The Review *of* Economics *and* Statistics

VOL. XCVI | MARCH 2014 | NUMBER 1

LIAR'S LOAN? EFFECTS OF ORIGINATION CHANNEL AND INFORMATION FALSIFICATION ON MORTGAGE DELINQUENCY

Wei Jiang, Ashlyn Aiko Nelson, and Edward Vytlacil*

Abstract—This paper presents an analysis of mortgage delinquency between 2004 and 2008 using a loan-level data set from a major national mortgage bank. Our analysis highlights two problems underlying the mortgage crisis: a reliance on mortgage brokers who tend to originate lower-quality loans and a prevalence of low-documentation loans—known in the industry as "liar's loans"—that result in borrower information falsification. While over three-quarters of the difference in delinquency rates between bank and broker channels can be attributed to observable loan and borrower characteristics, the delinquency difference between full- and low-documentation mortgages is due to unobservable heterogeneity, about half of it potentially due to income falsification.

I. Introduction

A decade-long boom in the housing market and related financial sectors was followed in 2007 by falling house prices and a rapid increase in mortgage defaults and foreclosures. The crisis that began in the mortgage market quickly spread to other financial markets and throughout the economy. In this study, we use the experience of a major national mortgage bank to uncover the determinants and the evolution of the mortgage crisis at a microlevel.

Our sample bank provides an ideal context for the study: its experience presents a representative yet amplified version of the boom-and-bust cycle that occurred in the national mortgage sector over the past 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; it issued a majority of its loans for low- and no-documentation mortgage products (nicknamed "liar's 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 fifty states. Therefore, lessons from this 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. In the data set are 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), but also about borrower demographic characteristics (for example, race, age, gender) and economic conditions (including income, cash reserves, and employment status). Finally, we are able to use the property address information to match loans to community attributes, such as demographics and employment opportunities, at a narrow geographic level.

We divide our sample into six 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. Third-party originators may be correspondent brokers (brokers with long-term and often exclusive business relations with the bank, referred to henceforth as "correspondents") or noncorrespondent brokers (brokers who work with multiple originators on a commission basis, referred to henceforth as "brokers"). 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 six subsamples as Bank/Full-Doc, Bank/Low-Doc, Correspondent/Full-Doc, Correspondent/Low-Doc, Broker/Full-Doc, and Broker/Low-Doc.

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 third-party originators (especially brokers), driven by the credit expansionary policies pursued by many large lending institutions. We find

Received for publication March 18, 2010. Revision accepted for publication January 3, 2013.

* Jiang: Columbia Business School; Nelson: Indiana University; Vytlacil: New York University.

We 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 and conference participants at the 2009 Annual Meeting of the Association for Public Policy Analysis and Management, Columbia, Federal Reserve Board of Governors, George Mason, Georgia State, Kansas City Federal Reserve Bank, the NBER 2009 Summer Institute, the FDIC 2009 Mortgage Symposium, the 2009 Philadelphia Federal Reserve Conference on Recent Developments in Consumer Credit and Payments, the 2010 Sixth Annual Credit Risk Conference by Moody's & NYU Stern School of Business, and the 2010 AEA have contributed to this draft. We also thank Erica Blom, Guojun Chen, Vyacheslav Fos, Andres Liberman, Sunyoung Park, and Mike Tannenbaum for excellent research assistance.

that loans issued by brokers have delinquency probabilities that are 3.7 percentage points higher than those issued by correspondents; in turn, delinquency probabilities for correspondent-issued loans are 10.6 percentage points higher than those issued by the bank. A decomposition exercise conditional on loan documentation level attributes up to three-quarters of the bank-broker delinquency gap, and almost all the bank-correspondent delinquency gap, to differences in observable borrower characteristics. Hence, the higher delinquency rates among third-party originated loans are largely explained by loan issuance to borrowers of observably lower quality (as measured by, for example, credit score, loan-to-value ratio, or income) as compared to borrowers with bank-issued loans. High delinquency rates among broker loans also reflect incentive structures that compensate brokers primarily based on origination volume rather than loan performance.

The Low-Doc subsample also exhibits worse performance than the Full-Doc subsample, and the difference in delinquency is 5 to 8 percentage points depending on origination channel. However, the same decomposition method reveals that characteristics observed by the bank account for little or none of the delinquency difference between Full-Doc and Low-Doc loans. Thus, nearly 100% of the delinquency difference must be attributed to unobserved heterogeneity—that is, differences in characteristics across borrowers that are not observed by the bank at the time of loan origination. In contrast to the Bank-Correspondent/Broker comparison, Low-Doc loans do not necessarily compromise lending standards along observable metrics as compared to Full-Doc loans; rather, low-documentation mortgages suffer from adverse selection along unobservables. We argue that an important source of unobserved heterogeneity in loan quality is due to less careful verification of low-documentation borrowers' reported information, notably income, and less diligent screening of financial conditions that are difficult to verify, 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. These agency problems are exacerbated in the broker channel, resulting in the highest delinquency rates among Broker/Low-Doc mortgages.

We also provide detailed evidence of borrower income falsification among low-documentation loans and assess its impact on loan performance. By decomposing reported income among Low-Doc loans into predicted income (a proxy for true income) and residual income (a proxy for income falsification), we resolve the perverse relation (a positive correlation) between reported income and delinquency probabilities within the Low-Doc sample. The positive relation is driven entirely by residual income, indicating that higher income exaggeration levels (as compared to true income) are associated with lower propensities to repay. This pattern is especially strong in the Broker/Low-Doc subsample, where, all else constant, a 1 standard deviation increase in residual income is associated with a direct increase in delinquency of 1.4 percentage points (significant at less than the 1% level). Once we account for the indirect effect by effects on loan contract terms (such as higher reported income allows borrowers to qualify for larger loans), the total effect increases to 4.0 percentage points—about half the total difference in delinquency rates between Broker/Full-Doc and Broker/Low-Doc subsamples.

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 8 to 30 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, risk of negative publicity stemming from banks 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 and most closely relates to a few recent empirical papers exploring the causes of the mortgage crisis using large sample microlevel archival data.[1] Mian and Sufi (2008) identify the effects of the increase in the supply of mortgage credit on the housing bubble between 2001 and 2005. Demyanyk and Van Hemert (2008) and Keys et al. (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 they include only securitized loans. Some earlier papers (for example, Munnell et al., 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 borrower 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 unobserved heterogeneity. Such decomposition provides us with an accurate calibration of the information

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

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 retained 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 omitted variable bias.

