, an out-of-sample prediction method is more appropriate for this particular research purpose because it avoids look-ahead bias. With these two issues in mind, we develop the following "excess percentage of correct predictions" measure to assess the predictive power of the probit model.

Let P_i denote the predicted probability of delinquency for the *i*-th observation, where the prediction is made out-of-sample using only information known at the time of the observation. Let Y_i denote an indicator variable for delinquency, and let $\overline{p}$ denote a cutoff value. Then the objective to maximize "correct predictions" can be expressed without loss of generality as maxmizing:

$$S = \omega S_1 + (1-\omega) S_2 - \alpha = \omega \Pr(P_i \geq \overline{p} \mid Y_i = 1) + (1-\omega) \Pr(P_i < \overline{p} \mid Y_i = 0) - \alpha \qquad (4)$$

for some $\omega \in (0,1)$, which reflects the relative importance of a type-I error S_1 (failure to predict a delinquent loan) and a type-II error S_2 (mistakenly predicting that a non-delinquent loan will be delinquent); α, equal to max(ω, 1−ω), represents the maximum weighted probability of obtaining a correct prediction with a random guess.

Our choice of ω is 1/2, which weighs the two types of prediction errors equally. The corresponding choice for $\overline{p}$ becomes $\overline{p} = E(Y)$; the empirical analog is the sample frequency of delinquency revealed at the time of the evaluation. We classify a loan as "predicted to be delinquent" if the out-of-sample predicted probability exceeds the time-adapted sample frequency of delinquency. This rule satisfies the *ex post* likelihood ratio rule; that is, $P_i > E(Y)$ ($P_i < E(Y)$) implies that the *i*-th observation is more (less) likely to be drawn from the subsample of *ex post* delinquent loans than from that of the *ex post* non-delinquent loans. Finally, α = 1/2 corresponds to ω = 1/2 . (Proofs of these assertions are in the Appendix II.)

Therefore, our measure of the model's predictive power becomes:

$$S = \ = \frac{1}{2}\Pr(P_i \geq E(Y) \mid Y_i = 1) + \frac{1}{2}\Pr(P_i < E(Y) \mid Y_i = 0) - \frac{1}{2}. \qquad (5)$$

We use the following empirical procedure to calculate the out-of-sample excess percentage of correct predictions. First, we divide each of the four subsamples into semi-year segments by the loan origination date, and pick one semi-year segment at a time to measure the accuracy of the model predictions. We call this the "test sample/period." Second, for each "test period," we use all information available up to just before the test period to estimate the model in equation (1) without the year dummy variables; we call this the "estimation sample/period." It is important to emphasize that not only do the loans in the estimation sample have to be originated before the test period, but their delinquency status must also be assessed at the beginning of the test sample/period. Third, we apply the predictive model using the coefficients estimated from the estimation sample on the test sample to form the predicted probability of delinquency. Finally, equation (5) formulates the calculation of the final measures.

Figure 3 depicts model predictive power by plotting the time series of the excess percentage of correct predictions (*S*) by loan type. It reveals several patterns. First, loan documentation type—not loan origination

channel—is the key determinant of the model's predictive power. The model's predictive power in the Bank/Full-Doc and Broker/Full-Doc subsamples is indistinguishable; the same can be said about the Bank/Low-Doc and Broker/Low-Doc subsamples. More importantly, the model's predictive power in the Full-Doc subsamples is substantially higher than for the Low-Doc subsamples. The across-time averages are as follows: Bank/Full-Doc (15.9%), Bank/Low-Doc (5.4%), Broker/Full-Doc (16.4%), and Broker/Low-Doc (6.6%).

[Insert Figure 3 here.]

Second, the predictive power of the model—especially for the Full-Doc subsamples—declined from 2005 to 2006, before rebounding slightly in late 2007. This trend suggests that loans originated during the boom period experienced positive shocks in delinquency that could not be predicted by their characteristics based on information available at the time of loan origination. Rajan, Seru, and Vig (2009) also find that the predictive power of credit score and LTV deteriorated during the high securitization period. The difficulty in predicting loan performance based on observed characteristics for loans originated in 2006 indicates the bank may not have been aware that it was originating highly delinquent loans during that time period; this explains why the bank did not tighten its lending standards until 2007, when large losses began to emerge.

There are two possible explanations for why low-documentation loans exhibit significantly lower predictive power for delinquency using information known to the bank at loan origination. First, low-documentation loans were subject to looser certification processes, allowing some borrowers to falsify information in order to qualify for higher loan amounts or to obtain more favorable loan terms. As a result, some of the variables in the regressions may contain measurement error, thereby compromising their predictive power. Second, low-documentation products might attract borrowers whose financial well-being varies more over time or with general economic conditions. For example, self-employed borrowers (for whom we control in our regressions) are significantly more likely to choose low-documentation loans (see Table 3), and their incomes tend to be more volatile and cyclical than the incomes of other employees. We next provide micro-level evidence to support the first hypothesis.

B. Evidence of Borrower Information Falsification from Individual Variables

B1. Overview

Due to both incentives and the underwriting system, falsification is most likely to occur among the following variables. First, borrowers purchasing a second home or investment property could falsely claim that the property will be owner-occupied and used as a primary residence, thereby securing a lower interest rate. While lenders are often able to verify occupancy status for refinance loans by requiring the borrower to submit proof of residence (such as utility bills), it is difficult to verify occupancy status for home purchase loans at origination. Occupancy fraud is often cited as a major contributor to the surge in delinquencies, as borrowers became over-leveraged from holding multiple mortgages.

Second, low-documentation loans enabled borrowers to falsify employment information—including employment tenure and self employment status—as well as income, assets, expenses, liabilities, and debt information. For many low-documentation loans, lenders do not verify borrowers' financial conditions by requiring a history of bank statements, W-2 forms, asset documentation (such as retirement, savings, or investment account information), or outstanding debt documentation (including student loan information, mortgage statements, credit card statements, and information on judgments/liens resulting from legal action). Borrowers who want to qualify for higher loan amounts or more desirable loan terms through a lower reported debt-to-income ratio could overstate their income and assets and/or understate expenses and other debt liabilities.

The analysis that follows focuses exclusively on income falsification for the following reasons. First, there is a strong *a priori* reason to believe that the income variable is most susceptible to falsification: both borrowers and brokers have better information about how income (rather than cash reserves, etc.) affects loan qualification and pricing. Second, asset is often more straightforward to verify than income because asset statements are usually more readily available than proof of income, especially among borrowers who are self-employed or cash-compensated. Finally, though several coefficients in Table 2 vary between the Full-Doc and Low-Doc subsamples, the coefficients on *Income* are markedly different, exhibiting a perverse relationship to delinquency in the Low-Doc subsample, for which we argue that income falsification is the most plausible explanation.

B2. Income Falsification

Tables 2 and 5 show that in the Full-Doc subsamples, higher income[25] is negatively associated with delinquency; however, the sign on the *Income* coefficient switches in the Low-Doc subsamples. Moreover, the coefficients are particularly strong in the Broker/Low-Doc subsample, where higher income is associated with a significantly higher propensity for delinquency. The most plausible explanation for this contrast is that, when income is not verified, higher income (conditional on all other attributes) may reflect exaggeration rather than financial strength. Reported income will have a positive sign in the delinquency prediction regressions if the incentive to exaggerate income is negatively correlated with individual credit quality. Moreover, borrowers who are more likely to falsify income may adversely select into the broker channel; or brokers may encourage borrower income falsification.

The dummy variable for missing income information, *IncomeMiss*, offers corroborative evidence. In the Bank/Full-Doc and Broker/Full-Doc subsamples, only 0.6% and 0.9% of the observations have missing income information, and in these subsamples missing income information does not predict loan performance. Thus, in the Full-Doc subsamples, the sporadic cases of missing income information most likely result from data recording error and not from falsification. In contrast, income is missing for 10.3% and 9.2% of the observations in the Bank/Low-Doc and Broker/Low-Doc subsamples, respectively. Missing income information significantly predicts

[25] In the regression, the *Income* variable is coded as zero when it is missing, and the dummy variable for missing income information, *IncomeMiss,* is set equal to one.

higher delinquency propensity in the Broker/Low-Doc subsample, where missing income information is associated with a 5.5 percentage point increase in the probability of delinquency. The same effect is present but not significant in the Bank/Low-Doc subsample. Thus, purposefully not reporting income information is a low-documentation-only phenomenon. Presumably, these borrowers are more likely to have irregular incomes and are more likely than comparable Full-Doc borrowers to exaggerate or omit their incomes on the loan application.[26]

A comparison of Table 2 and Table 5 reveals that the various perverse relations discussed above for broker-originated loans are mostly driven by non-correspondent brokers. This evidence suggests that correspondents are far less likely to encourage or accommodate borrower information falsification than non-correspondents; correspondent brokers have stronger reputation concerns due to their exclusive or long-term relationships with the bank.