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

II. 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.[2] 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 presence; product category, such as whether the interest rate is fixed or adjustable; loan purpose, such as home purchase or refinance; origination channel (that is, bank, correspondent, or broker originated); and documentation requirements. Among third-party originators, brokers act as matchmakers and submit loan applications to a variety of banks for competitive pricing; in contrast, correspondents 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.

Property data used in our analysis include the property address, whether the property will be owner occupied as a primary residence or used as an investment property or second home, and home appraisal value. Borrower data include 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 or foreclosure status at loan origination, credit score,[3] 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 foreclosure or short sale (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).

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.[4] Using ArcGIS geocoding software and Decennial Census geographic boundary files, we match each property address to its census tract, postal 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 postal-code-level average household income information from the Internal Revenue Service 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,

[2] Interested researchers may contact us for information on accessing the data set. External researchers who would like to access the data must obtain IRB clearance regarding human subjects research at both their home institution and Columbia University.

[3] The credit score the bank used is the median score obtained from the three major credit-reporting bureaus—Equifax, Experian, and Trans Union—and is numerically comparable and analytically equivalent to the Fair Isaac Corporation's FICO score.

[4] Approximately one-quarter of the property addresses were unmatched, mostly due to variations in address recording (such as nonstandard abbreviations) and, in some cases, recording errors. Regressions that require community attributes exclude observations where property addresses were not matched.

FIGURE 1.—NUMBER OF LOANS AND COMPOSITION BY SEMIYEAR, 2004–2008

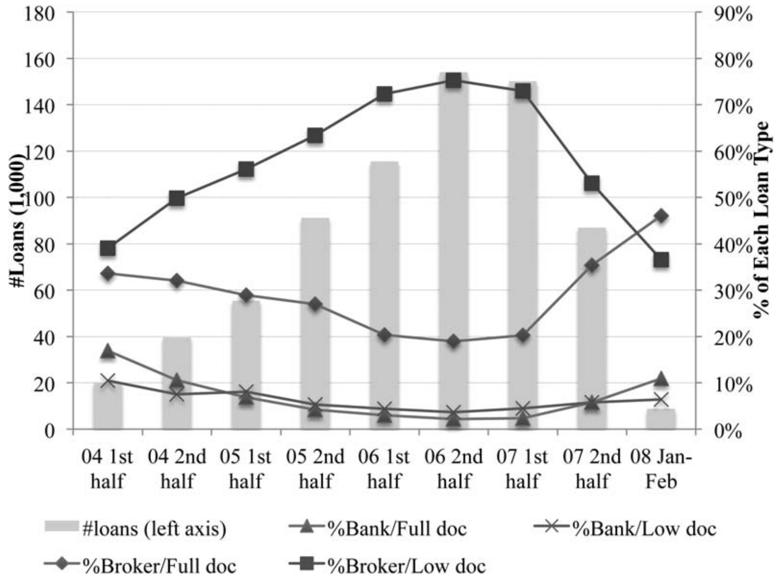
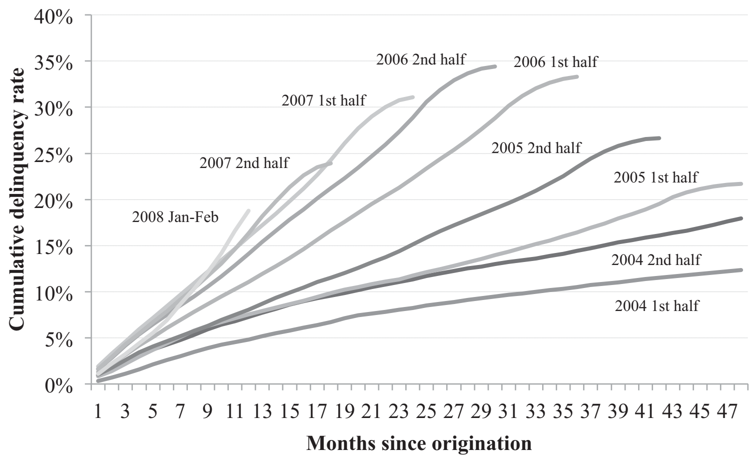


FIGURE 2.—DELINQUENCY RATES SINCE LOAN ORIGINATION BY SEMIYEAR, UPDATED TO JANUARY 2009



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.

Figure 1 also shows that the rapid expansion in loan production was driven almost exclusively by increased loan originations via third parties and in particular by the expansion of low-documentation loans using the broker channel. Third-party-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 2004, they were 59% of originations by late 2006.

Cumulative delinquency rates increased progressively and substantially over the time period in our sample (shown in figure 2). At eighteen months after origination, only 6.7% of loans originated in the first half of 2004 were ever more than sixty days delinquent, as compared to 23.9% of loans originated in the second half of 2007. Demyanyk and Van Hemert (2008) document a similarly deteriorating trend for subprime loans from 2001 to 2006 using the LoanPerformance database.

We define all variables used in this paper in the appendix and report their mean and standard deviation values by origination year in table 1. 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 in 2007. For example, 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. The housing boom welcomed many first-time home buyers to the mortgage market. In 2004, only 9.7% of borrowers in the sample were first-time home buyers, a figure that climbed to 17.5% by 2006 before falling to 15.5% by 2008. During the sample period, black and

TABLE 1.—SUMMARY STATISTICS OF MAJOR BORROWER CHARACTERISTICS, 2004–EARLY 2008

	2004	2005	2006	2007	2008 (January-February)
Age (years)	44.46	43.54	42.92	44.46	45.42
	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 and Hispanic who 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%

The mean is reported in the first line of each variable and the standard deviation in the second line.

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.[5]

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 fifty 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 originated by third parties and 72% of loans were originated by noncorrespondent brokers. These figures are considerably higher than those for mortgage banks with more traditional models.[6] 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 national securitization rate of 75% to 91% reported in *Inside Mortgage Finance* during the same period for subprime and nonconforming loans.[7]

We further compare our 2004–2008 sample average statistics to those covered by McDash Analytics, the most comprehensive commercial database on mortgage performance.[8] 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).[9] Finally, low-documentation loans represent 70% of the loans in our sample due to the lender's specialization in low-documentation products, but just 20% of all loans in the McDash database.

Finally, subprime loans, which constitute 14% to 15% of our sample, are not overrepresented.[10] Nationally, 18% to 21% of loans originated from 2004 to 2006 were subprime.[11] Our sample affords analyses on the full spectrum of the market, thereby complementing prior research focusing on the subprime sector (Keys et al., 2008; Demyanyk & Van Hemert, 2007) and highlighting the widespread crisis beyond the subprime sector.