We now examine the magnitude of income exaggeration by borrowers who self-report income. While we are not able to obtain precise estimates at the individual level, we some conservative estimates for the average extent of income falsification based on the following identifying assumption:

$$E(Income^* \mid X = x, LowDoc) \leq E(Income^* \mid X = x, FullDoc), \qquad (6)$$

where *Income** denotes the borrower's true income, and *X* denotes a vector of borrower characteristics. Formally, equation (6) is implied by the condition that *Pr(FullDoc|X, Income*)* is non-decreasing in *Income**.

All that is required for equation (6) to hold is a relative preference ordering: if Borrower A's true income is more favorable than that of a similarly situated Borrower B, then on average Borrower A should not have a stronger preference for low-documentation loans than Borrower B. Such an assumption is plausible because a high certified income is more likely to result in lower interest rates or more favorable loan terms on full-documentation loans; some of these benefits are forfeited in low-documentation loans because the sensitivity of loan pricing to uncertified income is lower. Self-reported income could still materially affect the loan qualification, providing an incentive for falsification.

The only group for whom equation (6) may not hold is the self-employed. Self-employed borrowers disproportionately choose low-documentation loans (as shown in Table 3), not necessarily because they want to exaggerate their income but because their income is often difficult to certify (e.g., they do not have W-2 forms) or they do not wish to reveal their true cash flows for tax reasons. We therefore exclude the self-employed from our income exaggeration estimations.

Our first income exaggeration estimate simply compares borrower income (at the household level) to the average income of the neighborhood where the property is located. We obtain the average per capita adjusted gross income information at the zip code level from the Internal Revenue Service's Individual Master File (IMF) system for the years 2004, 2005, and 2006. A zip code area has 2,326 households on average, and the average

[26] Some high-income borrowers may also have an incentive to hide income information when applying for "no ratio" mortgages (a type of low-documentation loan). By not stating their income, ratios such as debt-to-income would be left unreported. Such an omission allows a borrower to achieve higher leverage through multiple mortgages.

household size is 3.3 people. We use 2006 data for loans originated in the post-2006 years. The average ratios of borrower household income to neighborhood average per capita income are 3.6 and 3.3 for the two Full-Doc subsamples, and are considerably higher at 4.3 and 3.8 for the two Low-Doc subsamples. Thus, assumption (6) implies that the average degree to which low-documentation borrowers exaggerate their incomes is at least 16%-19%, if the ratio of their true income to the neighborhood average is no higher than that of their full-documentation counterparts.

A more refined estimate incorporates borrower demographics in addition to neighborhood attributes to proxy true income (*Income**). Suppose a borrower's *Income** can be expressed as a linear function of borrower characteristics, neighborhood characteristics, year dummies, and an error term, where the error term is mean independent of covariates conditional on documentation status. Then such a function may be estimated reliably using the sample of full-documentation loans, because there should be no systematic bias in certified income; hence, conditionally average *Income* ≈ average *Income**. Below we report the regression output for full-documentation loans, where the dependent variable is the reported (and certified) household monthly income in $1,000 units, and the t-statistics are reported below the coefficients.

$$
\begin{aligned}
Income = {} & \underset{[18.01]}{0.014 * CreditScore} - \underset{[-16.49]}{0.846 * Female} + \underset{[13.31]}{0.651 * \ln(Age)} - \underset{[-1.92]}{0.416 * Hispanic} \\
& - \underset{[-4.31]}{0.430 * Black} + \underset{[5.04]}{0.575 * Asian} + \underset{[4.40]}{0.051 * AvgIncome} - \underset{[-2.15]}{0.030 * Unemprate} \\
& + \underset{[2.58]}{0.131 * Y2005} + \underset{[5.40]}{0.373 * Y2006} + \underset{[4.76]}{0.299 * Y2007} + \underset{[0.096]}{0.010 * Y2008}
\end{aligned}
\tag{7}
$$

R-squared: 6.9%; number of observations: 138,514.

All coefficients in the above regression are intuitive. Older borrowers and borrowers with higher credit scores tend to have higher incomes. Female borrowers have lower income on average.[27] Black and Hispanic borrowers have lower income on average than white borrowers, and Asian borrowers as a group have the highest income. In addition, borrower income is significantly and positively correlated with the zip-code area average income (*AvgIncome*) and negatively correlated with the census tract unemployment rate (*Unemprate*). Finally, overall borrower income grew from 2004 (the omitted year in the regression) to 2006, and then decreased after.

The identifying assumption of (6)—which presumes that the error term from regression (7) is not positively correlated with Low-Doc status—provides the upper bound for the expected true income of low-documentation borrowers by applying the estimated coefficients from (7) to the covariates of these borrowers. We generate an "income exaggeration" variable to capture the difference between the reported *Income* and the estimated *Income**. In dollar terms the estimated average (median) income exaggeration is $1,830 ($753) per month; in percentage terms, the average (median) low-documentation borrower reports income that is 28.7%

[27] This gender effect is not primarily due to the male-female wage gap but rather to the fact that a female being listed as the sole borrower is a proxy measure for a female head of household; female-headed households have lower income on average than male-headed households.

(20.0%) above their estimated true income level. These being conservative estimates, the data suggest serious income falsification among low-documentation borrowers using the benchmark of full-documentation borrowers.

The correlations between estimated true income, estimated income exaggeration, and loan performance are all highly statistically significant, and reveal more about the incentives for and consequences of income falsification. First, the correlation between the estimated true income and estimated income exaggeration in percentage terms is -7.9%, indicating a stronger incentive to inflate income when the true income is lower. Second, the correlation between the estimated true income and *ex post* delinquency is -23.5%, recovering the normal inverse relationship between income and delinquency in the Low-Doc subsample that was perverted using reported income. Finally, as expected, the correlation between estimated income exaggeration and *ex post* delinquency is positive at 8.2%. In other words, delinquency risk increases when borrowers inflate income, presumably in order to obtain a loan beyond their true means.

V. Loan Types and Loan Pricing

The previous sections show that loans originated via the broker channel and with low-documentation requirements are of lower quality due to heterogeneity in both observable and unobservable characteristics, including likely information falsification. An important question thus arises regarding whether market participants were aware of such differences *ex ante* and whether loan pricing (the interest rate) adequately reflects the additional risk associated with brokered and low-documentation loans.

Table 6 addresses this question by analyzing the determinants of interest rates with loan types as the key regressors of interest. Our full sample consists of a mixture of fixed-rate loans (46.6% of the sample) and various adjustable rate loan products. The interest rates on many of the adjustable rate loans had not yet aged out of the initial rate period by the end of the sample period. Therefore, neither the initial (*InitialRate*) nor the current interest rates (*CurrentRate*) are comparable across loans. For this reason we conduct our analyses using several separate subsamples; we examine fixed- and adjustable-rate loans separately (columns (1) and (2)) and focus on the current rate of early-period adjustable-rate loans (column (3)), which are more likely to have reset after the expiration of the introductory period. Control variables include all regressors that appear in the delinquency prediction analysis as reported in Table 2.

[Insert Table 6 here.]

Most of the coefficients on the control variables are intuitive. That is, variables that are associated with higher delinquency rates also tend to command higher interest rates. The only notable exception is loan size (*Loan*) which predicts higher delinquency rates but is associated with lower interest rates. Such a relation can be interpreted as reverse causality: borrowers tend to borrow more when facing low interest rates.

It is worth noting that federal law prohibits loan pricing based on demographic information including race/ethnicity. Although some coefficients on these variables are statistically significant, the economic magnitude is quite small. For example, after controlling for observable information, black borrowers pay 10-17 additional

basis points in the interest rate as compared to white borrowers. Yet given the average (median) unpaid balance on loans among black borrowers of $185,000 ($150,000), the estimated black-white difference in interest rate amounts to an additional monthly payment of $16-$27 (or $13-$22) using the mean (or median) balance. Such a payment difference should not be pivotal in loan delinquency.

Most important are the coefficients on loan type. With the Bank/Full-Doc category serving as the benchmark (the omitted category), column (1) of Table 6 shows that among fixed-rate loans, low-documentation loans command interest rates that are 24 basis points higher on average, while the same premium for brokered loans is a very insignificant 3 basis points. The rate premium is very modest for low-documentation loans ($46 more in monthly payments for a median sized loan) considering the adverse selection involved. As Table 4 shows, 100% of the incremental delinquency rates among low-documentation loans is due to unobserved heterogeneity. The same rate premium is non-existent for brokered loans, implying that the bank did not price loans to account for the incremental delinquency between bank and broker originations that are attributed to unobservables.

Adjustable-rate loans are about equally represented across origination and documentation channels. Brokered adjustable-rate loans are associated with initial interest rates that are 37 basis points *lower* than those for bank-issued adjustable rate loans (column (2)). However, the broker effect is both economically and statistically insignificant for the subsample of early-period (2004-2005) loans (column (4)), for which the interest rate is more likely to have been re-set. Combined, the two results suggest that brokers are more likely to sell products that carry a more lucrative "teaser rate." The same contrast is non-existent for low-documentation loans.

The overall evidence suggests that the loan pricing for brokered and low-documentation loans did not compensate for their additional risk. In comparison, LaCour-Little (2009) shows that brokered loans tend to have interest rates that are 20 basis points higher than loans available directly from retail lenders. More similar to our finding, Alexander, Grimshaw, McQueen, and Slade (2002) show that the agency risk associated with brokers was not priced during the period of low default. However, they also show that the rate premium surged to more than 200 basis points after such risk was recognized. The same figure is much lower in our study: Column (4) shows that even among the fixed-rate loans issued after July 1st, 2007 (right after two Bear Sterns hedge funds disclosed colossal losses due to their subprime exposure, an event considered to signal the start of the financial crisis), the interest rate premium for brokered and low-documentation loans remained modest at 12 and 30 basis points, respectively.

In addition to relying on a different sample, our study controls for more detailed borrower characteristics, which decrease the coefficient on *Broker* in the loan pricing equation because brokered loans tend to have lower observable quality as measured by borrower characteristics (see Tables 3 and 4). In fact, if we drop the borrower characteristics that were not included in previous studies, the rate premium for the broker channel more than doubles our current estimates. Moreover, our sample bank took measures to ensure they had a one-door pricing policy (i.e., no differential pricing treatment based on origination channel) in compliance with fair lending regulations.

Our analysis nevertheless raises the question of why this major mortgage bank—as well as other market players—issued lower-quality loans without adequately pricing the additional risk, and allowed the deterioration in borrower and loan quality to persist before tightening its lending standards. One possible explanation is information. While it was obvious in 2009 that brokered and low-documentation loans were more delinquency-prone, the differences were not significant in the early part of the sample period when overall delinquency rates were low. The delinquency rates for all outstanding loans were no higher for brokered loans than for bank-originated loans until 2006; 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, and not until 2008 that the mortgage crisis began to unfold on a full scale.

Another plausible explanation concerns incentives. The expansion of the secondary mortgage market and the ease of loan securitization weakened the bank's incentive to screen borrowers by allowing the bank to offload risk. We refer the readers to Keys, Mukherjee, Seru and Vig (2008) for an analysis on the relation between loan performance and the *ex ante* probability of loan securitization, and to Jiang, Nelson, and Vytlacil (2010) for a contrast between the *ex ante* and *ex post* relations between the two.

VI. Conclusion

This paper uses a unique, proprietary data set from a major national mortgage bank to examine how mortgage loan performance relates to loan origination channel, documentation level, and borrower demographics. Our research aims to identify and quantify the micro-level fundamental causes of the mortgage crisis and highlights two major problems. The first problem arises between the bank and its mortgage brokers, who originate observably lower quality loans. We find that brokered loans are more than 50% more likely to be delinquent than bank-originated loans, and up to three-quarters of this difference can be attributed to lower borrower/loan quality based on observable risk factors. The second problem lies between lenders and borrowers in the form of borrower information falsification among low-documentation loans, especially when issued through a broker. We find poor model predictive power and strong evidence of information falsification among low-documentation loans. Finally, we show that loan pricing did not adequately compensate for the additional risk of brokered and low-documentation loans.

Appendix I: Definition of main variables

AddLTV	The ratio of additional loans (including from other banks) secured to the property
Age	Borrower age
ARM	Dummy variable = 1 if the mortgage is adjustable-rate (excludes option ARM and interest-only mortgages)
Asian	Dummy variable = 1 if the borrower is Asian
Avgincome	Average income per capita in the census tract where the property is located
Black	Dummy variable = 1 if the borrower is black
Cashresv	Cash reserves, in multiples of monthly mortgage payments
Delinquency	Dummy variable for delinquency, defined as being at least 60 days behind in payment
Female	Dummy variable = 1 if the borrower is female
CreditScore	Borrower credit score
CurrRate	Current interest rate (updated in February 2008) on the loan
FirstTimeOwner	Dummy variable = 1 if the borrower is a first-time mortgage borrower
Hispanic	Dummy variable = 1 if the borrower is Hispanic
HPI6MAfter	State level housing price change during the six months post-origination (Federal Housing Finance Agency home price index)
HPI6MBefore	State level housing price change during the six months prior to origination (Federal Housing Finance Agency home price index)
Income	Monthly income of the borrower, in thousands
IncomeMiss	Dummy variable = 1 if the income information is missing
InitialRate	Initial interest rate on the mortgage
IO	Dummy variable = 1 if the mortgage carries an interest-only feature.
Loan	Total loan amount
LTI	Loan-to-income ratio, the percentage of monthly gross income that is used to pay for the mortgage
LTV	Loan-to-value ratio
Medage	Median age of residents in the census tract where the property is located
OneBorrower	Dummy variable = 1 if there is only one borrower on the mortgage
OptionARM	Dummy variable = 1 if the mortgage is an option ARM but does not carry an interest-only feature
OwnerOccupied	Dummy variable = 1 if the property is the owner's primary residence
Pctblack/Pcthisp	Percent 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
PrepayPenalty	Dummy variable = 1 if there is hard prepayment penalty in the loan contract
Refinance	Dummy variable = 1 if the mortgage purpose is for refinance (rather than initial purchase)
Secondlien	Dummy variable = 1 if the mortgage is a second lien against the property
SelfEmploy	Dummy variable = 1 if the borrower is self-employed
Subsample1	Bank/Full-Doc subsample
Subsample2	Bank/Low-Doc subsample
Subsample3	Broker/Full-Doc subsample
Subsample4	Broker/Low-Doc subsample
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

Appendix II. Proof of Statements in Section IV.A.

1. $\overline{p} = E(Y)$ is the unique solution to the maximization of Equation (6) when ω = 1/2.

We first show that the maximization of (6) has a unique solution of $\overline{p}$ as a function of ω:

$$\overline{p} = \frac{(1-\omega)E(Y)}{\omega\left[1-E(Y)\right]+(1-\omega)E(Y)} . \tag{8}$$

Let f^D (f^{ND}) be the probability density functions of the predicted probability of delinquency for the subsample of loans that are *ex post* delinquent (non-delinquent), and f be the probability density function for the combined sample.

Suppose the model is correctly specified, i.e., equation (1) holds with the residual ε normally distributed. We have *E(P) = E(Y)* by the Law of Iterated Expectations. By Bayes Rule and the Law of Iterated Expectations we have:

$$f^D(v) = \frac{vf(v)}{E(Y)};\ f^{ND}(v) = \frac{(1-v)f(v)}{1-E(Y)}, \tag{9}$$

for all $v \in [0,1]$.

Equation (9) implies:

$$\Pr\left(P \geq \overline{p} \mid Y=1\right) = \int_{\overline{p}}^{1} f^D(v)dv = \int_{\overline{p}}^{1} \frac{vf(v)}{E(Y)}dv = \frac{\left[1-F(\overline{p})\right]}{E(Y)}E(P \mid P \geq \overline{p}) . \tag{10}$$

Similarly,

$$\Pr\left(P < \overline{p} \mid Y=0\right) = \int_{0}^{\overline{p}} f^{ND}(v)dv = \int_{0}^{\overline{p}} \frac{(1-v)f(v)}{1-E(Y)}dv = \frac{F(\overline{p})}{1-E(Y)}E(P \mid P < \overline{p}) . \tag{11}$$

We obtain (10) by substituting (12) and (13) into (4), and arrive at $\overline{p} = E(Y)$ by setting ω = 1/2 in (10).