[5] According to HMDA data on home purchase loans (http://www.ffiec.gov/hmdaadwebreport/NatAggWelcome.aspx), 6.6% (10.8%) of borrowers were black (Hispanic) in 2004; the percentages increased to 8.7% (14.4%) in 2006.

[6] For example, a 2007 *Wall Street Journal* article estimated that brokers originate around 60% of all home loans. See James Hagerty, "Mortgage Brokers: Friends or Foes?" *Wall Street Journal*, May 30, 2007.

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

[8] The comparison data set is used in recent studies including Piskorski, Seru, and Vig (2010). We thank Amit Seru for providing the summary statistics for this data set.

[9] Part of the difference can be attributed to the overrepresentation of prime loans in the McDash database. McDash covers about 60% of the entire mortgage market but only 30% to 40% of subprime originations.

[10] Despite its wide use, 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.

[11] The source of information is Joint Center for Housing Studies (2008). This report mostly relies on the credit score cutoff at 640 for subprime classification (available at http://www.jchs.harvard.edu/publications/markets/son2008/son2008.pdf).

In summary, the bank in our analysis pursued an aggressive expansion strategy relying heavily on third-party 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%.[12]

Overall, the sample bank experienced a representative and yet amplified version of the boom-bust cycle that occurred 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.

III. Prediction of Loan Delinquency

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 sixty days behind in payment or in a more serious condition of default (such as short sale or foreclosure). Our model of loan delinquency is a critical input into our analysis in section IV, which decomposes the differences in delinquency rates across loan types into differences due to observed characteristics of borrowers and loans versus differences due to unobserved characteristics. In addition, our finding of a perverse relationship between reported income and delinquency among low-documentation loans motivates our analysis in section V, in which we investigate more thoroughly evidence of income falsification among borrowers of low-documentation loans.

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 to account for common shocks to real estate markets in the same MSA.[13] 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.

Our main analysis applies the standard probit method:

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

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 sixty days behind payment) by the end of our sample period. In our sample, 25.6% of the loans are delinquent: 11.0% are sixty or more days behind in payments, 4.7% are in a state of short sale, and 9.9% are in a state of foreclosure.[14]

The set of covariates X includes the following categories. The first category comprises loan contract terms and product categories:[15] loan-to-value ratio (*LTV*); additional leverage on the same property (*AddLTV*); loan size (*LoanAmt*); second-lien status (*SecondLien*); refinance status (*Refinance*); and variables indicating whether the mortgage interest terms have 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 Option ARM 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 1.

Borrower characteristics comprise the second category of covariates. They include whether the property is owner occupied (*OwnerOccupied*); whether there is only one borrower on the loan application (*OneBorrower*); and borrower income (*Income*), cash reserves (*CashResv*), credit score (*CreditScore*), gender (*Female*), ethnicity (*Hispanic, Black*, and *Asian*), age (*Age*), job tenure (*Tenure*), self-employment status (*SelfEmploy*), and whether the borrower is a first-time homeowner (*FirstTimeOwner*). The final category includes housing price changes at the state level during the six months before and after loan origination (*HPI6MBefore* and *HPI6MAfter*).[16] In addition, all regressions include a

[12] Source of information: Loan Processing Services, http://www.lpsvcs.com/NewsRoom/IndustryData/Pages/default.aspx.

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

[14] Prepaid mortgages remain in the sample. If the loan is prepaid after a short sale, the loan is considered to be delinquent. If a loan is paid off in full, it is considered nondelinquent.

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

[16] Like other covariates, the housing price changes are measured around loan origination. Their impact on delinquency is in addition to that from the housing price evolution later on. Strictly speaking, housing price changes post-origination (*HPI6MAfter*) are not known at loan origination. For this reason, this variable is not included in some of the later analyses that rely strictly on information obtained at origination. We conduct a further sensitivity analysis by excluding loans originated in the hot markets of California and Florida. Results are qualitatively indistinguishable from those of the full sample. Finally, we obtain similar results using three-month or twelve-month windows.

set of state dummies (*State*$_j$) to control for unobserved heterogeneity in regional property markets.[17]

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 for 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 reset interest rate terms at different stages during the life of the loan. In section VI, we analyze interest rates in detail by examining specific subsamples in which the rate information is comparable across observations.

We conduct the analysis separately for each of the six 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 $\phi(g)$ is the standard normal probability density function, such that the empirical analog to the average partial effects (APE, or $E(\partial Pr(Delinquency_i = 1|X_i)/\partial X_i)$) can be calculated as $\hat{\beta}\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$.

While the estimated coefficients vary considerably across the subsamples, most coefficients are intuitive. Note that the relationships evident in the correspondent subsamples always fall between those in the bank and broker subsamples; furthermore, relationships in the correspondent subsample tend to resemble those in the bank subsample more closely due to the alignment of incentives between the bank and its correspondent brokers. Perhaps most interesting, we find that, as expected, higher income in the Full-Doc subsamples is associated with lower probability of delinquency, while we find a perverse relationship in the Low-Doc subsample, with higher reported income associated with a higher probability of delinquency. These results provide suggestive evidence of systematic income falsification in the Low-Doc subsample, a hypothesis we investigate further in section V.

We conduct two sensitivity analyses capturing the timing information from origination to delinquency. First, we employ a hazard model to analyze the per-period "failure (delinquency) rate." Second, we separate early (within twelve months of loan origination) and later delinquencies. The analyses mostly confirm the patterns revealed in table 2 but contribute two additional insights (results are available on request). First, higher recent past housing price run-ups (*HPI6mBefore*) are associated with higher eventual delinquency rates but lower early delinquency rates. Presumably areas with high recent past housing price appreciation had more appraisal inflation and more borrowers who hurried to buy without careful calculation, yet borrowers in these markets were less likely to enter early delinquency due to persistence in housing price appreciation. Second, option ARM loans are associated with high delinquency rates only after twelve months post-origination. Due to artificially lower teaser rates, these loans are no more delinquency prone in the initial period following loan origination.

IV. Loan Types and Attribution of Differences in Delinquency

Table 2 reveals that third-party- (and especially broker-) originated loans exhibit much higher delinquency rates than bank-originated loans: 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 to 10 percentage points higher than for full-documentation mortgages. This section discusses 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 loan type. We initially model the binary choice of originating a mortgage through the bank or through a third party. Later in this section, we use an ordered choice model with three options for loan origination: bank, correspondent, or broker. Results are presented in table 3. The first three columns use only loan and borrower characteristics as regressors; the next four add neighborhood characteristics to the list of covariates. The sample size for the regressions including neighborhood characteristics is about 25% smaller due to the additional data requirement.

The following variables predict a higher likelihood that a borrower will obtain a loan from a correspondent or 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 nonwhite borrowers favor third-party loan origination relative 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 by phone or Internet. The sorting of less

[17] Thanks to our large sample and meaningfully large number of loans in almost all states, the state dummies do not cause incidental parameter problems.

Table 2.—Delinquency Prediction: Probit Analysis

	Bank/ Full-Doc (1)	Bank/ Low-Doc (2)	Correspondent/ Full-Doc (3)	Correspondent/ Low-Doc (4)	Broker/ Full-Doc (5)	Broker/ Low-Doc (6)
LTV	1.597***	2.387***	2.212***	3.391***	2.133***	3.100***
	[13.68]	[17.61]	[12.67]	[18.47]	[16.78]	[18.25]
AddLTV	1.314***	1.530***	2.020***	3.574***	1.747***	3.026***
	[6.37]	[7.43]	[9.13]	[24.47]	[17.40]	[25.87]
LoanAmt (log)	0.106***	0.175***	0.053	0.151***	0.170***	0.167***
	[3.95]	[7.27]	[1.23]	[4.69]	[6.94]	[6.97]
SecondLien	0.354***	0.836***	0.208	0.099	0.552***	0.311***
	[2.59]	[7.46]	[1.50]	[1.23]	[10.28]	[4.68]
Refinance	−0.036	−0.001	0.011	0.111***	−0.055***	0.078***
	[−0.94]	[−0.03]	[0.33]	[3.59]	[−2.81]	[6.86]
PrepayPenalty	0.055	−0.039	0.001	0.014	−0.048**	0.027**
	[1.02]	[−0.95]	[0.03]	[0.64]	[−2.18]	[2.34]
ARM	0.223***	0.137***	0.138***	0.151***	0.218***	0.159***
	[6.94]	[5.64]	[3.92]	[6.73]	[13.61]	[11.66]
OptionARM	0.180***	0.303***	0.238***	0.243***	0.199***	0.222***
	[3.36]	[9.92]	[3.99]	[8.22]	[5.75]	[7.58]
IO	0.180***	0.184***	0.150***	0.114***	0.117***	0.156***
	[6.35]	[9.47]	[4.29]	[6.70]	[8.32]	[13.63]
FirstTimeOwner	−0.146***	−0.031	−0.140***	−0.098***	−0.007	−0.051***
	[−3.25]	[−0.55]	[−4.48]	[−4.69]	[−0.45]	[−4.22]
OwnerOccupied	−0.216***	−0.225***	−0.348***	−0.247***	−0.330***	−0.279***
	[−4.99]	[−7.01]	[−7.59]	[−7.33]	[−12.71]	[−12.34]
OneBorrower	0.250***	0.340***	0.205***	0.286***	0.280***	0.300***
	[11.90]	[17.58]	[7.99]	[17.89]	[25.63]	[23.67]
Income (log)	−0.109***	0.006	−0.056*	0.031*	−0.066***	0.047***
	[−7.23]	[0.33]	[−1.89]	[1.67]	[−4.73]	[7.16]
IncomeMiss	−0.028	−0.026	−0.095	0.099**	−0.168***	0.199***
	[−0.24]	[−0.52]	[−0.50]	[2.29]	[−3.26]	[11.44]
CashResv	−0.050***	−0.030***	−0.101***	−0.093***	−0.089***	−0.068***
	[−6.24]	[−3.85]	[−9.72]	[−15.72]	[−18.32]	[−14.95]
CreditScore	−0.008***	−0.008***	−0.009***	−0.007***	−0.008***	−0.007***
	[−47.90]	[−35.48]	[−30.11]	[−55.74]	[−44.50]	[−67.40]
Female	−0.037	−0.017	0.008	0.009	−0.020	−0.006
	[−1.47]	[−0.83]	[0.34]	[0.70]	[−1.64]	[−0.68]
Hispanic	0.235***	0.163***	0.264***	0.307***	0.283***	0.181***
	[5.24]	[2.99]	[6.89]	[9.69]	[9.24]	[8.58]
Black	0.122***	0.143***	0.181***	0.128***	0.169***	0.134***
	[2.86]	[3.18]	[5.48]	[4.66]	[6.32]	[5.38]
Asian	−0.052	−0.057	0.000	0.106***	0.005	−0.009
	[−0.50]	[−1.12]	[0.00]	[4.23]	[0.18]	[−0.41]
Age (log year)	−0.088***	0.002	−0.034	0.022*	−0.021*	0.003
	[−3.67]	[0.11]	[−1.21]	[1.94]	[−1.96]	[0.30]
Tenure(log month)	−0.016*	−0.037***	0.005	−0.010*	−0.012*	−0.034***
	[−1.75]	[−3.91]	[0.56]	[−1.84]	[−1.84]	[−11.22]
TenureMiss	−0.068	−0.137***	0.007	−0.059**	−0.278***	−0.253***
	[−1.09]	[−2.86]	[0.11]	[−1.97]	[−9.36]	[−11.35]
SelfEmploy	0.002	0.069***	0.077	0.052***	0.082***	0.022**
	[0.03]	[3.84]	[1.43]	[3.04]	[3.45]	[2.51]
HPI6MBefore	0.247	0.208	−0.024	−0.026	0.091	−0.098
	[0.53]	[0.62]	[−0.05]	[−0.14]	[0.47]	[−0.59]
HPI6MAfter	−0.279	−0.235	0.098	−0.259	−0.239	−0.223**
	[−0.80]	[−0.87]	[0.26]	[−1.45]	[−1.48]	[−2.40]
2005	−0.016	0.082**	0.063	0.131***	0.023	0.105***
	[−0.40]	[2.19]	[1.25]	[3.51]	[0.81]	[4.29]
2006	0.003	0.095**	0.080	0.220***	0.082*	0.230***
	[0.07]	[2.48]	[1.58]	[4.59]	[1.80]	[5.94]
2007	−0.183***	0.051	0.009	0.154***	−0.051	0.126***
	[−3.09]	[0.96]	[0.15]	[3.09]	[−1.05]	[3.23]
2008	−0.203***	−0.040	−0.050	−0.005	−0.107**	0.083*
	[−2.61]	[−0.47]	[−0.32]	[−0.03]	[−2.18]	[1.66]
Observations	31,405	35,552	25,666	88,778	140,735	336,398
Pr(Delinquency)	0.132	0.180	0.189	0.293	0.246	0.331
$\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$	0.164	0.222	0.220	0.283	0.251	0.300
Pseudo R^2	0.235	0.155	0.186	0.177	0.198	0.159