2. $\overline{p} = E(Y)$ satisfies the likelihood ratio property:

Using equation (9) and the fact that *Var(Y) = E(Y)[1-E(Y)]*, we have:

$$f^D(v) - f^{ND}(v) = f(v)Var(Y)[v - E(y)] . \tag{12}$$

Thus,

$$\begin{aligned} f^D(v) - f^{ND}(v) &> 0 \text{ if } v > E(y), \\ &= 0 \text{ if } v = E(y), \\ &< 0 \text{ if } v < E(y). \end{aligned} \tag{13}$$

Therefore, $f^D(v)$ and $f^{ND}(v)$ cross once at $v = \overline{p} = E(Y)$. With such a choice of $\overline{p}$, we classify a loan as "predicted to be delinquent" if and only if it is more likely to be from the distribution of *ex post* delinquent loans than from that of the *ex post* non-delinquent loans. Hence the classification satisfies the likelihood ratio rule.

References

Alexander, William, Scott Grimshaw, Grant McQueen, and Barrett Slade, 2002. Some Loans are More Equal than Others: Third-party Originations and Defaults in the Subprime Mortgage Industry. *Real Estate Economics* 30(4): 667-697.

Ashcraft, Adam and Til Schuermann, 2008, Understanding the Securitization of Subprime Mortgage Credit, *Federal Reserve Bank of New York Staff Report,* no. 318.

Ben-David, 2008, Manipulation of Collateral Values by Borrowers and Intermediaries, working paper, Ohio State University.

Calem, Paul, Kevin Gillen, and Susan Wachter, 2004, The Neighborhood Distribution of Subprime Mortgage Lending, *Journal of Real Estate Finance and Economics,* 29, 393-410.

Chomsisengphet, Souphala and Anthony Pennington-Cross, 2006, The Evolution of the Subprime Mortgage Market, *Federal Reserve Bank of St. Louis Review*, 88:1, 31-56.

Cox, D. R., 1972, Regression Models and Life-Tables, *Journal of Royal Statistical Society. Series B (Methodological)*, 34:2, 187-220.

Dell'Ariccia, Giovanni, Deniz Igan and Luc Laeven, 2008, Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market, Working Paper.

Demyanyk, Yulia and Otto van Hemert, 2007, Understanding the Subprime Mortgage Crisis, working Paper, New York University.

Deng, Yongheng, John Quigley and Robert Van Order, 2000, Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options, *Econometrica,* 68 (2), 275-307.

Fine, Jason, and Robert Gray, 1999, A Proportional Harzards Model for the Subdistribution of a Competing Risk, *Journal of the American Statistical Association,* 94 (446), 496-509.

Haughwout, Andrew, Christopher Mayer, and Joseph Tracy, 2009, Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing, *Federal Reserve Bank of New York Staff Reports,* No. 368.

Jiang, Wei, Ashlyn Nelson, and Edward Vytlacil, 2009, Mortgage Securitization and Loan Performance: A Contrast of Ex Ante and Ex Post Relations, Working Paper, Columbia Business School.

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru and Vikrant Vig, 2008, Securitization and Screening: Evidence from Subprime Mortgage Backed Securities, *Quarterly Journal of Economics*, forthcoming.

LaCout-Little, Michael, 2009, The Pricing of Mortgages by Brokers: An Agency Problem? *Journal of Real Estate Research* 31, 235-264.

Mayer, Christopher, Karen Pence, and Shane Sherlund, 2008, The Rise in Mortgage Defaults: Facts and Myths, *Journal of Economic Perspectives*, forthcoming.

Mian, Atif and Amir Sufi, 2008, The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis, working paper, University of Chicago.

Munnell, Alicia, Geoffrey Tootell, Lynn Browne, and James McEneaney, 1996, Mortgage Lending in Boston: Interpreting HMDA Data, *American Economic Review* 86, 25-53.

Nelson, Ashlyn, 2009, Credit Score, Race, and Residential Sorting, Working Paper, Indiana University.

Oaxaca, Ronald, 1973, Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review,* 14, 693-709.

Petersen, Mitchell, 2004, Information: Hard and Soft, working paper, Kellogg School of Management.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig, 2009, Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis, working paper, Columbia Business School.

Rajan, Uday, Amit Seru, and Vikrant Vig, 2009, The Failure of Models that Predict Failure: Distance, Incentives, and Defaults, working paper, London School of Business.

Rosen, Richard, 2007, The Role of Securitization in Mortgage Lending, *Chicago Fed Letter,* No. 244.

Figure 1. Number of Loans and Composition by Semi-Year: 2004-2008

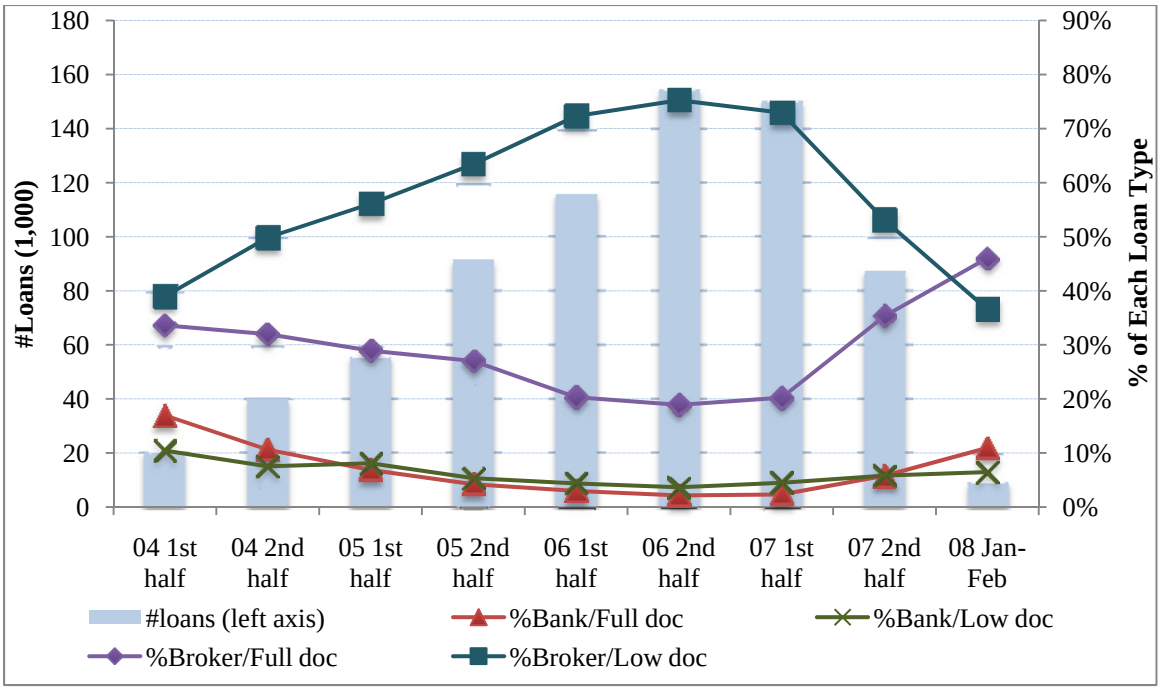

Figure 2. Delinquency Rates since Loan Origination by Semi-Year: Updated to January 2009

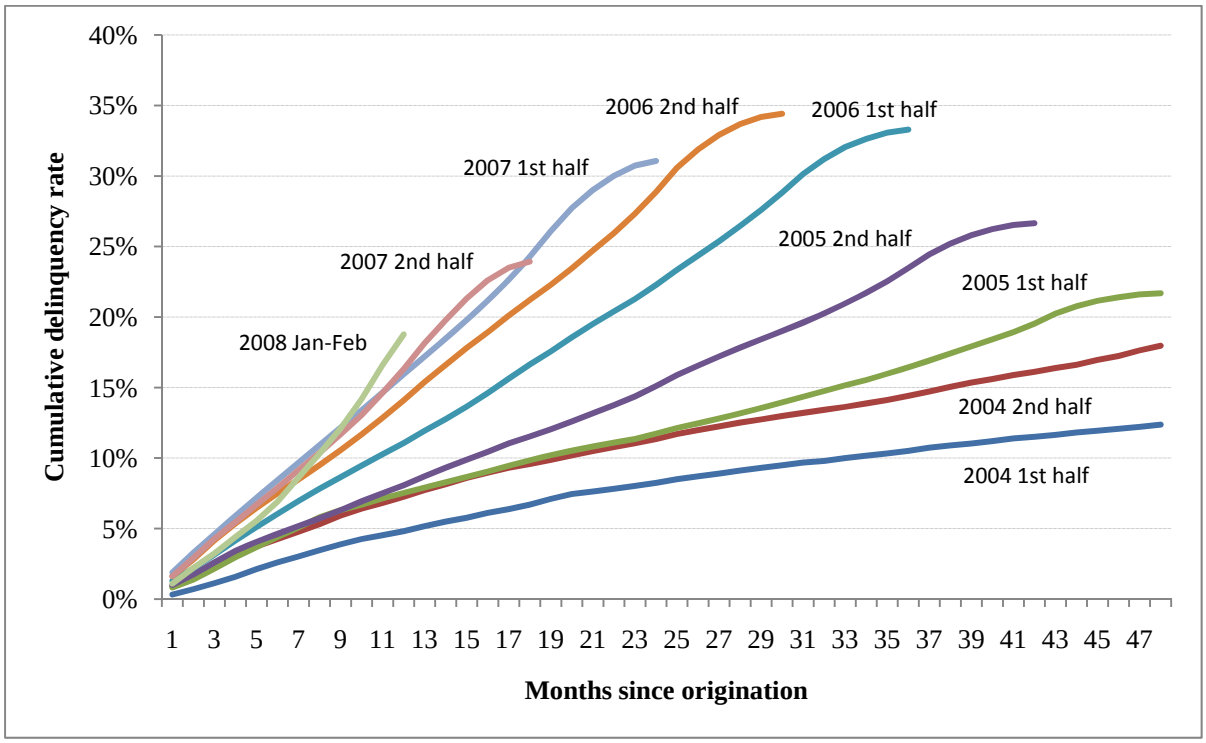

Figure 3. Time Series of Out-of-Sample Model Predictive Power by Loan Type

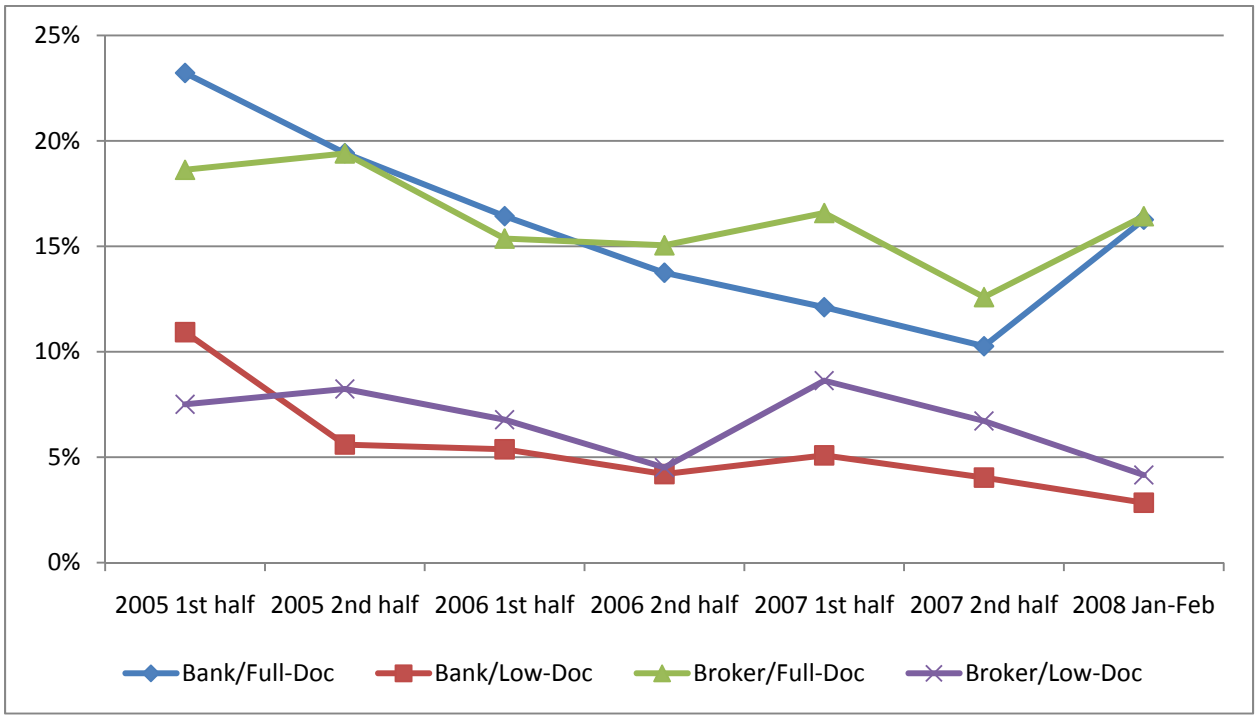

Table 1. Summary Statistics

This table reports the mean and standard deviation (the first and second line of each variable) values of the major borrower characteristics by year from 2004 to early 2008.

	2004	2005	2006	2007	2008 (Jan-Feb)
Age (years) (average)	44.46	43.54	42.92	44.46	45.42
(std. dev.)	12.21	12.33	12.62	12.69	12.57
Credit score	699.36	697.60	693.82	696.28	699.51
	60.91	56.99	53.39	57.21	62.21
Income ($1,000, monthly)	6.67	6.64	7.38	7.49	7.21
	7.45	7.86	8.19	9.54	8.58
Initial interest rate	5.43	5.23	6.43	7.19	6.84
	2.14	2.62	2.90	1.87	0.76
Loan size in $1,000	232.33	259.84	266.20	282.16	281.23
	161.34	177.04	193.57	209.71	165.42
Loan-to-income	2.92	3.03	2.72	2.96	3.73
	1.66	1.71	1.67	1.80	1.83
Loan-to-value	0.70	0.71	0.67	0.69	0.77
	0.18	0.18	0.22	0.23	0.17
Tenure (months)	75.56	65.73	56.08	66.78	92.89
	91.28	84.88	78.89	85.41	91.23
%Asian	5.5%	5.7%	5.1%	5.1%	4.2%
%Black	5.7%	7.0%	8.3%	9.2%	10.3%
%Black & Hispanic that are first-time owners	12.1%	17.0%	23.6%	21.5%	20.3%
%Female	31.3%	31.9%	33.6%	35.0%	36.0%
%First-time owner	9.7%	13.4%	17.5%	15.7%	15.5%
%Hispanic	9.6%	14.7%	19.6%	22.8%	23.5%
%Owner occupied	84.8%	84.5%	85.8%	84.0%	88.3%
%Refinance	61.2%	56.5%	55.0%	61.7%	65.3%
%Self-employed	18.4%	18.1%	19.9%	21.8%	20.5%

Table 2. Delinquency Prediction: Probit Analysis

The dependent variable is loan delinquency, and the estimation method is probit as specified in equation (1). The definitions of all covariates (X) are given in the Appendix. Reported are the coefficients ($\hat{\beta}$) and t-statistics (in the brackets) that adjust for clustering at the MSA level. At the bottom of the table, we report the sample frequency of delinquency, the pseudo R-squared, the number of observations, and the sample average of the probit density function ($\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$) that can be used to construct the average partial effect $\hat{\beta}\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$.