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. We report the coefficients ($\hat{\beta}$) and t-statistics (in brackets) that adjust for clustering at the MSA level. Dummy variables for states are included, but the coefficients are not reported. At the bottom of the table, we report the sample frequency of delinquency, the pseudo R^2, the number of observations, and the sample average of the probit density function $\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)\right)$ that can be used to construct the average partial effect $\hat{\beta}\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

Table 3.—Choice of Loan Origination Channel and Documentation Level

Dependent Variable	*Third Party* (1)	*Low Doc* (2)	*ThirdParty& LowDoc* (3)	*Third Party* (4)	*Low Doc* (5)	*ThirdParty& LowDoc* (6)	*Broker/ Correspondent* (7)
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]
LoanAmt (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]
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)	89.8%	70.0%	64.6%	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^2	0.149	0.265	0.201	0.145	0.267	0.201	0.0604

The dependent variable is the choice of origination channel, low documentation, and the combination of the two. We employ probit estimation for columns 1 to 6 and ordered probit estimation for column 7, where the choice of broker, correspondent, and bank are assigned as the highest, medium, and lowest outcomes, respectively. The definitions of all variables are given in the appendix. We report the coefficients ($\hat{\beta}$) and t-statistics (in 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^2 the number of observations, and the sample average of the probit density function $\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)\right)$ that can be used to construct the average partial effect $\hat{\beta}\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right)$. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

financially sophisticated borrowers into the third-party channel is compatible with two (nonmutually exclusive) explanations. On one hand, borrowers may select correspondents and brokers because they believe that third-party originators 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 it did expand branch operations in some regions in the past decade. As a result, and as indicated by the empirical results in table 3, the bank relied on third-party originators for the majority of its origination volume, especially as it expanded rapidly into underserved communities. The bank relied on correspondents and brokers both to increase origination volumes in the absence of high visibility as a depository institution and 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 third-party originator. 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 credit scores and reported income tend to choose low documentation, and age is not correlated with documentation level. Finally, black borrowers do not appear disproportionately among 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—information that is quantifiable and could potentially be verified, such as LTV, prior mortgage experience, high income, and high credit score—sort into low-documentation mortgages.

Prior research has shown that lending practices and borrower characteristics are correlated with neighborhood characteristics (Calem, Gillen, & Wachter, 2004; Nelson, 2010). Columns 4 to 6 of table 3 report the relation between neighborhood characteristics and the respective likelihoods that a borrower will obtain a third-party-originated loan or a low-documentation loan. The model's regressors include average per capita income (*AvgIncome*) at the postal code level, as well as the following regressors at the census tract level: log population size (*Population*),[18] percentage of residents who are black (*PctBlack*) and Hispanic (*PctHisp*), median age (*MedAge*), unemployment rate (*UnempRate*), and the state-level change in housing prices during the six-month period preceding loan origination (*HPI6MBefore*).

Third-party-originated loans predominate in neighborhoods with low minority representation and young residents. The combination of results from earlier columns indicates that minority households in nonminority neighborhoods are the prime clients of correspondents and brokers. Low-documentation loans are significantly more popular in minority neighborhoods and in booming neighborhoods (with low unemployment rates, high recent past housing price appreciation, and young populations).

Table 2 indicates that results for the correspondent channel fall between those reported for the bank and broker channels. We therefore supplement the channel choice prediction with an ordered probit analysis, in which the high-, middle-, and low-outcome values are assigned to the broker, correspondent, and bank channels, respectively. Results reported in column 7 of table 3 confirm that noncorrespondent brokers, more so than correspondents, issued mortgages to borrowers (as measured by higher leverage, first-time home buying status, lower credit scores, and minority status) and in neighborhoods (as measured by lower average income and age, and lower recent past housing price run-ups) with lower average credit quality.

B. *Decomposition of Pairwise Subsample Differences in Delinquency*

The analyses in this section attempt to decompose the difference in loan performance across loan types into two components: one that can be predicted based on borrower and loan characteristics that are observable to the lender at 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.[19]

We apply a nonlinear version of the Blinder-Oaxaca (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 compare loans from the Bank ($D = 0$) and Correspondent/Broker ($D = 1$) channels, controlling for documentation level, and we also compare Full-Doc ($D = 0$) and Low-Doc ($D = 1$) loans, controlling for origination 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.

[18] The average and median population size of a census tract is between 5,000 and 6,000 residents.

[19] While an earlier study by Alexander et al. (2002) also documents higher delinquency rates among brokered loans, the study does not contain the level of borrower detail used in our study and hence cannot decompose the difference into differences due to characteristics that are observable versus unobservable to the bank.

Table 4.—Nonlinear Blinder-Oaxaca Decomposition of Differences in Delinquency Rates

	Bank			Correspondent			Noncorrespondent Broker		
	Difference	*t*-statistic	Percentage	Difference	*t*-statistic	Percentage	Difference	*t*-statistic	Percentage
A: Comparison of Full-Doc and Low-Doc subsamples									
Endowment effect	0.43%	0.62	8.7%	0.92%	1.08	8.7%	−0.42%	−0.62	−5.3%
Coefficient effect	4.39%	7.66	91.3%	9.63%	13.57	91.3%	8.15%	16.33	105.3%
Total	4.81%	5.13	100%	10.55%	8.25%	100%	7.74%	7.65	100%

	Full-Doc			Low-Doc		
	Difference	*t*-statistic	Percentage	Difference	*t*-statistic	Percentage
B: Comparison of bank and correspondent subsamples						
Endowment effect	5.02%	7.95	89.5%	10.9%	10.17	96.5%
Coefficient effect	0.59%	1.50	10.5%	0.44%	0.62	3.5%
Total	5.61%	8.96	100%	11.3%	8.76	100%

	Full-Doc			Low-Doc		
	Difference	*t*-statistic	Percentage	Difference	*t*-statistic	Percentage
C: Comparison of bank and noncorrespondent broker subsamples						
Endowment effect	8.46%	10.02	75.4%	10.6%	17.64	74.6%
Coefficient effect	2.76%	7.72	24.6%	3.5%	8.1	25.4%
Total	11.22%	13.21	100%	14.2%	20.45	100%

This table reports the nonlinear Blinder-Oaxaca (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 equation (2).

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

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

or as

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

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. To economize on space, we present results using equation (2) in table 4, where the subsample with higher delinquency rates serves as the base sample for covariate weights.

The left sides of the equations represent the difference in the expected value of the outcome variable (delinquency) between 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, essentially a residual term, is called the coefficient effect because it is equal to 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.

Panel A of table 4 reveals that almost 100% of the 5 to 10 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 indistinguishable from 0 both statistically and economically. We thus conclude that Low-Doc loans are just as good on paper as Full-Doc loans, but they encompass more adverse selection along unobserved dimensions. In other words, 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).

The comparison between bank and correspondent/broker loans conditional on documentation level offers a different picture, as shown in panel B of table 4. Here, the endowment effect accounts for a great majority (90% to 97%) of the 6 to 11 percentage point total difference in delinquency rates between bank and correspondent loans, and exactly three-quarters of the 11 to 14 percentage point total difference between bank and broker loans. Put differently, if the third-party originators were to serve borrowers of the same observable quality as did the bank, then the performance of correspondent loans would be almost identical to that of bank loans, and three-quarters of the difference in the delinquency rates between bank and broker loans would have disappeared. In other words, correspondents seem to serve a clientele that was underserved by the bank, and any adverse selection due to unobserved borrower quality is minimal. In contrast, brokers not only issue loans to borrowers of observably lower credit quality, but they also may attract borrowers of worse quality along unobservable dimensions. Pure brokers have the weakest incentives to screen borrowers diligently and may even endorse borrower behaviors that are correlated with high delinquency, such as exaggerating income in order to qualify for a larger loan.

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. 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 (including information falsification) that are unobserved by the bank. Nevertheless, we emphasize that the message from table 4 could be made stronger: if brokers originate loans to borrowers with worse measured characteristics *and* facilitate falsification of those characteristics, then differences in true characteristics between the borrowers from the bank channel and those from the broker channel can only be larger.

V. Liar's Loan: Borrower Information Falsification

The previous section showed that nearly 100% of the difference in delinquency rates between Full-Doc and Low-Doc loans is due to unobserved heterogeneity. In this section, we provide suggestive evidence that a significant part of the unobserved heterogeneity results from income falsification among low-documentation loans—the "liar's loan" problem. Despite ample anecdotes, there are no formal empirical analyses of borrower information falsification and its impact on loan performance. Our paper fills this void.

A. Borrower Information Falsification: Overview

The term *liar's loan* refers to mortgages that allow borrowers to falsify loan application information, possibly at the encouragement of brokers who have stronger incentives to close deals than to screen applicants. The common perception is that such falsification appears primarily among low- or no-documentation loans, where much of the recorded information is self-reported without strict verification. Due to both financial incentives and the underwriting system, anecdotal evidence suggests that the following falsifications are among the most common.[20] 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 overleveraged 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 or liens resulting from legal action). Borrowers seeking 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 or understate expenses and other debt liabilities—or both.

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 or something else) affects loan qualification and pricing. Second, assets are 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.

B. Identification of Income Falsification

Table 2 shows that in the Full-Doc subsamples, higher income is negatively associated with delinquency; however, the sign on the *Income* coefficient switches in the Low-Doc subsamples.[21] Moreover, the coefficients are particularly strong in the Broker/Low-Doc subsample, where higher reported 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

[20] See, for example, Edmund Andrews, "My Personal Credit Crisis," *New York Times*, 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.

[21] In the regression, the *Income* variable is coded as 0 when it is missing, and the dummy variable for missing income information, *IncomeMiss*, is set equal to 1.

reported 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. Correspondents, as compared to brokers, appear to have stronger reputational concerns due to their exclusive or long-term relationships with the bank.

The dummy variable for missing income information, *IncomeMiss*, offers corroborative evidence. In the Full-Doc subsamples, only 0.8% of the observations have missing income information; moreover, 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 9.4% of the observations in the Low-Doc subsamples. Missing income information significantly predicts higher delinquency propensity in the Broker/Low-Doc subsample, where it is associated with a 6.0 percentage point increase in the probability of delinquency. The same effect is present but less significant in the Bank/Low-Doc and Correspondent/Low-Doc subsamples. 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.[22]

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

$$E(Income^*|X = x, LowDoc) \leq E(Income^*|X = x, FullDoc), \tag{4}$$

where *Income** denotes the borrower's true income and *X* denotes a vector of borrower characteristics. Formally, equation (4) is implied by the condition that *Pr(FullDoc|X, Income*)* is nondecreasing in *Income**.

All that is required for equation (4) 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 than borrower B for low-documentation loans. 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 of a lower sensitivity of loan pricing to uncertified income. Self-reported income could still affect the loan qualification materially, providing an incentive for falsification.

The only group for which equation (4) may plausibly 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 (for example, 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 postal code level from the Internal Revenue Service Individual Master File (IMF) system for the years 2004, 2005, and 2006. A postal code area has 2,326 households on average, and the average 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 (4), with *X* denoting local average per capita income, implies that the average degree to which low-documentation borrowers exaggerate their incomes is at least 16% to 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, for the Full-Doc subsample, average reported income conditional on covariates should be approximately equal to average true income. We report in the heading of table 5 the regression output for full-documentation loans, where the dependent variable is the reported (and certified) household monthly income and the regressors include borrower information (*CreditScore, Female, Age, Hispanic, Black, Asian*), neighborhood economic conditions (*AvgIncome, UnempRate*), and year dummies. In this regression, we include only variables strictly exogenous to individual borrowers, which are also considered to be free from falsification.

[22] 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.

TABLE 5.—DELINQUENCY AND INCOME DECOMPOSITION

	All Covariates			No Contract Terms		
	Bank (1)	Correspondent (2)	Noncorrespondent Broker (3)	Bank (4)	Correspondent (5)	Noncorrespondent Broker (6)
Predicted income	−0.085***	−0.056***	−0.050***	−0.093***	−0.037***	−0.040***
	[−8.40]	[−9.68]	[−14.48]	[−8.74]	[−5.52]	[−10.28]
	−4.45%	−4.31%	−3.96%	−5.28%	−2.86%	−3.30%
Residual income	0.002	−0.004	0.008***	0.006**	0.016***	0.021***
	[0.71]	[−1.53]	[4.39]	[2.53]	[5.56]	[12.31]
	0.30%	−0.65%	1.44%	0.99%	2.98%	3.98%
Observations	12,326	30,391	125,351	12,326	30,391	125,357
Pseudo-R^2	11.1%	10.1%	9.4%	6.1%	6.0%	5.3%