	(1) Bank/Full-Doc	(2) Bank/Low-Doc	(3) Broker/Full-Doc	(4) Broker/Low-Doc	(5) Bank/Full-Doc	(6) Bank/Low-Doc	(7) Broker/Full-Doc	(8) Broker/Low-Doc
LTV					1.585***	2.389***	2.004***	2.960***
					[14.51]	[18.81]	[18.42]	[19.13]
AddLTV					1.276***	1.484***	1.627***	2.965***
					[6.19]	[7.23]	[16.30]	[22.06]
Loan (log)					0.100***	0.154***	0.194***	0.212***
					[3.71]	[6.22]	[8.49]	[8.28]
SecondLien					0.374***	0.837***	0.594***	0.377***
					[2.71]	[7.57]	[10.62]	[5.37]
Refinance					-0.024	0.007	-0.055**	0.082***
					[-0.58]	[0.25]	[-2.42]	[4.72]
PrepayPenalty					0.058	-0.038	-0.025	0.056***
					[1.10]	[-0.92]	[-1.17]	[4.19]
ARM					0.221***	0.145***	0.209***	0.182***
					[6.90]	[6.05]	[12.37]	[11.56]
OptionARM					0.183***	0.337***	0.239***	0.292***
					[3.21]	[9.84]	[6.23]	[11.44]
IO					0.182***	0.201***	0.133***	0.201***
					[6.08]	[11.23]	[7.74]	[13.77]
FirstTimeOwner	0.028	0.08	0.191***	0.171***	-0.147***	-0.038	-0.007	-0.049***
	[0.67]	[1.39]	[7.77]	[6.47]	[-3.33]	[-0.64]	[-0.44]	[-3.59]
OwnerOccupied	-0.233***	-0.246***	-0.150***	0.011	-0.203***	-0.233***	-0.338***	-0.288***
	[-4.33]	[-5.51]	[-4.63]	[0.29]	[-4.48]	[-7.04]	[-13.50]	[-11.91]
OneBorrower	0.225***	0.329***	0.232***	0.258***	0.255***	0.341***	0.288***	0.295***
	[10.92]	[15.38]	[14.34]	[11.63]	[12.23]	[15.33]	[19.41]	[16.96]
Income (log)	-0.039**	0.120***	0.080***	0.180***	-0.112***	0.014	-0.065***	0.045***
	[-2.44]	[5.73]	[3.59]	[11.32]	[-7.41]	[0.78]	[-4.73]	[5.32]
IncomeMiss	0.05	0.093*	0.064	0.192***	-0.053	-0.024	-0.157***	0.182***
	[0.42]	[1.74]	[1.04]	[3.99]	[-0.45]	[-0.51]	[-2.94]	[8.17]
CashResv	-0.039***	-0.020**	-0.088***	-0.073***	-0.046***	-0.022***	-0.090***	-0.069***
	[-5.04]	[-2.50]	[-17.10]	[-16.84]	[-5.45]	[-3.13]	[-18.18]	[-16.62]
CreditScore	-0.008***	-0.007***	-0.007***	-0.006***	-0.008***	-0.008***	-0.008***	-0.007***
	[-44.99]	[-28.15]	[-42.52]	[-45.07]	[-47.73]	[-34.24]	[-51.86]	[-73.43]

	(1) Bank/Full-Doc	(2) Bank/Low-Doc	(3) Broker/Full-Doc	(4) Broker/Low-Doc	(5) Bank/Full-Doc	(6) Bank/Low-Doc	(7) Broker/Full-Doc	(8) Broker/Low-Doc
Female	-0.063**	-0.02	-0.007	-0.007	-0.039	-0.015	-0.003	0.002
	[-2.54]	[-1.09]	[-0.48]	[-0.84]	[-1.55]	[-0.79]	[-0.22]	[0.21]
Hispanic	0.260***	0.226***	0.435***	0.312***	0.272***	0.218***	0.387***	0.269***
	[4.81]	[3.88]	[8.36]	[10.75]	[5.53]	[3.77]	[7.85]	[10.66]
Black	0.188***	0.199***	0.219***	0.192***	0.129***	0.161***	0.167***	0.120***
	[3.90]	[3.32]	[6.25]	[7.37]	[2.75]	[2.90]	[5.10]	[4.51]
Asian	-0.026	-0.039	0.057*	0.084**	-0.067	-0.049	0.021	0.032
	[-0.28]	[-0.79]	[1.69]	[2.47]	[-0.66]	[-0.98]	[0.68]	[1.13]
Age (log year)	-0.204***	-0.101***	-0.107***	-0.121***	-0.087***	0.015	-0.017	0.005
	[-7.28]	[-4.36]	[-9.06]	[-10.25]	[-3.51]	[0.81]	[-1.43]	[0.53]
Tenure(log month)	-0.023***	-0.054***	-0.028***	-0.049***	-0.016*	-0.041***	-0.01	-0.034***
	[-2.79]	[-6.61]	[-3.87]	[-8.91]	[-1.74]	[-4.71]	[-1.56]	[-6.84]
TenureMiss	-0.120**	-0.327***	-0.296***	-0.364***	-0.06	-0.143***	-0.228***	-0.229***
	[-2.10]	[-8.00]	[-8.04]	[-15.98]	[-0.96]	[-3.18]	[-7.18]	[-9.99]
SelfEmploy	-0.014	0.031*	0.022	-0.014	-0.004	0.060***	0.060***	0.013
	[-0.28]	[1.76]	[1.00]	[-1.07]	[-0.08]	[3.22]	[2.87]	[1.22]
HPI6MBefore	0.407	0.459	0.21	-0.088	0.237	0.347	0.043	-0.008
	[0.88]	[1.32]	[1.03]	[-0.52]	[0.51]	[1.00]	[0.22]	[-0.05]
HPI6MAfter	-0.416	-0.393	-0.392**	-0.294**	-0.275	-0.312	-0.270*	-0.417***
	[-1.23]	[-1.41]	[-2.51]	[-2.10]	[-0.79]	[-1.15]	[-1.83]	[-3.90]
2005	0.000	0.187***	0.057*	0.194***	-0.011	0.088**	0.011	0.103***
	[-0.01]	[4.94]	[1.85]	[6.98]	[-0.26]	[2.37]	[0.39]	[4.22]
2006	0.024	0.276***	0.106**	0.393***	-0.007	0.105***	0.041	0.215***
	[0.59]	[6.10]	[2.19]	[9.24]	[-0.16]	[2.59]	[0.92]	[5.44]
2007	-0.151**	0.218***	-0.057	0.284***	-0.197***	0.054	-0.071	0.115***
	[-2.55]	[3.96]	[-1.04]	[6.06]	[-3.28]	[1.03]	[-1.37]	[2.68]
2008	-0.195***	0.038	-0.132**	0.013	-0.230***	-0.028	-0.104*	0.066
	[-2.64]	[0.40]	[-2.29]	[0.25]	[-3.02]	[-0.33]	[-1.94]	[1.22]
Constant	5.534***	4.149***	4.470***	3.313***	2.646***	0.453	1.059***	-1.028***
	[33.46]	[20.69]	[25.78]	[29.09]	[7.64]	[1.42]	[2.84]	[-3.31]
Observations	31,406	35,552	166,412	425,187	31,405	35,552	166,401	425,176
Pr(Delinquency)	0.132	0.180	0.236	0.316	0.132	0.180	0.236	0.316
$\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$	0.172	0.238	0.259	0.323	0.166	0.226	0.250	0.302
Pseudo R-squared	0.197	0.0945	0.156	0.0862	0.224	0.141	0.184	0.149

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Table 3. Choice of Loan Origination Channel and Documentation Level

The dependent variable is the choice of broker channel, that of low documentation, and that of the combination of two. The estimation method is probit for columns (1) to (6) and is ordered probit for column (7), where the choice of non-correspondent broker, correspondent, and bank are assigned as the highest, medium, and lowest outcome. The definitions of all variables are given in the Appendix. Reported are the coefficients ($\hat{\beta}$) and t-statistics (in the brackets) that adjust for clustering at the MSA level. At the bottom of the table, we report the sample frequency of delinquency, the pseudo R-squared, the number of observations, and the sample average of the probit density function ($\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$) that can be used to construct the average partial effect $\hat{\beta}\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Variable	Broker	Low-Doc	Broker &Low-Doc	Broker	Low-Doc	Broker &Low-Doc	Broker /Correspondent
LTV	0.372***	-0.764***	-0.500***	0.365***	-0.766***	-0.502***	0.595***
	[5.15]	[-6.79]	[-5.16]	[4.89]	[-6.61]	[-4.98]	[8.19]
AddLTV	3.730***	0.401***	1.081***	3.681***	0.412***	1.089***	1.488***
	[14.93]	[3.25]	[8.14]	[14.71]	[3.29]	[8.19]	[10.65]
Loan (log)	0.088***	0.221***	0.171***	0.091***	0.219***	0.170***	-0.008
	[3.02]	[11.51]	[7.70]	[3.11]	[11.53]	[7.62]	[-0.36]
SecondLien	-1.895***	-0.148**	-0.490***	-1.864***	-0.154**	-0.494***	-0.587***
	[-10.85]	[-2.03]	[-5.85]	[-10.23]	[-2.08]	[-5.67]	[-6.01]
Refinance	-0.146***	-0.052**	-0.091***	-0.135***	-0.042**	-0.079***	0.036
	[-5.26]	[-2.16]	[-5.25]	[-4.83]	[-1.97]	[-4.84]	[1.27]
FirstTimeOwner	0.331***	-0.045***	-0.003	0.330***	-0.046***	-0.003	0.138***
	[16.23]	[-2.61]	[-0.18]	[16.82]	[-2.68]	[-0.18]	[7.36]
OwnerOccupied	0.126***	-0.045***	0.084***	0.124***	-0.041***	0.088***	0.009
	[3.25]	[-2.94]	[3.18]	[3.22]	[-2.87]	[3.59]	[0.28]
OneBorrower	0.217***	0.508***	0.449***	0.222***	0.516***	0.453***	0.168***
	[18.39]	[37.57]	[39.44]	[16.41]	[39.11]	[39.95]	[15.59]
Income (log)	-0.038***	0.241***	0.218***	-0.038***	0.237***	0.216***	-0.002
	[-3.49]	[14.97]	[12.54]	[-3.48]	[14.55]	[11.67]	[-0.21]
IncomeMiss	0.128***	2.272***	1.606***	0.122***	2.286***	1.593***	-0.115***
	[3.28]	[56.38]	[34.11]	[3.31]	[56.40]	[34.73]	[-4.48]
CashResv	-0.015*	0.003	-0.003	-0.013	0.003	-0.003	-0.036***
	[-1.84]	[1.07]	[-0.73]	[-1.52]	[0.82]	[-0.69]	[-4.92]
CreditScore	-0.001***	0.002***	0.001***	-0.001***	0.002***	0.001***	-0.001***
	[-14.19]	[14.03]	[9.15]	[-13.42]	[14.07]	[9.42]	[-9.84]
Female	0.027***	0.150***	0.126***	0.025***	0.151***	0.124***	0.018***
	[3.66]	[11.14]	[10.64]	[3.48]	[11.43]	[10.75]	[3.02]
Hispanic	0.448***	0.433***	0.476***	0.448***	0.437***	0.479***	0.200***
	[13.52]	[6.67]	[8.65]	[13.08]	[6.58]	[8.60]	[4.62]
Black	0.439***	-0.030	0.059**	0.444***	-0.026	0.064**	0.210***
	[15.57]	[-1.15]	[2.14]	[15.58]	[-0.92]	[2.19]	[9.89]