This table reports the relation between delinquency and predicted versus residual income in the subsample of low-documentation loans excluding self-employed borrowers. In the first stage, a predictive regression of *Income* (monthly, in thousands of dollars) using the full-documentation sample excluding self-employed borrowers yields the following outcome:

$$\begin{aligned} \textit{Predicted Income} = {} & 0.015 * \textit{CreditScore} - 0.899 * \textit{Female} + 0.688 * \ln(\textit{Age}) - 0.460 * \textit{Hispanic} - 0.458 * \textit{Black} + 0.524 * \textit{Asian} \\ & [17.91] \qquad [-16.98] \qquad [13.39] \qquad [-2.09] \qquad [-4.37] \qquad [4.49] \\ & + 0.054 * \textit{AvgIncome} - 0.031 * \textit{UnempRate} + 0.146 * Y2005 + 0.398 * Y2006 + 0.334 * Y2007 + 0.039 * Y2008 \\ & [4.42] \qquad [-2.17] \qquad [2.80] \qquad [5.60] \qquad [5.15] \qquad [0.35] \\ & - 5.589 \\ & [-9.66] \qquad R^2\text{: 6.9\%; number of observations: 138,515.} \end{aligned}$$

We impute *Predicted Income* for borrowers of low-documentation loans using the coefficients in the equation above; *Residual Income* is the difference between reported income and *Predicted Income*.

The second stage involves running a probit regression of *Delinquency* on *Predicted Income* and *Residual Income* separately for each of the three origination channels (bank, correspondent, and noncorrespondent broker). Columns 1 to 3 include all covariates used in table 2 except those used in the first-stage regression displayed above; columns 4 to 6 further exclude all covariates that represent loan contract terms (*LTV*, *AddLTV*, *Loan*, *SecondLien*, *Refinance*, *PrepayPenalty*, *ARM*, *OptionARM*, and *IO*). The table reports the probit coefficients, *t*-statistics, and marginal effects associated a 1 standard deviation change in *Predicted Income* and *Residual Income*. Coefficients on other covariates are suppressed for the economy of space. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels.

All coefficients in the first-stage regression are intuitive. Older borrowers and borrowers with higher credit scores tend to have higher incomes. Female borrowers have lower incomes on average.[23] Black and Hispanic borrowers have lower incomes on average than white borrowers, and Asian borrowers as a group have the highest income. Borrower income is significantly and positively correlated with the postal 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 equation (4), which presumes that the error term from the income regression (reported in the heading of table 5) 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 the regression to the covariates of these borrowers. We generate a residual income variable (a proxy for income exaggeration) to capture the difference between the reported *Income* (as reported by the borrower) and *Income** (predicted 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% (20.0%) above their estimated true income level. As these are conservative estimates, the data suggest serious income falsification among low-documentation borrowers using full-documentation borrowers as a benchmark.

C. The Impact of Income Falsification on Delinquency

Our analysis also shows that the correlations of estimated true income, estimated income exaggeration, and loan performance are all highly statistically significant; these relationships are informative about the incentives for and consequences of income falsification. First, the correlation between estimated true income and estimated income exaggeration in percentage terms is −7.9%, indicating a stronger incentive to inflate income when true income is lower. Second, the correlation between estimated true income and ex post delinquency is −23.5%, which recovers 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.

We further calibrate the effect of income falsification on delinquency in the subsample of low-documentation loans (excluding self-employed borrowers) by running the delinquency probit regressions with *Predicted Income* and *Residual Income* as regressors of key interest (replacing *Income*) and separately for each of the three origination channels (bank, correspondent, and broker). For each regression, we adopt two specifications. The first specification contains all loan and borrower covariates except the covariates appearing

[23] This gender effect is not primarily due to the male-female wage gap but rather to the fact that a female head of household is usually associated with lower household income.

in the first-stage income predictive regression.[24] It uncovers a direct (or marginal) effect of *Predicted Income* and *Residual Income* on delinquency, conditional on all other covariates being constant. Only the broker channel exhibits a significant positive relation between *Residual Income* and *Delinquency*, where the marginal effect of a standard deviation variation is 1.4 percentage points. This, however, does not represent the total effect if a primary motive to falsify income is to qualify for more favorable loan terms. For example, suppose income exaggeration qualifies a borrower for a loan beyond the borrower's true means. In this case, the falsification behavior does lead to higher delinquency probabilities once the loan amount is controlled for.[25]

For this reason, we adopt a second specification that excludes loan contract variables from the set of covariates used in the first specification. The remaining covariates presumably should not be affected by reported income. In this specification, the indirect effect of income misreporting via loan contract terms is reflected in the coefficient on *Residual Income*. After accounting for the indirect effects, the total effect of *Residual Income* becomes significantly positive across all origination channels. The strongest effect remains with the broker subsample, in which a 1 standard deviation increase in *Residual Income* is associated with a 4.0 percentage point increase in delinquency probability.

As a placebo test, we run the delinquency regression on *Predicted Income* and *Residual Income* using the sample of full-documentation loans and including the same set of covariates as the regressions in columns 1 to 3 of table 5. We find that both *Predicted Income* and *Residual Income* are significantly (at less than the 5% level) negatively associated with delinquency. More specifically, a 1 standard deviation increase in *Predicted Income* (*Residual Income*) is associated with an 11.9 (0.51) percentage point decrease in delinquency probability. The contrast in the effect of *Residual Income* on delinquency between full- and low-documentation loans suggests that income exaggeration prevails in the low-documentation subsample only, especially among brokered loans.[26]

[24] These covariates are excluded because they are almost collinear with *Predicted Income*. If we keep them in the regression, the coefficient on *Residual Income*, the variable of key interest, is little affected. However, the coefficient on *Predicted Income* would be rendered insignificant (and small in terms of economic magnitude). If we exclude a subset of the covariates used in the first-stage regression to relax the collinearity, we obtain results similar to those reported in table 5.

[25] In a similar vein, we do not find any significant relation between *LowDoc* and loan terms, indicating that low documentation status per se (holding reported borrower information constant) does not result in more favorable loan terms. The adverse selection effect comes from the indirect effect of information falsification facilitated by low documentation requirements.

[26] When comparing *Residual Income* across the origination channels, we do not find a significant difference. Given that high *Residual Income* is associated with high delinquency rates only in the broker channel when all covariates are controlled for, *Residual Income* likely proxies for income exaggeration in the broker channel, but mostly reflects income idiosyncrasies in the bank and correspondent channels.