Dep. Variable	(1) Broker	(2) Low-Doc	(3) Broker &Low-Doc	(4) Broker	(5) Low-Doc	(6) Broker &Low-Doc	(7) Broker /Correspondent
Asian	0.486***	0.367***	0.442***	0.492***	0.372***	0.448***	0.182***
	[18.38]	[18.62]	[25.88]	[16.98]	[16.35]	[21.16]	[5.18]
Age (log year)	-0.039***	0.000	-0.013*	-0.039***	0.005	-0.010	-0.072***
	[-3.73]	[0.06]	[-1.87]	[-3.82]	[0.56]	[-1.51]	[-8.50]
Tenure(log month)	-0.017***	-0.055***	-0.055***	-0.019***	-0.055***	-0.055***	-0.012***
	[-4.57]	[-9.56]	[-9.51]	[-4.68]	[-9.56]	[-9.39]	[-2.80]
TenureMiss	0.540***	-0.348***	-0.174***	0.527***	-0.328***	-0.157***	0.718***
	[13.62]	[-9.87]	[-4.85]	[13.13]	[-9.31]	[-4.39]	[18.23]
SelfEmploy	0.208***	1.036***	0.775***	0.210***	1.046***	0.779***	0.095***
	[8.80]	[48.49]	[27.57]	[9.12]	[51.18]	[29.09]	[8.25]
PctBlack				-0.077***	0.053***	0.027	-0.080***
				[-4.73]	[2.94]	[1.42]	[-4.88]
PctHisp				-0.056*	0.177***	0.144***	-0.081**
				[-1.67]	[8.98]	[6.41]	[-2.30]
MedAge				-0.002***	-0.001***	-0.002***	-0.002***
				[-3.00]	[-2.88]	[-3.22]	[-3.92]
AvgIncome				-0.000	0.000	0.000	-0.000***
				[-0.49]	[0.40]	[0.63]	[-2.91]
UnempRate				-0.000	-0.006**	-0.007**	0.004
				[-0.17]	[-2.34]	[-2.45]	[1.45]
HPI6MBefore	0.003	0.932***	0.658***	-0.017	0.974***	0.680***	-0.202**
	[0.04]	[9.47]	[6.71]	[-0.18]	[10.87]	[7.16]	[-2.15]
2005	0.339***	0.261***	0.305***	0.349***	0.254***	0.303***	0.166***
	[12.99]	[14.03]	[16.67]	[14.23]	[14.11]	[16.36]	[7.22]
2006	0.444***	0.594***	0.581***	0.444***	0.577***	0.563***	0.159***
	[12.56]	[32.64]	[28.14]	[12.54]	[34.06]	[25.61]	[5.86]
2007	0.420***	0.344***	0.375***	0.420***	0.319***	0.352***	0.290***
	[17.93]	[18.88]	[18.99]	[16.99]	[19.14]	[16.53]	[11.18]
2008	0.196***	-0.213***	-0.151***	0.217***	-0.228***	-0.165***	0.407***
	[5.19]	[-7.40]	[-5.04]	[5.08]	[-7.81]	[-5.18]	[9.27]
Constant	0.187	-4.303***	-3.640***	0.213	-4.284***	-3.598***	-1.330***
	[0.57]	[-19.10]	[-15.62]	[0.62]	[-18.88]	[-15.17]	[-5.42]
Constant 2							-0.601**
							[-2.55]
Observations	658,534	658,534	658,534	491,772	491,772	491,772	491,772
E(Dep Var)	0.898	0.700	0.646	89.9%	69.9%	64.6%	
$\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$	0.153	0.255	0.295	0.155	0.254	0.295	
Pseudo R-squared	0.149	0.265	0.201	0.145	0.267	0.201	0.0604

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Table 4. Non-Linear Blinder-Oaxaca Decomposition of Differences in Delinquency Rates

This table reports the non-linear Blinder-Oaxaca (1973) decomposition to the probit model. The total difference in delinquency rates between two subsamples is decomposed into an "endowment effect" and a "coefficient effect" using equations (2). "Relatively exogenous covariates" correspond to the short regressions of Table 2 columns (1)-(4), while "All observable covariates" correspond to the long regressions of Table 2 columns (5)-(8).

Panel A: Comparison of Full-Doc and Low-Doc subsamples

	Bank			Broker		
	Difference	t-stat	Percentage	Difference	t-stat	Percentage
Relatively exogenous covaraites only						
Endowment Effect	0.18%	0.29	3.71%	0.32%	0.73	4.58%
Coefficient Effect	4.62%	8.07	96.29%	6.67%	10.83	95.42%
All observable covariates						
Endowment Effect	0.43%	0.62	8.92%	-0.40%	-0.67	-4.99%
Coefficient Effect	4.39%	7.66	91.08%	8.42%	14.34	104.99%
Total	4.81%	5.13	100%	8.00%	8.06	100%

Panel B: Comparison of Bank and Broker subsamples

	Full-Doc			Low-Doc		
	Difference	t-stat	Percentage	Difference	t-stat	Percentage
Relatively exogenous covaraites only						
Endowment Effect	5.38%	6.23	52.13%	7.18%	11.31	53.10%
Coefficient Effect	4.94%	16.20	47.87%	6.34%	13.77	46.90%
All observable covariates						
Endowment Effect	8.10%	8.10	78.19%	7.44%	13.65	55.15%
Coefficient Effect	2.26%	8.11	21.81%	6.05%	11.12	44.85%
Total	10.32%	10.51	100%	13.51%	13.98	100%

Table 5. Delinquency Analysis: Correspondent and Non-Correspondent Brokers

This table repeats the analysis in Table 2 columns (5) to (8) using loans originated by brokers only, where the Broker channel is decomposed into Correspondent and Non-Correspondent channels.