The magnitude of the total effect from potential income falsification among low-documentation loans is economically meaningful. A 1 standard deviation increase in the proxied income exaggeration accounts for about half of the delinquency difference between the Broker/FullDoc and Broker/LowDoc subsamples and is comparable to decreasing credit score by 25 points or changing the *IO* (interest-only) status from 0 to 1. Moreover, the majority of the indirect effect (i.e., the difference between 4.0 and 1.4 percentage points) comes from *LoanAmt*. Such a relation reflects underwriting guidelines that allow borrowers with higher reported incomes to qualify for larger loans without the constraint of carrying higher proportional leverage (for example, *LTV* or *AddLTV*). The cross-channel pattern is also consistent with the hypothesis that third-party originators, especially brokers, are less diligent in screening borrowers and perhaps are more likely to endorse or even assist income falsification.

VI. Loan Types and Loan Pricing

Given that loans originated by third parties and with low-documentation requirements are of lower quality, 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.[27] 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.

Most of the coefficients on the control variables are intuitive; that is, variables associated with higher delinquency rates also tend to command higher interest rates. The only notable exception is loan size (*LoanAmt*), 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

[27] We acknowledge that the reported interest rates may not fully reflect compensation to the lender (or broker) because we do not observe the prices at which the loans were purchased by investors, nor do we observe borrowers' paid points or origination fees.

TABLE 6.—DETERMINANTS OF INTEREST RATES

Sample	Fixed Rate Only	Adjustable Rate Only	Adjustable Rate, 2004–2005	Fixed Rate, July 2007–2008
Dependent Variable	Initial Rate	Initial Rate	Current Rate	Initial Rate
	(1)	(2)	(3)	(4)
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]
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]
ThirdParty	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^2	0.766	0.851	0.571	0.447

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

pricing based on demographic information including race/ethnicity. Although some coefficients on these variables are statistically significant, the economic magnitude is quite small.[28]

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 third-party issued 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. Even more surprising, the same rate premium is nonexistent for brokered loans.

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 3), for which the interest rate is more likely to have been reset. Combined, the two results suggest that brokers are more likely to sell products that carry a more lucrative teaser rate. The same contrast is nonexistent for low-documentation loans.

If we run a delinquency probit regression with the same set of regressors, the dummy variables *LowDoc* and *Third-Party* command coefficients (in marginal probability terms) of 9.4% and 2.7%, respectively; both are significant at less than the 1% level. Therefore, the overall evidence in table 6 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 et al. (2002) show that the agency risk associated with brokers was not priced during periods 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 fixed-rate loans issued after July 1, 2007 (right after two Bear Stearns 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 decreases 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.

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. We offer three possible explanations with some factual support.

The first is information. We believe that the bank was aware that low-documentation borrowers were qualifying for loans larger than what they were able to sustain. If we add a *LowDoc* $\times$ *LoanAmt* interaction to the regressions in table 6, we find that the coefficients in the two adjustable-rate regressions, at 24 (17) basis point for initial (current) rates, are significantly positive (at less than the 1% level). This interactive relation indicates that the bank priced large loans more aggressively for low-documentation borrowers. But it was not obvious during most of the sample period that low-documentation loans were more delinquency prone (controlling for loan contractual terms) when overall delinquency rates were low. 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.

The second explanation relates to regulations. The bank also should have been aware of the observable differences in borrower credit quality across origination channels, yet it relied on third-party originators for boosting volume and earning credit under the Community Reinvestment Act. The bank also adopted strict measures to ensure a one-door pricing policy (no differential pricing treatment based on origination channel) in order to comply with fair lending regulations. Yet the bank may well have underestimated the difference, as evidence of higher delinquency rates among third-party-originated loans did not surface until the housing market softened in 2007.

The final 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 readers to Keys et al. (2010) for an analysis of the relation between loan performance and the ex ante probability of loan securitization, and to Jiang, Nelson, and Vytlacil (forthcoming) for a contrast between the ex ante and ex post relations.

VII. Conclusion

This paper uses a unique proprietary data set from a major national mortgage bank to examine how mortgage

[28] For example, after controlling for observable information, black borrowers pay an additional 10 to 17 basis points on the interest rate as compared to white borrowers. The estimated black-white difference in interest rate amounts to an additional monthly payment of $16 to $27 (or $13 to $22) using the mean (or median) balance, which should not contribute to a significant difference in loan delinquency rates.

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 third-party-originated loans are more than 50% more likely to be delinquent than bank-originated loans, and about three-quarters of this difference can be attributed to lower borrower and 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 strong evidence of information falsification among low-documentation loans, especially among broker-issued loans. Finally, we show that loan pricing did not adequately compensate for the additional risk of brokered and low-documentation loans.

REFERENCES

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

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

Calem, Paul, Kevin Gillen, and Susan Wachter, "The Neighborhood Distribution of Subprime Mortgage Lending," *Journal of Real Estate Finance and Economics* 29:4 (2004), 393–410.

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

Dell'Ariccia, Giovanni, Deniz Igan, and Luc Laeven, "Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market," Centre for Economic Policy Research discussion paper DP6683 (2008).

Demyanyk, Yulia, and Otto Van Hemert, "Understanding the Subprime Mortgage Crisis," New York University working paper (2007).

Jiang, Wei, Ashlyn Aiko Nelson, and Edward Vytlacil, "Mortgage Securitization and Loan Performance: A Contrast of Ex Ante and Ex Post Relations," *Review of Financial Studies* (forthcoming).

Joint Center for Housing Studies, *The State of the Nation's Housing 2008* (Cambridge, MA: Harvard University, 2008).

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *Quarterly Journal of Economics* 125:1 (2010), 307–362.

LaCour-Little, Michael, "The Pricing of Mortgages by Brokers: An Agency Problem?" *Journal of Real Estate Research* 31:2 (2009), 235–264.

Mayer, Christopher, Karen Pence, and Shane Sherlund, "The Rise in Mortgage Defaults: Facts and Myths," *Journal of Economic Perspectives* 23:1 (2009), 27–50.

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

Munnell, Alicia, Geoffrey Tootell, Lynn Browne, and James McEneaney, "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review* 86:1 (1996), 25–53.

Nelson, Ashlyn Aiko, "Credit Scores, Race, and Residential Sorting," *Journal of Policy Analysis and Management* 29:1 (2010), 39–68.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review* 14:3 (1973), 693–709.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig, "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis," *Journal of Financial Economics* 97 (2010), 369–397.

Rosen, Richard, "The Role of Securitization in Mortgage Lending," *Chicago Fed Letter*, no. 244 (November 2007).

APPENDIX

Definition of Main Variables

AddLTV	The loan-to-value 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 postal code 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 sixty days behind in payment
Female	Dummy variable = 1 if the borrower is female
CreditScore	Borrower credit score
CurrRate	Current interest rate on the loan as of February 2008
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 after 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
LoanAmt	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 a 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
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