	(1)	(2)	(3)	(4)
	Correspondent/Full-Doc	Correspondent /Low-Doc	Non- Correspondent /Full-Doc	Non- Correspondent /Low-Doc
LTV	2.000***	3.118***	1.975***	2.904***
	[14.84]	[17.08]	[18.03]	[20.39]
AddLTV	1.746***	3.385***	1.655***	2.888***
	[7.87]	[18.88]	[17.28]	[23.49]
Loan (log)	0.111***	0.231***	0.213***	0.213***
	[2.75]	[5.93]	[8.88]	[8.82]
SecondLien	0.351**	0.232***	0.599***	0.395***
	[2.51]	[2.61]	[10.15]	[5.51]
Refinance	0.005	0.103***	-0.066***	0.074***
	[0.15]	[2.91]	[-2.78]	[4.73]
PrepayPenalty	0.013	0.035	-0.041*	0.051***
	[0.28]	[1.59]	[-1.91]	[3.61]
ARM	0.139***	0.175***	0.220***	0.183***
	[4.03]	[8.74]	[12.18]	[10.10]
OptionARM	0.247***	0.330***	0.236***	0.283***
	[4.09]	[11.93]	[6.03]	[10.46]
IO	0.163***	0.185***	0.137***	0.207***
	[4.18]	[11.11]	[8.78]	[13.07]
FirstTimeOwner	-0.117***	-0.093***	0.006	-0.044***
	[-3.68]	[-3.96]	[0.36]	[-3.47]
OwnerOccupied	-0.358***	-0.262***	-0.335***	-0.290***
	[-7.97]	[-7.34]	[-12.83]	[-12.17]
OneBorrower	0.206***	0.280***	0.296***	0.296***
	[8.15]	[14.29]	[20.54]	[17.31]
Income (log)	-0.058**	0.022	-0.066***	0.047***
	[-2.13]	[1.12]	[-4.39]	[5.89]
IncomeMiss	-0.042	0.077*	-0.162***	0.201***
	[-0.22]	[1.67]	[-2.87]	[8.98]
CashResv	-0.097***	-0.094***	-0.087***	-0.063***
	[-9.42]	[-15.29]	[-17.34]	[-14.00]
CreditScore	-0.008***	-0.007***	-0.008***	-0.007***
	[-29.97]	[-57.63]	[-48.41]	[-68.53]
Female	0.014	0.013	-0.007	-0.002
	[0.56]	[1.02]	[-0.49]	[-0.18]
Hispanic	0.321***	0.358***	0.374***	0.247***
	[9.97]	[11.20]	[7.67]	[10.18]

	(1)	(2)	(3)	(4)
	Correspondent/Full-Doc	Correspondent /Low-Doc	Non- Correspondent /Full-Doc	Non- Correspondent /Low-Doc
Black	0.151***	0.105***	0.166***	0.126***
	[3.95]	[3.59]	[5.12]	[4.80]
Asian	0.066	0.134***	0.011	0.008
	[1.40]	[4.41]	[0.33]	[0.27]
Age (log year)	-0.016	0.023*	-0.016	0.005
	[-0.54]	[1.80]	[-1.35]	[0.48]
Tenure(log month)	0.008	-0.013*	-0.015**	-0.039***
	[0.84]	[-1.83]	[-1.99]	[-8.59]
TenureMiss	0.004	-0.077**	-0.282***	-0.268***
	[0.07]	[-2.42]	[-7.46]	[-10.86]
SelfEmploy	0.049	0.040**	0.064***	0.011
	[0.92]	[2.06]	[2.85]	[1.08]
HPI6MBefore	0.041	0.022	0.026	-0.012
	[0.08]	[0.13]	[0.13]	[-0.07]
HPI6MAfter	0.010	-0.351**	-0.267*	-0.428***
	[0.03]	[-2.04]	[-1.70]	[-3.54]
2005	0.034	0.119***	0.014	0.102***
	[0.66]	[3.40]	[0.49]	[4.26]
2006	0.022	0.194***	0.058	0.222***
	[0.45]	[4.18]	[1.18]	[5.52]
2007	-0.057	0.122**	-0.072	0.111***
	[-0.97]	[2.34]	[-1.33]	[2.63]
2008	-0.093	-0.045	-0.123**	0.048
	[-0.59]	[-0.30]	[-2.23]	[0.93]
Constant	1.994***	-1.436***	0.873**	-0.995***
	[3.74]	[-3.55]	[2.24]	[-3.14]
Observations	25,666	88,778	140,735	336,398
Pr(Delinquency)	0.189	0.18	0.246	0.331
$\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$	0.223	0.287	0.255	0.305
Pseudo R-squared	0.173	0.165	0.186	0.146

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Table 6. Determinants of Interest Rates

This table examines the determinants of interest rates (expressed in percentage points), with loan types as the main regressors of interest. The definitions of all covariates (*X*) are given in the Appendix. The samples for columns (1) and (2) are fixed rate and adjustable rate loans respectively. Column (3) examines adjustable rate loans issued in 2005-2006 while column (4) examines fixed rate loans issued after July 2007. The t-statistics reported in the brackets adjust for heteroskedasticity and clustering at the MSA level.

	(1)	(2)	(3)	(4)
Sample	Fixed rate only	Adjustable rate only	Adjustable rate 2004-2005	Fixed rate 200707-2008
Dep. Varaible	Initial Rate	Initial Rate	Current Rate	Initial Rate
LTV	1.2617***	1.4557***	1.2881***	1.3440***
	[29.62]	[16.76]	[30.37]	[32.49]
AddLTV	0.6663***	-1.1366***	0.1901**	0.1525*
	[8.09]	[-10.57]	[2.20]	[1.67]
Loan (log)	-0.2402***	-0.3973***	-0.3277***	-0.3402***
	[-18.64]	[-32.69]	[-21.48]	[-24.20]
SecondLien	3.1023***	4.3074***	3.0473***	2.3134***
	[29.80]	[25.24]	[39.52]	[32.70]
Refinance	-0.2719***	-0.3027***	-0.2106***	-0.1344***
	[-18.40]	[-17.65]	[-17.18]	[-8.35]
PrepayPenalty	-0.2232***	-0.0469***	-0.0855***	0.2198***
	[-17.15]	[-2.67]	[-5.91]	[4.25]
OptionARM		-5.0172***	1.6219***	
		[-124.98]	[66.22]	
IO		-0.5044***	-0.2719***	
		[-29.29]	[-16.47]	
FirstTimeOwner	0.0386***	0.0088	0.0756***	-0.0839***
	[3.65]	[0.87]	[6.32]	[-5.16]
OwnerOccupied	-0.5421***	-0.5413***	-0.5216***	-0.3941***
	[-23.71]	[-15.54]	[-17.75]	[-17.24]
OneBorrower	0.0695***	0.0084	0.0495***	0.0401***
	[10.50]	[1.01]	[8.01]	[4.79]
Income (log)	0.0049	0.0587***	0.0041	0.1089***
	[0.84]	[8.39]	[0.66]	[9.58]
IncomeMiss	0.1586***	0.1629***	-0.0480**	0.3973***
	[12.39]	[5.80]	[-2.08]	[15.64]
CashResv	-0.0403***	-0.0455***	-0.0747***	-0.0318***
	[-6.16]	[-8.29]	[-11.37]	[-6.85]
CreditScore	-0.0080***	-0.0068***	-0.0078***	-0.0056***
	[-35.03]	[-19.48]	[-23.26]	[-29.11]
Female	0.0106**	0.0150***	0.0186***	0.0012
	[2.53]	[3.75]	[3.78]	[0.15]
Hispanic	-0.0227	-0.0416**	0.0398*	0.0232
	[-1.30]	[-2.26]	[1.73]	[1.38]

	(1)	(2)	(3)	(4)
Sample	Fixed rate only	Adjustable rate only	Adjustable rate 2004-2005	Fixed rate 200707-2008
Dep. Varaible	Initial Rate	Initial Rate	Current Rate	Initial Rate
Black	0.1247***	0.0885***	0.1743***	0.1204***
	[9.13]	[4.90]	[7.13]	[9.55]
Asian	-0.0989***	-0.0639***	-0.0235*	-0.0011
	[-6.75]	[-5.06]	[-1.67]	[-0.07]
Age (log year)	0.0496***	0.0597***	0.1008***	0.0597***
	[13.57]	[9.24]	[12.70]	[8.02]
Tenure(log month)	-0.0061***	-0.001	-0.0005	-0.0123***
	[-2.72]	[-0.37]	[-0.16]	[-2.75]
TenureMiss	-0.1847***	-0.3368***	-0.3154***	-0.3197***
	[-7.82]	[-17.68]	[-10.01]	[-9.07]
SelfEmploy	0.0147*	0.0264***	0.0005	0.0223*
	[1.92]	[2.88]	[0.05]	[1.81]
2005	0.0984***	0.3837***	0.1315***	
	[6.03]	[14.17]	[8.71]	
2006	0.7938***	1.1456***		
	[47.04]	[31.52]		
2007	0.6712***	1.3568***		7.3844***
	[61.11]	[35.38]		[25.21]
2008	0.6210***	1.1722***		7.1963***
	[37.45]	[35.20]		[24.67]
Broker	0.0266	-0.3672***	-0.0436	0.1237***
	[1.27]	[-5.05]	[-0.88]	[8.92]
LowDoc	0.2382***	0.0805***	0.1556***	0.2971***
	[23.49]	[5.50]	[16.67]	[13.71]
Constant	14.2160***	15.5758***	15.2995***	6.6351***
	[68.85]	[45.79]	[44.20]	[16.37]
Observations	310,027	348,517	192,231	52,694
R-squared	0.766	0.851	0.571	0.447

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.